Università degli Studi di Parma

SCUOLA DI DOTTORATO IN INGEGNERIA E ARCHITETTURA
Dottorato di Ricerca in Ingegneria Civile
XXVIII CICLO - CURRICULUM: GEOMATICA (ICAR/06)

Semi-Global techniques in Image Matching and Change Detection with applications to Civil and Environmental Engineering

Dissertazione per il conseguimento del titolo di Dottore di Ricerca

Coordinatore: Prof. Ing. Gianfranco Forlani
Relatore: Prof. Ing. Riccardo Roncella
Tutor: Prof. Ing. Gianfranco Forlani

Dottoranda: Elisa Dall'Asta

Parma, Gennaio 2016

Alla mia famiglia …
calore e colore del mio cammino

Contents

CHAPTER 1

INTRODUCTION

Introduction and context

The history of photogrammetry is closely related, in his theoretical principles, to the history of descriptive geometry and in particular to the formulation of the theory of perspective. Differently, with regard to the applications, it is related to the history of optics, photography and related technological developments. Photogrammetry represents one of the most reliable approach for metric (and thematic) data acquisition; therefore, it is important to highlight that the purpose of reconstruction techniques, based on this approach, must be focused on the achievement of the most accurate, complete and detailed three-dimensional reconstruction.

In recent years, photogrammetry has experienced an important development and renovation, and the possibilities, now made concrete by the electronic and information technologies, offer new perspectives in all the detection and survey applications that make full use of three-dimensional digital contents.

Image matching, digital image correlation, stereo correspondences problem: all refer to the class of noncontact techniques that utilize two or more digital images depicting (at least partly) the same scene and provide a dense reconstruction of the shape and the spatial position of the analysed object. In other words, given two (or more) views of a scene, correspondence needs to be established among homologous features [160], which are projections of the same physical identity in space. A stereo algorithm is usually applied to compute the disparity map (i.e. a rastermap where the value of each cell is equal to the difference (also called disparity value) between the coordinates of corresponding pixel locations on the two images of the stereo pair).

It is possible to say that image matching represents a key component of many photogrammetric and computer vision tasks. However, right now, it is quite hard to identify a completely satisfactory taxonomy of the different techniques: encouraged by the growing interest in automatic 3D image-based reconstruction, a large number of image matching algorithms and methods have been developed and refined over the years leading to the complete automation of the entire 3D modelling process in several application areas (such as Remote Sensing data processing, Medical Image analysis, Computer Vision and Pattern Recognition [30]). A very good overviews on the topic can be found in literature (see for instance [13][157]).

An easy way to classify the image matching algorithms is to consider the matching primitives: in accordance with that condition it is possible to distinguish between Feature Based Matching (FBM) and Area Based Matching (ABM) methods.

FBM algorithms [58][59] extract features of interest from all the images (lines, edges, peculiar shapes, etc.) using specific analytical operators [116] and afterwards identify

correspondences between feature lists. These methods are reliable and fast, but they produce sparse disparity maps that are generally thickened in a second step by using area based matching or model-based interpretation step (e.g. the assumption that the intermediate space between the edges is occupied by planar surfaces). Differently, in ABM, starting from a point on a reference image, the algorithm identifies the most probable location of the homologous point on the other images; usually this operation is computed for all the points of a dense points grid defined on the image. Consequently, these methods are able to produce dense disparity maps, allowing the reconstruction of the parallax field in all the image areas.
Since the matching problem is an important research topic both in photogrammetry and computer vision, in a short time (the last decades) several methods and applications have been developed and are still developing. In this context, the thesis will discuss of some popular and widespread stereo matching algorithms (local and global) and techniques. In particular, the Semi-Global Matching (SGM) [82][84] will be deeply analysed, and a proprietary code implementation will be presented.
Being the final purpose of this dissertation, the description of the developed image matching software code, the current chapter will now present a brief analysis of the state of the art in image matching (and the related application fields) followed by a brief analysis of the objectives and motivations of this work.

1.1 State of the art in image matching algorithms and techniques

Over the past 60 years, important changes have characterized the advancements in image matching techniques and, according to Gruen [73], the chronology can be structured into the following four period: the "Early Years" (1960s – 1960s); the "New Approaches" (1980s); the "Time of consolidation and Extensions" (1990s) and, finally, the "Time of Acceptance" (2000s).
The earliest matching algorithms were developed in the photogrammetric community in the 1950s [88]. In the 1970s, the concepts of epipolar geometry and cross-correlation [77][78][79] for image matching were rapidly spread and, in the middle of the 1970s, the advent of digital images has allowed the researcher to start focussing on automatic processing of photogrammetric procedures (to offer the possibility of replace manual operator intervention and achieve more powerful and accurate performances).
By the early 1980s, the literature on image analysis and matching had grown extremely thanks to the important developments of new, more powerful and automatic, matching approaches. Precisely in the early 1980s we saw the most significant contributions: the

Least Squares Matching (LSM) technique [68], which is currently used in many digital photogrammetric tasks. Early investigations were reported by [58] [2] [135]. In [59] and [69] the multi-photo geometrically constrained (MPGC) extension were introduced. Indeed, the quality of a matching procedure depends mainly on the type and content of the image signal but very often there is, however, additional information available that can support the matching. These important categories of information are geometrical and radiometric conditions, which have to be added to the observation equations in the least squares context. Thanks to the use of all the geometric information available and the internal consistency of the algorithm, it offered considerable advantages with respect to precision and reliability of the results ([73]). Subsequently, the matching procedure was also generalised to object space through the introduction of the concept of the "groundel" or "surfel" ([51][80][200]).

As accurately described in [148] and [73], the 1990s was a time of consolidation for image matching and a large number of commercial photogrammetric systems appeared, in particular, for automated Digital Surface and Terrain model (DSM/DTM) generation from large blocks of near-nadir aerial images. The results of studies comparing the performance of different digital stations with respect to DTM generation are illustrated in [71]. The results of yet another empirical accuracy study were published in [70]. The matching software of three commercial systems was tested with three different aerial image data-sets of different image scales. The RMS (Root Mean Squares) errors achieved were worse than the theoretical expectations for manual measurements and this was largely due to very large numbers of blunders. Gong et al. [65] have also carried out an interesting assessment, obtaining low-quality results as well. Thus, the capabilities of existing matching algorithms had not been fully utilised but, in the early 2000s, significant improvements and much attention to the automatic DSM/DTM procedures were obtained. These years represented the period of major diffusion of digital photogrammetric workstations, in particular for aerial/satellite images processing. Differently, Close-Range applications were still considered an area of interests of research groups (due to the difficult in obtaining high-quality and accurate results). At the same time, new and rapid developments in this area were largely achieved in the computer vision community where, at first, the attention was focused on obtaining complete results in real-time. In this field, stereo matching was investigated as early as the mid-1970s [113] and developments continued in the 1980s mainly for terrestrial applications [5]. Then, in the 1990s, the focus moved to multi-view approaches [127][61], 3D imaging [15] and then, more recently, to field programmable gate array (FPGA) and graphics processing unit (GPU) developments in computer architecture [92].

1.2 Image matching applications

Over the years, a wide range of applications made full use of image correlation techniques for analysing various type of problems and data (close-range/archeologic survey, orthophotos generation, change detection, precision agriculture, Geographic Information System (GIS) services, and obstacle avoidance problems), making use of different type of data (such as Red-Green-Blue (RGB), Near-Infra-Red (NIR) and multispectral images, and so on). The possible products that can be obtained from matching process varying from three-dimensional surface and volumetric models, to depth maps and planar layers (such as orthoimages, panoramic layers and rectification of plane surfaces) and find suitable applications in many areas, including: monitoring activities, topographic mapping and GIS analysis, entertainment, automated systems, displacements/deformations measurements, high accurate 3D digital surface models, clinical studies, and many, many more. Scientific and applicative tasks for digital stereo vision include the extraction of information from aerial/satellite surveys, for contour maps calculation, volumetric surveys or geometry extraction for 3D building models and mapping. Other applications in robotics field include object recognition, where depth information allows the system to separate occluding image components which the robot may otherwise not be able to distinguish as a separate object by any other criteria.
Hereafter, a synthetic description of state of the art for some image matching applications is presented. It is important to highlight that the following brief examination will regard only the fields of study that have been investigated in more detail during the work of thesis. Such applications have, in fact, required the development of special functions and the obtainment of specific performance from the matching image implemented algorithm. For this reason, the peculiarity of each field of interest (in terms of algorithmic requirements and observed difficulties) will be following presented.

1.2.1 Dense image Matching for cartography and DEM

Digital Terrain Models and Digital Surface Models are classical and common products of a photogrammetric system and they have large relevance in many territorial/environment engineering applications such as modelling water flows or mass movements, topographic mapping, natural hazards analysis, spatial and temporal change detection, visualization and many others [47]. For decades terrestrial surveying techniques and aerial images from airborne/spaceborne sensors were the only

approaches available for generating DTM and DEM (Digital Elevation Model). With the launch of IKONOS on September 1999 [153], very high resolution (VHR) optical satellites have been introduced, leading to significant improvements and transformations in digital elevation model extraction pipeline.

At present, the data required for the generation of DSMs can be acquired by several sensors/techniques, among which airborne LiDAR, aerial photogrammetry, optical and radar spaceborne sensors play the major role. In modern map production, they have become an information source for scene analysis and understanding, for change detection, for GIS database updating, for cartographic 3D feature extraction and reconstruction. In particular, DSMs in suburban and urban areas are very useful for a great number of applications, which include the inter-visibility calculations for the optimization of the location of telecommunication antennas, risk mapping, (true) orthoimage generation, mission planning and rehearsal, virtual and augmented reality applications and urban planning.

Alongside the sensor-side developments (linear array sensors have being widely used to acquire panchromatic and multi-spectral imagery for photogrammetric and remote sensing applications), the algorithmic procedure for DEM extraction has been improved significantly. Indeed, the processing of these kinds of images has provided a challenge for image matching algorithmic redesign and this offers the possibility of reassessing and improving many photogrammetric processing components (like image enhancement, multi-channel colour processing, triangulation, orthophoto and DEM generation and object extraction). In fact, recently matching techniques have been improved by adapting the algorithms to aerial [207] and terrestrial photogrammetry [54], partly overcoming some of the matching problems that affect the quality of the final results (such as repetitive patterns, occlusions, shadows, reflections, etc., as regard, in particular, urban areas). For example, Zhang [207] has presented a matching procedure for automatic DSM generation from linear array imagery data. The proposed approach is able to produce reliable and accurate results by the use of an innovative procedure, which combines different image matching algorithms and automatic quality control systems; the resulting DSMs will be in turn a combination of matching results derived from different analysed primitives (feature points, grid points and edges). In recent years, others research efforts have been devoted to the efficient utilization of satellite images.

The recent developments and improvements of the matching techniques performances for dense point clouds generation has gained considerable research attention over recent years. In fact, over the last decade, LIDAR techniques have replaced traditional photogrammetric systems in many applications, because of their speed in point cloud

generation and the percentage of acceptable points. A traditional image matching survey, based on the use of stereo pairs, can acquire no more than 80% of the possible points ([125]). However, with regard to satellite/airborne applications for DSM/DTM production, laser scanning techniques have non-negligible limits due to the impossibility of directly obtaining radiometric information and the exact position of object breaklines. Furthermore, other limits for LIDAR terrestrial applications are the weight and the size of the instrument, the cost, the limited range of application and the output data volume, which often requires high performance workstations to be managed and analysed. Finally, when point clouds are produced, segmentation and classification procedures have to be applied in order to correctly interpret and model the surveyed object. For this reason, most LIDAR surveys are integrated by digital image acquisition. Digital photogrammetry directly associates a radiometric information to the acquired points and the use of a stereo pair allows a manual or semi-automatic survey of the breaklines, when automatic algorithms fail. For these reasons, a wide variety of approaches have been developed and automatic DTM generation packages ([184][142][102]) are nowadays commercially available for most digital photogrammetric workstations, focusing on improving the performances of matching techniques in order to generate dense point clouds from only images.

1.2.2 Geostructural surveys and monitoring activities

Environmental control and monitoring systems represent the most interesting and important fields of interest and application for geologists and geotechnical engineers: analysing the evolution of an unstable slope may provide important information to achieve a better knowledge of the active processes that could lead to failure and to forecast potential geo-disasters [110] [199]. The behaviour of different phenomena (in the time domain as well as in the space domain) depends on a lot of factors, which presents the geologists and the surveyors each time with different challenges. One of the most widespread earth surface hazard is due to slope processes such as falls of rock-weathered fragments, landslide in cliffs together with civil engineering structures stability checking (and maintenance) problems. Therefore, it is crucial to investigate, and to develop, prevention measures and monitoring systems that allows evaluating slope stability and predicting the failure risk (e.g. by using automatic measurements systems).

In principle, as far as a geometric survey is concerned, the main parameter when designing a measurement and control system is the accuracy needed to assess, with a given probability, the magnitude of the expected displacement. However, a number of

other issues influence the choice of the best monitoring system to use: the size of the area to control, the frequency of data acquisition, the time to deliver the results (alert time), the stability of the reference system, the influence of atmospheric parameters on measurement accuracy or operation, the site constraints, etc. In this context, photogrammetry is one of the many fields for three-dimensional measurement that can be taken into account. It has been used since long to periodically control the evolution of landslides, either from aerial images [33], as well as from ground [31]; in [117]the same technique has been used in combination with GPS surveys on the landslide body. Survey activities performed with photogrammetric methods allow assessing a regular basis monitoring in order to investigate landslide or slow movements, and to assess whether a movement is still active. The use of photogrammetry techniques and systems represents a low-cost alternative (w.r.t. other techniques such as total station/LIDAR monitoring, ground-based SAR (GB-InSAR), GNSS and so on) to have important information about the monitored object/area: indeed, thanks to image matching, it is possible to reconstruct a dense description of the object itself and, comparing the digital surface models reconstructed at different time instants, an accurate description of the object movements during time can be obtained. The systems/platforms used to acquire the data can be different: images can be taken using a UAS (Unmanned Aerial System) flight (aerial photogrammetry) or from the ground (close range photogrammetry) and several are the photogrammetric products that can be used to study the investigated geo-structural phenomena (DSMs, orthophotos, depth maps, displacement maps, etc.). The advantages of this method is that the acquisition systems and procedures are cheap and totally user-customizable: it is possible to decide the acquisition frame rate, the monitoring frequency, the precision and accuracy of the obtained photogrammetric products, considering just the system and data analysis costs. However, more recently, terrestrial (TLS) and aerial (ALS) laser scanning are also being used as alternative [23][1]; in this regard, on the contrary of LIDAR techniques, photogrammetric survey methods suffers from some disadvantages connected with the images acquirement settings: the presence of bad meteorological conditions (such as fog, wind, snow) can reduce (or make impossible), the survey quality, the reference system definition in hard-accessible areas can be difficult (in particular for Ground Control Points (GCP) measurements and monitoring), and so on. However, in particular in these context, the use of UASs is rapidly spreading in many applications and will surely become a promising alternative in the next future: their relatively low cost and their capability of acquiring concurrently geometric (usually producing Digital Surface Model) and thematic data (using RGB or NIR imaging systems), as well as a very good productivity rate, make the technology extremely appealing also for monitoring applications.

1.2.3 Planetary mapping

The importance of high resolution DTMs in the geomorphological studies of planets and asteroids has established stereo coverage and the photogrammetric reconstruction of the body surface as a standard among the scientific requirements of space missions [145][184] [163][146].

From the acquisition of stereo-images of planetary bodies, it is possible to reconstruct the DTM, that could also be derived, usually with lower resolution, from laser altimeter experiments. Nowadays, the acquisition of planetary bodies stereo-images has become (and will continue to be) one of the most important requirements for many planetary missions. In fact, the possibility to have high-precision topographic information is critical for planetary exploration and surface operations and the DTM generation is acquiring a central role in the programming and realization of the future space missions on the solar system planets. These products support the choice and the analysis of the landing sites and provide important information for geologists studying the geomorphology, structure and physics of the planets and asteroids. The great progress on last decades in high-resolution imaging of planetary surfaces allows providing detailed 3D geomorphological information over planetary surface and also the height variation with a fine spatial resolution. [145]. However, the process for deriving DTMs starting from these raw data is very complex because it has to consider two important aspects: it must create the highest accuracy results possible (considering that image data transfer is very costly in the mission budget), and it must be able to work with extremely large data volumes (since from a planetary mission the whole planet surface should be acquired and processed). Therefore, it is necessary to consider that a matching algorithm must be optimized for processing the high resolution images provided from the planetary missions (e.g. LROC mission produced more than 160 GB of image data so far). Efforts must be mainly directed at the improvement and investigation of the image correlation kernels/functions (for obtaining high performance in terms of precision and accuracy) and of the process optimization in terms of memory requests and computational time (implementing for example tiling and multi-resolution strategies) n fact, several aspects, such as occlusions, shadows, low image-texture, atmospheric dust, steep terrain and also the illumination conditions can be, in this context, more problematic for the image matching performances, and they must be taken into consideration during the development/adaptation of the matching algorithms or in the case of its simple application.

Nowadays the most important institutes involved in the planetary mapping (University College of London (UCL), German Aerospace Center (DLR), U.S Geological Survey

(USGS), Ohio State University (OSU)) are working on developing strategies to fulfil these requests.

1.2.4 Digital Image Correlation analysis

Accurate measurements of full-field strain components are required in experimental solid mechanics testing. In this context, it is usual to speak of Digital Image Correlation (DIC), which refers to a class of non-contacting techniques that utilize two or more digital images and, performing image analysis, are capable of extracting a full field description of the deformations of the object itself [31]. The basic principle of DIC is the tracking of the same points (or pixels) between two consecutive images (i.e. before and after loading.) and the feature tracking is usually achieved using Area Based Matching (ABM).

The improvement in image processing by means of microcomputers has fostered non-contact measurement techniques to become more and more popular in the experimental mechanics community [37] in comparison to some full-field measurement techniques like Moiré, interferometry or photoelasticimetry. In the last decade, the opportunity to obtain displacement or strain contours directly (by automatic image processing), with low costs and high accuracy, has determined an increased use of optical devices, digital cameras, algorithms and software. DIC was originally proposed in the 1980's [38] to study 2D solid mechanics problems, such as resin films [119], fiber reinforced polymer composites [120] and concrete [36]. In asphalt pavement technology field, where our research group has important co-operation agreement, Kim and Wen [98]first proposed the use of DIC technique as a possible displacement/strain measurement method for asphalt mixture. Seo et al. [166] utilized a DIC technique to investigate the size and shape of the fracture process zone for asphalt mixtures. Masad et al. [114] used both digital imaging and X-ray computed tomography techniques to evaluate the microstructure of hot mix asphalts in terms of aggregate orientation and air voids concentration, as well as strain distribution. Birgisson et al. [21] used DIC to validate the theory at the base of the visco-elastic fracture mechanics-based crack growth, which identifies a fundamental crack growth threshold as the key element in defining the cracking mechanism and fracture resistance of asphalt mixtures. Nowadays, DIC strain measurement is a continuously improving technique: for instance, Nashon et. al. [120]have analysed the ductile fracture of aluminum panels, using the obtained results also as calibration and validation data for the numerical modeling of ductile fracture in large structures. Strain and displacement analysis of specific and unusual materials were recently tested by using DIC: Makki et. al. [112] have presented the stress localization

and concentration for isotropic and orthotropic materials with holes and, in Petrikova et. al. [136], the big deformations of a hyperelastic material are investigated. Recent advances and perspectives in DIC and related methods for accurate, full-field deformation mappings have been described in [176] and in [46] the application of different algorithms for DIC applications is presented.
The use of matching algorithms to obtain a full-field descriptions of the analysed specimens and their mechanic characteristics, requires the use of some precautions, in order to provide accurate results. To ensure a successful imaging acquisition, the specimen surface must show a random grey intensity distribution (i.e. random speckle pattern) that it is not always guaranteed by the natural texture of the material (asphalt mixture, aluminium panels, plastic samples, etc.). Therefore, the speckle pattern is usually artificially made, resulting in a homogeneous and well-contrasted randomly oriented texture, paying attention to the used paint, which mustn't affect the real material behaviour. The correlation results, in term of accuracy of points displacement measurements, will depend on the resolution of the used digital cameras, the quality of the speckle pattern on the specimen surface and, not least, the conditions and setting of material testing system (climatic conditions inside the testing chamber, sources of illumination, etc.). The choice of the correlation algorithm represents also an important task: recently, in [46], the use of a Semi-Global algorithm in a 2D disparity search range space has been tested, comparing it with traditional Least-Squares matching method. The use of a semi-global (or global) technique for this kind of applications, where the continuity of the displacement field through the image sequence is requested, can be an appropriate method since the hypothesis at the base of the algorithm (i.e. the regularity of the displacement field) perfectly describe the physical geometry of the problem.

1.3 Motivations and Objectives

Since the early 60s, the introduction of digital cameras and imagery started the development of automation in the photogrammetric processing; in the last decades, advanced computer technologies has enabled the processing of digital images, regarding in particular the automatic recognition and measurement of image features and the matching process (research matter of this work of thesis) involved in the generation of the final dense point cloud that describes the surveyed scene. By the early 1980s, the literature on image analysis and matching had grown extremely: the methods and achievements obtained in the photogrammetric community are well-described in [7] [111]. The most significant contribution was probably the Least Squares image Matching technique [68] which, thanks to its adaptability and accuracy, has been widely

researched and explored in the scientific community and in many image matching tasks. Simultaneously, Computer Vision scientists has also had a significant impact in the development of matching techniques (as an example in [5]). Obviously, in the current scientific scenario, the boundary between the two disciplines is increasingly blurred: image matching products (such as three-dimensional digital surface and terrain models, displacement/deformation maps, image blocks orientation, image registration, etc.) are essential component for both the disciplines; at the same time, currently, the application and interest fields are very different. In photogrammetry the image matching products represent crucial data for monitoring activities, Geographic Information System (GIS) development, topography surveys and cartography production, image registration, points tracking for deformation studies in engineering applications and three-dimensional reconstruction and modelling. Differently, in CV the interest is shifted towards automatic navigation systems, robotic automation, object recognition, entertainment, security and intrusion detection systems, motion tracking and 3D modelling. Moreover, in the last decades a common cliché stated that photogrammetry was focused primarily on image metrology, i.e., on the precision and reliability of the results [10], while CV was primarily aimed at the results completeness and the computational efficiency of the process (in order to allow the three-dimensional models generation with low computationally efforts). However, it is important to notice that, in the current scenario, these differences have almost disappeared and, as initially said, the boundary between the two disciplines (both in term of application fields (CV-based techniques and software packages are more and more being used for engineering [11][180], geology [191], geoscience [90][55] and cultural heritage [4][72] applications) and method major objectives) is increasingly blurred. If, on one hand, in CV the achievement of high accuracies is becoming an important and required tasks, and their software application are already capable of guaranteeing such results, on the other, in photogrammetry, the process speed and results completeness are becoming key factors. Therefore, it is important to understand and identify the successful strategies and find the interferences, rather than the differences, between these two approaches, integrating the qualities of both disciplines in order to solve each other's weak points.

This work of thesis wants to present a dissertation of the wide range of modern dense matching algorithms, which are spreading in different application and research fields, with a particular attention to the innovative "Semi-Global" matching techniques. the choice of develop a semi-global numerical code was justified by the need of getting insight on the variables and strategies that affect the algorithm performances with the primary objective of maximizing the method accuracy and efficiency, and the results level of completeness.

The Semi-Global Matching stereo method realizes a pixel-wise matching and relies on the application of consistency constraints during the matching cost aggregation. Combining many 1D constraints realized along several paths, symmetrically from all directions through the image, the method performs the approximation of a global 2D smoothness constraint which allows detecting occlusions, fine structures and depth discontinuities. The regularity constraints allow using very small similarity windows (usually 1÷5 pixel) making the method particularly robust where shape discontinuities arise; on the other hand, traditional area based (template) matching techniques, using bigger templates to achieve good accuracies, are more prone to such issues.

The dissertation will consist in the metrological characterization of the proprietary implementation of the semi-global matching algorithm, evaluating the influence of several matching variables and functions implemented in the process and comparing the accuracy and completeness of different results (digital surface models, disparity maps and 2D displacement fields) obtained using our code and other commercial and open-source matching programs in a wide variety of application fields.

In Chapter 2, an attempt of classification and evaluation of the stereo Area-Based matching algorithms and all the parameters and processes strictly connected with the image correlation workflow, will be illustrated. Focusing on the essential component of the images correspondence problem, in the first part of the thesis, stereo methods that operate on two frames (stereo pair) will be described, and a general way of characterizing the stereo algorithms workflow will be presented. Multi-Image methods will be then investigated and the more popular techniques will be described.

After this general overview, in Chapter 3 the Semi-Global proprietary implementation will be presented: the chapter will face, at a first level, the classical stereo matching implementation (i.e. that operate on two frame); after that, the chapter will deal with the algorithm developments of multi-image approaches and the innovative extension to bi-dimensional research of correspondences (e.g. temporal point tracking).

Next, in chapter 4, the performances evaluation of the algorithm in surface reconstruction applications is presented. First, the results of some tests performed with other stereo matching software package and algorithms will be discussed, with the aim of inspecting the accuracy and completeness of the obtained three-dimensional digital models and of analysing the influence of several process variables. Second, several application fields and case studies will be presented, to illustrate the algorithm capabilities in real world scenarios.

Finally, the results of the novel extended semi-global code (for 2D displacement search domain) will be presented in Chapter 5, together with an explanation of the different areas of interests and the capabilities of the method.

In a concluding chapter, an overview of the all the aspects covered in the dissertation, with particular regard to the emerged applications and other possible interest areas, will be discussed, highlighting the elements of originality conducted and possible future developments.

CHAPTER 2

AREA-BASED IMAGE MATCHING ALGORITHMS & TECHNIQUES

Introduction

Area-based methods perform the correlation among brightness (intensity) patterns in the local neighbourhood of a pixel in one image, with brightness patterns in a corresponding pixel neighbourhood in the other image. In other words, they make a one by one search of the homologues: each image point to be matched is the centre of a small window of pixels (template) in a reference image (master image). The grey values of the template are (statistically) compared with those of an equally sized window of pixels (patch) in a search image (slave image), which corresponds to the second image of the stereo pair. A simple explanation in described above in Figure 2.1.

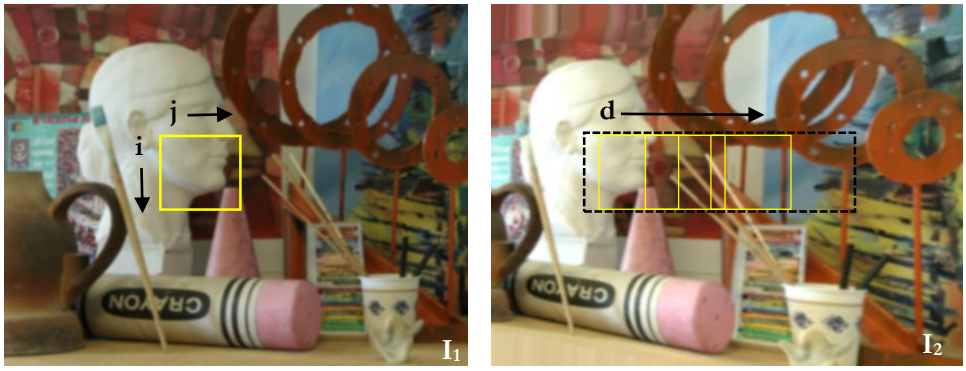


Figure 2.1: ABM example. On the left: the template extracted on the reference image; On the right: the patch on the slave image and the search area in black. For each template defined in the reference image, the best corresponding region in a second image is searched and identified.

The centre of the patch, that gives the highest similarity with respect to the template (the maximum correspondence between the grey values of the two windows), represents the homologous point (the two correspondent features are therefore considered the projection of the same physical entity in each view). However, grey value correspondences always differ since the patch is affected by both radiometric and geometric differences. Radiometric differences are due to sensor response, illumination changes, and object reflective changes. Geometric differences arise from camera movements (translations and rotations), and subsequent perspective effects. The continuous improvement of stereo matching algorithms, in terms of robustness against illumination differences, occlusions, repetitive pattern, and prospective variations, as well as efficiency and computation load, represents the most important requirement,

and some of the strategy developed to overcome these limits will be described in the following sections.

Focusing on the essential component of the images correspondence problem, in this first part of the thesis, stereo methods that operate on two frames (stereo pair) will be considered and a general way of characterizing a stereo algorithm workflow will be presented.

It is increasingly evident that matching problems are one of the most active research areas both in photogrammetry and computer vision and, therefore, very different algorithms and approaches have been proposed and investigated in the last decades by both communities. The wide range of these methods makes difficult to identify and illustrate a clear, unique and general algorithm workflow, along with their algorithmic components and processing steps. In this regard, a general taxonomy and categorization of stereo matching algorithms, where a set of well-defined processing steps has been presented in [157]. According to Scharstein, stereo algorithms generally perform the following four steps:

1. Compute a pixel-wise matching cost (a similarity measurement between a pixel
2. in the master image and a pixel in the slave image);
3. Aggregate cost spatially (e.g., by summing over a window/support, or by diffusion);
4. Compute disparity (with a local, global or semi-global strategy) and find the best match based on the aggregated support;
5. Compute a sub-pixel disparity refinement (optional) by interpolation or image filtering.

Although this classification proved to be suitable for a lot of image matching algorithms, it not always fit exactly all the approaches. Indeed, more commonly in photogrammetry, the distinction between steps 1 and 2 is not so clear: the costs calculation presumes the use of a window ("template" on the master image and "patch" on the slave image) within which the pixel costs are "aggregated" and this means that steps 1 and 2 collapse in a unique processing level. In other words, the similarity measurement between the images is not computed with a pixel-wise approach, but rather with a "window-wise" approach. However, the subdivision in the two first steps remains more attractive for applications where the computing efficiency is one of the primary tasks: compute a pixel-wise cost and then aggregate over a support is generally more efficient than the window-based approach. To provide a more general processing workflow we propose the following categorization scheme:

1. Image pre-processing procedures (optional);
2. Compute matching cost by using a window/support;

3. Disparity computation and optimization by implementing local, global or semi-global methods.
4. Compute a sub-pixel disparity refinement (optional).

In the following an accurate description of each step will be presented, identifying and structuring the individual components that characterize an image stereo matching algorithm and with the foresight to give a useful guide to offer the knowledge, according to this new interpretation, about existing stereo matching algorithms.
Finally, the multi-image matching process will be introduced and described.

2.1 Image pre-processing procedures

The essence of stereo matching is, given a point in one image, to find the corresponding point in another image, such that the two points are the projections of the same physical point in space. Therefore, at the base of any matching algorithm, a matching cost that measures the similarity of two locations on the images is necessary and the performance of all the existing different strategies rely on the radiometric characteristics of the images. In other words, in order to obtain reliable and precise results, the matched area must have high radiometric gradients (to facilitate a unique localization of the searched area within the image). In this regard, image pre-processing procedures can be used: these techniques allow optimizing the grey values distribution, obtaining better brightness and contrast values than the original ones.
The most famous pre-processing filter, used for providing better image information in shadowed and saturated areas, is represented by the Wallis filter [196]. Differently from global contrast filters (e.g. Linear, Normalized contrast stretch), the algorithm is locally-adaptive, and therefore allows to simultaneously enhances the contrast values at both end of the brightness range (bright and dark areas). The output image is a user-controlled weighted average of the Wallis filter output and the original image: the user must specify target mean and standard deviation values and the filter will adjust the local areas into the image in order to achieve the target values. The results will be a new transformed and optimized image.
There are many other filters created to reduce the effects of the radiometric problems, for example the adaptive smoothing filter proposed by Saint-Marc et. al. [154] which is applied to reduce the noise level and to sharpen edges and preserve fine detail such as corners and line end-points. Moreover, in [208], a new pre-processing method that combines the adaptive smoothing filter and the Wallis filter, was developed. Furthermore, it is important to mention the Laplacian of Gaussian (LoG) filter, which is often used for removing noise and brightness changes, in particular in local real-time

method [101], and the mean filter, which aims to compensate a change in bias by subtracting from each pixel the mean intensity of a certain neighbourhood area of pixels. Several authors have suggested to compute the correlation measures on transformed images instead of on the original intensities [109]. This transformation can be considered as a pre-processing step, which has the goal to increase independency in camera bias-gain.

In fact, patterns in images are corrupted by many non-Gaussian and non-additive phenomena; in order to minimize non-Gaussian noise many computer vision techniques use discrete approximations of image derivatives. In [109] it was shown that, by computing correlation measures on the gradient norm of each pixel, a higher robustness, with regard to illumination and geometric rotation insensitivity, can be obtained. Another example is represented by the histogram equalization methods [42]: these methods are based on the assumption that the image brightness of corresponding points (or windows) must be equal; however, this assumption is often false and the comparison of corresponding pairs of intensity histograms must be modelled. In [42] is demonstrated that a linear model does not adequately represent these deviations while, on the contrary, a non-linear monotonically increasing function is capable to solve the problem, correcting, or at least reducing, the errors introduced by the constant brightness assumption. This function tries to consider and model some of the problems that affect the images radiometric contents (variations in illumination, contrast and camera signal response) and allows mapping the intensity values in an image to intensity values in another one. Finally, instead of consider the original images, the intensity corrected histograms of the two images will be matched. Roy and Cox [42] have proposed an original solution to perform the local comparison of histograms based on a dynamic histogram warping that, instead of locally compare histograms, works directly on the intensity histograms by expanding or compressing intensity bins (one-to-one and one-to-many mapping are allowed). This method, avoiding intensity derivatives computation, seems superior to histogram specification procedure[174]. Similarly, in [64] a model for image brightness transformation between images is also proposed, introducing a spatially varying multiplier to relate images brightness.

Other transforms, that try to create transformed images which can better respond to the needs of matching algorithm are the non-parametric rank and census transforms. These measures allow to generate new image pairs where the radiometric values of the image pixels were replaced with a different information, which gives indication, and takes into account, of variations in brightness between neighbouring pixels.

The Rank transform [204] defines the number of pixels p' in a square window W whose intensity is less than the luminosity value of the central pixel p:

$$R(p) = \|\{p' \epsilon\, W(p) |\, I(p') < I(p)\}\| \tag{2.1}$$

In practice, this rank is determined by counting the number of neighbours with a smaller grey value than the reference pixel. In this case the function output is not an intensity value but an integer and the image correspondence can be realized with any correlation method on the transformed image. In other words, the images can be pre-filtered with the rank transform and then compared. Although this method allows reducing sensitivity to radiometric gain and bias, at the same time it also reduces the discriminatory power of the matching procedure since (usually) some information are lost (e.g. the new "colour" depth is lower than the original).

In rank transform approaches, the relative ordering of all the pixels surrounding a given location is encoded in a single value, therefore, Zabih and Woodfill [204] have also proposed a variation of the rank transform, called census transform, that preserves the spatial distribution of ranks by encoding them in a bit string. In fact, it can be considered as a fairly new area-based approach to the correspondence images problem [204] since it realize a non-parametric summary of local spatial structure followed by a correlation method using, for example, an Hamming distance metric (see following sections). Census measure tries to map the intensity values of the pixels within a square window W to a bit string where pixels intensities are compared to the window central pixel intensity p. The boolean comparison returns 1 when the pixel intensity is less than the central pixel, or 0 otherwise. That is:

$$R(p) = \otimes_{p' \in W}\ \xi\left(I(p'), I(p)\right) \quad \text{where} \quad \begin{matrix} \xi\,(i,j) = 1, i < j, \\ \xi\,(i,j) = 0, i > j \end{matrix} \tag{2.2}$$

where $\otimes$ represent the concatenation operator.

Census transform has been proved to be one of the most robust image transform for stereo vision [85][82]. Stein et al. [171] used the census transform as an efficient descriptor for structure matching in driver assistance systems and Fröba and Ernst [60] used a modified census transform for face recognition.

The advantage of the above schemes is that correlation rely on the relative ordering of local intensity values (and not on the intensity values themselves). Therefore, measures are independent of absolute intensity scale and invariant to monotone transformations of intensity values like gamma variation between images. Correlation using such

transforms can tolerate a significant number of outliers. An example of the two non-parametric cost metrics is shown in Figure 2.2.

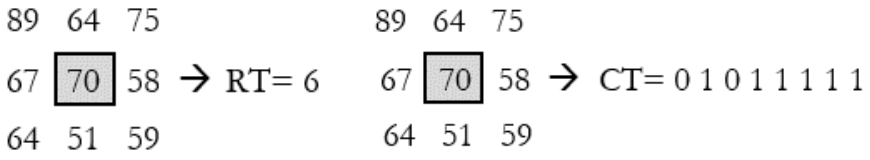


Figure 2.2: Rank (left) and Census (right) transform of a 3 x 3 neighbourhood patch (the reference pixel is marked in grey).

However, one drawback of these approaches is that the local measures depend quite heavily on the centre pixel intensity. In this context, other methods have been developed [158] to match local intensity gradients instead of raw intensity values but it is worth noting that their performance can be poor when gradient information is not reliable.

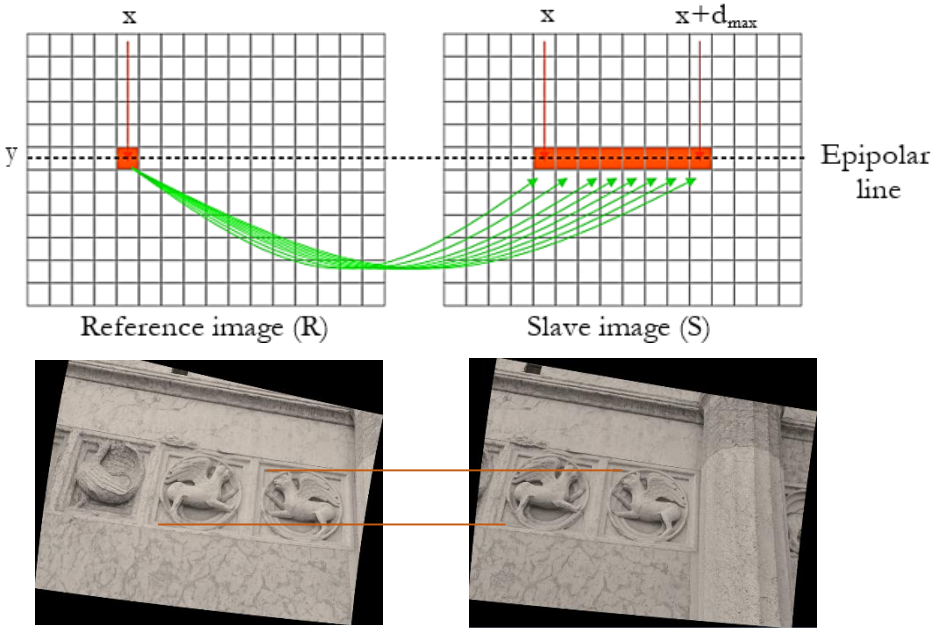


Figure 2.3: Epipolar constraint introduction during the disparity optimization process and an example of an epipolar rectified image pair: epipolar lines are highlighted in the above pictures.

The epipolar constraint enforcement can be also considered pre-processing step: indeed, this means that images are firstly rectified (using known interior and exterior orientation parameters) in order to obtain a new image pair where correspondent image

elements appear on the same line (called as epipolar line) of pixels on the images (see Figure 2.3). The use of rectified images allows to considerably reducing the stereo matching computation effort and time and to improving matching reliability (more details about epipolar image rectification are provided in section 3.1.2).

2.2 Matching cost computation

All image matching algorithms rely on a quantification of the similarity of image pixels: such metric can be calculated by computing a pixel-wise matching cost, followed by costs aggregation over a support region, or with windows-based approach where costs are directly calculated on the aggregation region (e.g. as done by Normalized Cross Correlation (NCC) similarity function). In other words, once the pixel-wise costs are calculated within a finite area, they must be "aggregated": these measures must provide a unique value which will indicate the similarity information of each evaluated window of pixels. According on the chosen similarity function, the costs aggregation step can be performed in several ways: costs can be simply summed (as in the case of Sum of Absolute/Squared Differences (SAD/SSD) cost functions) or can be related to a statistic indicator (as it occurs in the Normalized Cross Correlation indicator and in the LSM process). Common pixel-based matching costs include absolute differences, squared differences and the interesting (and more robust) sampling-insensitive absolute differences criteria developed by Birchfield and Tomasi [19].
In [9], a similarity measure taxonomy is presented, analysing the different measures quality and properties. The most traditional and frequently applied matching cost, used both in photogrammetry and computer vision, are SAD/SSD [85], NCC [106][73][77] and Mean Absolute Difference (MAD), while, in the recent past, Mutual Information (MI) [195] and the non-parametric measures based on Census transform [204], were introduced. The following section wants to provide a brief summary of the most popular similarity cost functions, usually implemented in photogrammetry to perform a dense stereo reconstruction.

2.2.1 Sum of Absolute/Squared Differences

Sum of Absolute Differences is one of the simplest similarity measures commonly implemented for image similarity analysis. It performs the absolute difference between each pixel of the master image and the corresponding pixel in the slave image, using a search window W to realize the comparison. Similarly, in Sum of Squared Differences the differences between corresponding pixels are squared.

In a subsequent step, these differences are summed (cost aggregation).
SAD and SSD formulations have the following expression:

$$SAD = \sum_{(i,j)\epsilon W} |f(i,j) - g(i+\Delta y, j+\Delta x)| \tag{2.3}$$

$$\mathrm{SSD} = \sum_{(\mathrm{i,j})\epsilon \mathrm{W}} (\mathrm{f(i,j)} - \mathrm{g(i+\Delta y, j+\Delta x)})^2 \tag{2.4}$$

where $f(i,j)$ is the gray scale pixel intensity at location (i,j) in the master image and $g(i+\Delta y, j+\Delta x)$ the gray scale pixel intensity at location $(i+\Delta y, j+\Delta x)$ on the slave image, considering Δx and Δy the corresponding pixel disparity. The quantity SSD (or SAD) measures the squared Euclidean distance between (f, g), and a value close to zero indicates a strong match.

2.2.2 Normalized Cross Correlation

Normalized Cross Correlation is more complex than both SAD and SSD but it is invariant to linear changes in image intensity amplitude. Normalizing features vectors to unit length, the similarity measures between the features becomes independent to (linear) radiometric changes [106].
The NCC finds matches of a reference template $f(j,i)$ over an area *W* with respect to an equally sized area on the slave image $g(j,i)$ and it is defined as:

$$NCC = \frac{\sum_{(i,j)\in W}[(f(i,j)) \cdot (g(i+\Delta y, j+\Delta x)]}{\sqrt{\sum_{(i,j)\in W}[\left(f(i,j)\right)^2 \cdot \left(g(i+\Delta y, j+\Delta x)\right)^2]}} \tag{2.5}$$

As it can be seen, the cross-correlation between f and g is normalized by the L2 norms of the two vectors, in order to render the measure robust to any spatially constant multiplicative bias. By subtracting the mean intensity value of the two image windows, we can get an even more robust matching measure:

$$ZNCC = \frac{\sum_{(i,j)\in W}[[(f(i,j) - \overline{f}) \cdot (g(i+\Delta y, j+\Delta x) - \overline{g})]}{\sqrt{\sum_{(i,j)\in W}[\left(f(i,j) - \overline{f}\right)^2 \cdot (g(i+\Delta y, j+\Delta x) - \overline{g})^2]}} \tag{2.6}$$

where $\bar{f}$ and $\bar{g}$ represent the corresponding sample means. This measure is referred to as Zero-mean NCC (ZNCC) and it is robust to spatially constant affine variations of the image intensities.
A unitary value of the coefficient indicates a perfect match, although, ideally, a unit value can only be obtained by a noise-free image pair of a flat surface with images acquired in the so-called normal case of photogrammetry.
NCC and ZNCC appear more accurate and reliable then the other costs function (which don't relate the radiometric information within the window, to a statistic indicator) and show good robustness with respect to linear brightness and contrast variations. However, all these metrics are not the ideal approach to feature tracking since it is not invariant with respect to imaging scale, rotation, and perspective distortions. SSD can be influenced moreover from radiometric changes, but it is computationally very attractive and shows good insensibility toward noise [109]. Further, by definition, they are not suitable in the presence of nonlinear intensity variation at corresponding pixels.

2.2.3 Sum of Humming Distance

The Humming distance is a metric expressing the distance between two objects by evaluating the number of mismatches among their pairs of variables. It is mainly used for string and bitwise analyses, but can also be useful for numerical variables.
Sum of Hamming Distances (SHD) is normally employed for matching census-transformed images by computing bitwise-XOR between the values extracted within a square window W on the master and slave images. The two bit strings (which must have the same length) are evaluated identifying the number of pixels that change from one string to the other; in other words, a bit-counting operation, which results in the final Hamming distance score, is performed:

$$SHD = \sum_{(i,j)\in W} f(i,j)\ bitwiseXOR\ g(i+\Delta y, j+\Delta x) \\ = f(i,j)\hat{}g(i+\Delta y, j+\Delta x) \tag{2.7}$$

where SHD is the humming distance between the strings f and g and W is the window used to extract the bit-values to compare. In Census metric, a window size of three or five pixels is generally used to Census-transform the images.

2.2.4 Other cost functions

As already mentioned, stereo matching systems often fails when the radiometric conditions between images change. For example, Birchfield and Tomasi have proposed a matching cost that is insensitive to image sampling [19] where, instead of just comparing pixel values shifted by integral amounts (which may miss a valid match), the comparison of each pixel in the reference image against a linearly interpolated function of the other image is computed. More recently, with respect to the previously analysed cost metrics, the Mutual Information stereo similarity metric [52][83], which has been introduced in CV by Viola and Wells [195], has been proposed for matching images with complex relationship of corresponding intensities and reflection problems. Mutual information depends upon the entropy and the joint entropy of two random variables that, in the case of stereo matching, are represented by the information content of the image pixels taken from each image of a stereo pair. MI can be easily incorporate into an area-based algorithm with fixed size windows centered at each point (as it has been doing by Egnal [52]). However, this approach suffers from the classical problems of windows-based local methods as regard discontinuities, occlusions and low-textured regions on the images (problems that can be solved using energy minimization approaches).

Considering two images I_1 and I_2, we can express Mutual Information MI at a given disparity d, as:

$$MI(I_1, I_2, d) = H(I_1) + H(I_2, d) - H(I_1, I_2, d) \tag{2.8}$$

Then entropies $H(I_1)$ and $H(I_2)$ are calculated from the probability distributions P of the associated images intensities:

$$\begin{aligned} MI(I_1, I_2, d) = &- \int_0^1 P_{I_1}(i) log P_{I_1}(i) di - \int_0^1 P_{I_2,f}(i) log P_{I_2,f}(i) di \\ &+ \int_0^1 \int_0^1 P_{I_1,I_2,f}(i_1, i_2) log P_{I_1,I_2,f}(i_1, i_2) di_1 di_2 \end{aligned} \tag{2.9}$$

Since, usually, the matching algorithms tries to minimize the similarity cost, while a high mutual information rate corresponds to high information (since that well-registered images produce an high joint entropy because one image can be predicted by the other), the formulation is commonly changed in sign.

A new and recent approach to compute the stereo cost computation, is that proposed by Zbontar and LeCun [205]. They have solved the problem of computing the stereo matching costs by learning a similarity measure on pairs of small image areas/patches using a convolutional neural network. The output of the convolutional neural network is used to initialize the stereo matching cost; then, they proceed to a series of post-processing steps, necessary to achieve good results.
The key of the method relies in the identification and construction of the optimal network architecture for solving a specific problem: in [205] the described neural network has produced disparity maps with lower error rates than any other previously published method, on four different high-resolution scientific image dataset extracted from the Middlebury Stereo Evaluation Dataset. In particular, the 2014 high-resolution datasets has been considerate; such results are presented (and can be download) on the Middlebury Stereo Matching performance Evaluation web page [115].

2.2.5 Cost area supports

As previously illustrated, window-based methods perform the aggregation of the matching cost by summing or averaging the pixels cost over the support region (as happens in SAD, SSD, MAD), possibly using suitable well-identified weights or statistic indicators referring to the radiometric data within window (as performed in NCC similarity function).
However, implicitly all the previous metrics assume equivalent depth values for all pixels of the correlation window even if this hypothesis is violated at depth discontinuities, strong perspective changes between the images and whenever the object surface is not parallel to the image plane (slanted/non-planar surfaces). This assumption produces a smoothness effect of the pixels data within the window, especially if the matching window covers pixels which lie at different depths: this leads to wrong values in the matching cost. On one hand, small windows can lead to noisy, low precision results because consider just few information to evaluate the template similarity and might not cover enough intensity variations [94] but are less prone to such issues (and consequently can improve the restitution level of detail). On the other hand, larger windows give better support for the matching but makes constant depth hypothesis more inadequate. In fact, big window sizes usually produce smoother surfaces, losing information near object boundaries and where small object shape details are present. The larger the window, the larger will be the loss of object information in presence of depth variations, in particular for wide baselines and large angular images motions (image neighbourhood of corresponding pixels will look different in shape and size in

the stereo images pair). The ideal block size to perform the stereo matching depends on the chosen function and the evaluated object: in analogy with other techniques for DTM generation in close range (e.g. LSM), there seems to be an optimal range for template size value according to object features [146]. Nevertheless, several approaches face this issue in different ways, trying to solve the just mentioned correspondences difficulties connected to the use of a fixed window. Some rely on the concepts of a variable support employment [3][29][62][50] and some others try to set up an iterative stereo correspondence method that employ a variable shape window which changes adaptively according to the information content of the analysed image (e.g. LSM [68]).

The basic concept at the base of the variable support methods is the following: instead of using a fixed size correlation window for each image point of the reference image, a more generic support, whose shape and dimensions vary by adapting themselves according to image characteristics, is used. These methods want to improve the accuracy and reliability of depth information. Indeed considering a fixed support region, the results reliability, when the patch is located near border and discontinuities, are severely compromised.

Variable support methods rely, generally, on a fixed set of rectangular or squared window pairs which can dynamically change. On the basis of a specific condition on the image (such as a border, a discontinuity, a repetitive pattern, etc.), each window can modifies its dimension and/or can be centred on different neighbouring points in order to finally identify the support (extracted from the initial set of windows) which better identifies the searched corresponding point. After that, the disparity value determination of the evaluated point is performed.

The most famous variable support methods use multiple windows anchored at different points, i.e. shiftable windows [24], windows with adaptive sizes [94][126][193][95], and windows based on connected components of constant disparity [29]. Three-dimensional support functions have also been proposed and include limited disparity difference [67], limited disparity gradient [140], and Prazdny's coherence principle [144].

For example, multiple windows method [3][24] considers multiple square windows centred at different locations and selects the one that gives the higher similarity cost ([24][62]), allowing the determination of the best window that maximizes the similarity measure. An extension has also been proposed in [82], which picks several windows from surrounding square windows and computes the sum of their average costs. Multiple windows algorithm can employ, at every pixel location, 5 or 9 squared windows (clear examples can be seen in [82]), in symmetrical positions with respect to the central point (see Figure 2.4). Some windows are designed so that they will match to the left,

some are designed to match to the right, some are designed to match towards the top, and so on. Near an occlusion boundary, at each pixel, only the best result from matching using all 9 windows is stored while the bad matches resulting from occlusion tend to be discarded.

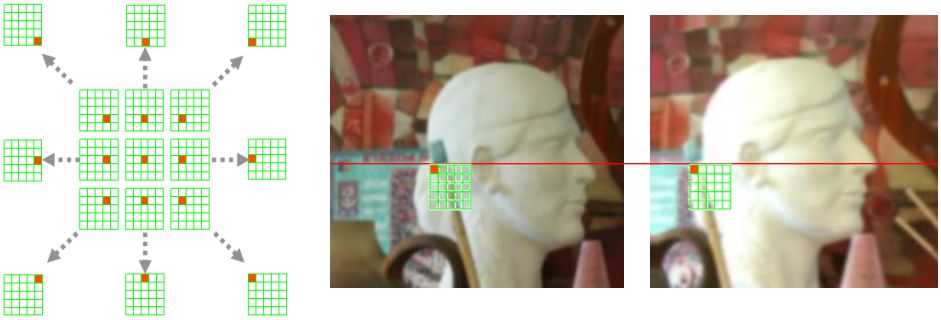


Figure: 2.4: Multiple windows algorithm example: using 9 matching windows, border localization and noise problems in the disparity space image are avoided. The windows center (marked in red) are shifted, not constraining the support to be centered on the central position. Finally, just the position with the best score is selected.

However, for both square-fixed-window and multiple-window approaches, the size of the support window is fixed and it should be large enough to include sufficient intensity variation data and, at the same time, small enough to avoid depth variation inside the window.

An alternative is to change the window dimension, by maintaining it centred on the same point [126]. This is useful, in particular, in order to decrease the ambiguity of the correspondences determination within uniform textured areas. Several approaches has been developed in order to calculate the matching costs with adaptive windows: for example, window size may vary within a single image: larger ones for areas with weak textures and smaller ones for areas with geometry details. At the same time, the shape of the support can also automatically change. To these ends, several adaptive-window approaches have been proposed, which select the optimal window sizes and shapes automatically by evaluating the image local information. For example Kanade and Okutomi [94] have proposed an adaptive window solution. In this case (and other variable support methods based on not-fixed windows shape such as [35][47]) the variable support can have different shapes. They modified the window size and shape adaptively depending on the local intensity and disparity variations for each pixel. The selected window is optimal in the sense that it produces the disparity estimate having the least uncertainty. By evaluating both the intensity and the disparity variations within

a window, we can compute both the disparity estimate and its uncertainty which can then be used for selecting the optimal window. Although this algorithm produced better results than the standard single-window algorithms (in particular regarding accuracy at object boundaries), its final output depended on the choice of the initial disparity estimate.

A more recent method has been proposed in [203]: it is an adaptive-weight algorithm that adjusts the support-weight of each pixel in a fixed-sized square window. The support-weight of each pixel in the window is calculated based on both the colour similarity and the Euclidean distance with respect to the central pixel. However, the computation cost with this method is high, making the method usually computationally unfeasible.

Lastly, the Least Squares Matching method [68] can be also considered a variable support method. In fact, it implements an adaptive support whose shape can change on the slave image iteratively according to the radiometric information within the support itself. In other words, starting from an initial solution, characterized generally by a plane and squared or rectangular support, the methods modifies the patch shape (on the evaluated slave image) on the basis of a solution that is calculated iteratively. The aim is minimizing the implemented similarity cost function that represents the better solution to solve the system (the method details are described in paragraph 2.3.1.2).

The similarity function gives a cost for each pixel at each disparity and these data can be stored in the Disparity Space Image (DSI) $C(x, y, d)$, where (x, y) represents the pixel location and d the considered disparity value. In general, a DSI is any image or function defined over a continuous or discretized version of disparity space (x, y, d). In practice, for cost computation, each DSI element $C(x, y, d)$ represents the radiometric similarity cost of the correspondence between the reference (or master) image pixel/window $I_R(x,y)$ and the slave image pixel/window $I_S(x +d, y)$. It is easy to note that a 1D disparity search domain makes the complexity of the cost computation proportional to $W \times H \times D$, where W and H are the image pixel resolution and D is the disparity search range (as it is represented in Figure 2.5). In other word, all costs for all pixels and all possible disparity values are evaluated up front in a big tensor of these dimensions

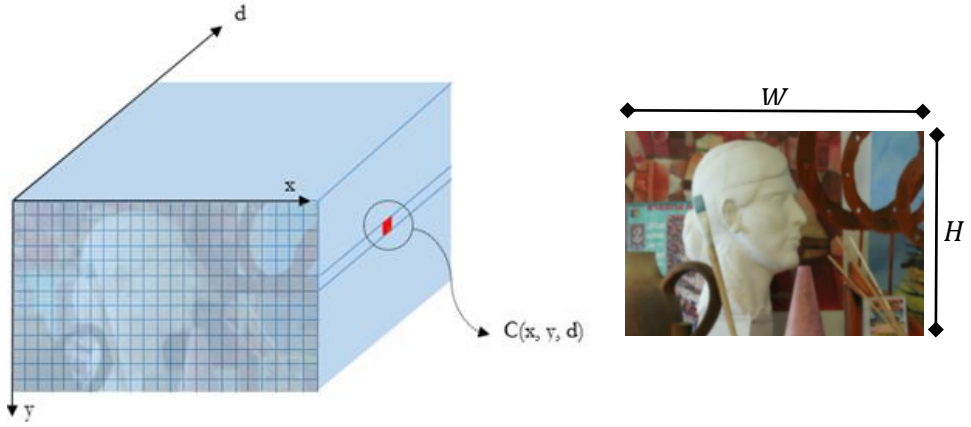


Figure 2.5: DSI three-dimensional representation.

2.3 Disparity computation

Once the costs are calculated, it is necessary to determine which discrete set of disparities represents the scene surface optimally. Three different stereo matching approaches will be described in the following sections: local, global and semi-global. Every approach performs the disparities computation and optimization steps (the latter will be separately discussed in paragraph 2.4) identifying the optimal disparity in an image matching problem. However, Local, Global and Semi-Global methods implement very different algorithms and optimization functions (or strategies): the main distinction between these algorithms is the minimization procedure used to compute the final disparities. On one hand we have the local methods, where computing the final disparities simply means choosing for each point the disparity associated with the minimum radiometric similarity cost value: each correspondences is evaluated one point at a time, not considering neighbouring points/measures; on the other hand, global and semi-global techniques make explicit assumptions on the regularity of the results. The use of regularity constraints provide an additional support for the solving process and allows obtaining results even with areas which may be difficult to compute with local methods (e.g. with no texture variations).

2.3.1 Local methods

Local methods evaluate the correspondences of one point at a time (i.e. point-wise approach) by analysing the radiometric intensities within the finite support window, not considering neighbouring points/measures. This approach is very efficient, making it

particularly suitable for real-time applications; however, the disparity computation at a given point will depend only on intensity values within the finite window.
Following, two typical local strategies are presented.

2.3.1.1 Correlation based method

Correlation methods realize a similarity measurement between a template in reference image and a patch extracted from a larger search area within the slave image. The matching process moves the patch window to all possible locations in the slave image and computes a numerical index that indicates how well the template matches the image in that location. The position of best agreement is assumed to be the location of the corresponding point.

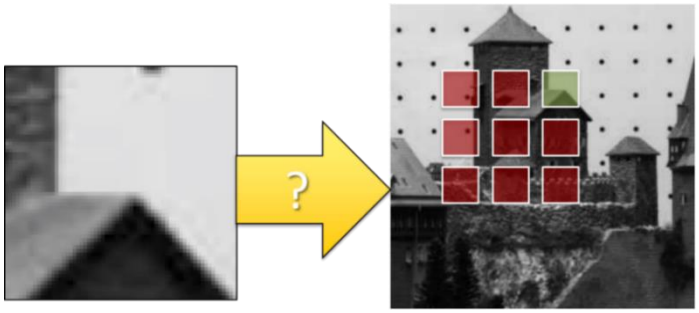


Figure 2.6: The extracted template in the reference image (on the left) is moved to different locations in the slave image (on the right), in order to identify the corresponding area.

The similarity between the two areas can be evaluated using simple SAD/SSD matching cost or the more rigorous NCC coefficient. Finally, the easiest way of choosing the best disparity is to select, at each pixel, the minimum aggregated cost across all disparities under consideration, in other words, performing a local "Winner-Takes-All" (WTA) optimization at each pixel (see Figure 2.6 and 2.7).
It is interesting to notice that local matching strategies suffer of important problems connected to the method assumptions: for instance, considering individually each point neighbourhood, low textured and repeated patterns (Figure 2.8) represent a problem, leading to error and mismatches (i.e. the algorithm can find multiple correspondences that seem to be correct). These problems can be overcome by introducing some constraints to the matching process (as it performed by global matching methods, see paragraphs 2.3.2 and 2.3.3). On the other hand, problems often arise at depth discontinuities (an identical correspondent window on the slave image cannot be found - see for instance Figure 2.9). The algorithms are considerably sensitive to the presence of image regions characterized by sudden depth variations

as far as the content of the two templates differs: such problems can be overcome using variable supports as illustrated in paragraph 2.2.5.

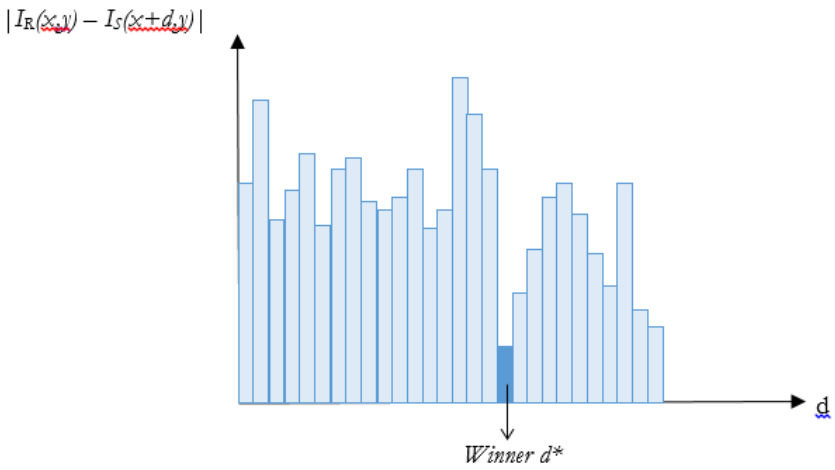


Figure 2.7: Identification of the best disparity value through the Winner-Take-All (WTA) strategy.

Finally, correlation methods, which usually assume a constant disparity value inside the matching template, are influenced by geometric transformations such as rotations, perspective changes or scale variations. When the template and the patch are affected by some geometric variation, the similarity values, measured by simple correlation metrics, are low, even when the homologous points are well-identified. In this context, several authors suggested to use, in the matching process, not the original image but that obtained by means of appropriate convolutions [109]: for example, images derived from the gradient norm of each point proves to be invariant under rotations.

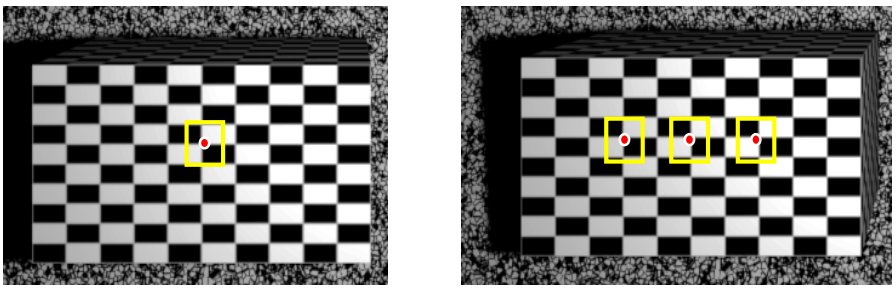


Figure 2.8: Images with repeated texture patterns (which represent a problem for image matching algorithms).

Figure 2.9: The correlation window contains element at different depth in the image scene and some elements within it are occluded.

2.3.1.2 Traditional Least Squares Matching

As mentioned above, correlation methods are not robust if geometric variations (different from rigid translation) occur between the images; in order to overcome this problem, different algorithms were developed in the last 30 years, allowing to consider more general and complex geometric difference. The famous Least Squares Matching technique [68] is one of those: in this case the matching process consists in the estimation of a transformation parameter vector that makes the similarity between the template and the patch the highest considering a parametric deformation of the patch. LSM is an optimization method that employs an iterative geometric and radiometric transformation between reference image and search image in order to minimize the least-squares sum of pixel value differences (SSD) between the images. To keep the overall model simple but still effective, in most implementations the radiometric parameters allow a linear intensity stretch while the geometric parameters allow an affine transformation between the two images (interesting is the recent investigation of different and more complex geometric transformations between the images, described in [146] and [92]). The 6-parameter affine geometric model is an approximation of the actual perspective transformation between two images; when the object surface in the search area can be assumed to be planar, and for small patches and not too convergent images it works generally well. The six parameters are estimated by least-square adjustments using the template pixel values as observations.

Given two image points, LSM considers the two conjugate image regions as discrete two-dimensional functions: the template $f(x_1, y_1)$ and the patch $g(x_2, y_2)$. The patch is transformed applying both radiometric and geometric adjustments to obtain a new patch $g'(x_2, y_2)$. The matching process establishes a correspondence minimizing the L2-norm of the residual vector $e(x_1, y_1)$:

$$f(x_1, y_1) - g'(x_2, y_2) = e(x_1, y_1) \tag{2.10}$$

Radiometric changes (due to contrast and brightness variations of intensity values in the slave image) are modelled in the patch function as:

$$g'(x_2, y_2) = r_0 + (1 + r_1) \cdot g(x_2, y_2) \tag{2.11}$$

where r_0 and r_1 are two parameters accounting for brightness and contrast changes in the slave image, respectively.
Geometric corrections are considered by means of a geometrical parametric transformation:

$$g(x_2, y_2) = g(x_2(x_1, y_1), y_2(x_1, y_1) \tag{2.12}$$

The following affine transformation model is the most commonly used in image matching applications:

$$\begin{cases} x_2 = a_1 x_1 + a_2 y_1 + a_3 \\ y_2 = b_1 x_1 + b_2 y_1 + b_3 \end{cases} \tag{2.13}$$

where (a_1, a_2 , b_1, b_2) model shape differences between patch and template, while (a_3, b_3) are the shift (translation) parameters.
Radiometric and geometric correction parameters are then estimated solving, for $||e(x_1, y_1)||$ =min, the following least squares system, obtained by substituting the transformed functions in eq. (2.11):

$$\begin{aligned} f(x_1, y_1) + e(x_1, y_1) &= r_0 + r_1 \cdot g(x_2(x_1, y_1), y_2(x_1, y_1)\,) \\ &= \bar{g}(r_0, r_1, a_1, a_2, a_3, b_1, b_2, b_3) \end{aligned} \tag{2.14}$$

Linearization of the function $\bar{g}(r_0, r_1, a_1, a_2, a_3, b_1, b_2, b_3)$ with respect to the parameters (a_1, ..., b_3) requires the computation of the derivatives of the grey values with respect to (x_2, y_2) using numerical approximations.

$$\begin{aligned} f(x, y) = \; & g^0(x_1, y_1) + g_x da_1 + g_x x_1 da_2 + g_x y_1 da_3 + g_y db_1 + g_y db_2 \\ & + g_y y_1 db_3 + r_0 + r_1 g^0(x_1, y_1) \end{aligned} \tag{2.15}$$

The partial differentials are given by the pixel value gradients:

$$g_x = \frac{\partial g^0(x_1, y_1)}{\partial x_1} \qquad g_y = \frac{\partial g^0(x_1, y_1)}{\partial y_1} \tag{2.16}$$

If the transformation parameters are written as an unknown vector $\hat{x}$, the partial derivatives coefficients are stored in the design matrix A, and the pixel value differences between reference and search images are inserted in the vector of observations l, then the linearized correction equation is given by:

$$l + v = A\,\hat{x} \tag{2.17}$$

where

$$\hat{x}^T = [da_0, da_1, da_2, db_0, db_1, db_2, r_0, r_1] \tag{2.18}$$

$$A = \begin{bmatrix} {}^1g_x & {}^1g_x \cdot {}^1x & {}^1g_x \cdot {}^1y & {}^1g_y & {}^1g_y \cdot {}^1x & {}^1g_y \cdot {}^1y & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^ng_x & {}^ng_x \cdot {}^nx & {}^ng_x \cdot {}^ny & {}^ng_y & {}^ng_y \cdot {}^nx & {}^ng_y \cdot {}^ny & 1 \end{bmatrix} \tag{2.19}$$

$$\hat{x} = (A^T P\, A)^{-1}(A^T P\, l) \tag{2.20}$$

For usual cases, the weight coefficient matrix P is set as identity.
The adjustment equations must be solved iteratively (as is shown in Figure 2.10), and at every iteration the unknowns are corrected. This leads to new pixel value differences at each iteration between search and reference image, until the least-squares sum of the corrections is less than a predefined threshold or a specific number of iterations is reached.

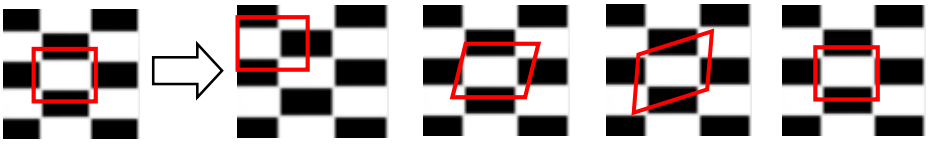


Figure 2.10: An example of the patch transformation during the iterations.

2.3.2 Global methods

Global methods, try to overcome local methods limits: neighbours points correlation measures affect each other, leading to a global simultaneously investigation of the whole image. They therefore result less sensitive to ambiguous disparity values as occurs in presence of occlusions, repeated patterns and uniform texture regions. While local techniques perform almost all of their work during the radiometric similarity cost computation and the cost aggregation steps, in global methods, the emphasis is on the disparity computation phase, skipping frequently the aggregation step (i.e. the cost is evaluated pixel-wise). Such algorithms seek a disparity solution (step 3) that minimizes a global cost function that combines data (step 1) and smoothness terms (the main distinction inside these class of algorithms is generally the implemented minimization technique).

The problem of finding the correspondences deals with the minimization of a global cost function extended (usually) to all image pixels: since individual pixels do not contain enough information for unique matching solution, global methods add smoothness constraints that represent a further support to solving the process. This is typically formulated in a cost function:

$$E(d) = E_{cost}(\boldsymbol{p}, d(\boldsymbol{p})) + \alpha E_{smooth}(\boldsymbol{p}, d(\boldsymbol{p})) \tag{2.21}$$

where $\boldsymbol{p}$ is the location vector of each point in the master image, $\boldsymbol{d}$ is the disparity value and $\boldsymbol{\alpha}$ represents the disparities penalization variable (or weight) introduced for computing the disparity regularization enforcement. The first term, therefore, represents the image consistency/similarity cost, whereas the second term evaluates the regularity of the disparity field. Several approaches rely on simple pixel-based cost functions, but support aggregation strategies have been successfully adopted as well.

The data term $E_{cost}(d)$ sums all pixel-wise/aggregated matching costs, associated with a disparity value of$\boldsymbol{d}$, over the whole image:

$$E_{cost}(\boldsymbol{p}, d) = \sum_{p} C(\boldsymbol{p}, d_p) = \sum_{x,y} (C((x, y), d_p) \tag{2.22}$$

The smoothness term $E_{smooth}(d)$ encodes the smoothness/regularization assumptions made by the algorithm:

$$E_{smooth}(\boldsymbol{p}, d) = \sum_{p,q \in N_p} \alpha P(|d_p - d_q| \geq \gamma) \tag{2.23}$$

This term adds a penalty P for all pixels that have neighbours at a different disparity: γ represents the disparity difference of neighbour pixels beyond which the regularization term intervenes, penalizing the cost value depending on the extent of the estimated difference. In this way, discontinuities are permitted if pixel-wise matching data is stronger than the penalty, i.e. if the texture indicates a discontinuity. The second term indirectly connects all pixels with each other in the image and makes the function global. In this formulation, the disparity image $\boldsymbol{d}$ that minimizes the equation must be identified performing a 2D-optimization. Unfortunately, this is an NP-hard problem [28] [192], which means that…. In other words, the major difficulty with energy minimization with these approaches lies in the enormous computational costs. Moreover, typically, the energy function presents many local minima.

Finding the minimum of the global energy function can be performed with different approaches: famous approximate solutions to this problem are Dynamic Programming algorithm [20], Graph Cuts [28] and Belief Propagation [56]. A detailed comparison of significant energy minimization methods can be found in [178].

2.3.3 Semi-Global methods

Semi-Global Matching [82][83][84] successfully combines concepts of global and local stereo methods allowing to obtain accurate results at low runtime. The algorithm works with a pair of images with known interior and exterior orientation parameters – and consequently epipolar geometry (i.e. corresponding points lie on the same horizontal image line) - and performs a pixel-wise matching: it considers both the image similarity and the displacement continuity, by the concurrent application of regularization constraints (in terms of adjacent pixels displacement).

Semi-Global techniques realize the minimization of a global cost function, combining matching costs along independent one-dimensional paths from all directions through the image. The costs extracted by each path, referred to a particular displacement value, are summed for each pixel and possible displacement (also referred as "disparity") value. Finally, the algorithm choses the disparity solution with the minimum cost, using algorithms like Dynamic Programming, Graph Cuts or Belief Propagation (which will be described in the following paragraphs). As in Global methods, it is possible to say that they combine step 1 and step 3, identifying a disparity assignment that minimize the global cost function (which in turn combines data in step 1).

The first developments, which exploit a single global matching cost for each individual image line ("scanline"), were prone to streaking effects (see Figure 2.11), being the optimal solution of each line not connected with the neighbouring ones. SGM algorithm allows to overcome these problems thanks to the innovative idea of symmetrically compute the pixel matching cost through several paths (N scanlines) in the image (as is shown in Figure 2.12).

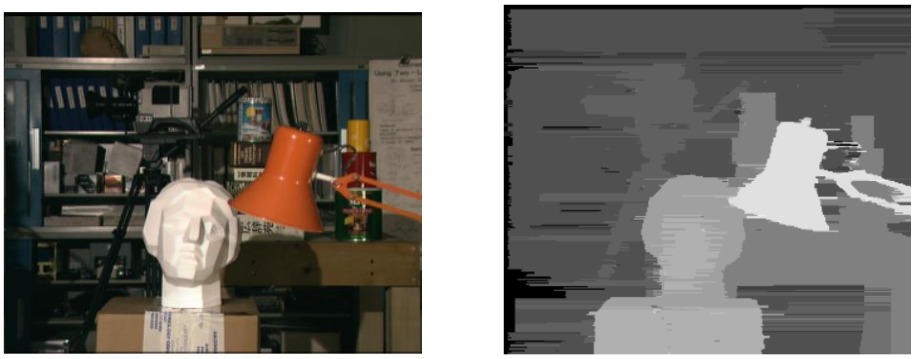


Figure 2.11: An example of the misalignment of the disparity maps scanlines obtained with scanline optimization process, which does not allow inter-scanline consistency in the 2D image.

With a known disparity value, the costs extracted by each path are summed for each pixel and disparity value. Finally, the algorithm choses the pixel matching solution with the minimum cost (most commonly using a dynamic programming approach).
According to the original Hirschmuller paper [82], the cost $L'_r(\boldsymbol{p}, d)$ of the pixel $\boldsymbol{p}$ at disparity $\boldsymbol{d}$, along the path direction $\boldsymbol{r}$ is defined as:

$$\begin{aligned} L'_r(p,d) = C(\boldsymbol{p},d) + \min(&L_r(\boldsymbol{p}-\boldsymbol{r},d),\\ &L_r(\boldsymbol{p}-\boldsymbol{r},d-1) + P_1,\\ &L_r(\boldsymbol{p}-\boldsymbol{r},d+1) + P_2,\\ &L_r(\boldsymbol{p}-\boldsymbol{r},d+1) + P_2,\\ &\min_i L_r(\boldsymbol{p}-\boldsymbol{r},i) + P_2) - min_k L_r(\boldsymbol{p}-\boldsymbol{r},k) \end{aligned} \tag{2.24}$$

where the first term is the similarity cost associated with a disparity value of $\boldsymbol{d}$, whereas the second term evaluates the regularity of the disparity field, adding a penalty term P_1 for little changes and P_2 for all larger disparity changes with respect to the previous point in the considered matching path. The two penalty values allow to describe non-frontal surfaces and to preserve disparity discontinuities, respectively. Since the cost

gradually increases during cost aggregation along the path, the last term allows reducing the final value subtracting the minimum path cost of the previous pixel from the amount.

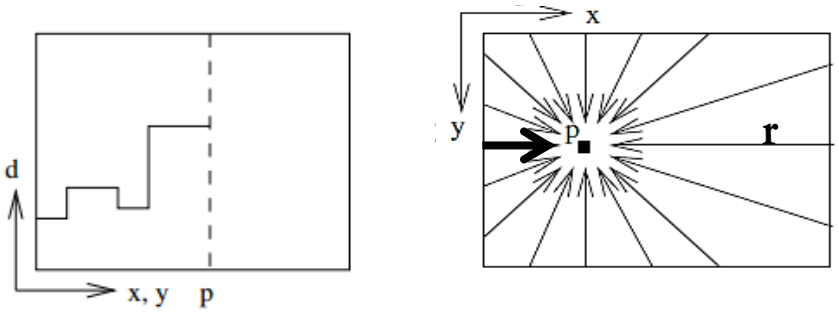


Figure 2.12: On the left: matching solution with the minimum cost; on the right: symmetrical computation of the pixel matching cost through 16 paths from all directions r in the image.

Summing the path costs in all directions $\boldsymbol{r}$ and searching the disparity with the minimal cost for each image pixel $\boldsymbol{p}$, the final disparity map was obtained. The aggregated cost is defined as:

$$S(\boldsymbol{p}, d) = \sum_r L_r(\boldsymbol{p}, d) \tag{2.25}$$

and, for sub-pixel estimation of the final disparity solution, the position of the minimum is generally calculated fitting a quadratic curve through the cost values of the neighbours pixels (see next section). This may shift the actual minimum by a value smaller than 1 and, therefore, increase the disparity resolution.

Similar approaches, where the surface reconstruction is solved through an energy minimization problem, has been evaluated in [139], where a Semi Global matching-like method was implemented identifying the formulation of an energy function $E(Z)$ described as:

$$E(Z) = \sum A\big(x, y, Z(x,y)\big) + \alpha * F(\vec{G}(Z)) \tag{2.26}$$

where

- $E(Z) = \sum A\big(x, y, Z(x,y)\big) + \alpha * F(\vec{G}(Z))$ is the disparity function;
- $A\big(x, y, Z(x,y)\big)$ represents the similarity term;

- $F\left(\vec{G}(Z)\right)$ is the positive function expressing the initial parameters which characterize the surface regularity;
- α represents the weight to permit the data adaptation to the image content (i.e. the weight of disparity regularization enforcement).

2.4 Disparity optimization

The disparity values optimization in a matching problem can be performed using very different algorithms and functions. In particular, according to the disparity computation method used, i.e. Global or Semi-Global, the optimal strategy for computing the final disparity optimization process can be performed with different techniques: some of them are Dynamic Programming, Graph Cuts and Belief Propagation.

2.4.1 Dynamic Programming

A different class of global optimization algorithms are those based on Dynamic Programming (DP) techniques [20][123][41][87]. While the 2D-optimization of eq. (2.21) can be shown to be NP-hard for common classes of smoothness functions [190], dynamic programming method can reduce the computational complexity of optimization problems by decomposing them into smaller and simpler subproblems. Each subproblem is solved only once in order to reduce the number of computations and, once the solution to a given subproblem has been computed, it is stored (in DP the particular data storage is called "memoization"). The next time the same solution is necessary, it will be simply extracted from the stored data. Finally, subproblems solutions are combined to reach an overall solution. This approach is especially useful when the number of repeating subproblems grows exponentially as a function of the size of the input.
The term "Dynamic programming" is not indicative of computer programming, but it rather denotes the "solution planning" of the problem; synthetically, the steps for solving a DP problems are:

1. define the problem structure subdividing it in optimal substructure, which are solved sequentially one stage at a time;
2. write-down the recurrence that relates subproblems; although each one-stage problem is solved as an ordinary optimization problem, its solution helps to define the characteristics of the next ("overlapping subproblems");
3. implement memoization (i.e. remember and re-use solutions to subproblems for solving the problem).

Optimal substructure means that the solution to a given optimization problem can be obtained by the combination of optimal solutions to its subproblems. Consequently, the first step for planning a DP problem resolution, is to check whether the problem exhibits such optimal substructure. Such optimal substructures are usually described by means of recursion. For example, in term of mathematical optimization, given a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, the shortest path p from a vertex u to a vertex v exhibits optimal substructure. Take into consideration any intermediate vertex w on p: p can be split into subpaths p_1 from u to w and p_2 from w to v:these (the subpaths p_1 and p_2), in turn, are the shortest paths between the corresponding vertices. Hence, one can easily formulate the solution for finding shortest paths in a recursive manner, which is what the Bellman algorithm does (see the recursive relationship called the Bellman equation [16].

Overlapping sub-problems means that the space of sub-problems should be small: any recursive algorithm solving the problem should solve the same sub-problems over and over, rather than generating new sub-problems. For example, consider the recursive formulation for generating the Fibonacci series: $F_n = F_{n-1} + F_{n-2}$ with base case $F_1 = F_2 = 1$; the goal is to compute F_n. In the following, the naïve recursive relationship and the optimized DP approaches are explained (Figure 2.13 shows a numerical basic example of the Naïve Fibonacci Algorithm).

- **Naive recursive algorithm**:

$$\begin{aligned} &\text{fib}(n): \\ &\text{if } n \leq 2 : \text{return } f = 1 \\ &\text{else: return } f = \text{fib}(n-1) + \text{fib}(n-2) \\ &\Longrightarrow \quad T(n) = T(n-1) + T(n-2) + \ominus(1) \geq F_n \approx \varphi^n \\ &\qquad \geq sT(n-2) + \ominus(1) \geq 2^{n/2} \end{aligned} \tag{2.27}$$

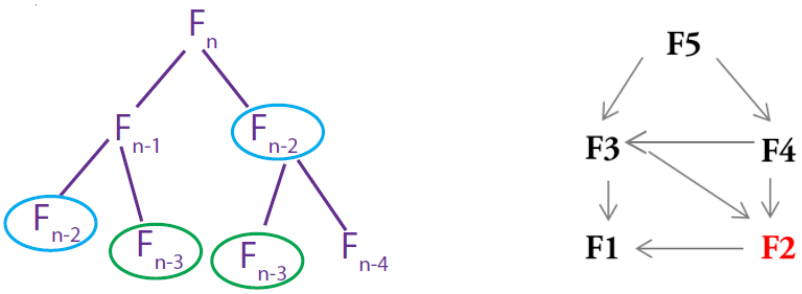


Figure 2.13: Naïve Fibonacci Algorithm, with a numerical basic example.

The computational time $T(n)$, of the $\text{fib}(n)$ function, exponential increases. An optimized approach for computing and optimize the recursive problem, can be achieved in either of the two, following ways.

- **Top-down memoized algorithm**: This is the direct fall-out of the recursive formulation of any problem. If the solution to any problem can be formulated recursively using the solution of its sub-problems, and if such sub-problems are overlapping, then one can easily memoize (i.e. store) the solutions to the sub-problems in a table. Whenever he attempts to solve a new sub-problem, he first checks the table to see if it has been already solved. If a solution has been recorded, we can use it directly, otherwise he solves the sub-problem and adds its solution to the table. Referring to the Fibonacci example:

$$\begin{aligned}&\text{memo} = \{\,\}\\&\text{fib}(n):\\&\qquad \text{if } n \text{ in memo: return memo}[n]\\&\qquad \text{else: if } n \leq 2 : f = 1\\&\qquad \text{else: } f = \text{fib}(n-1) + \text{fib}(n-2)\\&\text{memo}[n] = f\\&\text{return } f\end{aligned} \tag{2.28}$$

In this case, once $fib(n)$ is computed, it can be stored in a "memo-dictionary". Thereafter, every time $fib(n)$ will have to be calculated, it can be just extracted In this case, once $fib(n)$ is computed, it can be stored in a "memo-dictionary". Thereafter, every time $fib(n)$ will have to be calculated, it can be just extracted from the memory (if such solution has already been previously calculated) instead of recalculating it at each iteration. In this way, $fib(k)$ only recurs the first time it's called, and only n non-memoized must be compute $(k = n, n-1, \dots, 1)$:memorized calls requests $\ominus(1)$ computational time. Therefore, this non-recursive method demands $\ominus(1)$ time per call (ignoring recursions) and the algorithm works with a polynomial time. In other words, the running time can be expressed as:

Running time = n° of subproblems to solve • time/subproblem.

- **Bottom-up DP algorithm**: Once we formulate the solution to a problem recursively as in terms of its subproblems, we can try reformulating the problem in a bottom-up fashion: try solving the subproblems first and use their solutions

to build-on and arrive at solutions to bigger subproblems. This is also usually done in a tabular form by iteratively generating solutions to bigger and bigger subproblems by using the solutions to small subproblems. In this case, algorithm example can be schematize as:

$$
\begin{aligned}
&\text{fib} = \{\,\} \\
&\text{for } k \text{ in } [1,2,\dots,n]: \\
&\quad \text{if } k \leq 2 : f = 1 \\
&\quad \text{else: } f = \text{fib}[k-1] + \text{fib}[k-2] \\
&\quad \text{fib}[k] = f \\
&\text{return fib}[n]
\end{aligned}
\qquad
\begin{cases}
\ominus(1) \\
\ominus(n)
\end{cases}
\tag{2.29}
$$

with exactly the same running time as the previous top-down DP algorithm.

Leading back to the stereo correspondence problem, we can treat such task as a minimization problem, finding an optimal set of disparities that minimizes the global energy function (in eq. (2.21)). As previously mentioned, founding the disparity map that minimizes the global energy function is a NP-complete problem but minimization along individual image rows (also defined as "scanlines") can be efficiently performed in polynomial time using DP [20][190]. These approaches work by computing the minimum-cost path through the matrix of all pairwise matching costs between two corresponding scanlines that usually correspond to epipolar lines of rectified images. In this case, the correspondence can be expressed as a disparity value: if $m(x,y)$ and $s(x,y)$ are corresponding pixels in the master and slave image respectively, then the disparity d of $m(x,y)$ and $s(x,y)$ is defined as the difference of their horizontal image coordinates, i.e., $d = x_m - x_s$. Note that we have $\mathrm{y_m} \equiv \mathrm{y_s}$ since corresponding pixels must be on the same scan line in rectified image pairs.

However, it is important to consider that Scanline optimization of depth map does not allow inter-scanline consistency in a 2D image, leading to the misalignment of the disparity maps scanlines (Figure 2.11 for example). Several approaches have been proposed to address this issue (e.g., [194][20]), and, more recently, new algorithms have been proposed to perform DP on a tree constructed from the matching cost volume [82]. In this regard, as previously disclosed in paragraph 2.3.3, the Semi-Global Matching method allows to efficiently overcome the streaking effect problems: combining scanline disparity computation method with a cost aggregation strategy, it is possible to obtain accurate results avoiding some of the problems that affect local and global methods.

2.4.2 Graph Cuts

Graph cuts can be employed to efficiently solve a wide variety of low-level computer vision problems, such as image smoothing, stereo correspondence problem, and many other problems that can be formulated in terms of energy minimization.
The basic technique is to construct a graph for the energy function to be minimized such that the minimum cut on the graph also minimizes the energy (either globally or locally). The minimum cut, in turn, can be computed very efficiently by max flow algorithms. Several visual correspondences problems are based on it. For example, stereo and motion [18][97], image synthesis [105], image segmentation [26], voxel occupancy [170] and multi-camera scene reconstruction [100], showing good results, with higher accuracy and lower errors, than standard methods such as local normalized correlation.
Many vision problems require the estimation of some spatially varying quantity (such as intensity or disparity) from noisy measurements. The major restriction is that the energy function smoothness term must only involve pairs of pixels which intensity (or disparity) can vary smoothly (at most points) or can change suddenly at object boundaries.
In the classical approach to stereo, the goal is to compute, for each pixel in the left image, a label d_p, which denotes a disparity value for a pixel p. Therefore, let M be the set of pixels in the master image, let S be the pixels in the slave image, and let P be the set of all pixels ($P = M \cup S$), the pixel p will have coordinates (p_x, p_y) and, a characterization of the energy functions that can be minimized by graph cuts can be:

$$E(f) = \sum_{p \in M} D_p(f_p) + \sum_{p,q \in N_p} V_{p,q}(f_p, f_q) \tag{2.30}$$

The goal is to find a labelling that minimizes some energy function. N is the set of pairs of adjacent pixels, $D_p(f_p)$ is a function derived from the observed data that measures the cost of assigning the label f_p to the pixel p, $V_{p,q}(f_p, f_q)$ measures the cost of assigning the labels f_p, f_q to the adjacent pixels p, q and is used to impose spatial smoothness. . It's worth noting that eq. (2.30) is a generalization of eq. (2.21). At the borders of objects, adjacent pixels should often have very different labels and it is important that E not over-penalize such labelling. This requires that V be a non convex

function[1] of $|f_p - f_q|$ because there can be multiple locally optimal solutions. Unfortunately, in this case, the identification of the true global solution generally takes a lot of time. In the last few years, however, efficient algorithms have been developed for these problems based on graph cuts.

Let $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ a weighted graph with two distinguished vertices called the terminals, a set of nodes $\mathcal{V}$ and a set of directed edges $\mathcal{E}$ that connect them. Take into consideration, henceforth, in the context of image matching, that nodes correspond to the selected matching representation (generally pixels, but also voxels or other features) and have an associated cost; terminals, correspond to the set of labels that can be assigned to pixels. Each edge has an associated flow capacity that is defined as a function of the costs of the adjacent nodes it connects.

Considering a graph with two terminal, namely source s and sink t (Figure 2.14 shows a simple example of a two terminal graph as is described in [152]).

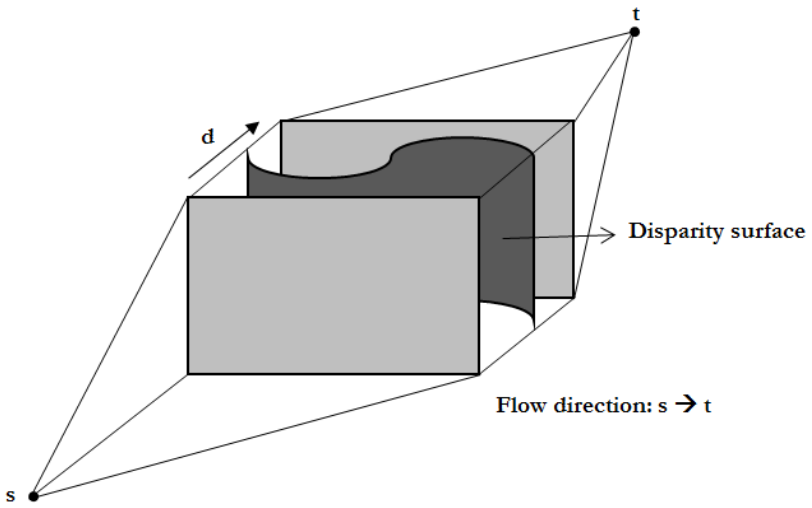


Figure 2.14: Maximun flow representation of disparity estimation (with a six-connected 3D mesh) as described by Roy and Cox [152].

[1] A convex optimization problem maintains the properties of a linear programming problem and a non convex problem the properties of a non linear programming problem. The basic difference between the two categories is that a) in convex optimization there can be only one optimal solution, which is globally optimal or you might prove that there is no feasible solution to the problem, while b) in nonconvex optimization may have multiple locally optimal points and it can take a lot of time to identify whether the problem has no solution or if the solution is global. Hence, the computational effort in convex optimization problem is usually much lower.

A s/t cut C on the graph will be a partitioning of the nodes in the graph into two disjoint subset S and T such that $s \in S$ and $t \in T$. The cost of the cut C is defined as the sum of the costs of all edges that go from S to T: it has also possible to say that the capacity of the cut is the sum of the edge capacities making up that cut. Then, the minimum cut problem is to find the cut that has the minimum cost among all cuts (as mentioned in the theorem of Ford and Fulkerson [57], this is equivalent to computing the maximum flow from the source to sink).
Consider the example in Figure 2.15: the graph corresponds to a 3x3 image; s/t cut partitions the nodes into disjoint groups which contain one terminal each one.

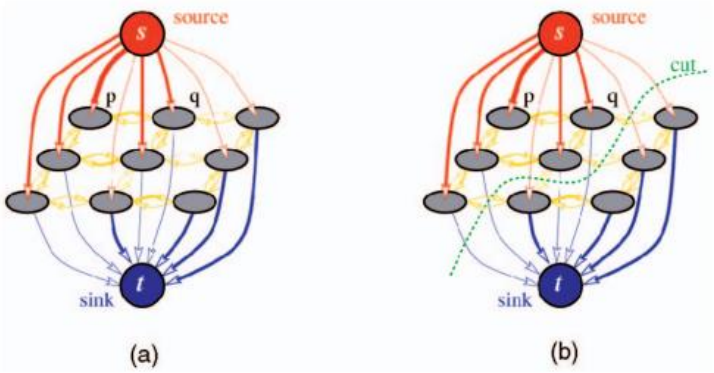


Figure 2.15: Example of a directed capacitated graph. Edge costs are reflected by their thickness. A similar graph-cut construction was first used in vision by Greig et al. [66] for binary image restoration. In (a) and (b) are shown respectively a graph $\mathcal{G}$ and a cut on $\mathcal{G}$.

Therefore, any cut corresponds to some assignment of pixels (nodes) to labels (terminals). If edge weights are appropriately set based on parameters of an energy, a minimum cost cut will correspond to a labelling with the minimum value of this energy. An explicative description and representation of the stereo correspondence problem by using graph cuts is described in Roy and Cox [152] (see Figure 2.14). The graph in this case is composed by the image pixels (nodes) and the disparity values (edges).
They searched a different solution for solving the image stereo (and also multi-image) correspondences problem, by transforming it into a maximum-flow problem. In particular, the authors have focused on the simplification and effects that the use of epipolar lines generates into the matching process. Indeed, these lines reduce the stereo matching problem to a 1D-search domain, operation that is fundamental for computational process efficiency. However, the use of epipolar lines can vary the accuracy of the method in particular near object boundaries (i.e. boundary that are perpendicular to the epipolar lines) and can create artefacts and misalignment problems.

Therefore, in [152], the traditional approach to stereo matching, which correlate a single pair of epipolar lines at a time, is extended to the whole image at once by matching all pairs of epipolar lines simultaneously. In this way, every minimum-cost path that defines the matching cost of an epipolar line is integrated into a single minimum cost surface, that contains all the disparity data of the master (or reference) image. The traditional ordering constraint (which states that the points order along corresponding scanline is preserved) is thus replaced with the more general local coherence constraint, which suggests that disparities data tend to be locally very similar in any directions. Since this property regards naturally the points ordering constraint along epipolar lines, but it is also true for all directions (and thus across epipolar lines), it becomes possible to take advantage of local coherence for putting all the epipolar lines together and solving globally for a disparity surface. This means that we must consider a global 2D stereo correspondences global problem that, in this case, is transformed in a maximum-flow problem in which the minimum-cut of a graph can be interpreted as a disparity surface. Naturally, many graph cut methods have been developed through the years. To mention some: Thomos et al. [181] have developed an efficient data structure that reduces the memory requirements, making this algorithm more manageable for large data sets; Kolmogorov et. al., [99] has presented a new graph cut algorithm that handles occlusions properly, while maintaining the key advantages of graph cuts; Boykov and Kolmogorov [27] have developed an approximate Ford-Fulkerson style augmenting paths algorithm, which they show being much faster in practice than standard push-relabel approaches; Boykov et al. [28] propose expansion move and swap move algorithms that can simultaneously modify labels of arbitrarily large pixel sets and, finally, in [140] the MICMAC dense point clouds and orthophotos generation tool is presented.

2.4.3 Bayesian Networks and Markov Random Fields in graphical models

State-of-art computer vision algorithms use prior knowledge about the statistics of typical surfaces to infer the three–dimensional (3D) shape of a scene from ambiguous, local image measurements. Probabilistic graphical models provide a powerful, general framework for designing systems like these. In this context, graphs are used to decompose joint distributions into a set of local constraints and dependencies. Such modular structure provides an intuitive language for expressing domain–specific knowledge, and facilitates transfer of modelling advances to new applications.

Several methods have been proposed for representing probability distributions with the use of a graph. For example, directed graphical models, i.e. Bayesian networks or undirected graphical models, i.e. Markov random fields (MRFs). This two methods allows to model the disparity estimation problem as in image labelling problem and finally, once the problem has been formulated using the graphical model, the energy minimization task can be performed with global optimization methods, such as the above described Graph Cut or the following presented Belief Propagation technique.
First of all, a synthetic dissertation of the two most popular graphical models is presented.

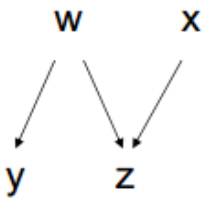
Bayesian Networks

Bayesian networks are probably the most popular type of probabilistic graphical model [134] [91] that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). They are frequently used for medical diagnosis problem: for example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of various diseases presence.
Formally, Bayesian graphical models are graphs whose nodes represent random variables that can be in a discrete number of possible states. Set as x_i the variable that represents the different possible states of node i. In addition to the qualitative dependencies described by the Bayesian network graph, there are quantitative statistical relationship associated with each edges. Edges represent conditional dependencies (or probability). Nodes that are not connected represent variables that are conditionally independent of each other, while connected nodes are defined as "parent". Each node is associated with a probability function that takes, as input, a particular set of values for the node's parent variables, and gives (as output) the probability of the variable represented by the node. Nodes that have more than one parents are described by conditional probability in terms of all their parents.
Generally, a Bayesian networks graph of N random variables x_i is described by a joint probability function

$$p(x_1, x_2, \dots, x_N) = \prod_{i=1}^{N} p(x_i | Par(x_i)) \tag{2.31}$$

where $Par(x_i)$ denotes the states of the parents of the node i.
If i has no parents, $p\left(x_i \middle| Par(x_i)\right) = p(x_i)$.

To give a simple example of a Bayesian net see Figure 2.16.



$$P(w,x,y,z) = P(w)P(x)P(y|w)P(z|w,x)$$

Figure 2.16: An easily understood representation of a simply Bayesian network.

The algorithm goal is to compute a certain marginal probability within the variables introduced in the graph and, in fact, the term "inference" simply wants to indicate the computation of these marginal probabilities.
Mathematically, marginal probabilities are defined in terms of sums over all the possible states of all the other nodes in the system. For example, in order to obtain the marginal probability of the last node N, we have to compute

$$p(x_N) = \sum_{x_1}\sum_{x_2}\dots\sum_{x_{N-1}} p(x_1, x_2, x_3, \dots, x_N) \tag{2.32}$$

Marginalization sums can easily do directly but the number of terms in the sums grows exponentially with the number of hidden nodes in the network. Indeed, if information about some of the nodes are available, the corresponding variable can be fixed and we will not have to sum over the unknown states of the nodes.

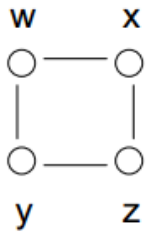
Pairwise Markov Random Fields

Pairwise MRF's is a powerful theoretical model to solve many computer vision and image matching problems [63]. To take an example [202], suppose that we want to infer the distance of the objects in a scene from the viewer. The problem, given the image, is to infer the distance values d_i corresponding to intensity values I_i, where i ranges over all the possible pixel positions within the image.
More in general, some quantities about the image y_i are observed, and some other quantities x_i about the underlying scene are investigated. The indices i could represent single pixel positions, or it might represent the position al a small patch of pixels.
Assuming that there are statistical dependencies between x_i and y_i, at each position i, it is possible to write the overall joint probability function as

$$p(x,y) = \frac{1}{Z}\prod_{(i,j)}\psi_{i,j}(x_i,x_j)\prod_{i}\phi_i(x_i,y_i) \tag{2.33}$$

where the first product is over connected pairs of nodes, while the second term $\phi_i(x_i,y_i)$ represents a joint compatibility function which assumes, and describes, that some statistical dependency between x_i and y_i exist. It is also called as the "evidence" for x_i. Z is a normalization constant.

This structure can be better understood assuming that the nodes i are arranged in a two-dimensional grid and considering the scene variables x_i are compatible with neighbour scene variables x_j, as represented by the compatibility function $\psi_{i,j}(x_i,x_j)$. In this example model the product over (i,j) will be over nearest neighbours on a square lattice.



$$P(w,x,y,z) = \frac{1}{Z}f_{wx}(w,x)f_{xz}(x,z)f_{yz}(y,z)f_{wy}(w,y)$$

Figure 2.17: Example of a MRF graph.

The MRF is defined as "pairwise" because the compatibility functions only depend on pairs of sites (i,j). In contrast to Bayesian networks, this graphical model is undirected therefore, there is no notion that the variable at one node x_i represents a causal parent of its neighbour x_j. Thus, undirected compatibility function $\psi_{i,j}(x_i,x_j)$ are used instead of conditional probability functions $p(x_i|x_j)$.

An example of a MRF graph is shown in Figure 2.17.

2.4.4 Belief Propagation

Belief propagation (BP), also known as max-product message passing, is a message passing algorithm for performing inference on graphical models, such as Bayesian networks [134][63] and Markov random fields [89] that represent statistical dependencies of variables by a graph. In other words, the key idea of BP is simplified Bayesian networks: it allows propagates information throughout a graphical model using a series of messages sent between neighbouring nodes iteratively.

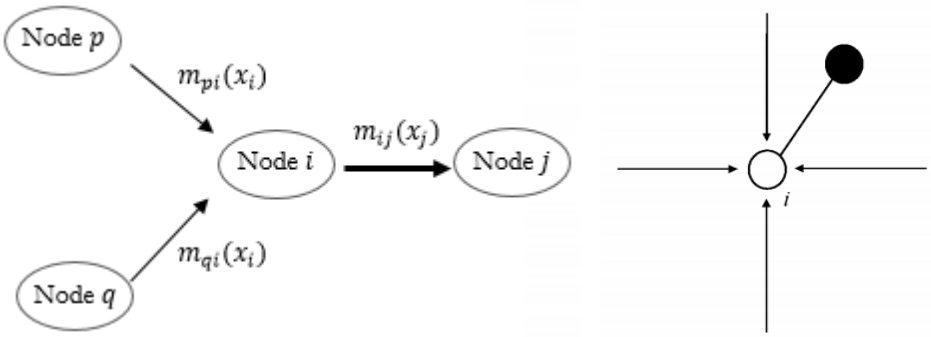
The algorithm was first proposed by Judea Pearl in [133] who formulated this algorithm on trees. The tree representation insists that only one variable can be considered a cause for any other variable. This restriction simplifies computations and avoids the problem of maintaining consistency among interrelated variables. At the same time, using the Pearl approach, many real problems cannot be modelled naturally. Therefore, the method was later improved to poly-trees by Kim & Pearl in [96] who extend the hierarchical tree representation by allowing multiple causes to a given manifestation. Overall, BP has since been shown to be a useful approximate algorithm on general graphs [134].

The BP algorithm works by passing real valued functions, called messages, along the edges between the model/network nodes. These messages contain the "influence" that one variable exerts on another. In other words, BP is an iterative process in which neighbouring variables "talk" to each other, passing messages (such as "X variable think that Y variable belong in these states with various likelihoods …"). After enough iterations, this series of conversations is likely to converge to a consensus that determines the marginal probabilities of all the variables. The estimated marginal probabilities[2] are called beliefs.

Since undirected graphs are more useful in stereo mathing applications, a standard Belief Propagation process (focusing on pairwise MRF's for simplicity) will be described. There are two types of BP methods for MRFs: max-product and sum-product. The max-product BP algorithm finds a labelling with maximum posterior probability, or equivalently with minimum energy. It works by passing messages around the graph defined by the four-connected image grid. The method is iterative, with messages from all nodes being passed in parallel. The sum-product BP algorithm can be used to approximate the posterior probability of each label for each pixel. As with the max-product algorithm, the sum-product method works by passing messages around the graph.

In order to describe a BF process, the "message" from node i to node j must be introduced. In the BP algorithm, the variable $m_{ji}(x_i)$ represents a message from a hidden node i to the hidden node j about the state of node j (see Figure 2.18 (a)) and it is a vector of the same dimension as x_i.

[2] In probability theory and statistics, the marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of the values of the variables in the subset without reference to the values of the other variables. This contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables.



(a) (b)

Figure 2.18: a) An example of a messages network (with node and message vectors). b) Representation of a node neighbourhood.

The belief at a node i is proportional to the product of the local evidence at the node ($\phi_i(x_i, y_i)$), and all the messages coming into node i:

$$b_i(x_i) = k\phi_i(x_i, y_i) \prod_{j \in N(i)} m_{ji}(x_i) \tag{2.34}$$

where k is a normalization constant (messages are similar to likelihoods: non-negative, don't have to sum to 1) and $N(i)$ denotes the nodes neighbouring i (see Figure 2.18 (b)).

The messages are determined through the message update rules:

$$m_{ij}(x_i) \leftarrow \sum_{x_i} \phi_i(x_i, y_i)\psi_{i,j}(x_i, x_j) \prod_{j \in N(i) \setminus j} m_{ki}(x_i) \tag{2.35}$$

On the right-hand side the algorithm takes the product over all messages that goes into node i except for the one coming from node j. These rules give beliefs that are exact (or exact marginal probabilities) if the pairwise MRF is "single connected" and there are no loops in the pairwise MRF, i.e. for all the nodes in any singly-connected graph (as is demonstrated in [202]).

In a practical computation, one starts with the nodes at the edge of the graph, and only computes a message when one has available all the messages necessary. It is important to note that each message must be computed once for single-connected graphs. Therefore, the whole computation takes a time proportional to the number of graph

links, which is less than the exponentially big time that would be required to compute marginal probabilities.

More information are presented in [56][174][197].

Now, try to examine in depth how the max-product algorithm can be used to find an approximate minimum

cost labelling of energy functions in the form of eq. (2.21).

The max-product BP algorithm works by passing messages around the graph defined by the four-connected image grid. Each message is a vector of dimension given by the number of possible labels [56].

Consider m_{pq}^{t} as the message that node p sends to a neighbouring node q at time t: at each iteration new messages are computed in the following way:

$$m_{pq}^{t}(f_q) = min_{f_p}\left\{V(f_p, f_q) + D_p(f_p) + \sum_{s\in N(p)\backslash q} m_{sp}^{t-1}(f_p)\right\} \tag{2.36}$$

where $N(p)\backslash q$ represents the neighbours of p other than q. After T iterations a belief vector is computed for each node as:

$$b_q(f_q) = D_q(f_q) + \sum_{p\in N(q)} m_{pq}^{T}(f_q) \tag{2.37}$$

Finally, the algorithm will allow to select the label f_q^* that minimizes $b_q(f_q)$ individually at each node.

Unfortunately, the standard implementation of this message passing algorithm on the grid graph runs in $\ominus(nk^2T)$ time, where n is the number of pixels in the image, k is the number of possible labels for each pixel and T is the iteration number. It can be also explained as $\ominus(k^2)$ time to compute each message and $\ominus(n)$ messages per iterations. Several techniques for reducing the time needed to compute the messages updates are described in [56]: combining all these methods together we obtain an $\ominus(nk)$ algorithm that is very fast in practice. Moreover, they shown that the achieved results are as accurate as those obtained when using standard max-product BP or graph cuts algorithms to minimize energy functions of the form in eq. (2.21).

Concluding, as regard the Bayesian networks goal, the advantage of BP algorithm use is that it is possible to compute marginal probabilities in a time that linearly increases with

the number of system nodes; at the same time, BP is very useful to solve MRFs problems, where the computation of marginal probabilities takes exponential time.

2.5 Disparities Refinement

Since the disparity maps are typically computed at discrete pixel level, more accurate disparity estimation (and also outliers detection) would be desirable: in fact, several photogrammetric and computer vision applications, require high levels of accuracy.
Generally, sub-pixel disparity estimation can be computed by fitting a curve to the matching costs at the discrete disparity levels [181][182] [94] and then searching for the extremum; typically it is obtained fitting three matching costs with a second degree function (see for instance Figure 2.19).

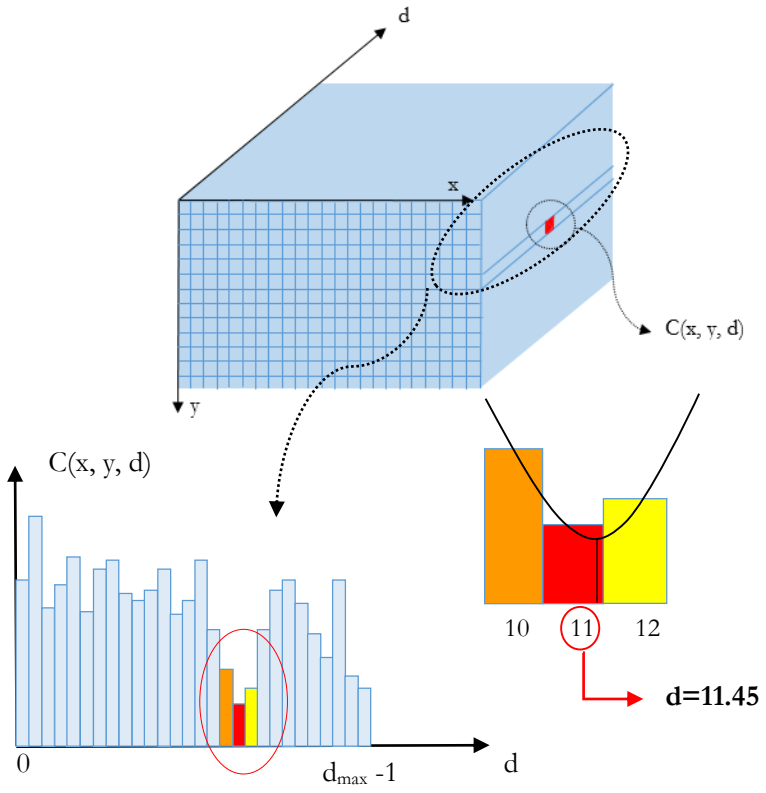


Figure 2.19: Sub-pixel disparity estimation by fitting a curve to the matching costs at the discrete disparity levels.

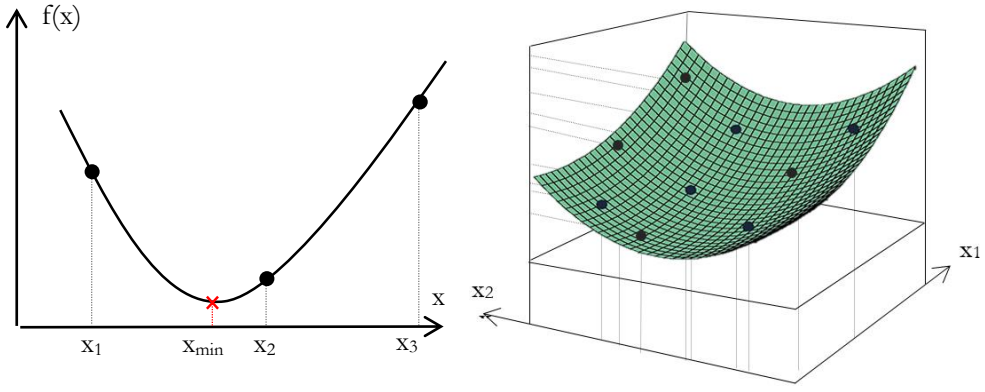
The position of the minimum point for this function will represent the subpixel shift. This provides an easy way to increase the resolution of a stereo algorithm with little additional computation.
Fitting a curve through the matching costs can be performed, in the easiest way, using a parabola fitting method (in the variable x) of the form:

$$f(x) = ax^2 + bx + c \tag{2.38}$$

that will exactly fit a parabola to three neighbour matching costs (see Figure 2.20 (a)).
If the order of the equation is increased to a fourth degree polynomial (Figure 2.20 (b)), the following quadric function (with two variables x_1, x_2) is obtained:

$$f(x) = ax_1^2 + bx_2^2 + cx_1^2x_2^2 + dx_1^2 + ex_2^2 + f \tag{2.39}$$

that will fit a surface to the nine values represented by a 3 x 3 pixels matching costs area. By equating the two expressions to zero the coefficients are characterized and the minimum of the polynomials functions can be determined.



a) b)

Figure 2.20: a) A second degree polynomial of form of eq. (2.38). b) A quadratic polynomial of form of eq. (2.39).

However, estimating a curve fitting to the cost function used for correspondences data between the master and slave images can lead to a systematic biasing towards integer values of displacements the corresponding points: this is a numerical problem also known as "pixel-locking (or peak-locking) effect". According to the work of [124], it is possible to avoid the peak locking bias refining both the reference and the matching

image coordinates simultaneously, in a symmetric way, thus the refined matches coordinates will have subpixel precision in both the images. In other words, the fundamental idea is to find the subpixel refinement by considering a 2-D neighbourhood of the matching cost around each matching pair and finding the optimal cut in the direction of the symmetric lines of the matching cost. The authors have also shown that the symmetry of the matching cost function can be described by a continuous surface that can be well-represented by a Gaussian Cylinder (GC) representation (more details are illustrated in paragraph 3.1.9.1).
At the same time, there are other methods for computing the disparities refinement, which include the post-processing of the computed disparities data in terms of image filtering. In this case, the disparity maps can be simply refined by means of image filtering techniques without enforcing any constraint about the underlining disparity image. Common image operators are median filtering, and bilateral filtering which can be applied to "clean up" spurious mismatches; holes due to occlusion can be filled by surface fitting or distributing neighbouring disparity estimates [19][159] and, finally, occluded areas can be detected using cross-checking (comparing left-to- right and right-to-left disparity maps) [39][61].

2.6 Multi-Image methods

The term "multi-frame" or "multi-image" refers to image matching techniques that have the purpose of reconstructing a complete 3D object model from a collection of images (more than two) taken from known camera viewpoints. The advantages of performing matching across multiple images have been pointed out in different studies where multi-image matching techniques were shown more robust and accurate than pairwise matching algorithm, thanks to the redundancy of multiple view points. This is true in particular for 3D reconstruction tasks, where the need of the highest robustness/accuracy is more important than the achievement of real-time results.
The term "multi-image" can refer to different matching approach; it is used generically to compute the correlation to any number of images greater than two. However, these images can be matched in pairwise mode or simultaneously, and namely all images can be used both as template/reference and search images [132].
Over the last few years, a number of high-quality algorithms have been developed, and the state of the art is improving rapidly. Many approaches try to solve the multi-image problem as a repeated binocular processing but, in order to perform a real simultaneous matching across multiple images it is necessary to generalize the two-image epipolar relations to some multi-linear geometric relation between the different views. In this

regard, for example, Shashua [167] presented a trilinear constraint where points in three images can be the projections of a single 3D scene point by following an algorithmic rule. Hartley [75] also conceived a similar constraint for lines in three view, whereas Triggs [186] provided a quadrilinear relation for points in four views. Okutomi and Kanade [127] described a multi-baseline stereo method for producing a dense depth map from multiple images (having different baselines) by performing pairwise stereo matching on all image pairs and combining the results. In other words, the matching technique is based on the idea that global mismatches can be reduced by adding the sum of squared difference values from multiple stereo pairs (i.e., the SSD values are computed for each pair of stereo images.) Thanks to the integration of information given from multiple images, the method appeared accurate, reducing matching ambiguity and improving precision values. However, the main disadvantage was the time efficiency and the method assumption that all cameras must be located on the same line. For this reason, the strategy was later changed, allowing to combine the matching results between one reference image and all the other sequence images. A generalized version of the Okutomi technique is the disparity normalization method proposed by Sakai et. al. [155]: it can integrate the correlation functions calculated from stereo image with different viewpoints even if the cameras are not located on the same line. Finally, multi-baseline matching (as proposed e.g. in [83][84]) performs stereo matching by SGM between a base image and all the other matching images. In most of the developments, the processing pipeline removes invalid disparities by consistency check (left-right check) and combines all stereo matching results by selecting the median value of all disparities for each pixel.

Although being popular approaches, there is an inherent flaw in this style of processing. Considering a matching algorithm that select one image as a "reference", features in that image are extracted, and then the other images in the dataset are searched for correspondence matches, typically using epipolar constraints between the reference image and each other image in turn. However, if an important feature is missing in the reference image due to misdetection or occlusion, it will not be present in the 3D reconstruction even if it has been detected in all the other views, because the system won't know to look for it [40] (for this reason many multi-image matching techniques that only operate in image-space were found less efficient with respect to multi-image methods that reconstruct scene structure in object space). Moreover, the combinatorial exploration of all possible correspondences could be computationally very expensive using these methods. A possible solution is to make full and efficient use of the geometric relationship between multiple images and the scene [162][40][131][128][17]. Briefly, it is possible to claim that: (i) a true multi-image matching technique should be

applicable to any number of images n greater than 2, (ii) that the algorithmic complexity should be O(n) in the number of images, and (iii) that the images should all be treated in an equal manner. In other words, the method it is not simply a repeated application of binocular - trinocular image matching techniques (i.e. approaches that process all image pairs and then fuse the results), and the information content of each image must be treated equally.

As is described in [165], multi-view stereo algorithms can be roughly categorized into four classes.

The first class operates by first computing a cost function on a 3D volume, and then extracting a surface from this volume. A simple example of this approach is the voxel colouring algorithm and its variants [164][185] which make a single sweep through the volume, computing costs and reconstructing voxels with costs below a threshold in the same pass. Other algorithms differ in the definition of the cost function and the surface extraction method. A number of methods define a volumetric MRF (i.e. a volumetric object reconstruction method using the three-dimensional Markov Random Field model-based-segmentation) and use max-flow [151] or multi-way graph cut [183] to extract an optimal surface.

The second class of techniques works by iteratively evolving a surface to decrease or minimize a cost function. This class includes methods based on voxels [53], level sets, and surface meshes. Space carving [103] and its variants (such as [169][104]) progressively remove inconsistent voxels from an initial volume. Other variants of this approach enable adding, as well as deleting, voxels to minimize an energy function [168][201]. Level-set techniques minimize a set of partial differential equations defined on a volume. Like space carving methods, level-set methods typically start from a large initial volume and shrink inward; unlike most space carving methods, however, they can also locally expand (if needed) to minimize an energy function. Other approaches represent the scene as an evolving mesh [206] that moves as a function of internal and external forces.

In the third class are image-space methods that compute a set of depth maps. To ensure a single consistent 3D scene interpretation, these methods enforce consistency constraints between depth maps [100][177], or merge the set of depth maps into a 3D scene as a post process [122].

The final class consists of algorithms that first extract and match a set of feature points and then fit a surface to the reconstructed features [179].

In Schluter [162], a new method for high resolution surface reconstruction by multi-image matching in object space is formulated with regard to the integration of a general 3D surface model in object space. Following, the concept of facets stereo vision in

[104] has outlined how it is possible to established a relationship between every pixel position in image space and the corresponding position on the surface in object space, providing the successfully estimation of a 3D surface description from multiple images. Collins has presented a multi-image matching algorithm that simultaneously determines image correspondences and 3D scene locations of point-like features across multiple views. The method assumption is that the areas of space where image feature viewing rays intersect are likely to be the 3D locations of observed scene features. Performing a volume partition of space into voxels, the back projection of each image can be guessed as a ray through this volume, allowing to record how many rays pass through each voxel (see Figure 2.21, from Schluter [162]).

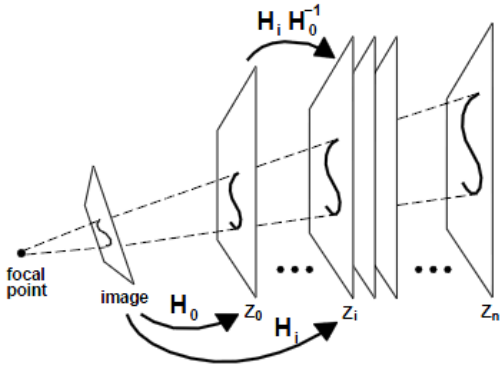


Figure 2.21: Illustration of the space-sweep method. Features from each image are backprojected onto successive positions $Z = z_i$ of a plane sweeping through space.

Recently, important advances are developing thanks to an innovative object-based multi-image Semi-global Matching (OSGM) that has been proposed by Bethmann in [17]. OSGM differs in two major aspects from standard Semi-global Matching because it is mainly characterized by transferring the process of cost calculation and path-wise cost aggregation from image space into object space. First, the cost calculation is formulated in object space within a dense voxel raster by using the grey- (or colour-) values of all images (the central coordinate of each voxel is reprojected into all images by using the collinearity equations) instead of pairwise cost calculation in image space. Then, the semi-global (path-wise) minimization process is transferred into object space as well, so that the result of semi-global optimization leads to index-maps (instead of disparity maps) which directly indicate the 3D (height) positions of the best matches.
OSGM appears a very promising method because it maintains the benefit of SGM method, adding several advantages. Opposite to most multi-baseline or multi-view stereo approaches, the new approach works without rectified images and therefore

reduces the efforts for pre-processing (no need for image rectification) and for post-processing (no need for the fusion of disparity maps). For example, in a general SGM pairwise multi-image matching algorithms, for a bundle of n images, $(n \cdot (n-1))$ images have to be rectified to create $(n \cdot (n-1))/2$ image pairs, and each image pair should be processed (e.g. with n=5, twenty images have to be rectified and ten different pairwise matching have to be performed). This, obviously, increases the computation time significantly. Besides, the image rectification process induces always a loss of information due to the need of grey- (colour-) value resampling. Further on, for all subsequently steps the results of pair-wise image matching (which are disparity maps) cannot directly be joined together but have to be fused.

CHAPTER 3

SEMI-GLOBAL MATCHING IMPLEMENTATION

Introduction

Stereo matching algorithms are used to correlate points from one digital image of a stereo pair with the corresponding points in the second image. However, finding the overall best algorithm and processing parameters, is usually difficult, since different aspects can be considered: accuracy, completeness, robustness against radiometric and geometric changes, occlusions, computational efforts, etc. The Semi-Global matching is, actually, one of the best matching strategies used both in photogrammetry and computer vision, offering good results with low runtime.

As it was previously explained, many image matching methods have been developed in the past and in recent years implementing both well-established local method (like Least Square Matching), and other innovative global and semi-global methods. The latter consider both the image similarity and the disparity continuity. Therefore, several considerations about the implemented matching cost functions (used to realize pixels correlation), the aggregation step that combine these costs and, finally, the choice of penalty functions which enforce depth continuity, need to be evaluated. Indeed, the implementation of these methods requires the introduction of many parameters and their optimal combination is fundamental to have good performances and accurate results.

In order to provide an in-depth view of all these aspects, a proprietary image matching code, based on Hirschmuller Semi-Global Matching algorithm [84] and optimized with Dynamic Programming method, has been implemented, enabling the evaluation of the best variables combinations and the optimal formulation of the matching cost function. At the same time, SGM approach has been mainly applied in stereo-vision problem (e.g. for photogrammetric applications such as the generation of digital surface/terrain models). In such context, the image geometry allows simplifying the problem of homologous points identification between stereo images, reducing the parallax domain to a mono-dimensional space (i.e. along corresponding epipolar lines). However, the mono-dimensional research of correspondences, although it represents an advantage in terms of computational efficiency and memory requirement, represents also a limit when the correlation analysis must perform a bi-dimensional investigation of image correspondences (for example, point tracking or displacement analysis). Therefore, an extension of the SGM method is proposed and developed in this work, in order to extend the disparity domain to 2D.

The current chapter will describe the heart of this work of thesis: with the aim of maximize the algorithm metrological accuracy, computational time and fields of

application, the following sections will deal with all the functional aspect of the implemented solution. In the first part, the method workflow, together with all the implemented strategies, will be accurately described. Then, a description of the tests performed for the algorithm calibration and optimization will be presented.

3.1 SGM stereo software code

Semi-Global Matching or SGM is a method developed by Heiko Hirschmuller from the DLR. He first reported his method in 2005 [83], and then he elaborated and proposed further improvements in [84][85][86]. Since its original publication, the Semi-Global Matching technique has been implemented by many researchers and companies.
The numerical code developed in this work, written in a .Net environment (using C# language), performs the images correlation, both with stereo and multi-image approach, using a Semi-Global method. It can produce, as final output, the disparity map of the investigated stereo image pairs, the stereo image correspondences for 2D displacement fields or the raster DSM. The core of the algorithm considers pair of stereo images with known interior and exterior parameters: if standard stereo-matching applications are concerned, the software pre-process the images providing their idealization and epipolar rectification. Then, a series of procedures common to all the different operational modes of the software (stereo-matching, multi-image matching, point tracking, etc.), which represent the real core of the matching process are performed.
Further developments, as already said, have regarded the algorithm extension to a multi-image matching process. Similarly, as far as displacement/tracking analysis is concerned, an algorithm extension for the 2D disparity search space investigation has been developed. In this case, due to the increase requirements of the new algorithm in terms of computational and memory cost, optimization strategies were introduced in the algorithm pipeline.
Assuming a general stereo geometry of two or more images with known relative orientation, the general matching process evolves through the following steps:

1. Image idealization/rectification: in a first step the images can be idealized and rectified. In the first case, the process consists in removing distortions from the analysed images in order to obtain a new resampled image pair where the effects of the distortions, generally introduced by the camera lens system, are corrected. In the rectification process, the stereo images are resampled, producing new photograms where corresponding points lie on the same horizontal image line.

2. Processing parameters initialization: at this initial stage the operator can initialize the algorithm processing variables and functions (such as the similarity and regularization function to use, the template size, the penalty factors, the points grid step, the master image mask, etc.). This step is fundamental since the user can intervene refining the process variables, obtaining more accurate and reliable solution based on the specific analysed problem. However, if the user doesn't initialize the required data, default parameters are internally defined.
3. Costs calculation: firstly, the similarity cost of the evaluated stereo pair is computed, with the selected correlation window size and similarity cost function (SSD, SAD, NCC, Humming Distance).
4. Smoothness constraints enforcing: in this stage, being the pixel-wise cost computation generally ambiguous (wrong matches can easily have a lower cost than correct ones, due to noise, repeated patterns, etc.), additional constraints are added to constrain the regularity of the displacement field. In other words, these "smoothness" constraints allow penalizing neighbouring disparities changes (in terms of adjacent pixels displacement). The pixel wise cost and the regularization constraints are expressed by defining a global energy function that has to be evaluated for all possible disparities combinations of each pixel. The problem is that the asymptotic time for evaluating the global energy function is the number of possible disparities raised to the image width times the image height: this is a NP-complete problem that takes an almost infinite computational time. The employed DP algorithm realizes the minimization of this global cost function, combining matching costs along independent one-dimensional paths from all directions through the image. The costs extracted by each path, referred to a particular displacement value, are summed for each pixel and for each possible displacement (also referred as disparity) value.
5. Disparity computation (with sub-pixel accuracy): finally, the algorithm choses the disparity solution with the minimum cost, using a Dynamic Programming approach.

The algorithm general workflow (stereo matching) is schematize in Figure 3.1:

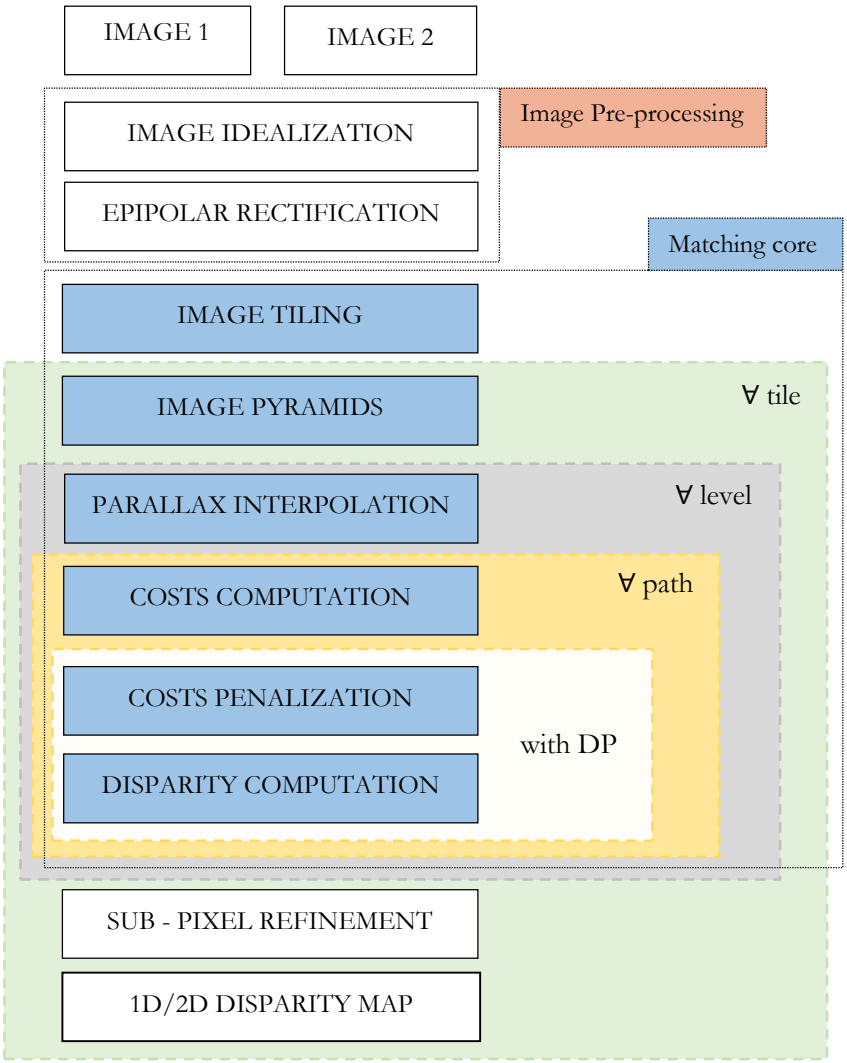


Figure 3.1: General SGM algorithm workflow (valid for 1D/2D disparity search) for stereo-pairs approach.

The workflow starts considering an image stereo pair and, being known their orientation parameters, the idealization and epipolar rectification processes are performed (see paragraphs 3.1.1 and 3.1.2 for more details). After that, in order to improve the computational performances, the program decompose the master image into tiles (the operator can choose the tile size and overlap). At this point, the determination of corresponding points will be done simultaneously (taking advantage of parallel computing capabilities of modern multi-core processors) for each tile. If the hierarchical image pyramid approach (see paragraph 3.1.5) is used, starting from the full resolution

of the image, each tile is down-sampled from the previous level, resulting in smaller images where the spatial sampling density (and, consequently, also the disparity search range) decreases level by level.

The correlation step can now start: the image matching is performed on the lower-resolution image pairs and, starting from a regular grid step defined on the reference image, the approximate correspondences on the slave image are computed by using one of the implemented parametric or non-parametric cost functions. At this time the process goes ahead performing the costs penalization and disparity optimization steps. The described workflow (and illustrated within the detached yellow frame in Figure 3.1) is repeated for each image level and, at the end of the image matching procedure, the results of each tile are joined and the disparity map of full-resolution images is determined. It is important to notice that the images rectification process is necessary if the research of image correspondences is performed in a mono-dimensional search domain. However, an extension of the algorithm to 2D search space has been implemented for tracking purposes. In this case, the rectification process (and often the idealization process, too) is not needed.

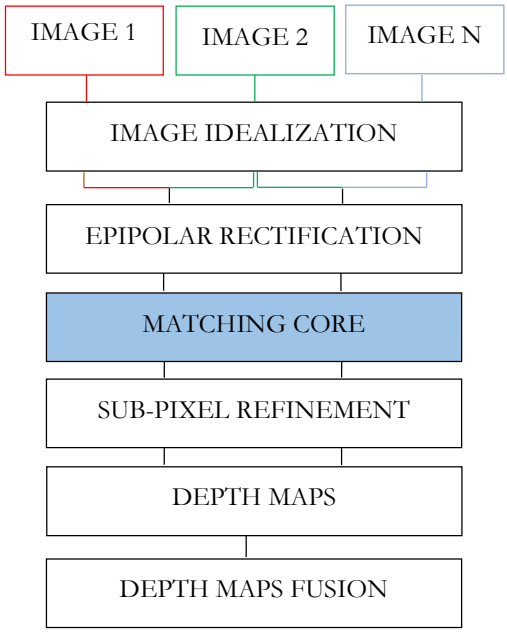


Figure 3.2: Algorithm workflow extension for pairwise multi-image approach.

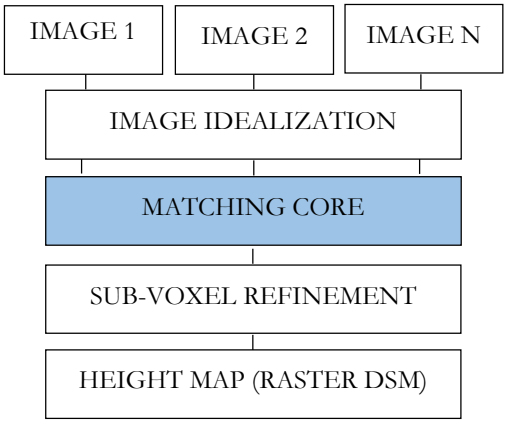


Figure 3.3: Algorithm workflow extension for object-based multi-image approach.

As far as the multi-image algorithmic extensions are concerned, two different methods have been developed: a pairwise-based and an object-based multi-image SGM algorithm. In Figure 3.2 and 3.3 the strategies for both approaches are presented (for a more clear comprehension, the "matching core" stage summarizes here all the matching steps detached in blue in the basic SGM workflow just described in Figure 3.1).
In the following paragraphs, all the strategies implemented in the method will be accurately described.

3.1.1 Image idealization

The image idealization process consists in removing distortion components (which are a consequence of the non rectilinear path of the projection ray through the camera lenses) resampling the analysed images and obtaining a new frame that geometrically fulfils the collinearity condition model.
The process is quite simple: in a first stage, for each image to resample, the correction to be applied on the image coordinates (i.e. the deformation field) is calculated with the Brown's distortion model (other custom distortion models can be specified by the user):

$$\begin{aligned} \Delta X_r &= \frac{X}{\rho}\Delta\rho = X\left(K_0 + K_1\rho^2 + K_2\rho^4\right) \quad \Delta X_d = P_1\left(\rho^2 + 2X^2\right) + 2P_2 X\eta \\ \Delta Y_r &= \frac{Y}{\rho}\Delta\rho = Y\left(K_0 + K_1\rho^2 + K_2\rho^4\right) \quad \Delta Y_d = P_2\left(\rho^2 + 2Y^2\right) + 2P_1 Y\eta \end{aligned} \tag{3.1}$$

where ΔX , ΔY represent the correction to apply in X and Y direction (r and d represent the radial and tangential distortion components) and K_0, K_1, K_2 P_1, P_2 are the radial and tangential distortion parameters, that can be obtained using a calibration procedure. In particular, on the original (distorted) image, a set of image points are selected on a regular grid. The point density (i.e. the grid step) is selected according to the magnitude of the distortion: for higher distorted images a denser grid is selected. For each point the corresponding, undistorted, position is computed using eqs. (3.1). The resulting point distribution, unfortunately, is not regular anymore and require to be re-interpolated: using a Delaunay triangulation the correction to pass from undistorted image position to the corresponding deformed location can be easily computed. In this way, a new resampled image, where the effects of the distortions are removed, can be now obtained by image resampling.

3.1.2 Epipolar rectification process

Consider (see Figure 3.4 below) two cameras, with their respective centres of projection points O_L and O_R, and a general object point P.

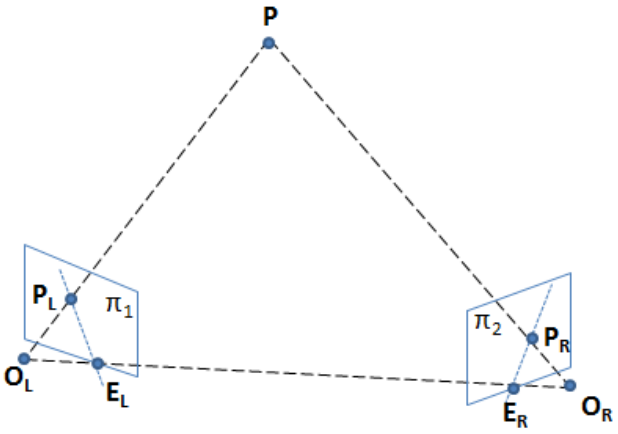


Figure 3.4: Epipolar geometry problem.

P_L and P_R denote the projection of P onto each of the image planes π_1 and π_2. Let's call epipolar plane the one defined by the point P, O_L and O_R. The epipolar plane intersects the image planes defining two epipolar lines, which contain the image points P_L and P_R. The images of O_L on π_2 and O_R on π_1 are the epipoles E_L and E_R.
The epipolar geometry can be summarized in two main statements:

- Every 3D point P together with the two camera centres defines a plane which intersects the image planes in two epipolar lines;
- Given an image point in one image, the epipolar constraint states that the corresponding image point on the other image plane belongs to the relative epipolar line.

Epipolar geometry can then be employed to simplify the problem of identification of homologous points between stereo images. To make the search of corresponding points along the epipolar lines faster and simpler, an image rectification algorithm can be used to resample the stereo images and to produce new photograms where corresponding points lie on the same horizontal image line.

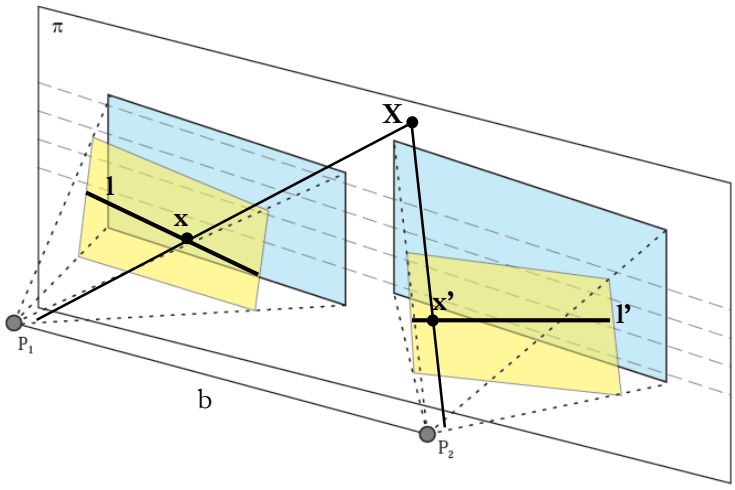


Figure 3.5: Epipolar rectification concept between two images (in yellow).

The epipolar image generation concept can be easily described by the previous Figure 3.5. The projective transformation that project the two original image planes (in yellow) to a new plane π parallel to the baseline P_1P_2 is estimated. Cropping adequately the two new images, correspondent points on the new images (in light-blue) lie on the same horizontal line, and the matching can be enforced to move along the same direction. In this way matching points search is greatly simplified, also by the near-correspondence (in terms of similarity) of the two images (e.g. rotation or other perspective deformations due to individual image poses are usually reduced).

The epipolar rectification requires that corresponding epipoles are mapped to a point at infinity (e.g., in homogeneous coordinates, to the point $(1,0,0)^T$) by a projective transformation (i.e. an homography). Referring, for clarity, to a single frame, it is possible to demonstrate that once the correspondence between the epipole and the

point at infinity has been determined, four degrees of freedom (one rotational, two translational and one for the plane scale) remain. If an unwanted homography H is chosen, severe projective distortion of the image can take place. Therefore, the four degrees of freedom must be chosen carefully.

A first condition that can be considered, is to minimize perspective deformations in the neighbourhood of a specific image point x_0 (for example an appropriate choice of x_0 may be the image centre). Let us suppose that the epipole of the considered frame lies on the x-axis (i.e. $e = (f, 0, 1)^T$). The homography should be identified by the following transformation matrix:

$$G = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{f} & 0 & 1 \end{vmatrix} \tag{3.2}$$

This transformation takes the epipole $(f, 0, 1)^T$ to the point at infinity $(f, 0, 0)^T$ as required. It is easy to prove that, if the point of interest is near the origin (0, 0, 1), the deformation in the point neighbourhood is null. In fact, a point of general coordinates $(x, y, 1)^T$ is mapped by G as follows:

$$x' = Gx = (x, y, 1 - x/f)^T = \left(\frac{x}{1 - x/f}, \frac{y}{1 - y/f}, 1\right)^T \tag{3.3}$$

Approximating these coordinates by means of a Taylor series expansion, assuming that the ratio $x/f \ll 1$ (i.e. the epipole must be at a sufficient distant from the frame format), it is possible to write

$$x' \cong (x(1 + x/f + \cdots), y(1 + x/f + \cdots), 1) \tag{3.4}$$

In other words, the transformation Jacobian is

$$J = \begin{bmatrix} 1 + 2x/f & 0 \\ y/f & 1 + x/f \end{bmatrix} \tag{3.5}$$

Please note that, if we consider the principal point $(x = y = 0)$, J degenerates to the identity matrix (i.e. the projective deformation is null). For an arbitrarily placed point of

interest $\boldsymbol{x_0}$ (not located in the origin) and epipole $\boldsymbol{e}$, the overall homograpy can be obtained as follows

$$H = GRT \tag{3.6}$$

where R is a rotation about the origin taking the epipole $\boldsymbol{e'}$ to a point $(f, 0, 1)^T$ on the x-axis, T the translation taking the point of interest $\boldsymbol{x_0}$ to the origin and G the mapping just considered taking $(f, 0, 1)^T$ to infinity.
After the epipole in one image is mapped to infinity, a map must be applied to the other image in order to allow the the epipolar lines matching. Two conditions must be taken into account in this case: the epipole in the second image must be also mapped to the same point at infinity, in order to obtain corresponding points on the same horizontal line. Then, since the previous condition is not able to fix all the degrees of freedom, a rule for fixing the residual free parameters is needed. The most obvious condition to apply could be to choose the transformation that minimize images parallaxes:

$$\sum_i d\,(Hx_i, H'^{x'}{}_i)^2 = min \tag{3.7}$$

A deeper description of this procedure can be found in [76], where the demonstration of this condition is presented in more details.

3.1.3 Processing parameters initialization

At the very beginning of matching processing, the code requires some important variables initialization. If it is not specified, default parameters are internally defined to allow the matching algorithm working; otherwise, the operator can intervene for refining and adjusting the processing variables with regard to the specific problem complexity.

- The processing parameters that can be set by the operator are:
- the region of interest (ROI): if it is not loaded or drawn a specific mask on the master image, the whole image will be considered as the matching area of interest;
- the disparity search range: i.e. the mean value and the disparity range which define the DSI depth (the values can be automatically evaluated using an initial points set obtained from the previous images orientation process, or using a SIFT feature matching);

- the image tiling procedure: if this process is selected, tiles number and tiles overlap can be defined;
- the image pyramiding option: if the procedure is selected, the image pyramid number of levels must be set; during the matching process, the disparity search range is adapted automatically taking into consideration the resolution level.
- the similarity cost function to use (with the relative image window size);
- the penalization function;
- the multi-processor option: the algorithm can run on multiple processing cores thanks to parallelization of several computation steps;

3.1.4 Matching by using image tiling

Due to the high memory requirements of SGM, especially if the DSI dimensions are relevant (i.e. using high resolution images and/or wide disparity search range), a tile-based strategy was adopted in order to increase the computational performances. The procedure allows dividing the whole master image into squared areas (partially overlapped) which are independently processed (see Figure 3.6). In case, a hierarchical approach, which is described in the following paragraph, can be applied further improving algorithm performances with high resolution images and allowing the use of bigger tiles.

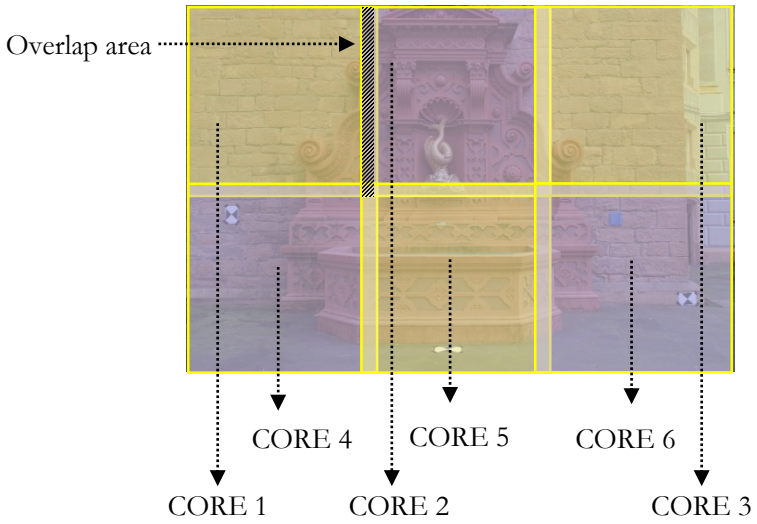


Figure 3.6: Image tiling concepts: in purple and yellow the image tiles, which are independently processed by a different processor core.

The problem of tiling is the definition of the tile size and their overlap. For well textured, rather flat scenes, small tiles (e.g. 100 x 100 pixels) and low overlaps (e.g. 15-20 pixels) are sufficient. However, low-textured scenes or high depth differences require bigger tiles and higher overlaps. The tiling procedure is widely used when high-resolution images must be correlated and represents an optimal strategy for maximizing the performance of a semi-global (or global) algorithm, where the efficiency in terms of memory requirements and time of process are determinant factors. In this context it is important to highlight that, although a global/semi-global algorithm requires to correlate neighbour pixels information, the effect on the matching operation (due to adjacent pixels cost contribute) becomes uninfluent at a certain distance from a considered pixel. Therefore, it is possible to process each tile individually without any problem connected to the loss of the costs information about the pixels processed in another tile.
In brief summary, the tiling procedure can be schematizes as follows: once the tile desired dimension is defined, the algorithm identifies the image tiles number and calculate an optimal tile dimension (i.e. making all the tiles, also on the edges, the same size); then, the master image is subdivided in n tiles that are matched individually (the number of processing cores determines the number of tiles that can be computed concurrently) with the slave image; the disparity computation/optimization (see section 3.1.8) of each tile is performed individually as well; finally, the matching results are merged together, blending the disparity results in the overlapping regions, in order to obtain the full disparity image.

3.1.5 Matching by using image pyramids

In order to perform an efficient stereo matching with high resolution images, the developed software implements also a multi-resolution approach using image pyramids and a coarse-to-fine disparity map evaluation.
Every level of the pyramid is down-sampled from the previous level (in particular the resolution is one-half of the previous level scale), resulting in smaller image. As this process proceeds, the result is a set of gradually smoother images, where the spatial sampling density (and, consequently, also the disparity search range) decreases level by level (see Figure 3.7). The idea is that performing image matching on the lower-resolution images allows performing the matching process for fewer pixels (and a smaller disparity search range) and to adaptively compute the matching parameters for the subsequent levels. In particular, the matching process has been modified to consider

an adaptively defined DSI so that the disparity range, except in the first pyramid level, is shrinked w.r.t. the full, user-defined, range.

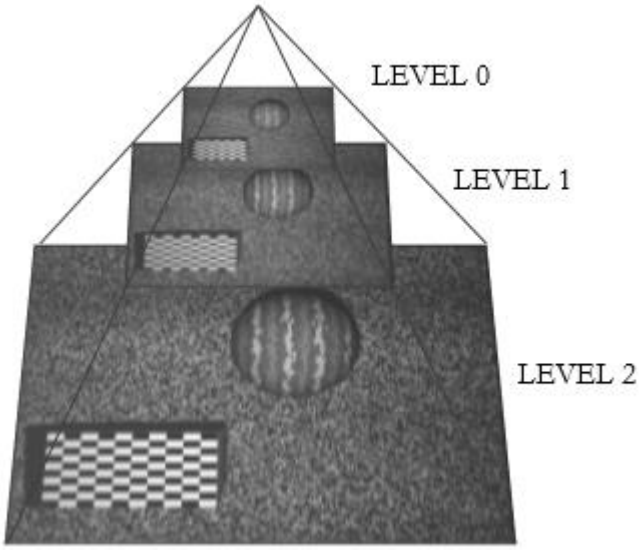


Figure 3.7: Disparity maps pyramid concept.

At the lower resolution level, the process allows to refine the initial approximate disparity values and, once an initial low-resolution parallax map has been obtained, a bilinear (super -sampling) interpolation of the parallax values is computed to provide an initial disparity best-guess for the next level. In this way, being available a pixel-wise approximate disparity value, a new matching process (with the new image pyramid pairs of the next level) using a reduced disparity search range and an optimized pixel-wise disparity offset can be performed. The reduced disparity search range can be set by the user in the initialization step.
If the tiling procedure is used, the images are processed iteratively using the pyramid strategy applied individually for each tile. This allows to considerably reduce the computational time of the correlation process. The process is repeated until it reaches the original, full resolution, image level. At this stage the final full-resolution disparity map can be obtained (see Figure 3.7).

3.1.6 Variables and implemented functions

As described in [84] and in 2.3.3 paragraph, the algorithm must consider some parameters that regard the similarity cost functions and others that allow the parallaxes regularity constraint (i.e. penalty functions).

In the developed proprietary SGM algorithm, a generalization of the Hirschmuller semi-global matching (already been presented in eq. (2.24) has been performed. The cost $L'_r(\boldsymbol{p}, d)$ of the pixel located in $\boldsymbol{p}$ at disparity d, along the path direction $\boldsymbol{r}$ is defined recursively as:

$$L'_r(\boldsymbol{p}, d) = C(\boldsymbol{p}, d) + min_i\{L_r(\boldsymbol{p} - \boldsymbol{r}, d + \Delta d_i) + P(\Delta d_i)\} \tag{3.8}$$

where the first term is the similarity cost associated with a disparity value of d, whereas the second term evaluates the regularity of the disparity field, adding a penalty term P, function of disparity changes (Δd_i) with respect to the previous point $(\boldsymbol{p} - \boldsymbol{r})$ *in* the considered matching path $\boldsymbol{r}$. Different cost and penalty functions have been implemented in the Semi-Global matching code, but the abstraction level of the source code allows also the user to provide his own functions.

First, as far as similarity cost computation is concerned (represented by the term $C(\boldsymbol{p}, d)$ in eq. (3.8)), we have implemented, the most common parametric and non-parametric correlation measures: Sum of Absolute/Squared Differences, Normalized Cross Correlation and Sum of Humming Distance:

$$SAD = \sum_{(i,j)\epsilon W} |f(i,j) - g(i + \Delta y, j + \Delta x)| \tag{3.9}$$

$$\mathrm{SSD} = \sum_{(\mathrm{i,j})\epsilon \mathrm{W}} (\mathrm{f(i,j)} - \mathrm{g(i} + \Delta\mathrm{y, j} + \Delta\mathrm{x)})^2 \tag{3.10}$$

$$NCC = \frac{\sum_{(i,j)\epsilon W}[(f(i,j)) \cdot (g(i + \Delta y, j + \Delta x)]}{\sqrt{\sum_{(i,j)\epsilon W}[\left(f(i,j)\right)^2 \cdot \left(g(i + \Delta y, j + \Delta x)\right)^2]}} \tag{3.11}$$

$$SHD = \sum_{(i,j)\in W} f(i,j)\ bitwiseXOR\ g(i + \Delta y, j + \Delta x) = f(i,j)\hat{}g(i + \Delta y, j + \Delta x) \tag{3.12}$$

An in depth descriptions of all these similarity functions has already been presented in paragraph 2.2.

As far as the disparity field regularization terms are concerned, two penalty functions are actually implemented: the one proposed by Hirschmuller, which uses two penalty parameters "P_1 , P_2" (P_1 - for small changes - permits an adaptation to slanted or curved

surfaces; , P_2 - for larger changes - preserves discontinuities), and a different penalization method, where penalty increases linearly with the disparity difference of neighbour pixels. In Figure 3.8 the two implemented functions are shown: P represents the penalization value and ΔD the neighbour pixels disparities value.

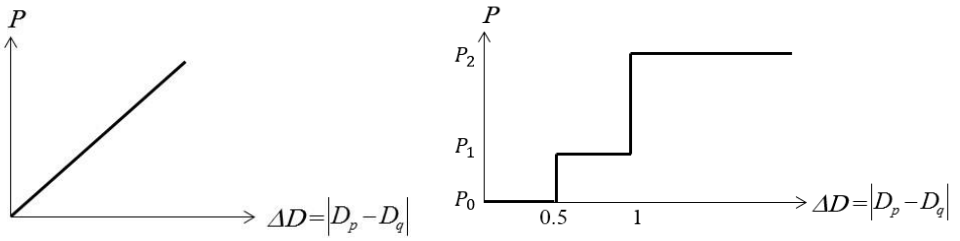


Figure 3.8: The two implemented penalty functions

Furthermore, as reported in [84], since that big disparity changes are usually indicated by significant intensity variations (as happens near objects borders and discontinuous areas), the penalty P_2 can be set not only as a constant value but it also possible to identify a different definition of the penalization parameter: in this case P_2 is adapted to the intensity gradient between the current and the previous pixel, according to the following equation:

$$P_2 = \begin{cases} \overline{P_2} & \Delta I > t \\ \dfrac{\overline{P_2}}{\Delta I + \varepsilon} & \Delta I \leq t \end{cases} \tag{3.13}$$

where $\Delta I = |I_L(\boldsymbol{p}) - I_L(\boldsymbol{p} - \boldsymbol{r})|$ represents the intensity registered difference value between adjacent pixels on the same path $\boldsymbol{r}$, ε a constant value and t is an intensity differences threshold.

3.1.7 Matching Core and disparity optimization

The minimization problem of the semi-global cost function expressed in eq. (2.24) is solved in our implementation using a DP algorithm combining matching costs along independent one-dimensional paths from all directions through the image (just like in the Hirschmuller implementation). Thus, for each pixel $\boldsymbol{p}$ and disparity $\boldsymbol{d}$, the cost is computed by the sum of the matching cost and the minimum path cost of the previous pixel $\boldsymbol{p} - \boldsymbol{r}$. The last term of eq. (2.24) subtracts the minimum cost at the previous pixel

from all costs of the current pixel. This is done for keeping the values L low and consequently using a more efficient (smaller) data type to store the matching cost. In fact, an arbitrary value could be chosen, as long as it is constant for all disparities: the minimum of the previous pixel is used, because it is already available and subtraction will never make the whole term negative.

In a DP framework (as the ones implemented in the code and in many SGM software packages [42]), the cost $L'_r(\boldsymbol{p}, d)$ is defined recursively for allowing the identification of the $min_i\{L_{\boldsymbol{r}}(\boldsymbol{p}-\boldsymbol{r}, d+\Delta d_i) + P(\Delta d_i)\}$. It is precisely in this moment that the "memoization" step (see paragraph 2.4.1), occurs. In other words, whenever a determinate sub-problem (whose solution contributes together with the other identified optimal sub-problems to the solution of the whole process) is proposed again, it will be not necessary to solve the problem, because its solution has already been calculated (i.e. it is available) and only have to be called back from the memory. The disparity solution of the previous pixels $L_{\boldsymbol{r}}(\boldsymbol{p}-\boldsymbol{r}, d+\Delta d_i)$ are necessary data to obtain the $L'_{\boldsymbol{r}}$ term but they must not be computed every time such sub-problems arise: the solutions have already been stored.

In the initial research stages, a first algorithmic development of the costs penalization step that provided more flexibility in terms of penalization paths w.r.t. the first Hirschmuller implementation of SGM, has been implemented and tested. Later, on the basis of such results, a different SGM solving approach, algorithmically more similar to the OpenCV Semi Global stereo Block Matching library [128], has been studied and employed, to improve computational efficiency of the algorithm. In the following two paragraphs, both implementation are described.

3.1.7.1 First implementation of SGM

In the first implemented SGM [84], the minimization of eq. (2.24) is done going along one dimensional paths at a time, usually considering four, eight or sixteen different directions $\boldsymbol{r}$. Anyway the user, in this implementation, can provide any number of paths indicating their direction vector (see further for details). In Figure 3.9 the paths direction are shown. The cost $L_{\boldsymbol{r}}$ are then summed over paths in all the directions . The final aggregated cost, that takes into account all the different $\boldsymbol{r}$ paths, is defined as:

$$S(\boldsymbol{p}, d) = \sum_r L_r(\boldsymbol{p}, d) \tag{3.14}$$

where S is the data volume from which the final disparity is selected.

The user can choose how many paths wants to use for the stereo correlation process; this is useful for analysing the influence of the each directions on the final results. Generally, the use of 16 paths (as is proposed in [83]) allows obtaining a detailed and more noise-free result (providing a good coverage of the 2D image) but, if a good trade-off between accuracy, completeness and computational time is requested, the matching using only 8 paths can also represent a good choice.

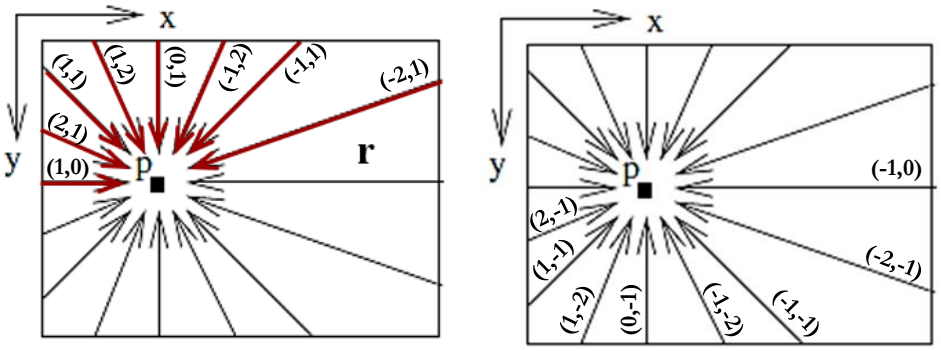


Figure 3.9: Representation of the 8 (on the left) and 16 paths (on the right) implemented in the software code.

Finally, the disparity solution can be determined by selecting for each pixel $\boldsymbol{p}$ the disparity $\boldsymbol{d}$ that corresponds to the minimum cost (which is set after all directions have been calculated), i.e. $D_{SGM}(p) = min_d S(\boldsymbol{p}, d)$.

In order to better understand and clarify how the program works when it process each scan line disparity costs with dynamic programming, a simple example is following presented.

Consider only a simple image row, composed by 5 pixels (see Figure 3.10), and suppose that we want to compute the stereo correspondence between the two pixel lines which ideally correspond to the first row of pixels of an images pair, considering the path corresponding to the direction r = (-1,0) (i.e. from right to left).

Master (or right) image	120	230	250	180	220
Slave (or left) image	140	250	200	240	250

Figure 3.10: Example of stereo correspondences between a simple image row composed by 7 pixels.

Now, starting from the last pixel (to be clear the pixel n° 5), the similarity and penalization step will be performed iteratively (thanks to the DP technique) for each pixel (which correspond at one pixel) of the disparity costs matrix filling the DSI. In this example, we suppose that the penalization linearly increases with respect to the

imposed disparity field, and the similarity term is calculated making the simple absolute value of the intensities difference and using a one pixel template size.
For the first pixel, the similarity cost $C(\boldsymbol{p}, d)$ function of the different disparity values results as in Figure 3.11:

					Disparity					
-5	-4	-3	-2	-1	0	1	2	3	4	5
1000	80	30	20	20	**30**	**1000**	**1000**	**1000**	**1000**	**1000**

Figure 3.11: Similarity costs of the first pixel by varying the disparity value.

Being the first pixel of the path, the similarity cost coincides with the total cost. The cost associated with locations in the slave image outside the boundaries are set to a very high value to avoid not congruent disparity solutions. For the second pixel, i.e. the second to last pixel, the similarity cost (Figure 3.12) is:

-5	-4	-3	-2	-1	0	1	2	3	4	5
1000	1000	40	70	20	**60**	**70**	**1000**	**1000**	**1000**	**1000**

Figure 3.12 Similarity costs of the second pixel by varying the disparity value.

Now, to implement the DP approach and fill the DSI with the matching costs expressed by eq. (3.8), it is helpful to create a penalty disparity matrix (Figure 3.13) that stores the matching cost of the pixel varying the disparity, and penalizing the costs considering the different parallax value w.r.t. the previous (first) pixel. For instance, in this case, the penalization has a value of 127 for every pixel of difference in the disparity between adjacent path pixels:

Pixel Index	Delta Disparity	Disparity								
1	**3**	2000	2000	1381	461	411	401	401	411	1381
2	**2**	2000	1254	334	284	274	274	284	1254	1254
3	**1**	1127	207	157	147	147	157	1127	1127	1127
4	**0**	80	30	20	20	30	1000	1000	1000	1000
5	**1**	157	147	147	157	1127	1127	1127	1127	1127
6	**2**	274	274	284	1254	1254	1254	1254	1254	2000
7	**3**	401	411	1381	1381	1381	1381	1381	2000	2000
Minimum values		80	30	20	20	30	157	284	411	1000

Figure 3.13: Matching costs penalization of the first pixel.

Finally, the minimum of each column of the disparity penalty matrix is selected according to eq. (3.8). At this time, the minimum vector will be summed to the previous (i.e. the fourth) line of the disparity costs matrix obtaining a new vector (shown in green in Figure 3.14) that will be penalized in the same above described.

80	30	20	20	30	157	284	411	1000
				+				
1000	40	70	20	60	70	1000	1000	1000
				=				
1080	70	90	40	90	227	1284	1411	2000

Figure 3.14: Aggregation of the first pixel penalized matching costs.

The process continues up to the location of the line considering that for all the pixels, different from the first, the minimums vector of the previous pixel must be subtracted. For sake of understanding, in Figure 3.15-16 the costs penalization matrix of the following pixel (the fourth) to penalize is shown.

Pixel Index	Delta Disparity								
1	**3**	3000	3000	3015	1095	85	105	55	105	242
2	**2**	3000	3010	1090	80	100	50	100	237	1294
3	**1**	3005	1085	75	95	45	95	232	1289	1416
4	**0**	1080	70	90	40	90	227	1284	1411	2000
5	**1**	197	95	45	95	232	1289	1416	2005	3005
6	**2**	344	50	100	237	1294	1421	2010	3010	3000
7	**3**	421	105	242	1299	1426	2015	3015	3000	3000
Minimum values		197	50	45	40	45	50	55	105	242

Figure 3.15: Matching costs penalization of the second pixel.

197	50	45	40	45	50	55	105	242
				+				
1000	1000	1000	90	20	30	10	20	1000
				-				
80	30	20	20	30	157	284	411	1000
				=				
1040	1070	130	60	120	60	50	80	170

Figure 3.16: Adjacent pixel aggregation.

In the end, once the DSI has been completely filled, the minima of the matching costs for each pixel can be searched, its location in the DSI can be memorized and the

solution parallax estimated (by reconstructing the route of the optimal indexes within the DSI matrix). An example of this step is shown in Figure 3.17nd 3.18.

Index	1	2	3	4	5	6	7	8	9
1° pixel	3	2	2	3	4	5	5	5	5
2° pixel	5	4	4	4	2	3	3	3	3
3° pixel	5	5	5	4	3	3	2	2	3
4° pixel	5	5	5	4	3	2	1	1	1
5° pixel	4	4	4	4	4	3	2	1	4

Figure 3.17: Optimal indexes matrix.

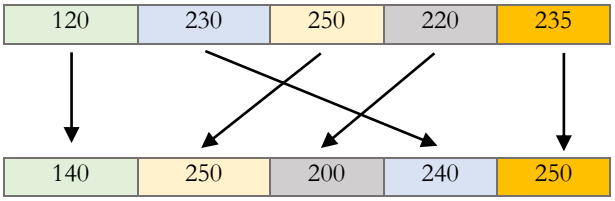


Figure 3.18: Optimal indexes route recover.

This example has simply considered an image pixels line, but a typical matching problem must evaluate the whole image. In this case, the process must continue, starting from the farthest image line pixel and repeating itself iteratively for each image scan line, from the top left pixel line wise through the image, computing the path for each defined direction (e.g. from left, diagonally from top-left, from top, diagonally from top, etc.). The Disparity Costs matrix becomes therefore a three-dimensional matrix where at each image pixel correspond a set of costs calculated by varying the disparity search range. In other words we computed a data storage $C(x, y, d)$, where (x, y) represents the pixel location and d the considered disparity value (as shown in Figure 3.19). In other word, each DSI element will contain the cost of the correspondence between the master image pixels and the slave image pixels, creating a big tensor of all the costs for all the pixels and all possible disparity values named as $C(\boldsymbol{p}, d)$ in eq. (3.8).

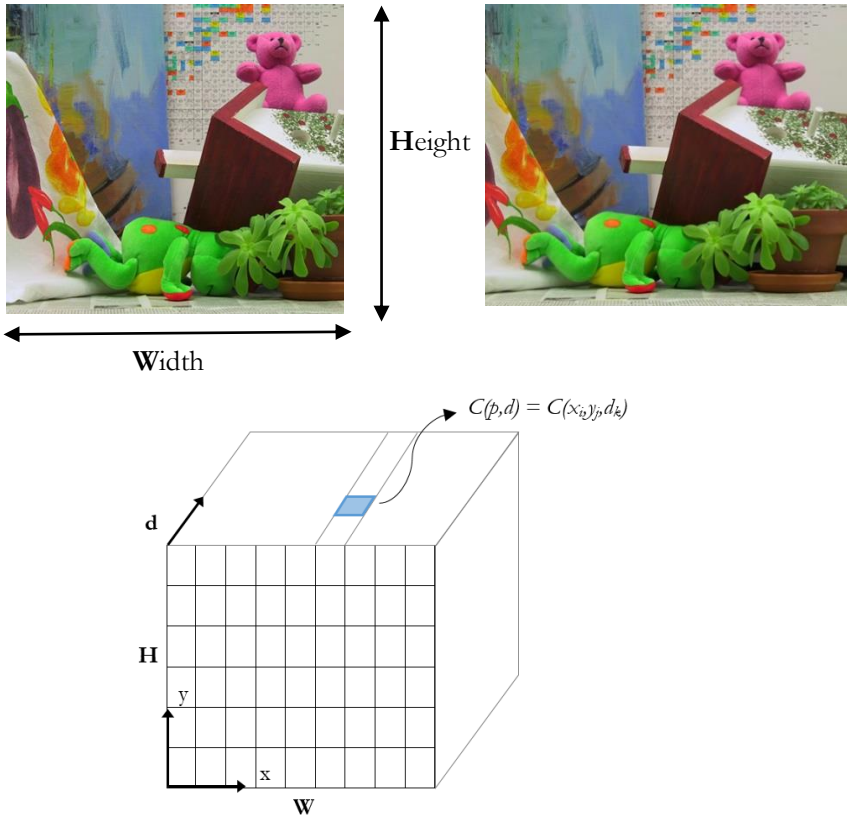
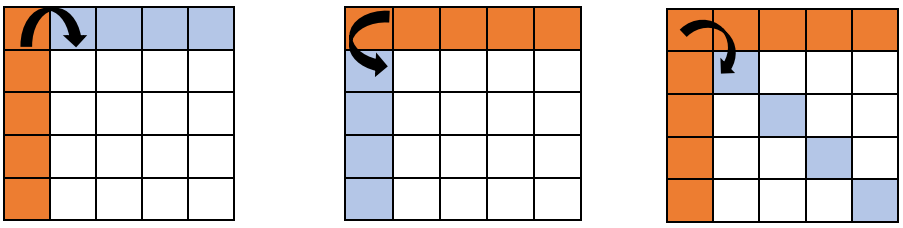


Figure 3.19: Disparity costs matrix for matching costs data storing.

Summarizing, finding the minimum path with a Dynamic Programming approach means iterate through the matrix left to right and take a rolling sum. The optimal cost of an element in the rolling sum vector, for the current pixel and disparity combination, is equal to the cost for the current location plus the lowest summed cost from the set of all possible disparities for the prior pixel locations. Simultaneously, the algorithm records the optimal indexes (i.e. the locations of the minimum values) and finally recovers the optimal route through the matrix that gives the minimum cost.

In this first SGM implementation the path directions are managed with an external algorithm cycle that takes one path at a time and performs the costs penalization. For each direction, first of all, the algorithm works by individuating the starting points from whom the next process of penalization should begin. For example, considering the horizontal and vertical directions, the algorithm will have to store the image (or mask if it has been defined) border pixels. From these, the penalization algorithm will start to perform the regularization step, paying attention to the component directions of the specified evaluated vector. A stylized starting points identification (in orange) is clarified

in Figure 3.20, where in blue are visualized the pixels exposed to the previously explained rolling sum procedure.



a) b) c)

Figure 3.20: In orange the identified starting points in the case of: a) Horizontal path direction; b) Vertical path direction; c) 45° path direction.

Finally, the information from all paths is fused for all pixels and disparities by eq. (3.14), and the optimal disparity for each pixel, which corresponds to the minimum cost, is determined.

The process produces a large amount of data that have to be stored. All costs for all pixels and all possible disparity values are stored in a big tensor that has a size of W * H * D, where W and H represents the image (or tile) resolution (W and H are equivalent to image/tile width and height) and D is the disparity search range. The high computation cost of the entire process, which in this implementation was not considered a key aspect, have made essential the investigation of a different data storage/access in order to obtain better performance for real case applications.

3.1.7.2 Memory efficient SGM implementation

The first tests performed on the initial implementation of the SGM algorithm showed that it's commonly useless to consider more than 16 path directions and the best performing penalization function is the one proposed by Hirschmuller, with two penalty factors that weights differently disparity differences equal to one pixel (i.e. object continuity of slanted surfaces) and higher differences (associated to object surface discontinuity). Limiting the context of the algorithm to such behaviour, some simplification can be provided, developing a much more efficient software code. In other words, in a second stage of the research, a new implementation of the matching algorithm was developed, obtaining a less general but much faster solving procedure. The drawback of the first SGM method is the temporary memory requirement that depends on the number of pixels (image resolution), the disparity range and the paths number. In fact, the base algorithm aggregates the disparity costs of each directions in

a data volume S_r, from the sum of which the final disparity is computed. This means that the costs have to be stored for each pixel and disparity value, and for each directions until the optimization process ends. In other words, the algorithm forces to memorize both the S total costs volume and one 3D costs matrix for each path direction. Moreover, the process allows to perform one direction at a time and then load the cost of the generic path inside S. Using this approach, the algorithm is not efficient because each pixel are evaluated N times (where N indicates the total number of paths) in the total matching process.

This paragraph describes a different SGM data storage method (similar to the one implemented in the OpenCV libraries [129]) where the matching costs penalization phase becomes faster and more memory efficient. The new algorithm, instead of storing the costs for all disparities, for each pixel and for N paths, performs the costs calculation and costs penalization steps simultaneously, considering two (or three if sixteen directions are considered) rows of image pixels at a time. The advantage of this approach is that the matching process doesn't need two data structure (the data volume for the generic path direction and data volume S from which the final disparity is selected) since it requires the memorization of just two (or three) image lines of pixels at a time: the present and the previous lines.

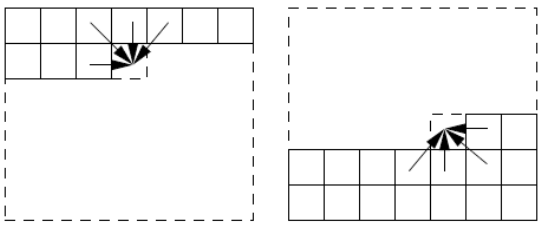


Figure 3.21: Calculation of the eight path directions in a top-down pass (on the left) and a bottom-up pass (on the right).

The algorithm follows partially the ideas described in [87] and shown in Figure 3.21: for each pixel, the costs of the previous pixel (on the basis of the considered path direction) are loaded. eq. (3.8) shows that, to compute the matching cost of one pixel, $L_r(\boldsymbol{p}-\boldsymbol{r},d)$ and $min(L_r(\boldsymbol{p}-\boldsymbol{r},d))$ should be considered beforehand: except for horizontal path direction, the previous line of pixels costs must be already computed. If $(\boldsymbol{p}-\boldsymbol{r})$ interests lines other than the previous one (i.e. diagonal direction that span 3 rows of pixels if sixteen directions are considered), also such lines must be memorized. The disadvantage of this method is that, being necessary the previous pixel costs data, it is not possible to perform all the optimization paths directions in one step unless they

all come from left or top of the current position. In all the other cases aggregation must be performed in two passes: the first pass goes from the top left pixel through the image and computes the path from left, diagonally from top-left, from top and diagonally from top-right. For each pixel, the four paths are continued from the previous pixels to the current pixel according to eq. (3.8) and as shown on the left in Figure 3.21. Thereafter, once the whole image has been analysed with the first pass, a second pass, that starts from the bottom, right pixel and goes upwards is performed in order to compute the remaining four paths.

Data are stored within an array which will contain the cost data of two row of image pixels (or three row if we consider eight directions instead of four) i.e. $\boldsymbol{L_r(p-r,d)}$ and $\boldsymbol{min(L_r(p-r,d))}$ for each pixel, direction and disparity value.

Figure 3.22 describes the data storage method: each $\mathbf{L_r}$ array contains, for each pixel, four memory locations (corresponding to the four evaluated paths); within them the cost for each disparity value under consideration is stored. At the boundaries of the image the method has no cost data, therefore, the costs information will be initialized to a *NaN* value (maintaining however the same data structure). Finally, the $\boldsymbol{min(L_r(p-r,d))}$ arrays are organized in the same way of the $\mathbf{L_r}$ arrays but contain less information since the minimum of the costs has just been calculated with respect to all the possible disparity values $\mathbf{d}$.

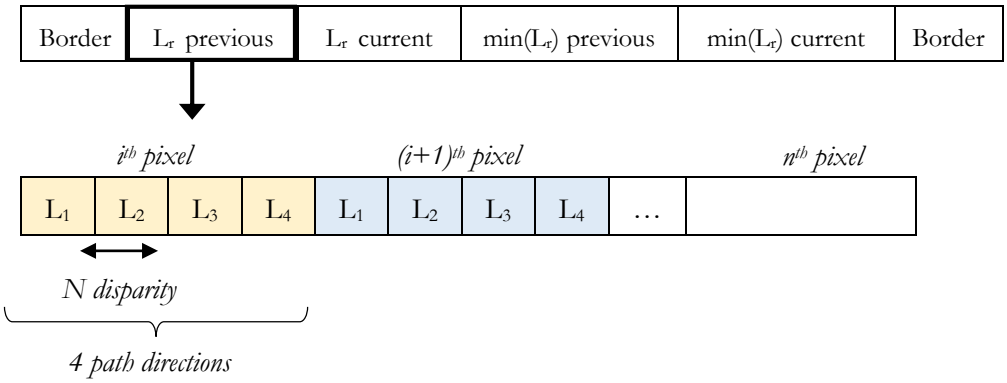


Figure 3.22: Data storage layout: for each pixel, four memory locations (corresponding to the four evaluated paths) are initialized and filled with the pixel costs for each disparity value under consideration.

The above illustrated data storage structure allows a much more efficient memory access, optimizing the processes and providing low computational time. In .Net environment, data structures are usually accessed in a "safe" behaviour, which means

that the compiler provide an internal system to avoid wrong memory access. The proposed data structure, however, allows to access the data using pointer algebra in "unsafe" conditions making the entire process extremely fast. Moreover, the proposed data structure allows to store the two lines of pixel data (the computing line and the previous) in a cyclic buffer: once the data of the previous image line are known, the second array is gradually filled; when this second image line is completely calculated the previous line data will be no longer useful, since its costs have just been transferred to the second image line (i.e. the "current" array), and can be overwritten. This means that, when the process switches to the third line, the pointer of the previous line becomes that of the second one (whose data must be used for updating the costs of the third) and the pointer of the computing line can point to the first line data (whose information are no longer needed), allowing to overwrite the data efficiently. After computing the four path-wise costs L_r, for the generic pixel and disparity, before shifting the pointers, the result is added to S_{tot}:

$$S_p(d) = L_1 + L_2 + L_3 + L_4 + S_p(d) \tag{3.15}$$

where $S_p(d)$ is the contribute that contains the data of the pixels calculated in the previous pass (i.e. the pass that goes from the top left pixel line of the image).

3.1.8 SGM algorithm extension for a bi-directional research of correspondences

The basic idea behind semi-global methods is to reduce the correlation problem complexity to a sub-set of image points (also called as "scanline"), reducing in this way the matching problem dimension.

The classical SGM approach is functional in stereo-vision problem and for photogrammetric applications such as the generation of digital surface models, where the images to be matched can be rectified (see images rectification in the paragraph 3.1.2), simplifying the problem of homologous points identification between stereo images. However, this represents also a strong limitation for traditional SGM implementation: there are a lot of applications that require a 2D disparity search and, moreover, real data always has imperfections and finding the true disparity should require searching in the Y direction even if an epipolar rectification process is performed. In other words, traditional SGM algorithms cannot be used when the displacements tracking of image points is investigated.

Given these considerations, and in order to create a software able to give a solution to many engineering applications, the proprietary SGM software code was improved, including the possibility to analyse a 2D disparity search domain.
The algorithm works within the same pipeline described in paragraph 3.1 but eq. (3.8) should be extended to consider the 2D search domain:

$$L'_{\boldsymbol{r}}(\boldsymbol{p},d) = C(\boldsymbol{p},dx,dy) + min_{i,j}\{L_{\boldsymbol{r}}(\boldsymbol{p}-\boldsymbol{r},dx+\Delta d_j,dy+\Delta d_i) + P(\Delta d_i,\Delta d_j)\} \quad (3.16)$$

where dx and dy, and Δd_i and Δd_j, are respectively the displacement components and the disparity changes along the principal image plane axis.
However, the high amount of memory and calculations required by the original SGM is further increased introducing the 2D disparity search option. It is easy to note that a 1D disparity search domain makes the complexity of the problem proportional to O(m · n · d), where m and n are the pixel resolution of the image and d is the disparity search range, while, in the novel approach, the complexity tends to O(m · n· d^2). In fact, the DSI matrix $C(\boldsymbol{p},d) = C(x,y,d)$ becomes a four-dimensional tensor $C(\boldsymbol{p},dx,dy) = C(x,y,dx,dy)$ where for each image pixel, the costs by varying the disparity search range in two direction are stored. Proportionally will also increase the dimension of the matrices that store the penalized costs and the optimal indices.
In particular, with the aim of solving these memory and computation issues, the implementation of the multi-resolution strategy (described in paragraph 3.1.5) was developed, in order to limit the disparity search range as much as possible.

3.1.9 Sub-pixel refinement

Most dense stereo correspondence algorithms (as the ones implemented in our SGM code) start by establishing discrete pixel matches (integer disparity map) and later refine these matches to sub-pixel precision, estimating a parabola fitting to the cost function. Once the analytical minimum cost is determined, the corresponding fractional offset is used to adjust the initial integer disparities (as described in paragraph 2.5).
This process is very simple and yet efficient, but many studies (such as [124][36][143] [172]) and tests reported that such strategy can lead to a systematic biasing towards integer values of displacements, also known as "pixel-locking (or peak-locking) effect". This bias produces random or coherent noise in the final reconstruction, introducing erroneous ripples or waves in the reconstruction of the disparity map.

As described in [172], considering a generic object and camera pose, the histogram of sub-pixel disparities should be approximately uniform (i.e. the fractional part of the disparity has a uniform probability distribution function); instead, using a simple parabola fitting over an initial integer disparity matrix, the histogram of sub-pixel disparities shows usually evident peaks (see for instance results in Figure 3.23).

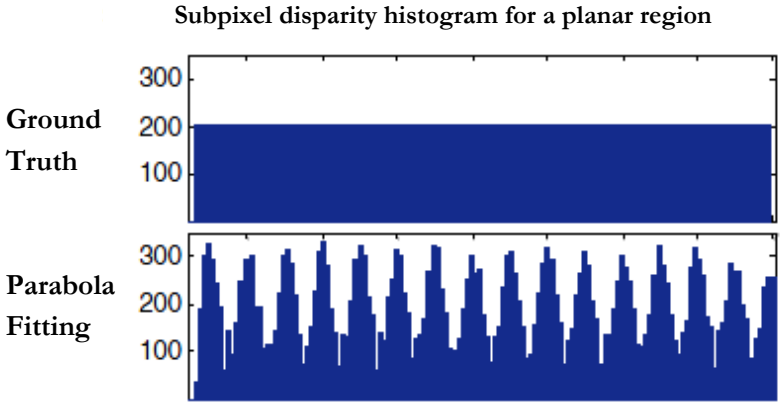


Figure 3.23: Example histograms of sub-pixel disparities for a planar region, illustrating the "pixel-locking" effect (based on [172] experiments). Using standard parabola fitting algorithms, peaks are clearly visible in the second histogram, while the ground truth is uniform.

In literature, several methods have been proposed to solve the pixel-locking problem. To avoid bias, and preserve details, Nehab et. al. [124] suggested a symmetric sub-pixel refinement strategy that improves both the reference and the matching image coordinates simultaneously, in a symmetric way, by looking for the minimum of the matching cost function along a direction that is insensitive to its confidence variations. Differently, Stein et. al. [172] approach refines initial integer disparities by using local estimation techniques based on the classical optical flow or the Lucas-Kanade (and Tomasi) tracker ([108][183]).

3.1.9.1 Symmetric sub-pixel refinement implementation

Since strong pixel locking effects occurred in most of the case studies performed to calibrate the algorithm, Nehab strategy [124] has been studied and later implemented in the proprietary SGM software code. This sub-pixel refinement approach can be used instead of the traditional, and primarily implemented, sub-pixel parabola fitting interpolation method. According to Nehab et al., the peak locking bias can be avoided not considering only one image as reference, i.e. refining both the matched images

coordinates simultaneously, in a symmetric way. They propose a symmetric approach in which both images are treated equal by working in image coordinates rather than in the disparity space. Looking for the minimum of the matching cost function along a direction that is insensitive to its confidence variations, neither camera is considered as reference, thus the refined matches coordinates will have subpixel precision in both the images. In other words, the fundamental idea is to find the subpixel refinement by considering a 2-D neighbourhood of the matching cost around each matching pair and finding the optimal cut in the direction of the symmetric lines of the matching cost. Consider two rectified cameras C_1 and C_2, producing images I_1 and I_2 of an object, such that the scan-lines in each image are corresponding epipolar lines, and defined the metric cost for comparing pixel values as M, we can define

$$F_y(x_1, x_2) = M(I_1(x_1, y), I_2(x_2, y)) \tag{3.17}$$

so that, instead of work in the disparity space, it is possible to work directly with image coordinates.

With the aim of capture the symmetry of the matching cost function (Figure 3.24 from [124] shows the matching ridge and how the cost functions are symmetric with regard to it), a 2D neighbourhood of matching cost values around (i_1, i_2) is considered, allowing to define a continuous surface $S(x_1, x_2)$ (see Figure 3.25). The surface can be well-represented by a Gaussian Cylinder (GC) representation (as described in [124]):

$$S(x_1, x_2) = G(D(x_1, x_2)) \tag{3.18}$$

$$G(d) = ae^{-d^2} + b \tag{3.19}$$

$$D(x_1, x_2) = s_1 x_1 + s_2 x_2 - p \tag{3.20}$$

since this surface enforces a ridge-like shape for the reconstruction. The parameters a, b, s_1, s_2, p can be determined by a non-linear least squares minimization system (composed by the derivatives of $G(d)$ in a, b, s_1, s_2, p) on the 3×3 neighbourhood around (i_1, i_2).Once the GC has been estimated the correct coordinate match pair (x_1, x_2) for each pixel can be located directly on the matching ridge.

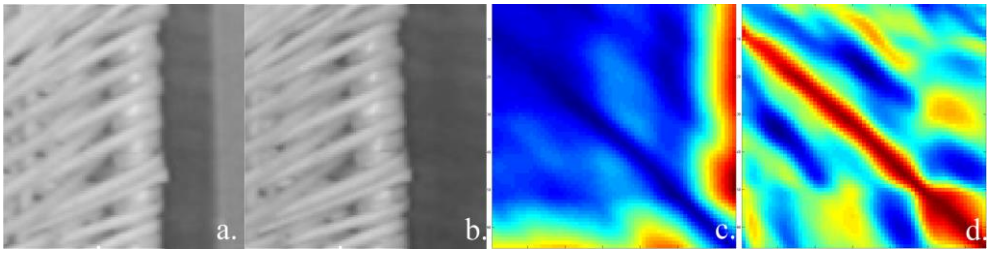


Figure 3.24: Examples of matching cost functions values and their symmetry w.r.t. the matching ridge. a. Master image; b. Slave image; c. SAD; d. NCC.

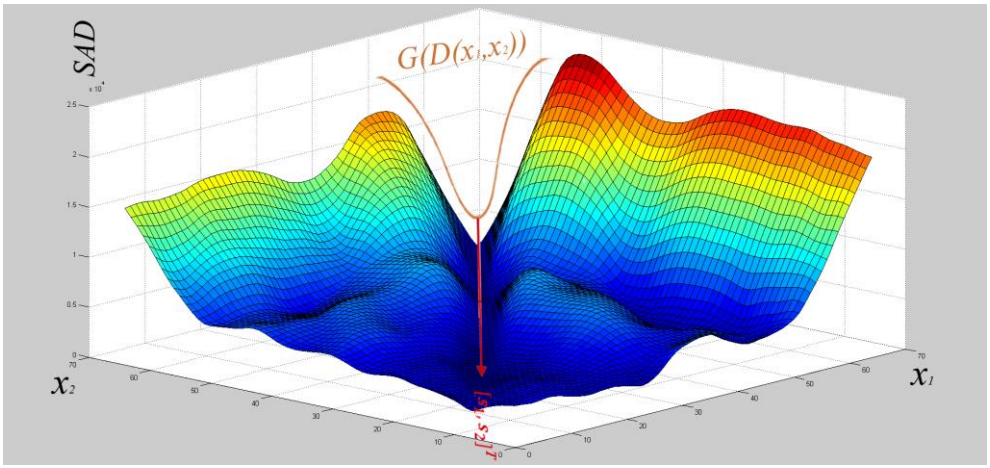


Figure 3.25: The continuous surface $S(x_1, x_2)$.

A good estimate for a cut through S is represented by the direction $[s_1 s_2]^T$ which follows the symmetric lines of the applied matching cost function: the cut direction can be directly identified. Then, the line $D(x_1, x_2) = 0$ will give the examined local approximation for the matching ridge. The system minimization, in our implementation is performed with Levenberg-Marquardt method [118], as suggested by Lourakis [107], which has demonstrated to be rather well robust.

3.1.10 Multi-Image SGM extensions

Originally developed for image pairs correlation, the proposed Semi-Global Matching algorithm has been also extended to cope with Multi-Image configurations. Indeed, although many applications do not use multi-images data, several tasks relies on accurate, reliable and complete 3D reconstruction of the scene. For these cases, stereo dense image matching processes have been adapted to perform the correlation process in multi-image conditions. These methods are called multi-baseline matching or multi-view stereo algorithms.

As described in paragraph 2.6, most of the implementations perform the image matching by analysing consecutive image pairs and combining the pairwise results afterwards, in order to generate the complete final solution of the matching process. However, according to [17], this methods work well, but often shows several disadvantages connected for example to points visibility problems between the stereo images, that can lead to a less accuracy and completeness of the final disparity solution and consequent 3D reconstruction. To test both methodologies and highlight their performances, a pairwise-based multi-image SGM implementation and an alternative Object Space based approach have been implemented. The latter differs from the original SGM method in two major aspects: firstly, the cost calculation is formulated in object space (instead of in image space) within a dense voxel raster, and uses the radiometric values of all images concurrently, instead of pairwise cost data. Similarly, the semi-global minimization process is performed into object space as well, since the result of the semi-global optimization directly represents the possible 3D positions of each points. Thus, the entire process leads to the generation of height-maps (instead of disparity maps) which indicate the 3D positions of the best matches.

3.1.10.1 Pairwise-based Multi-Image SGM implementation

The first algorithmic development in multi-image matching process has regarded the implementation of a multi-baseline stereo method for producing dense depth map by performing pairwise stereo matching on all image pairs and finally merging the results. The code initially identifies all the possible pairs of stereo images that build up the photogrammetric sequence and then performs the image correlation of each pair using epipolar rectified images.
Considering *n* pairs of stereo images, the correlation process produces *n* depth maps that are individually re-projected, in our implementation to a common image plane. This operation is necessary for combining the *n* depth maps into a unique reference frame. Once the *n* depth maps have been merged, a median filter is used to filter out the noise: for each pixel of the overlapping depth maps, the depth values are sorted and the median is selected.
In paragraph 4.3 the results of the application of this second multi-image algorithmic development are presented.

3.1.10.2 Multi-Image SGM algorithm in object space

Object-based Semi-Global matching works by subdividing the X-Y-Z object space into a raster volume subdivided into voxels (as described in Figure 3.26). Each voxel has precise dimension (ΔX, ΔY, ΔZ) which define the resolution of the object space (i.e.,

voxel size defines the resolution and level of detail of the final results). A good evaluation of the object space resolution can consider for example the mean Ground Sampling Distance (GSD) value, in order to ensure an adequate sampling rate of the space.

Traditional SGM method (see eq. (2.24) and (3.8)), compute the matching costs building up the structure $C = C(p,d) = C(x,y,d)$, in which x,y refer to the image pixels coordinates and d is the disparity value. Implementing a costs aggregation/penalization strategy, which search the minimum path costs, an optimal matching solution is computed. In the case of object-based SGM, the costs matrix $C(x,y,d)$ is modified to $C(X,Y,Z)$, in which the coordinates X,Y,Z indicate the 3D position of a voxel and the disparity search range D is replaced by a Z search range.

Once the user has defined the voxels size, the re-projection of each ground point (i.e. the centre of the voxel) into all the images is performed, by using the collinearity equations. Since windows-based cost functions are used, the matching window (defined parallel to X/Y plane) will be re-projected and then resampled into all the images. This allows to obtain a template within each image, from which the radiometric information will be stored and used for cost calculation (see Figure 3.26).

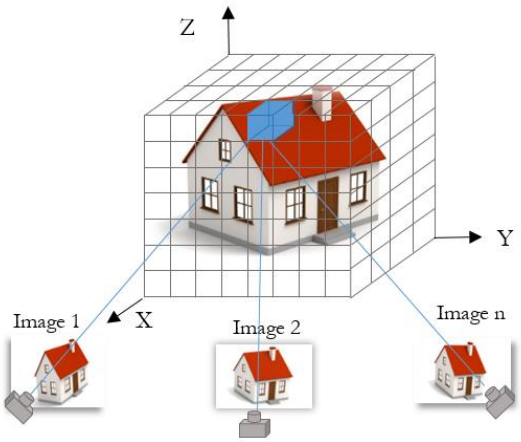


Figure 3.26: Calculation of matching costs for a voxel.

However, it is interesting to notice that, in order to perform a multi-image correlation process, it becomes necessary to investigate a method for extending the similarity functions and combining matching costs. According to [17], different solutions can solve this matter: the simplest way is represented by the cost calculation for all possible image pairs and the parallel initialization of the $C(X,Y,Z)$ structure where the cost value will be stored. The calculated value can be the minimum costs with respect to the costs

extracted with the correlation templates on the n images or, for example, the mean/median value of all these.
Finally, the global energy function, expressed in eq. (2.21), is modified to

$$E(Z) = \sum_{X,Y} C(X,Y,Z) + \\ + \sum_{q \in N_p} P_1 \cdot T\left[\left|Z - Z_q\right| = \Delta Z\right] + \sum_{q \in N_p} P_2 \cdot T\left[\left|Z - Z_q\right| > \Delta Z\right] \tag{3.22}$$

where the first term contains the matching costs for each voxel of the raster and the second and third term add penalties P_1 and P_2 in case of differences in Z-direction between adjacent voxels. Therefore, in the case of object-based SGM, is it possible to say that the smoothness constraints have the task to perform the "regularization" of the object Z-coordinates, imposing the continuity, at least along some specific plan direction of the object surface. Obviously, the energy minimization process must be transformed as well:

$$\begin{aligned} L_r(v,Z) = C(v,Z) + \min(&L_r(v-r,Z), \\ &L_r(v-r,Z-\Delta Z) + P_1, \\ &L_r(v-r,Z+\Delta Z) + P_1, \\ &min_i L_r\big((v-r,i\cdot\Delta Z) + P_2\big) \\ &-min_k L_r(v-r,k\cdot\Delta Z) \end{aligned} \tag{3.23}$$

where v represent the voxel (x,y) location.
The path-wise cost aggregation process can be done in analogy to the standard SGM separately for every path L_r (see eq.(2.21)) but, in this case, the minimum path costs will include added penalties P_1 and P_2 at the position of the previous voxel in path direction (instead of the previous pixel) and adds this minimum to the cost value $C(X,Y,Z)$ of the current voxel. The penalty P_1 is added if the difference in Z-direction between the current voxel and the adjacent voxel is equal to ΔZ (which is the height of one voxel) and P_2 is added if the difference in Z-direction is larger than ΔZ. The last term of eq. (3.23) subtracts the minimum path cost of the previous voxel to avoid very large values in L_r. Finally, as in the original SGM implementation, the results of the costs penalization through the *n* paths are fused into the matrix S and the optimal final Z-coordinates will be derived from $S(v,Z)$ by searching the minimum of S for each voxel v:

$$min_z S(v,Z) = min_z \sum_{r=1}^{n} L_r(v,Z) \tag{3.24}$$

The final Z values will be stored, for each v, in a 2D height.

3.2 Algorithm calibration

In order to get a better understanding of the relationships between all the algorithms variables (penalty functions, similarity cost functions, template size, etc.) a calibration of the algorithm's processing parameters was performed.
Several images datasets were used for studying the algorithm performances:

- Synthetics images: image rendering of simple three-dimensional computer generated objects, created in a 3D modelling software and draped with a texture (data are shown in Figure 3.27 (a –b – c - d)). For each dataset, images has been taken with different baselines, exposure and perspective effects and the matching performance has been evaluated by comparing the reconstructed parallax map with the original 3D model, considered as "ground-truth".
- Middlebury Stereo Evaluation Dataset: high-resolution scientific image dataset for stereo and multi-image matching applications (in particular, the 2014 high-resolution datasets has been used). As described in [161], the authors have created a structured lighting system for generating high-resolution stereo datasets of static indoor scenes with highly accurate ground-truth disparities (one case of study are visible in Figure 3.27 (e)). Each dataset consists of input images taken under multiple exposures and multiple ambient illuminations. They also provide each dataset with both "perfect" and realistic "imperfect" rectification, with accurate 1D and 2D floating-point disparities, respectively.
- CvLab dataset [173]: high-quality datasets consisting of distortion removed images, known interior orientation parameters and reference laser scanning DSMs. In particular, the Fountain- P11dataset has been used (see Figure 3.27 (f)). The performed tests will be presented in this case in the following Chapter 4, which considers the three-dimensional reconstruction algorithm capabilities. In fact, the image disparities comparison is not possible since the parallax fields are not directly provided: therefore, the influence of the matching process on the final 3D reconstruction (i.e. DSMs comparison) has been investigated.

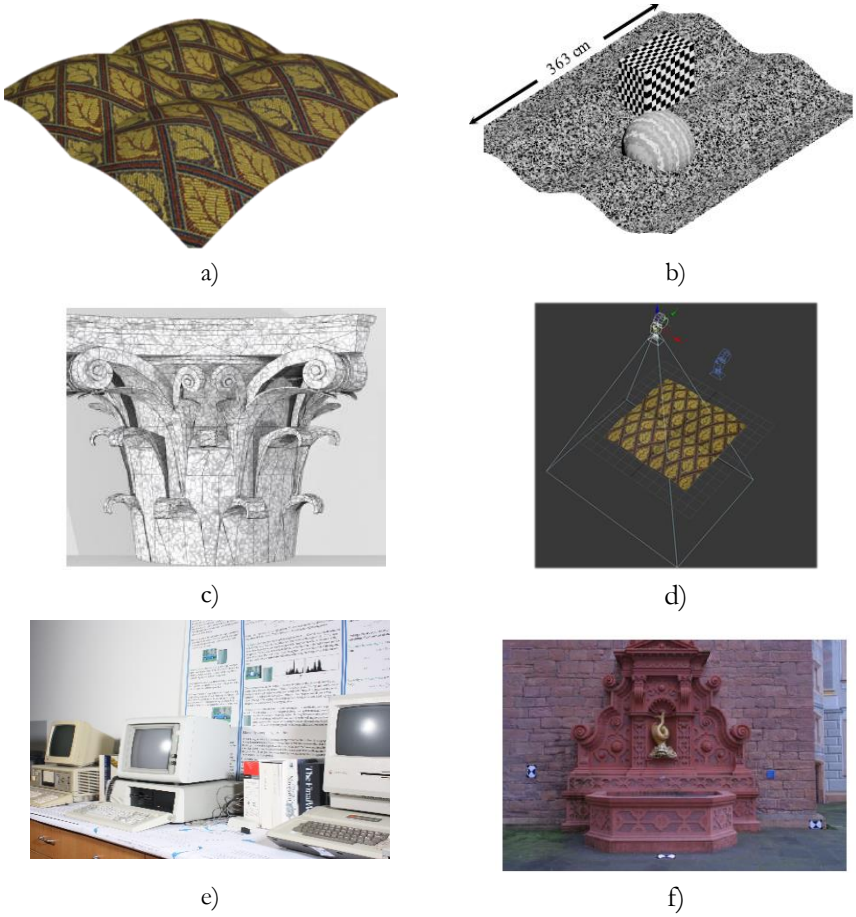


Figure 3.27. a)-b) Simple computer generated objects used as calibration dataset. c) Synthetic image of a 3D reference model. d) e) Example of Middlebury dataset. Real image of the P11-Fountain dataset.

The performed calibration tests have regarded in particular the following aspects:

- the algorithm accuracy: the SGM algorithm metric accuracy and reliability has been evaluated by comparing, for each case study, the reconstructed parallax map with the provided ground-truth disparities data (i.e. evaluating the algorithm accuracy in finding the corresponding point on image space). Information about the results quality has been acquired through the use of suitable statistical indicators (such as for example the mean, median and standard deviation values of the differences between the reconstructed and the ground-truth disparity maps). As far as the three-dimensional reconstructed models are concerned, accuracy evaluation has been done by computing the distances between the reconstructed and the ground truth DSMs;

- the results completeness and level of detail: the distribution of the reconstructed parallax values has allowed to evaluate the percentage of pixels whose errors are smaller than an established value: thus highlighting the results reliability in terms of "inlier and outliers percentage";
- the computational time of the whole matching process: since the computational efficiency of a matching algorithm is becoming a more and more influential factor, both SGM implementations (described in paragraph 3.1.7) are tested for analysing the time of the correlation process, in particular the time for computing the matching core step.

Many are the SGM processing variables that considerably intervene in the metrological algorithm characterisation; namely, the correlation window size, the regularization and similarity cost functions, the penalty values analysis and identification, the use of multi-resolution and image tiling approaches, etc. Therefore, the following paragraphs will describe how the algorithm performances (in terms of results accuracy, completeness and time of calculation) are influenced by the above parameters. Several parametric combinations have been tested, in order to analyse the process weaknesses and to highlight its performances. In the following, a summary of the more interesting results is presented (with the aim of identifying the best strategies and parameters combination that allows the most efficient description of different object typologies).

3.2.1 Algorithm accuracy evaluation

In this first stage, the simple computer generated object shown in Figure 3.27 (a), was used as calibration dataset and tests have been computed considering the SAD cost function and full-resolution images (i.e. no image pyramid and tiling processes). The achieved results will be following presented.

First, the calibration of the implemented two penalization methods and their regularization factors (e.g. P_1 and P_2 for Hirschmuller formulation) has been performed. In order to identify the best penalty values for allowing the correct application of semi global constraints, the accuracy of the method was tested by varying different combinations of penalty values (also with respect to the template size). The results are shown in the diagram in Figure 3.28, which has allowed the identification of the best values range that produces stable results.

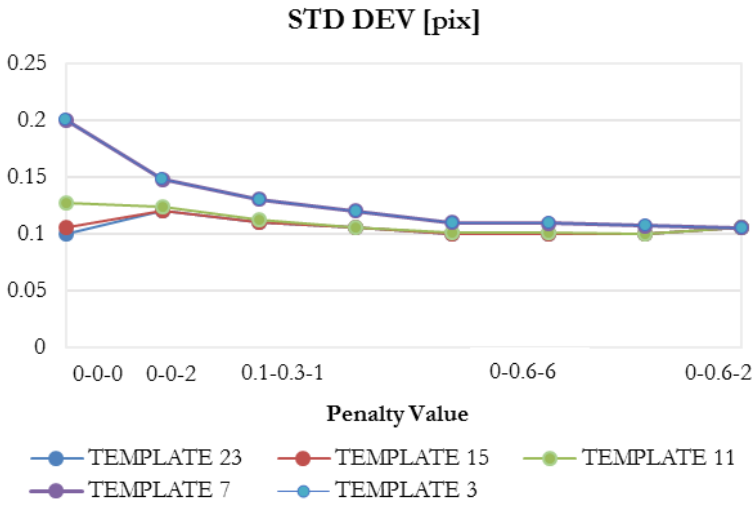


Figure 3.28: Penalty values calibration

The diagram above shows that the differences on varying the penalty values combination are light: standard deviation values vary in the range of 0.1 pixel. However, if small template size are used (e.g. 3 or 7 pixels window dimension) it is necessary to pay more attention on the penalty values to use in the correlation process. At the same time the results have allowed to assess a penalty value range that has been kept fixed, in the next tests, limiting the number of varying parameters.

As described in paragraph 3.1.6, two different penalization function has been developed in the proprietary software: one based on Hirschmuller formulation (which considers two penalty factors P_1 and P_2) and another one that linearly increase the penalty with the increase of the disparity changes between adjacent pixels. The two methods performances were tested and the results of the comparison are presented in the histogram of Figure 3.29: the accuracy of the two penalty implementations is practically the same for each examined template size. Thus, in all subsequent algorithm tests, the Hirschmuller defined penalty function was used.

The next testing stage considered the influence of the similarity functions on the final accuracy. For each function also the influence of the template size have been considered. The ideal block size to perform the stereo matching depends on the chosen function: the tests have been conducted using the same synthetic dataset which, not presenting discontinuities, allows to better analyse the influence of the template size on the final parallax accuracy: with bigger template sizes, in fact, object shape discontinuity are likely to produce gross errors.

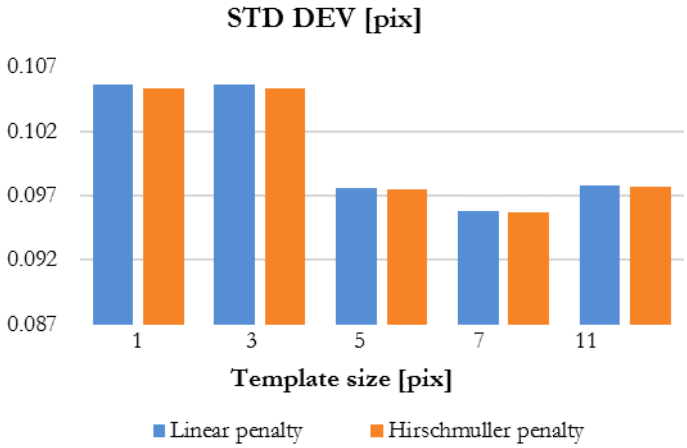


Figure 3.29: Comparison of the two implemented penalization costs functions accuracy.

In other words, in this initial stage we have focused the attention on the relationship between the template size and the cost function, not introducing more variability connected to the complexity of the analysed object shape. In this context, in analogy with other techniques for DTM generation in close range (e.g. LSM), there seems to be an optimal range for template size value according to object features. An extended evaluation is investigated in [146] and the diagram shown below (Figure 3.30) allows to observe the variations of the standard deviation w.r.t. the template size.

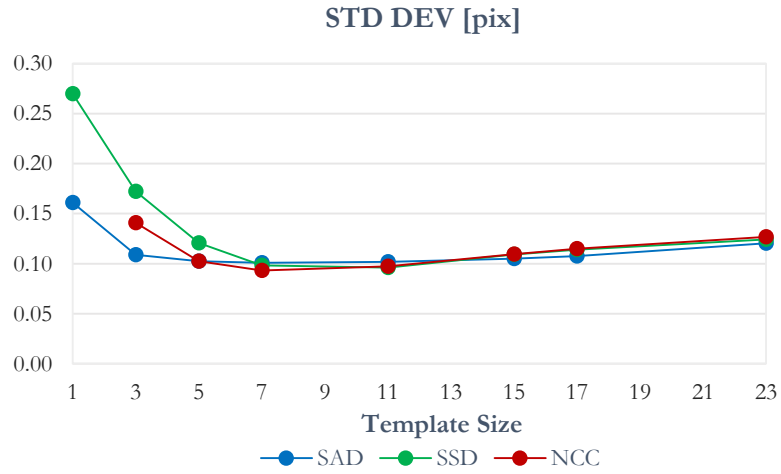


Figure 3.30: Accuracy of the similarity functions by varying the window template size.

In the diagram, the performances of three different cost functions changing the template windows size, can be observed. As expected, the enforcement of the disparity

continuity allows to obtain comparable results between the evaluated cost functions, showing a light improvements of the accuracy with the increase of the template size. Looking at the diagram, there seems to be an optimal windows size beyond which the cost functions seems to produce less accurate and reliable results. However, it's worth noting that the maximum differences in the registered standard deviation values are slightly more than one tenth of pixel. and the use of NCC with small block size has produced isolate outlier which were removed from the estimation.
Finally, the accuracy of the second memory efficient SGM algorithmic implementations has been investigated w.r.t. the first implemented method. The following diagram shown in Figure 3.31 presents an evaluation of each solution accuracy.

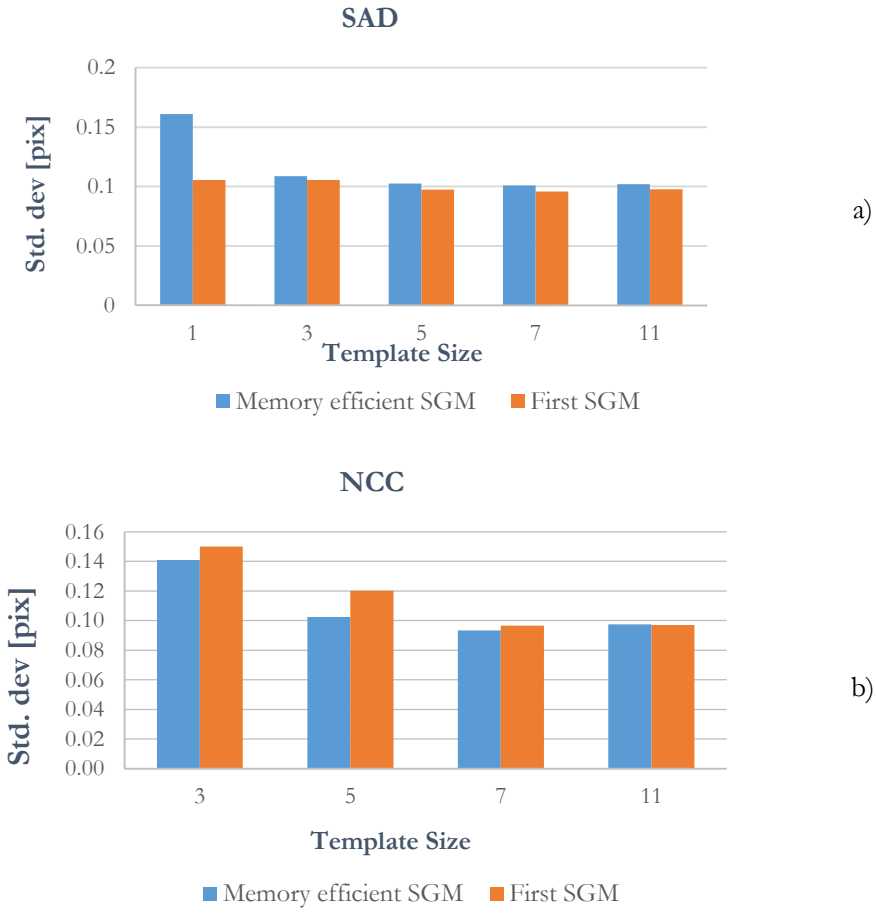


Figure 3.31: Comparison between the accuracy of the two implementation of SGM: the first implementation and the memory efficient variation. a) The results obtained using the SAD cost function; b) The results obtained using the NCC cost function.

The results show again the improvement of the accuracy with the increase of the window size: this is evident in the NCC behaviour while the same cannot be said for the SAD cost function, which maintains constant performances for template size greater than 3 x 3 pixels. As regard the accuracies comparison between the two strategies, the novel more efficient implementation seems to be as accurate as the first SGM implementation. Some lightly differences are just evident in particular for small template size: the memory efficient method seems to provide better performances with NCC w.r.t. to the first implementation, but with SAD cost function the same is not true. That's probably due to the fact that SAD with small template sizes relies heavily on a good disparity regularization: the first implementation, although much slower than the new one, performs a more rigorous penalization stage.

3.2.2 Algorithm completeness evaluation

A second calibration stage has regarded the study of the algorithm capability of obtaining complete results. The algorithm has been tested, in this context, on the Middlebury datasets. In fact, the synthetic images datasets were not appropriate for this type of examination since the computer generated object and the relative image renderings were generated considering optimal conditions for the image matching process (e.g. simple objects with no discontinuities, well-contrasted images texture and good illumination). Consequently, in all the synthetic datasets, the level of completeness of the reconstructed parallax is usually almost 100%. On the contrary, the Middlebury stereo datasets show characteristics that will likely cause problems at the matching algorithm in terms of reconstruction abilities.

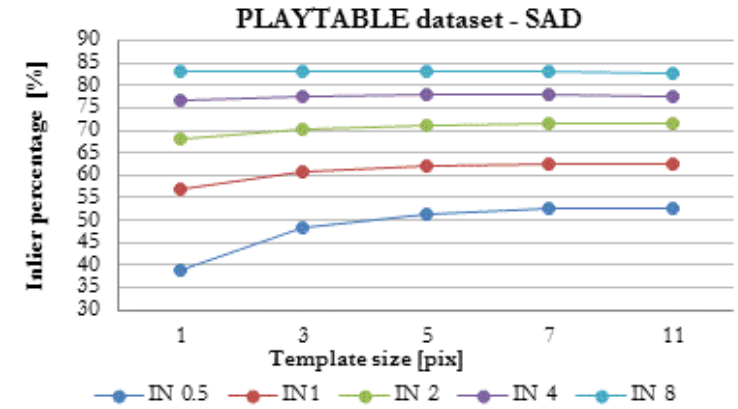
Figure 3.32: The two Middlebury datasets used for the algorithm calibration: on the left the "Playtable" case, on the right the "Vintage" case.

Tests were performed on the two dataset shown in Figure 3.32, using two different similarity functions, no image pyramiding and tiling, and the Hirschmuller penalization method with fixed penalty factors (the optimal set of parameters were previously

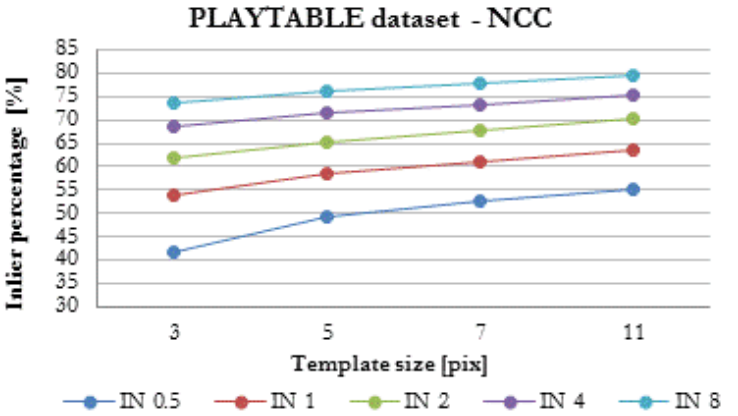
identified with the tests shown in the previous paragraph). The diagrams shown in Figure 3.33 and 3.34 present, for each case of study, the results of the algorithm completeness of reconstruction capability. The distribution of the parallax errors (evaluated with SAD, NCC, Rank and Census costs functions) has been reconstructed by highlighting the percentage of "inlier" pixels whose errors are lower than 0.5, 1, 2, 4 and 8 pixels respectively. Higher percentage (values) represent thus better results. As far as the two non-parametric Rank and Census transformation are concerned, images have been transformed applying the Rank metric on a 9 x 9 pixels template size and, after that, the correlation has been performed with NCC and SAD cost functions. On the other hand, for Census, the transformation has been performed using a 5 x 5 pixels window size and the matching process has been performed on the Census-transformed images with SHD metric.

The two evaluated cases of study have shown similar results which are presented in Figure 3.33(a) – (b) and Figure 3.34. Both for the SAD and NCC cost functions, the parallax completeness increase with the use of big template size but, the SAD similarity function shows better results since the inlier percentage values, for each considered inlier threshold, are greater than the NCC results. It is interesting to notice that, with the increase of the window size, NCC data tend more and more towards high accuracies, thus there is a continually improvement of the cost function performances (this is even more evident looking at the NCC results for the "Vintage" case (Figure 3.35).

Differently, SAD completeness levels becomes constant for template greater than 5 pixel. The NCC behaviour can be justified considering the peculiarity of the Middlebury images. In fact, in presence of low contrasted areas, repetitive patterns and constant brightness regions the NCC statistical indicator does not have enough information, with small templates, to produce reliable results. On the contrary, the formulation of the SAD cost function allows performing a better investigation of corresponding luminosity areas (even if low or no-texture are present), while it produces less reliable and accurate results in areas with high contrast and luminosity differences between the images. However, with the use of medium-large template size, the NCC performance can improve significantly thanks to the increase of the radiometric information within the correlation window.



a)



b)

Figure 3.33: Results of the "Playtable" dataset: Percentage distribution of the inlier parallax values whose errors are lesser than 0.5, 1, 2, 4 and 8 pixels respectively, by using the SAD (a) and NCC (b) cost functions.

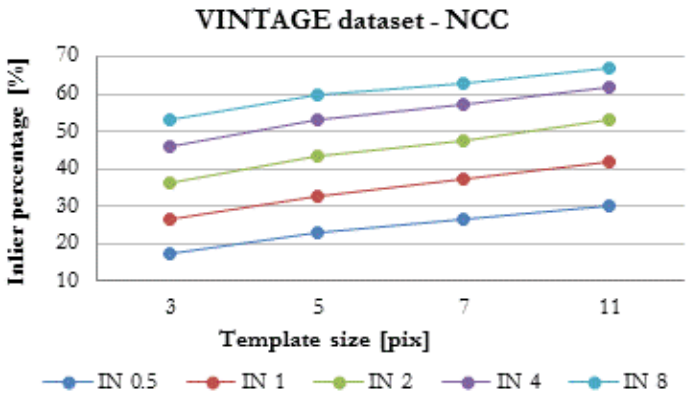
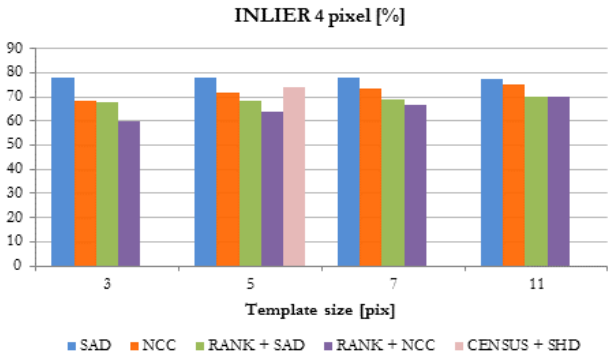
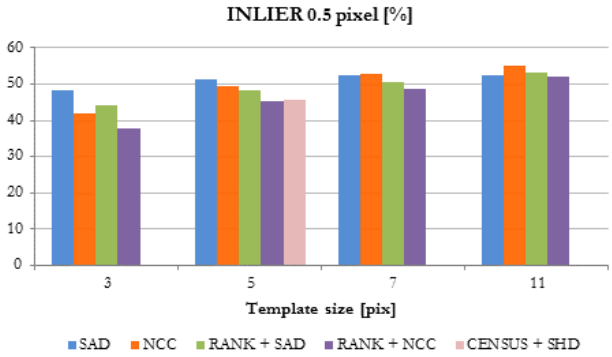


Figure 3.34: Results of the "Vintage" dataset: Percentage distribution of the inlier parallax values whose errors are lesser than 0.5, 1, 2, 4 and 8 pixels respectively, by using the NCC.

A more direct comparison between all the evaluated cost functions results , with regard to the 4 and 0.5 pixels "inlier" percentage (that threshold is considered as a good tolerance ranges for the completeness evaluation with these datasets) is shown in the histograms in Figure 3.35 and 3.36.



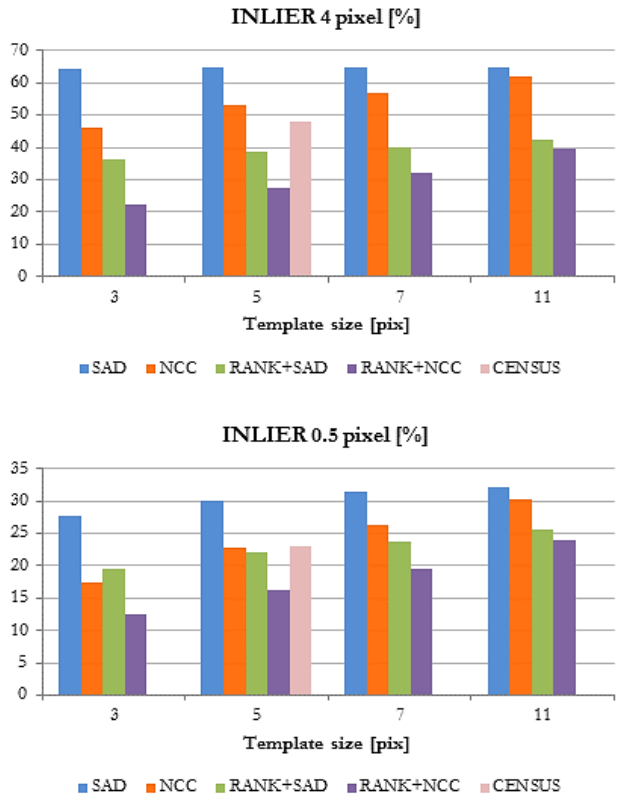
a)



b)

Figure 3.35: Results of the "Playtable" dataset. Comparison between the NCC, SAD, Rank and Census cost functions results, with regard to a) the 4 pixels "inlier" percentage; b) the 0.5 pixels "inlier" percentage.

Looking at the histograms in Figure 3.35 (a)-(b), the worse results are achieved by the Rank cost functions. Anyway, the results trend is similar to the previous SAD and NCC ones, since the completeness levels increase along with the template window size. On the other hand, the Census behaviour shows good results, which are slightly worse than the SAD cost function (that remains the "winner metric") but significantly better that NCC solution with regard to the 4 pixel inliers percentage (Figure 3.35 (a)). The metric shows thus a good robustness with respect to outliers identification but, at the same time, it proves to be less accurate (see Figure 3.35 (b)) than the other similarity measures. Following, the "Vintage" study case results, although very similar to the "Playtable", are illustrated.



a)

b)

Figure 3.36: Results of the "**Vintage**" dataset. Comparison between the NCC and SAD cost functions results, with regard to: a) the 4 pixels "inlier" percentage; b) the 0.5 pixels "inlier" percentage.

The completeness increases with the use of big template size but with lowest improvement degree than in "Playtable" (as already explained and shown in Figure 3.35). As regard the cost functions performances, the results improvement using the SAD metric is even more evident (than in the first dataset), both in the case of 4 and 0.5 pixels inliers percentages (see the histograms presented in Figure 3.36).

Differently from the first case results, the noticeably gap between the cost functions results is evident for each template size under evaluation and, at the same time, the depicted completeness levels are significantly lower. The images "Vintage" dataset has shown thus worsen radiometric characteristic for the matching algorithm capabilities, than those characterizing the first case. However, as it will be presented in 3.2.7 paragraph, the proprietary semi-global algorithm has reached up in this case comparable performance with respect to other semi-global strategies.

In order to better understand the algorithm robustness, the maps of the differences between ground truth and the reconstructed disparity maps (obtained using a correlation window size equal to 5 x 5 pixels and the SAD cost function) is presented. As visible in Figure 3.37, the parallax map suffers of lacks of completeness in correspondence of low textured regions, repetitive pattern and reflective materials.

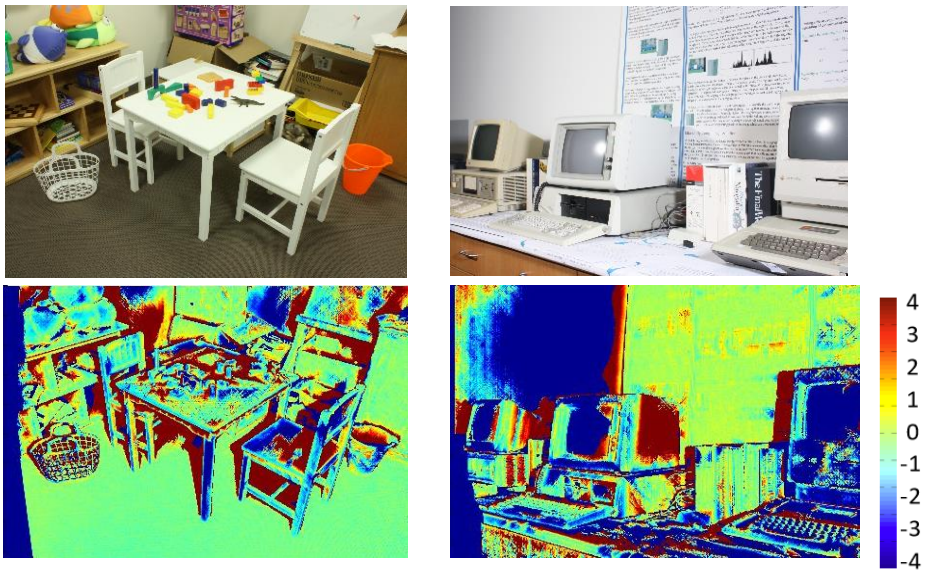


Figure 3.37: Maps of the differences between ground truth and the reconstructed disparity map considering a tolerance range of 4 pixel. On the left the "Playtable" case, on the right the "Vintage" case.

This occurs especially when the algorithm tries to find matching points on the white rooms walls, on the plastic surfaces of the workstation and the chair, which appear in the "Vintage" and "Playtable" cases respectively, and near the shiny terminals monitors of the "Vintage" case.
In paragraph 3.2.7 additional results on the Middlebury datasets are shown together with a comparison between the proprietary algorithm performance and other two SGM implementations. Moreover, further investigations about the algorithm reconstruction capabilities will be examined in Chapter 4.

3.2.3 Algorithm computational time evaluation

In this paragraph the process efficiency is considered. In particular, the second SGM implementation, which was developed for obtaining a much faster disparity penalization

procedure, has been tested and compared with the first algorithmic implementation. It is important in this analysis to distinguish between the three stages that characterize the matching solving process, namely: the costs computation, the disparity penalization and optimization steps. This proves to be important since the novel more efficient SGM strategy modifies, with significantly lower computational load, the cost data penalization processing stage. Differently, the matching costs analysis and the disparity minima search steps are unchanged between the two SGM approaches.
In the previous diagrams shown in Figure 3.31, the novel improved SGM algorithm has proved to be as accurate as the previous one. Figure 3.38 shows the time (in seconds) to compute the matching cost process for a single image pair, employing NCC and SAD similarity functions with different template sizes.

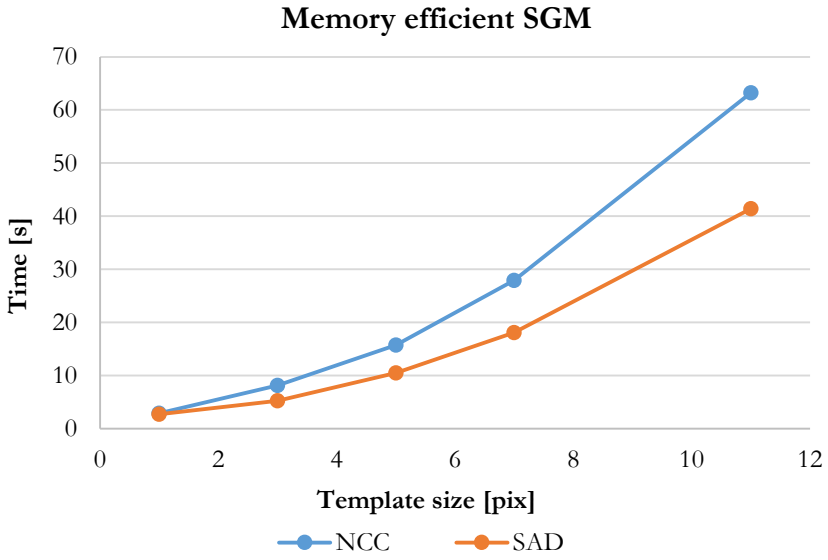


Figure 3.38: Computational time comparison of the costs computation process, by using NCC and SAD cost functions with different template size.

The computation time, in both cases, increase quadratically with the increase of the template size. Furthermore, in the histograms presented in Figure 3.39, an evaluation of the novel memory efficient SGM algorithm improvements (in terms of computational time) with respect to the oldest one is presented. The diagrams show a noteworthy enhancement of the algorithm efficiency.

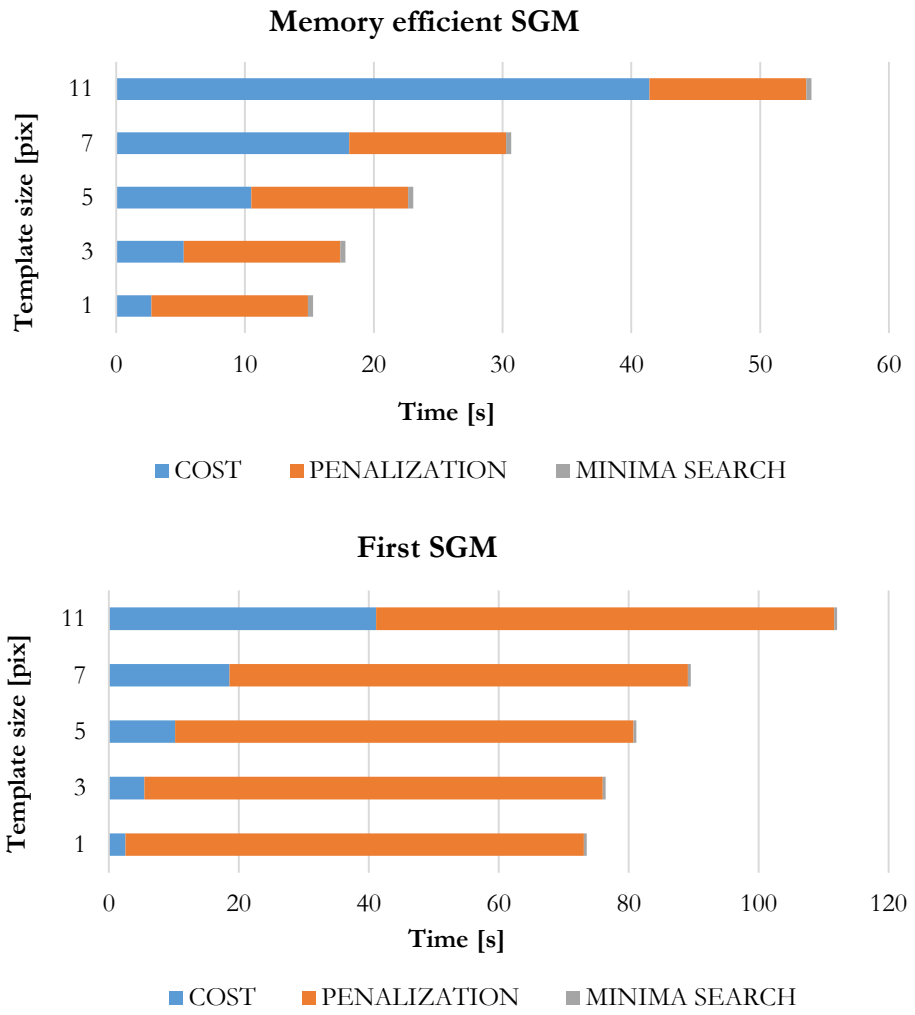


Figure 3.39: Computational time (in seconds) to process a single image pair, employing SAD similarity functions with different template sizes: a comparison between the novel memory efficient SGM algorithm and the first SGM implementation.

As previously proved in the diagram in Figure 3.38, the matching cost stage increase quadratically with the template size both for the two SGM approaches but it is evident the improvement of the novel strategy with respect to the old one in the costs penalization (about 12 seconds instead of the 70 seconds of the oldest method). It is also important noting that, in both cases, the disparities penalization and optimization steps (depicted in orange and grey respectively in the histograms) require the same computational time for each considered template, since the efficiency of these process do not depend on the size of the correlation window. At the same time, changing the

disparity search range, the total time for performing the penalization step is obviously higher with wider ranges, but the gain w.r.t. the old implementation is basically constant as shown in Figure 3.40.

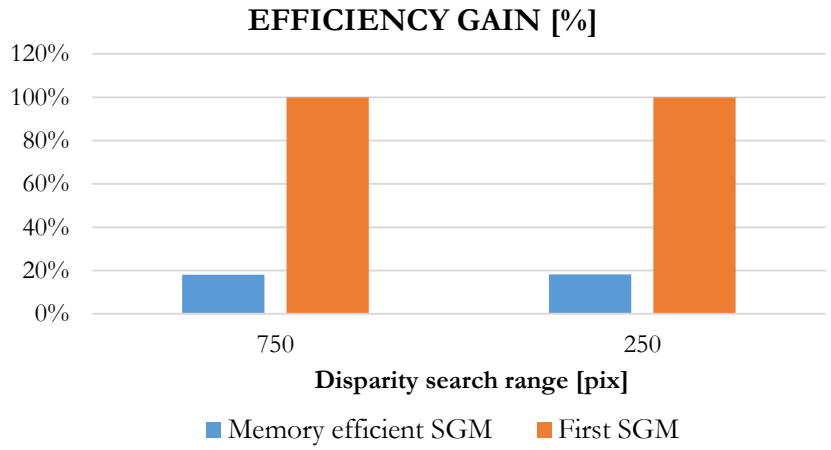


Figure 3.40: Penalization step efficiency gain of the memory efficient SGM approach w.r.t. the old implementation.

Concluding, both implementations produce accurate results, showing comparable deviation data when the comparison between the ground truth and the reconstructed parallax map is considered. For this reason, the novel SGM implementation seems to be more convenient, being much faster and more memory efficient.

3.2.4 Evaluation of the sub-pixel refinement accuracy

In all the previous examples, both the implementations have shown a random or coherent noise in the final reconstructed scenes. Such behaviour, which produces less accurate correspondences, is called "pixel locking". The following paragraph analyses the problem and showing its influence on the obtained results and proposing, finally, a method to avoid biasing and limit its influence.

The simple computer generated object used as calibration dataset, has allowed the clear identification of pixel locking phenomena. Indeed, comparing the "true" disparity map (or ground truth) with those obtained by the method using the three implemented cost functions, a sort of staircase pattern has been observed (see Figure 3.41).

The occurrence of peak locking becomes more visible in a histogram of the individual components of the displacement; another way to determine the degree of peak locking is to consider the histogram of the fractional part of the displacement in pixel units (i.e.

only the fractional part between –0.5 and +0.5 pixel units). An example of such fractional histogram is shown in in Figure 3.41.

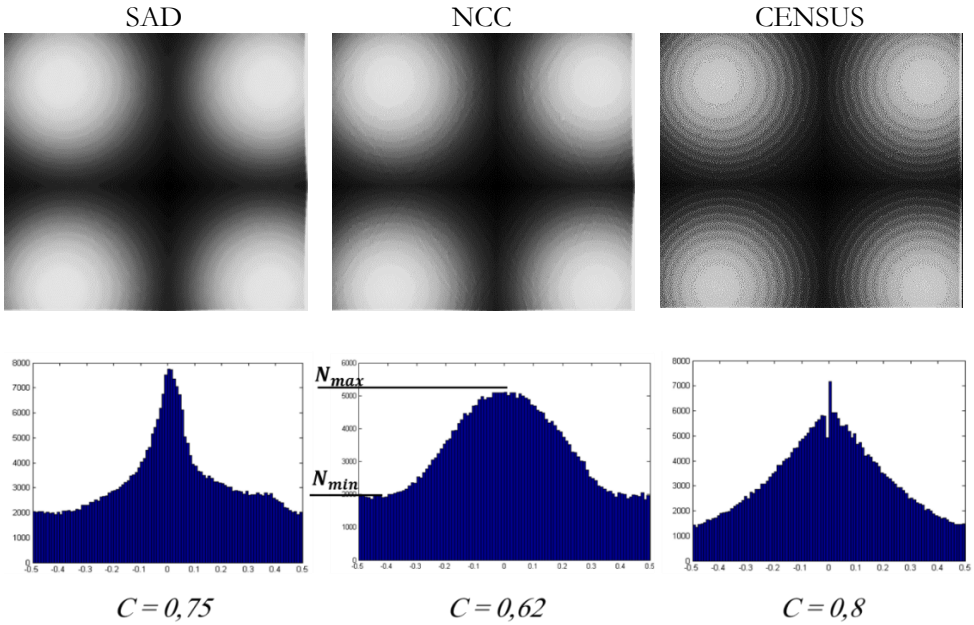


Figure 3.41. Top: Disparity maps of the computer-generated object, obtained with three different implemented cost functions. Bottom: Pixel-locking degree quantifications of each evaluated cost function (in pixel units).

The histograms of the fractional part of the displacement show clearly the pixel locking effect: it is possible to identify, for each case study, a high elevation near zero displacement and a depression near ±0.5 pixels displacement. Therefore, through the histograms and the parameter extracted from eq. (3.25), the identification of the results biasing, towards integer values of displacement, has been possible, showing a "severe" degree of the phenomena for all the presented case of study shown in Figure 3.41.
The degree of pixel locking can be quantified, as is described in [129], as:

$$C = 1 - \frac{N_{min}}{N_{max}} \tag{3.25}$$

where N_{min} and N_{max} are the lowest and highest number of counts in the fractional histogram (as it is illustrated in Figure 3.41). $C = 0$ indicates complete absence of pixel locking, while $C = 1$ indicates very strong pixel locking; Figure 3.42 (extracted from [128]) gives a useful indication of the fractional histogram shapes considering several

degree of pixel locking. Four different levels of pixel locking can be distinguished: (i) for C < 0.2 virtually no pixel locking occurs; (ii) 'mild' pixel locking occurs for 0.2 < C < 0.4, while (iii) 'strong' pixel locking occurs for 0.4 < C < 0.6, and (iv) C > 0.6 indicates 'severe' pixel locking.

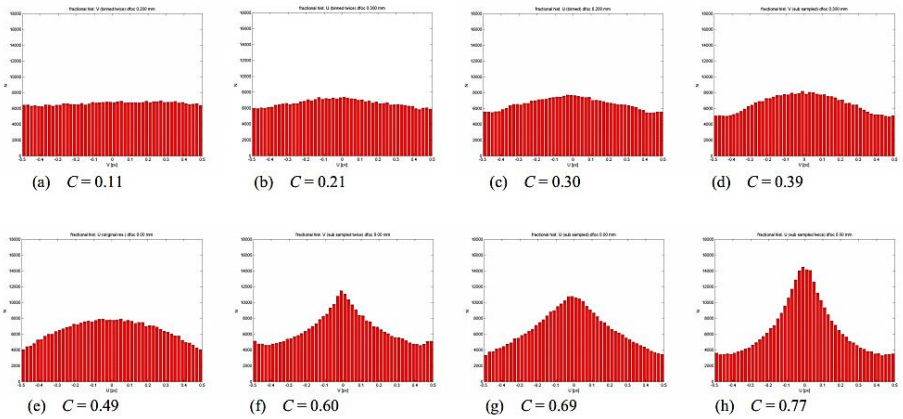


Figure 3.42: Fractional displacement histograms for various levels of the pixel locking parameter C defined in eq. (3.25).

Usually pixel locking is associated to numerical behaviour of the sub-pixel disparity refinement strategy at the end of the matching procedure.

For this reason, we chose to implement a different refinement strategy where the disparities sub-pixel investigation is performed not with the traditional parabola fitting method (see paragraph 3.1.9 for more details) but with the improved "sub-pixel symmetric refinement" method, as explained in the work of Nehab in [124]. Following, the result obtained with the novel refinement strategy is shown (Figure 3.43), considering the matching procedure applied on the same synthetic dataset and with the SAD similarity function.

The diagram shows the improvement of the results: the parallax map obtained with the new strategy is smoother and, concurrently, the staircase pattern (very evident in the case of the parallax maps obtained with parabola fitting method) is less marked; the histogram displays a more irregular distribution of the displacements fractional components, resulting in the improvement of the pixel locking degree index. Indeed, while C=0.75 can be quantified (according to [129]) as "severe" pixel locking, a degree of 0.46 reveals that the result is lightly affected by the locking effect (C=0.4 is classified as "mild" degree of pixel locking).

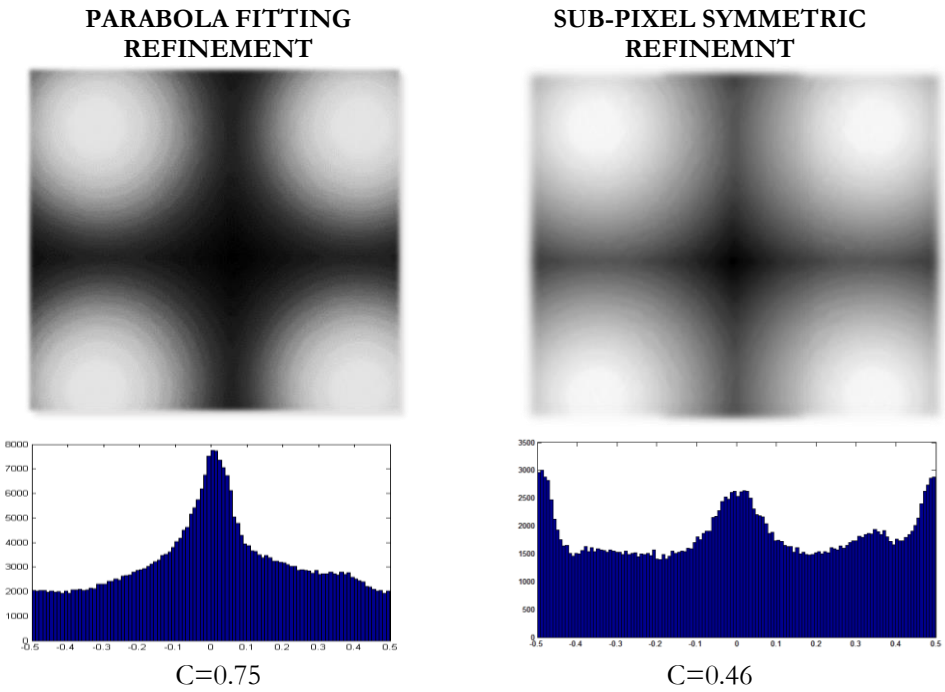


Figure 3.43: The parallax map and the histogram of the fractional part of the displacements: on the left the results obtained with parabola fitting sub-pixel refinement, on the right the results obtained with sub-pixel symmetric refinement method.

3.2.5 Tiling strategy performances evaluation

A final calibration stage has regarded the evaluation of the accuracy and reliability of the image pyramiding and tiling strategies, employed at the start of the matching workflow for improving the algorithm computational performance.

When high resolution images are used, the program can subdivide the master image into small, user defined size, tiles. In order to verify the procedure accuracy and reliability, the Middlebury datasets were processed both with and without the tiling strategy application and the obtained results were compared. As an example, Table 3.1 and Figure 3.44 show the results obtained for the "Playtable" dataset.

As expected, even if the tiling process allows reducing the computational efforts of the matching procedure, enabling the simultaneous processing of each tile in parallel, the final quality of the results is a little lower. This is due to the overlap size between adjacent tiles: for well textured scenes an overlap of 20-40 pixels, like the one used for the tests, is sufficient.

	WITH tile strategy	WITHOUT tile strategy
Median [pix]	0.537	0.479
INLIER 0.5 [%]	48.96	50.75
INLIER 1 [%]	58.26	61.24
INLIER 2 [%]	66.59	70.36
INLIER 4 [%]	73.25	76.86

Table 3.1: Distribution of the "inlier" pixels percentage by using the tiling strategy.

However, low-textured area or high depth differences in the scene require a higher overlap. The reason of this lies in the fact that the matching solution of the pixels within the overlapped areas can be not well-estimated since the matched points, within these areas, can have not according disparity values in the two different tiles, and the blending procedure can produce inaccurate results. Indeed, tiles borders (as well as image borders) represent the image regions where the semi-global matching process obtains the lower precision results, because lower information are coming from the neighbouring pixels.

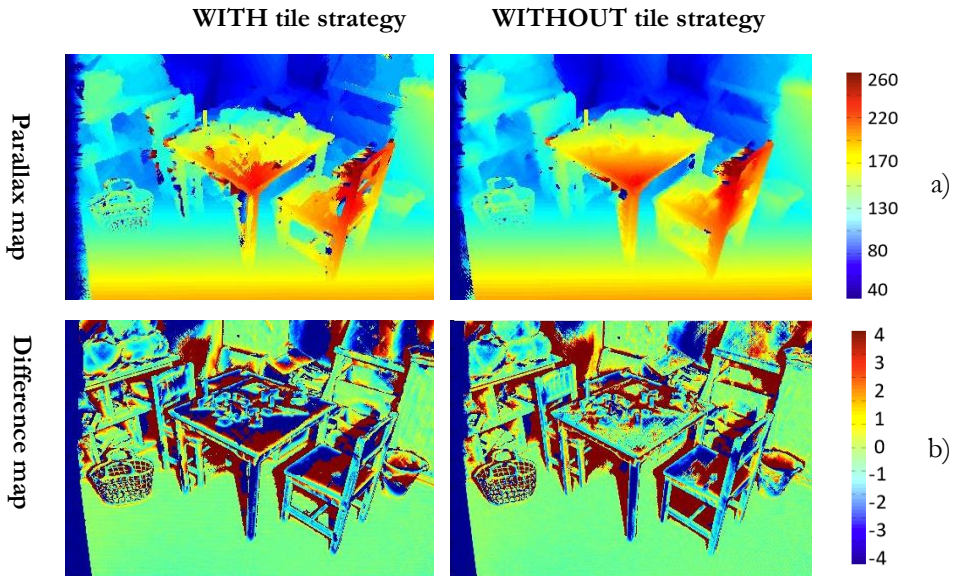


Figure 3.44: Comparison of the results obtained with the SGM algorithm with and without the use of the image tiling procedure. a) The obtained parallax maps. b) Maps of the differences of the reconstructed disparity maps and the reference disparity ground truth.

Therefore, scanline optimization procedure can generate erroneous or inaccurate correspondences data in these regions: if we consider that tiles matching solutions are

fused at the end of the whole process, the errors and the imprecisions at the boundaries of each solutions are added together, worsening the total accuracy of the results (Table 3.1) and increasing data noise. A good habit to follow may be to use not too small tiles, in order to reduce as much as possible the overlapping areas.

3.2.6 Multi-resolution approach (image pyramids) performances evaluation

The matching process using image pyramids is an efficient system for improving the algorithm performances in term of computation effort and memory requirements: as in the previous method, it is necessary to verify the method quality and reliability w.r.t. the single pyramid level approach. Thus, the computer generated object created for testing the semi-global proprietary method (see Figure 3.28 (a)) and the "Vintage" Middlebury image dataset (see Figure 3.28 (e)) have been used to compare the results of the algorithm by processing in one case the full-resolution image pair, in the other using the pyramid strategy (three image pyramids levels have been used). Tests were carried out to consider also the template size influence.
The accuracy of matching process considering one-level and three-levels image pyramids, is described in Figure 3.45 and Table 3.2, which show the increase of the standard deviation values of the differences between the ground truth disparities data and obtained final disparity map with the increase of the correlation window size. This is particularly true for medium-large template sizes. Indeed, the discrepancies between the two evaluated strategies are absent or quite low for small windows dimension (see the results for template sizes of three, five and seven pixels).

	Template size [pix]	**3 LEVELS**	**1 LEVEL**
STD DEV [pix]	*3 x 3*	0.109	0.109
	5 x 5	0.103	0.103
	7 x 7	0.101	0.101
	9 x 9	0.102	0.101
	11 x 11	0.104	0.102
	15 x 15	0.108	0.105

Table 3.2: Accuracy and efficiency evaluation of the multi-resolution approach by varying the template size.

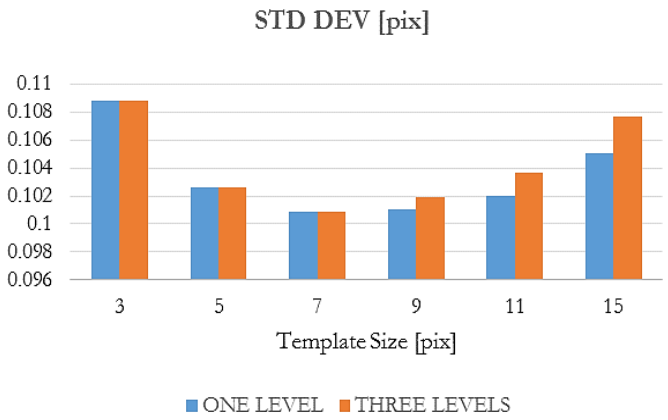


Figure 3.45: Comparison of the results obtained considering the image matching process with and without the application of the multi-resolution approach: evaluation of the accuracies by varying the window template size.

The algorithm behaves differently in the case of templates bigger than seven pixels. The use of big template produces the flattening of the images disparities data (the same can be said in terms of depth information), reducing the algorithm capabilities in reconstructing object fine structure and details and increasing the probability of passing wrong information from one level to the next. Furthermore, together with this issue, the image interpolation process intervenes: once the initial low-resolution parallax map has been obtained, a bilinear interpolation of parallax values between matched conjugate points is computed and utilized as approximate parallax map useful to perform the matching process with the new higher resolution image pairs. Thus, the inaccuracies introduced by the computation of image costs with big template size, together with the image pyramids resampling process, introduce errors that accumulate until the method reaches the original image level, producing lower accurate results.

As regard the computational efficiency improvements as concerned, the histogram presented in Figure 3.46 shows the significant gain in terms of process time using the multi-resolution approach.

It is also important considering that the image-pyramids employment offers considerable advantages (in process efficiency) the wider is the disparity search range. Figure 3.47 shows clearly the profits, in term of process efficiency, connected to the use of the multi-resolution strategy, using different disparity search ranges and different template sizes. In the case of a 250 and 750 pixels disparity ranges, the multi-resolution method is respectively 10 and 20÷25 times faster than the full-resolution approach applied in the same process conditions. At the same time, considering the window size

variation, the results highlight that the gain, in terms of process efficiency, lightly increases with the template dimension.

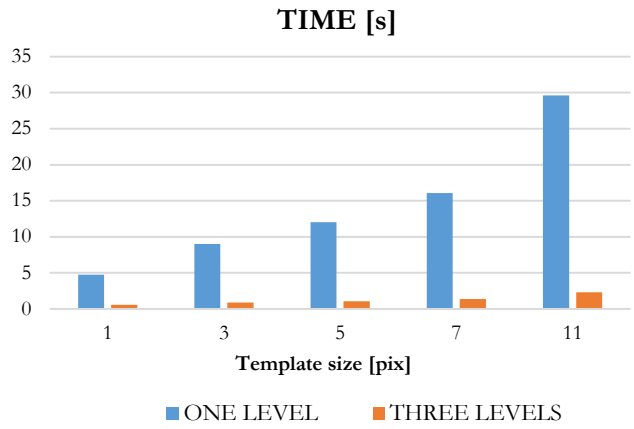


Figure 3.46: Comparison of the results obtained considering the image matching process with and without the application of the multi-resolution approach: evaluation of the computational time by varying the window template size.

It is possible to say that the use of image pyramids has undoubtedly a key role in correlation processes of high-resolution images (such as, for example, the planetary ones) where high disparity search ranges are needed. Moreover, this strategy has proved to be an essential step of the 2D SGM method workflow because of the quadratic increase of the image points search domain size.

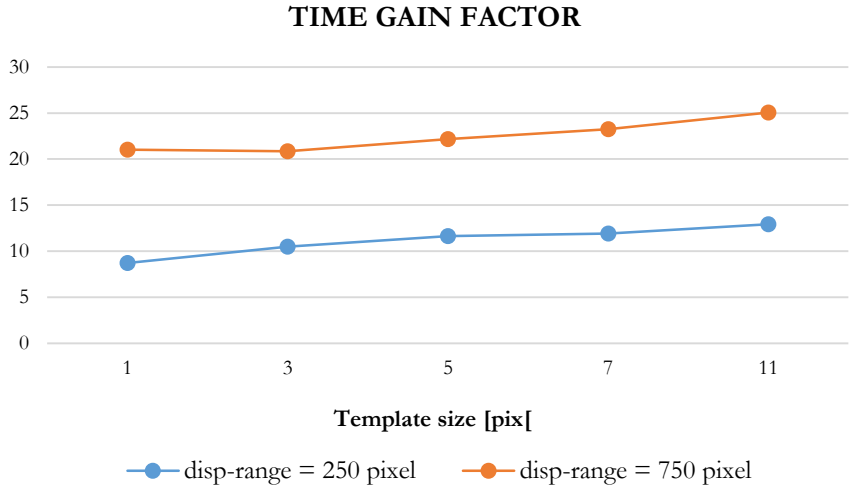


Figure 3.47: Computational time gain of the multi-resolution approach by varying the disparity search range and the template size.

3.2.7 Comparison with other SGM algorithm

A final validation stage has regarded the comparison of the proprietary semi-global algorithm performance with other SGM implementations considering some of the datasets extracted from the Middlebury Stereo Datasets.

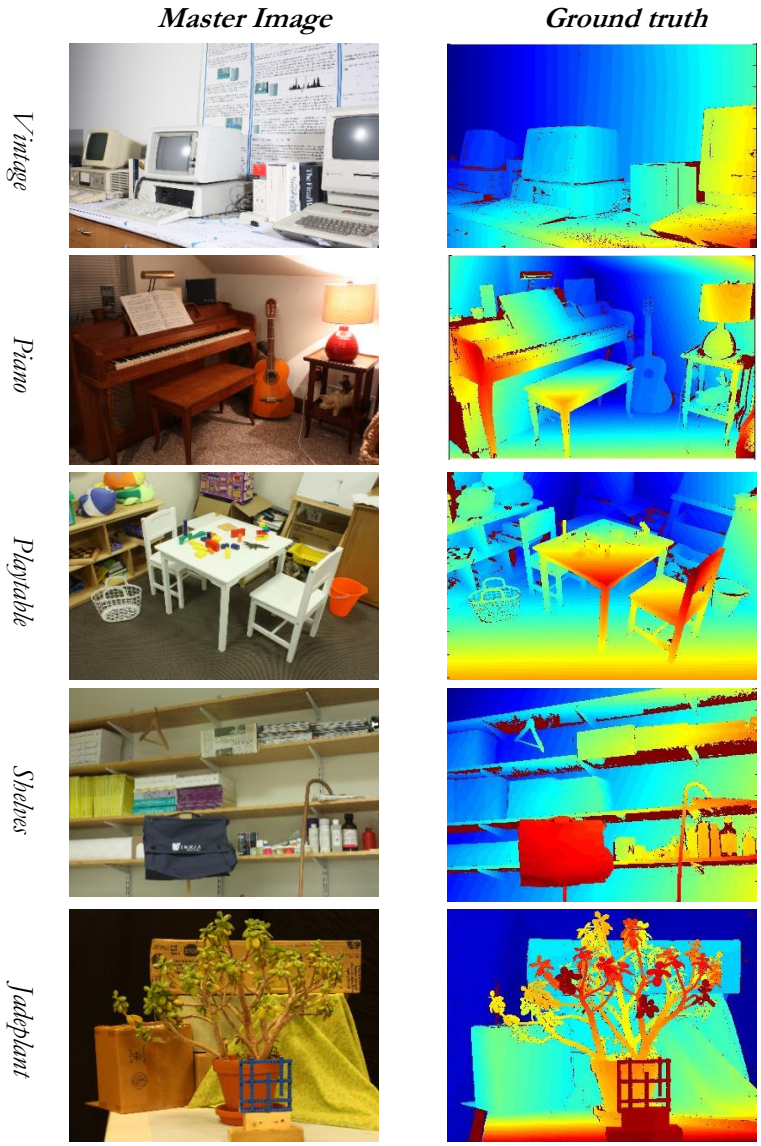


Figure 3.48: The evaluated Middlebury datasets with the ground truth disparities data.

In particular, the two compared SGM strategies are the Hirschmuller SGM algorithm [84] and the OpenCV Reimplementation of the latter [128].
Tests were performed with equal processing variables and input data, taking advantage of the considerations derived from the analysis of the previous calibration stages results:

- full resolution images;
- no multi-resolution processing (images were considered with their initial resolution and no image pyramids were created);
- size of the correlation window equal to 5 x 5 pixels;
- images similarity cost performed with the SAD cost function.

The considered Middlebury datasets are shown in Figure 3.48: different environments, illumination conditions, objects shapes and surface materials are represented in the stereo image pairs. Each dataset presents several difficulties that can affect the performances of the matching algorithms: the presence of image backgrounds characterized by constant radiometric contents, localized lighting sources that can create reflection phenomena on the objects surface (in particular for high-reflecting materials), occlusion problems and repetitive patterns.
The obtained results are presented in Table 3.3.

		Vintage	**Piano**	**Playtable**	**Shelves**	**Jadeplant**
MEDIAN [pix]	***SGM***	1.70	1.03	0.48	2.89	1.43
	SGBM [128]	2.37	0.68	1.96	2.64	0.98
	SGM [84]	2.53	0.56	5.12	2.12	0.74
BAD 0.5 [%]	***SGM***	69.8	63.8	49.2	77.2	67.4
	SGBM	72.7	59.1	72.9	73.1	60.7
	SGM	73.5	53.8	82.2	74.3	55.2
BAD 1 [%]	***SGM***	57.6	50.4	38.8	63.6	54.6
	SGBM	61.1	40.1	59.7	60.6	49.8
	SGM	62.3	33.8	71.4	60.7	42.2
BAD 2 [%]	***SGM***	47.6	42.0	29.6	53.8	46.6
	SGBM	52.2	31.1	49.7	52.8	42.4
	SGM	43.2	26.9	41.5	45.9	36.9
BAD 4 [%]	***SGM***	37.6	35.8	23.1	46.6	41.6
	SGBM	53.0	23.8	61.3	50.8	33.7
	SGM	44.3	18.2	52.7	40.7	28.7

Table 3.3: Comparison of the "bad" pixels percentage distribution between the proprietary SGM code, the Hirschmuller algorithm and the OpenCV variation.

For each case study, the distribution of the parallax values has been computed, highlighting the percentage of "bad" pixels whose errors are greater than 0.5, 1, 2 and 4 pixels respectively; the median value (50% error quantile) of each distribution is also calculated. Lower percentage (values) represent thus better results. The resulting parameters were compared with those uploaded online by the other authors that have implemented a Semi-Global matching method.

Table 3.3 shows the results obtained with the proprietary implemented **SGM** code and those obtained with other SGM versions uploaded on the Middlebury Evaluation page. Data demonstrate coherent results: in particular, two of the five cases of study show that the better solution was obtained with the proprietary semi-global algorithm; on the contrary, the remaining three cases show slightly worse results. Except in the "Playtable" case, where a great improvement of our result is evident, the parallax distributions obtained with the three different semi-global algorithms show similar behaviours on average. Nevertheless, it is important to notice that the proprietary software code does not implement pre or post-processing steps (such as hole-fitting, interpolation, median filtering, etc.) on the disparities data. On the contrary, both the compared strategies perform different post-processing methods to refine the disparity image and remove errors. In Hirschmuller SGM [84] disparity refinements strategies include peak filtering, intensity consistent disparity selection and discontinuity preserving interpolation processes; in OpenCV's "semi-global block matching" method [128], post-filtering and hole-filling methods are computed to obtain dense and smooth disparity image. It is therefore possible to say that the results of the comparison are very satisfying. In Figure 3.49 the reconstructed parallax map and the map of the differences, with regard to the ground truth disparities map, are shown for each considered dataset. Examining the parallax and difference maps, the main sources of error are represented especially by low textured areas and depth changes in the scene. The low radiometric content does not help the discontinuities and occlusions identification. For example, some mismatches are evident in "Vintage" case, at workstations boundaries and in correspondence of the terminals screen; also, the background on the right of the scene appears inaccurate, due to the black uniform area on the right side of the wall. The same stand for the other datasets: flattened colours of backgrounds and walls, reflective material surfaces, source of illumination and occlusions represent the major sources of errors and mismatches.

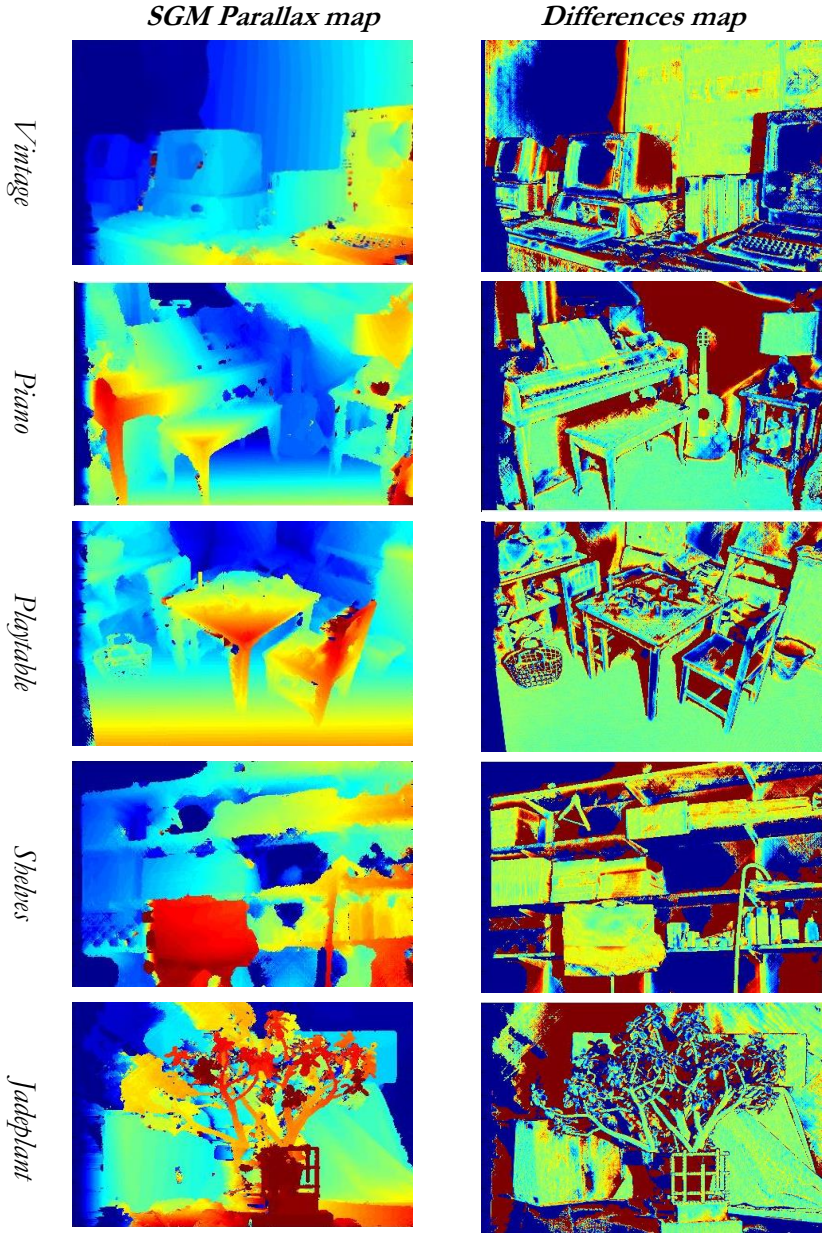


Figure 3.49: The reconstructed parallax map and the map of the differences with regard to the disparities ground truth considering a tolerance range of 4 pixel.

Comparing the obtained parallax maps with those performed with Hirschmuller and OpenCV semi-global strategies (shown in Figure 3.50), some considerations can be made.

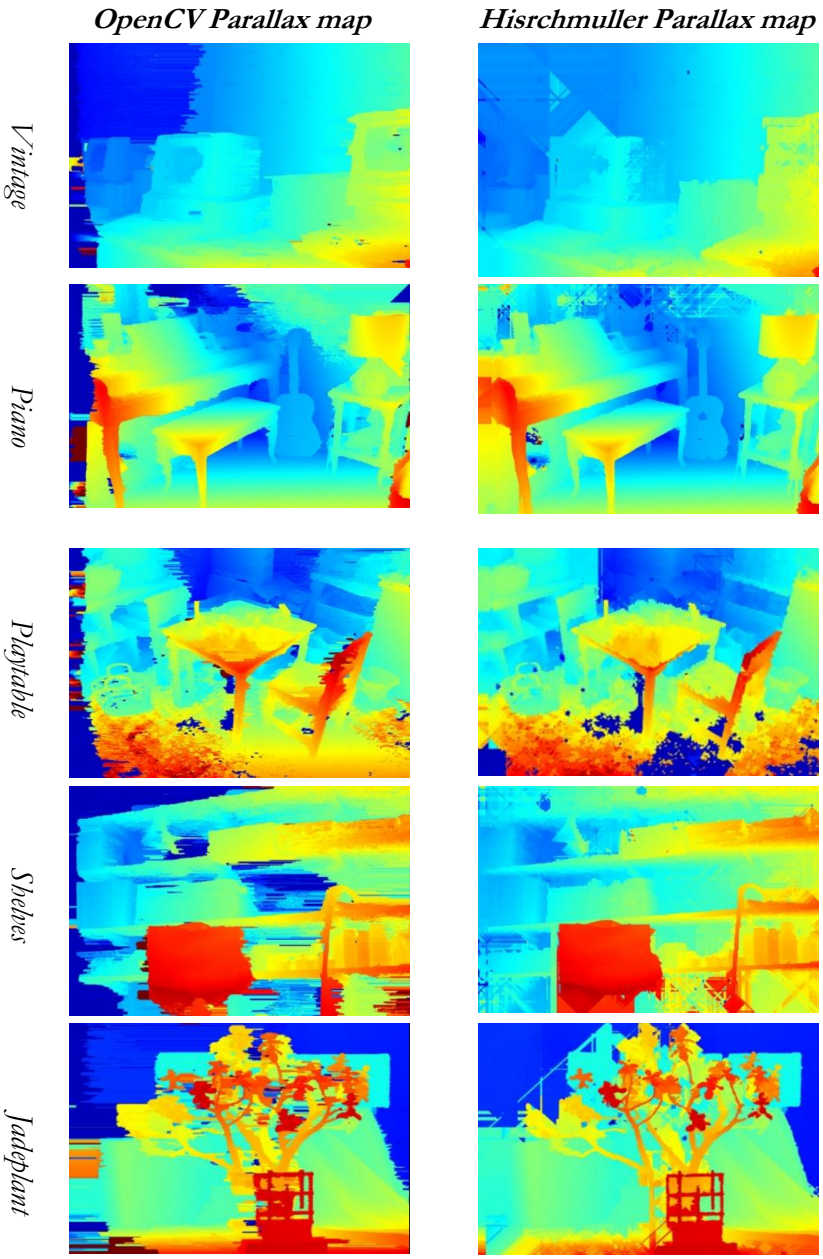


Figure 3.50: The reconstructed parallax map achived by Hirschmuller and OpenCV semi-global strategies.

Hirschmuller algorithm seems to better preserve objects discontinuities and boundaries (this is particular evident in "Jadeplant", "Piano" and "Shelves" cases) and, moreover,

it shows more robustness toward low textured areas, producing however an evident smoothing effect of the parallax maps data. Differently, OpenCV algorithm shows noisier parallax data and lower accuracies in identifying objects boundaries, probably due to misalignment problems of the disparity map scanlines (it is worth noting that in OpenCV method costs aggregation is performed by dynamic programming along only 5 paths). Overall, OpenCV parallax maps seem to be, at a glance, the most similar to the proprietary algorithm results. Just in the "Playtable" case our SGM algorithm has produced much better results, in term of accuracy, discontinuities preservation and it has correctly reconstructed the areas which are not well- identified by the other two strategies (such as the floor, the small objects on the table and the shelf).

CHAPTER 4

TESTS OF SEMI-GLOBAL MATCHING SOFTWARE CODE FOR SURFACE RECONSTRUCTION

Introduction

Encouraged by the growing interest in automatic 3D image-based reconstruction, the development and improvement of robust dense matching techniques is one of the most investigated research topic of the last years in photogrammetry and computer vision.
In order to understand the main performance differences between different strategies/implementation and the proprietary semi-global code, tests on real and synthetic images have been performed, evaluating in particular the accuracy of the obtained digital surface models. The comparison has basically two main objectives: first of all, checking the accuracy of the SGM proprietary implementation, and getting insight on the influence of some matching variables in the process. Second, comparing the accuracy and completeness of the digital models obtained using the proprietary code with other commercial and open-source matching programs [45] (in particular with respect to traditional local methods approach). The comparisons want also to consider the completeness and the level of detail within fine structures, and the reliability and repeatability of the obtainable data. All the codes/algorithms used in this first investigation considered only stereo pairs in order to identify the correlation performances of each strategy. In the final section of the chapter, the multi-image algorithmic extensions has been used and evaluated for analysing their performances in digital surface models reconstruction.

4.1 Comparison of dense matching algorithms for surface reconstruction

4.1.1 Image matching strategies description

Several algorithms and different implementation are considered in the comparison, using freeware software codes (e.g. MICMAC and OpenCV), commercial software (e.g. Agisoft PhotoScan (PS)) and proprietary codes implementing Least Squares and Semi-Global Matching algorithms.

4.1.1.1 DenseMatcher

DenseMatcher (DM) is a digital terrain model generation program developed at the University of Parma [146], based on classical Area Based stereo algorithm. It implements NCC, LSM and Multiphoto Geometrically Constrained Matching [69] correlation algorithms and it uses a multi-resolution scheme. The common processing

pipeline is as follow: using known interior and exterior orientation parameters, the algorithm perform the epipolar resampling of the image pair (see paragraph 3.1.2) improving the efficiency of the process; then, using tie-points information (or an initial approximate depth map), it realizes the disparity data optimization with an initial NCC matching step (optionally at each level of the pyramid). Finally, the LSM procedure is applied to obtain the final correspondences with a parallel dense matching procedure.
The program has been developed in C# object-oriented program language, it is able of multicore processing, and produces as the final output a point-cloud Digital Terrain Model.
The implemented code allows controlling several parameters such as the number of pyramid levels, the template size, the correlation coefficient threshold, the correlation algorithm and many others variables involved in the matching process.

4.1.1.2 MicMac

APERO and MICMAC are two open source tools realized at IGN (Institut National de l'information Géographique et Forestière, Paris) that allows to realize all the steps of a typical photogrammetry process, starting from Structure&Motion orientation up to dense point clouds and orthophotos generation. APERO is the orientation software, which uses both computer vision approach to estimate an initial solution and photogrammetry for a rigorous compensation of the total error [140]. It allows processing multi resolution images and, for each resolution level, it computes tie points extraction for all images pair performing finally a bundle block adjustment [187].
The final DSM generation phase is performed with the MICMAC tool which produces the depth maps, and eventually the 3D models, from the oriented images. This step is performed solving the surface reconstruction problem under the form of the minimization of an energy function [139].
MICMAC suite contains several photogrammetric tools for the automatic processing of aerial and terrestrial image blocks and the DSMs and orthophotos extraction.
The software is very interesting for the photogrammetric community because it provides statistical information of the data and allows detailed analysis of the photogrammetric processing results. Moreover, all the parameters and the results of the orientation and matching step are stored in XML files which can be adapted whenever the user needs to impose certain settings and values of the processing parameters.

4.1.1.3 Photoscan

Agisoft PhotoScan is a commercial software, developed by Agisoft LLC Company. It has a very simple graphical user interface and, as MicMac, it is able to perform both the

orientation and the following dense stereo matching steps using a multi-image approach.
Initially the software defines the images orientation and refines calibration camera parameters; in a second step, it proceeds to the DSM generation. Differently to MicMac, PhotoScan doesn't display the statistical results of the photogrammetric processing, being a sort of "black-box" software. All the photogrammetric process is performed with a high level of automation and the user can decide the desired points cloud density and the 3D modelling quality. The workflow is therefore extremely intuitive being an ideal solution for less experienced users.
Due to commercial reasons very few information about the used algorithms are available: some details can be recovered from the Photoscan User forum [138]. Apparently the depth map calculation is performed pair-wise (probably using all the possible overlapping image pairs) and merging all the results in a single, final, 3D model. A multi-baseline matching extension, like the one implemented, is more robust than pair-wise stereo correlation with regard to occlusions detection and wrong matches, realizing the fusion of disparity information given by all the matched images and producing smoother results.
Anyway, in all the following tests, only stereo-pair were considered in order to compare the base stereo matching approach of the proprietary semi global algorithm with respect to the other matching strategies.

4.1.1.4 OpenCV libraries

OpenCV (Open Source Computer Vision Library: [129]) is an open-source BSD-licensed library written in C, C++, Python and Java that offers high computational efficiency and a simple use of Computer Vision and Machine Learning infrastructures. The library, developed by Intel in 1998, is cross-platform, running on Windows, Linux, Mac OS X, mobile Android and iOS; it contains several hundreds of optimized algorithms for image processing, video analysis, augmented reality and many more, providing the tools to solve most of computer vision problems.
Using IPP (Intel Performance Primitives) it provides an improvement in processing speed and optimization that are very useful for real time applications [30]. In 2010 a new module that provides GPU acceleration was added to OpenCV and, right now, the library is expanding day-by-day.
In order to perform the image matching strategies comparison, the open library for computing stereo correspondence with semi-global matching algorithm was used. The method executes the semi-global block matching (SGBM) (adapted by the first Hirschmuller description of the algorithm [84]) on a rectified stereo pair, introducing

some pre and post processing steps (for example uniqueness check, quadratic interpolation and speckle filtering) of the data. Several matching parameters may be controlled and set to a custom value but, especially, in order to isolate only the matching step contribution, the additional pre-post processing parameters were disabled.
OpenCV version of the SGM strategy is focused on speed and, in contrast to our implementation of SGM, calculates the disparity algorithmic implementation described in [20], which uses less path directions to calculate the matching costs (8 paths, instead of 16).

4.1.2 Comparison datasets

In order to understand the main differences between the strategies/implementations, three tests on real and synthetic images have been performed. An exhaustive description about the different datasets and three-dimensional digital surface models used as reference data in the comparisons is presented in the next sections.

4.1.2.1 Synthetic images of 3D shapes

First of all, in order to evaluate the performance and the best parameters combination used by the different stereo matching approach, a simple 3D scenery was created using 3D modelling software. Spherical and rectangular objects were located on a wavy surface creating significant depth discontinuity in the scene (as it is visible in Figure 4.1).

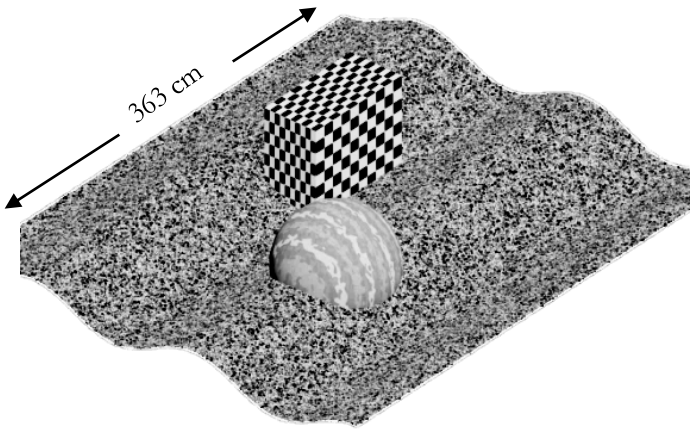


Figure 4.1: Computer-generated 3D primitives dataset.

An ideal, well-contrasted texture was draped on the objects favouring the matching algorithms. Two virtual cameras were located in the scene taking two nadiral synthetic

images. Optimum illumination conditions were simulated, producing light and shadows useful for a simpler (human) depth identification.

4.1.2.2 Synthetic images of a 3D reference model

Using the same 3D modelling program, a 3D model of an architectural element (Figure 4.2) was imported and draped with an ideal texture. The chosen object is a Corinthian capital 1.33 meter high and with a circular base of 90 cm diameter, characterized by complex architectural details.

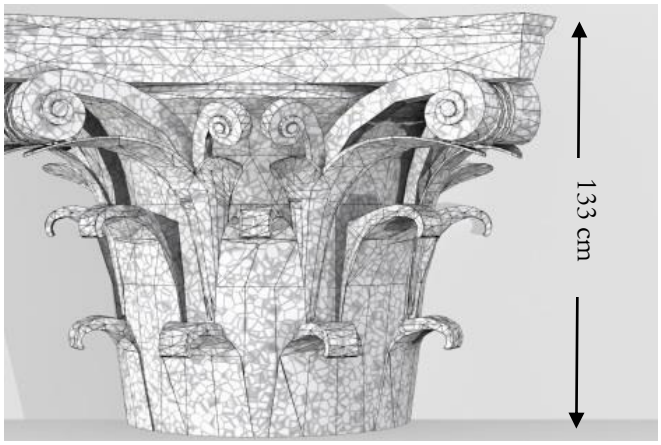


Figure 4.2: Synthetic image of a 3D reference model.

As in previous case, two virtual nadiral cameras were created, with known interior and exterior orientation parameters. To simulate a photorealistic scenario, illumination sources were located using photometric virtual lights. Finally, images rendered through a raytracing algorithm, were generated and exported (see Figure 4.2).

4.1.2.3 Real images and reference DSM

The third case is an image pair extracted from a sequence of a 5 meter high richly decorated fountain from the cvlab dataset [173]. The dataset consists of known interior orientation parameters, distortion removed images and a reference laser scanning DSM. The exterior orientation parameters were estimated through a Structure from Motion procedure, followed by a bundle adjustment step using some features extracted from the DSM as Ground Control Point. The availability of a laser scanning surface reference model of the fountain has allowed to validate the results of the surface reconstructions. Figure 4.3 illustrates one of the two convergent images used in the stereo matching.

Figure 4.3: Real image of the fountain.

4.1.3 DSM generation and comparison

The evaluated matching methods implement different strategies for generating the final DSM.

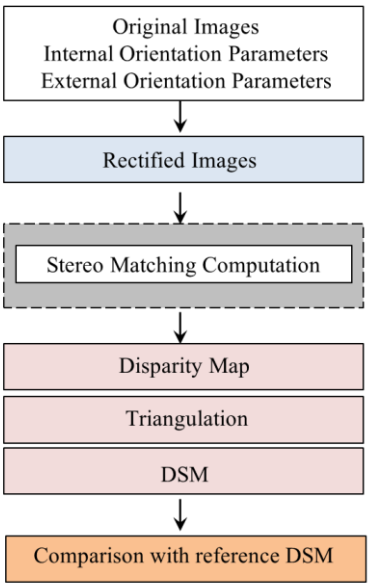


Figure 4.4: DSM generation and comparison workflow.

In order to obtain comparable solutions, the models were generated using the same known internal and external orientation parameters, making the results as independent as possible from the automatic orientation procedure. In fact, both PhotoScan and MicMac software are able to perform, beside the DSM generation, also the automatic

orientation of the image block; however, different orientation solutions can produce unwanted DSM deformation. On the other hand OpenCV library expects to work with rectified images (i.e. corresponding points in the stereo pair lay on the same horizontal image line) and produce a disparity map. Differently from MicMac and PhotoScan, a subsequent triangulation stage, to be carried out externally, was required to produce the final DSM.
Finally, the distances between the reconstructed and the reference DSM have been evaluated to perform the comparison. Figure 4.4 describes the comparison workflow.

4.1.4 Results: relative accuracy and reliability of the reconstructed DSM

The relative accuracy of the DSM reconstructed models with synthetic and real images is summarized in Table 4.1 where, for each test case, statistics of the distances between reconstructed and reference DSM are presented. Disparity map comparison (i.e. evaluating the algorithm accuracy in finding the corresponding point on image space) were considered as well but, for some software (e.g. Photoscan) the parallax field are not directly accessible and are hard to be computed. Also, the influence on the final reconstruction accuracy was considered more interesting.

	Test cases	DM	OpenCV	PS	SGM	MicMac
Normalized relative accuracy	3D Shapes	100 %	92 %	91 %	83 %	94 %
	Capital	100 %	85 %	83 %	71 %	81 %
	Fountain	90 %	97%	100 %	89 %	90 %
Inlier percentage	3D Shapes	86.9 %	91.4 %	88.7 %	83.2 %	80.9 %
	Capital	77.5 %	92.7 %	89.27	70.1 %	66.8 %
	Fountain	84.6 %	88.7%	97.5 %	86.5 %	88.1 %
Lacks of completeness	3D Shapes	*	*NO*	***	*NO*	*NO*
	Capital	*	*NO*	*	*NO*	*NO*
	Fountain	**	*NO*	*NO*	*NO*	*NO*
Pixel locking	3D Shapes	***	***	***	***	*NO*
	Capital	*	**	**	*	*
	Fountain	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*

Table 4.1: Statistics of the reconstructed DSM.

To make the results independent of the total size of the object all the distances standard deviations are normalized with respect to the best value ("*Normalized relative accuracy*"). At the same time, some methods present, in some areas of the model, very evident gross errors that must be removed from the relative accuracy computation (lack of

completeness problems are symbolized in Figure 4.5 with a red arrow on the incomplete area). On the other hand, it's important to highlight which algorithm produces more reliable results (in terms of "*inlier percentage*"): for each model, tolerance ranges were selected based on some assumption about image matching (and consequent reconstruction) a-priori precision and on the actual performance of the best method. In particular (1 cm for 3D Shape, 3 mm for Capital and 3 cm for the fountain case study ranges) are selected considering that at least one algorithm must produce a 90% in-tolerance 3D model: in this way reconstruction accuracy is related to a sort of quality completeness for each method.

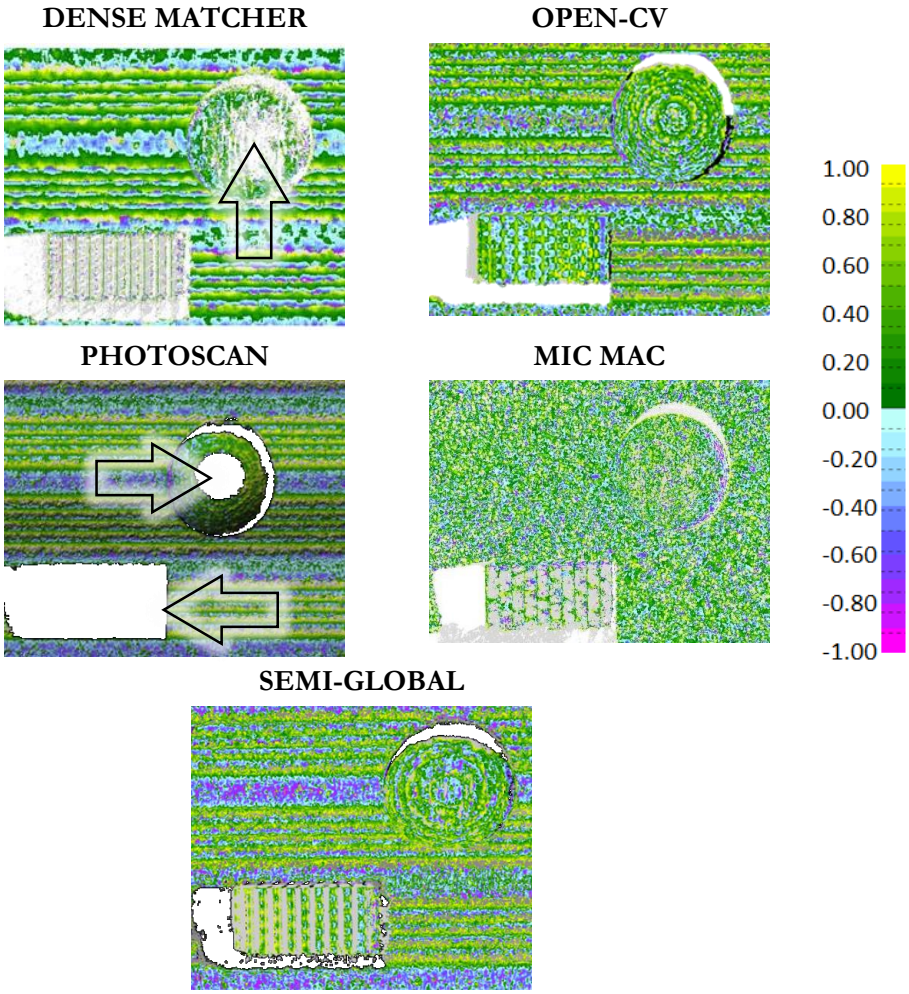


Figure 4.5: Map of the distances (in cm) between the reference and the reconstructed DSM for 3D Shape case

The table shows that, for each test case, the different matching algorithms produce results that are not dramatically different. The general trend is similar, though not

identical, in particular for computer-generated data. Analysing the first two tests, we can identify that the best solution were obtained by LSM with DenseMatcher, followed by OpenCV, PhotoScan and our implementation of Semi-Global matching.

For '*3D Shape*' case, looking at the error map presented in Figure 4.5, we can see two big holes (highlighted with the red arrows), in correspondence to high depth changes, in PhotoScan DSM. Therefore, it is worth noting that low values of discrepancies from the reference model, derived from standard deviation information, cannot be the only indication of the digital model reliability but we have to evaluate the completeness and surface distribution of the points.

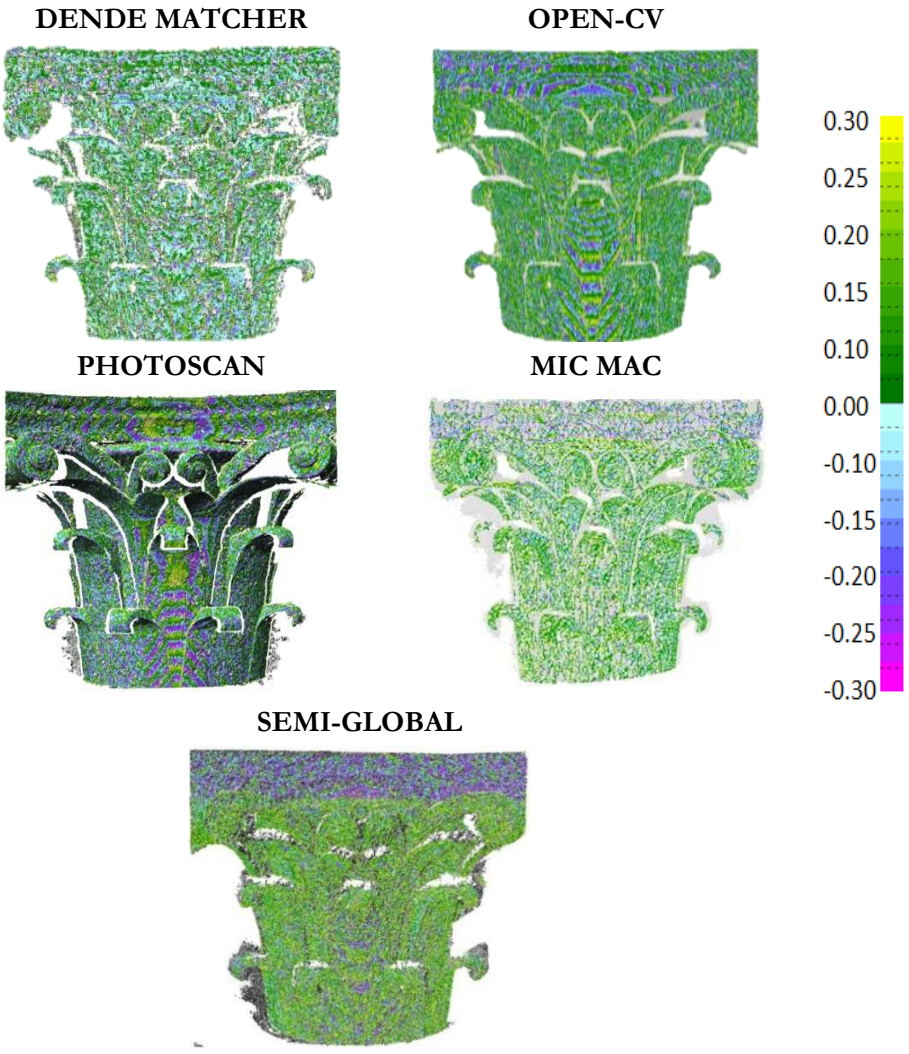


Figure 4.6: Map of the distances (in cm) between the reference and the reconstructed DSM for Capital case.

The capital test case, presents the higher base-length to distance ratio, and higher perspective effects can be expected: also in this case, unsurprisingly, the LSM achieve the highest accuracy. On the contrary, in the Capital and Fountain test cases some area of the images lack of a well-contrasted pattern: while the semi global methods achieve good quality results also in these areas, the LSM algorithm cannot always produce complete results.

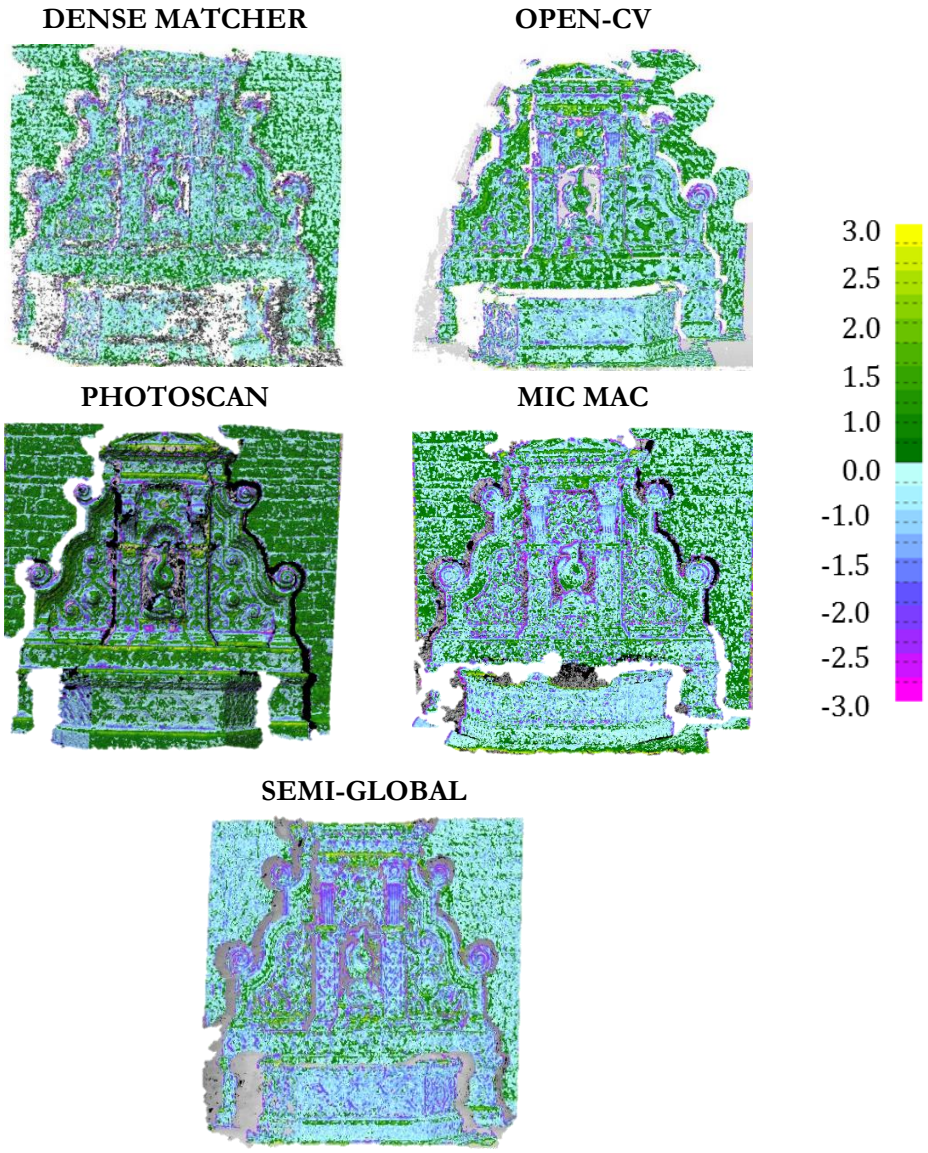


Figure 4.7: Map of the distances (in cm) between the reference and the reconstructed DSM for Fountain case.

In other words, we can say that the error map of DenseMatcher (see Figure 4.5, 4.6 and 4.7 for analysing the 3D Shapes, Capital and Fountain maps respectively) usually shows more noisy data, due to its pointwise estimation approach. While semi-global-like methods constrains (with different degree of enforcement) the regularity of the disparity field, every point in LSM methods are considered and evaluated individually. On the other hand, the reconstructed DSM reveals the LSM ability to produce accurate results, as shown not only from standard deviation (see Table 4.1) but also from spatial distribution of the distance values. The disparity regularity constraints (and smoothing filtering procedure – e.g. those implemented in the PS workflow) can generate erroneous systematic surface reconstruction, if image noise, occlusions, repeated pattern influence a whole matching path.

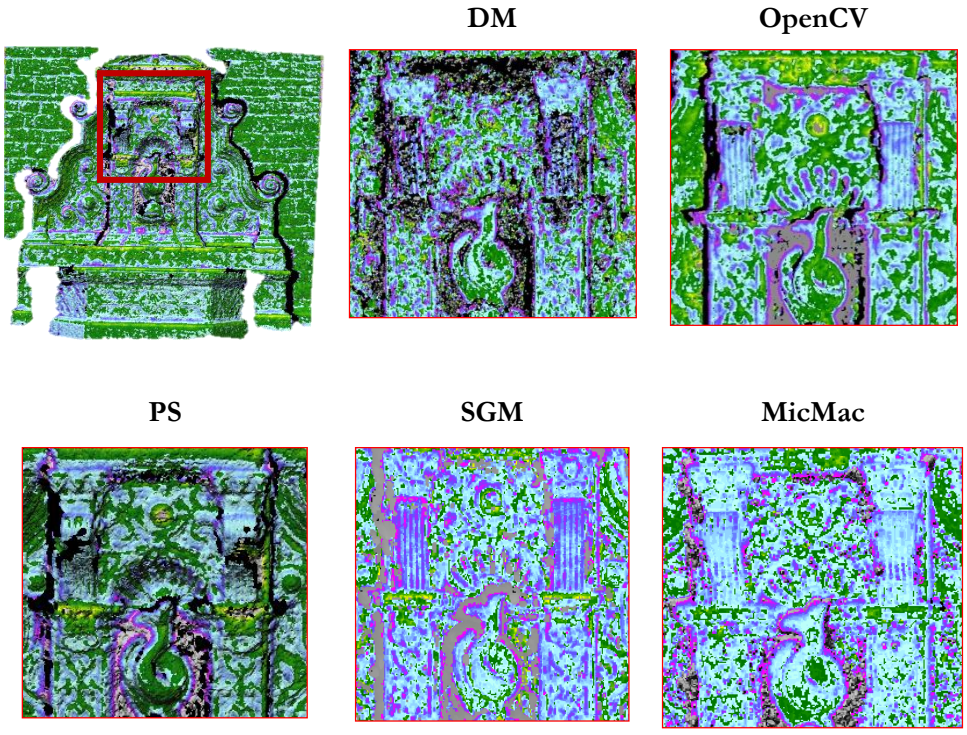


Figure 4.8: Fountain error map with a zoom on the red delimited area.

The same behaviour is clearly visible in the last case study, which considers real images. SGM and MicMac produced more noisy results (see Figure 4.7) but, at the same time, if the smaller object features are considered, they captured finer details; the smoothing effect implemented in PhotoScan, on the contrary, produced an apparently more

appealing results but with some local discrepancies (see for instance the zoomed areas extracted from '*Fountain*' case results – Figure 4.8), flattening some detailed areas; on flat, low contrasted areas, the smoothing and filtering procedures, probably allow acquiring better results with an overall higher completeness level.
Finally DM, suffers of lack of contrasted texture and higher noise due to the real image quality, producing a 3D model with higher error levels in some region.

4.2 Monitoring of an active landslide in Mont de la Saxe (AO)

Landslides control and monitoring systems face a large variety of cases and situations; indeed, the behaviour of a landslide (in the time domain as well as in the space domain) depends on many factors, which present the geologists and the surveyor each time with different challenges.
In principle, the main design parameter when installing a measurement and control system is the accuracy needed to assess with a given probability the magnitude of the expected displacement. However, a number of other issues influence the choice of the best monitoring system to use: the size of the area to control, the frequency of data acquisition, the time to deliver the results (alert time), the stability of the reference system, the influence of atmospheric parameters on measurement accuracy or operation, the site constraints and limitations for the installation of the system (visibility, energy consumption, data telemetry, safety, security), etc.
No single measurement system is therefore suited for any situation and the selection of the best solution for a specific site is not always straightforward; often, if the risk is high and the budget available allows for, more than a system is employed.
In the analytical photogrammetry era, detection of the movement was based on evaluation of DSM changes (measured manually along profiles) or on displacement of targets distributed over the area. Today automatic block orientation and DSM generation, with high resolutions and accuracies, makes photogrammetry an alternative to Terrestrial (TLS) and Aerial (ALS) Laser Scanning. When the requested monitoring frequencies are high, the low costs photogrammetric approaches and systems represents an efficient and optimal choice (in fact, photogrammetry has been used since long to periodically control the evolution of landslide and an accurate description of the state of the art has been provided in paragraph 1.2.2).
In this paragraph, a photogrammetric system to monitor landslides and, in general, changes in digital surface model, is presented. This study was carried out together with Fondazione Montagna Sicura (FMS) – a no-profit foundation promoted by the Valle

d'Aosta Region, which studies issues concerning the safety, rescue and life in high-mountain.

The study object is the landslide of Mont de la Saxe in Courmayeur (AO-Italy, Figure 4.9), which is considered one of the most complex landslides in Italy. The landslide is monitored since 2009. With an estimated volume of about 8 million cubic meters [43], the movements threatens the villages of Entreves and La Palud as well as the A5 motorway and the national road SS 26. A slow continuous movement alternates with sudden accelerations and, in the spring 2013, a sudden acceleration forced the local administration to evacuate the area and to close all the accesses to the Ferret Valley. This complex dynamic drew international attention, not only to assess the geological and geotechnical problems, but also to evaluate and compare different state-of-the-art monitoring techniques.

Figure 4.9: The Mont de la Saxe landslide (25 th June, 2013).

The most of the following arguments is taken from [150] where the authors have presented the first results of a fixed terrestrial stereo photogrammetric system based on a LSM algorithm and developed to monitor shape changes of the scene.

4.2.1 The photogrammetric system

The system is made of two single-lens reflex (SLR) cameras, each contained in a sealed box and controlled by a computer that periodically shoots an image and sends it to a remote computer (there is a programmable board for camera control and data transmission). The system automation is necessary because, being conceived for installation in hard-to-reach locations, the hardware must be remotely controlled and checked, reducing the on-site operator intervention to the minimum. Operating

outdoor temperatures well below −20 °C and exceeding +30 °C can be reached; as far as energy supply is concerned, a 50-W solar panel and a pack of three 17-Ah batteries ensure continuous operation (see Figure 4.10). Batteries are continuously recharged by the solar panel. In the case of persistent cloud coverage, the autonomy of the system is ca. 3 days. If available at a convenient distance, connection to a power outlet is possible. The acquisition frequency should be related to the attainable accuracy and matched to the expected magnitude of the displacement. In the Mont de la Saxe case, a Ground Sampling Distance (GSD) of ca. 20 cm and an accuracy of about 5 - 10 cm were foreseen: such values are compatible for a comparison of the reconstructed models on a monthly basis.

Two Canon EOS 5D Mark cameras (21 megapixels) with a 50 mm focal lens have been considered suitable for completely framing the area on interests for the survey. One of the box stations was installed near FMS facilities; the second box was installed at about 150 m from the first, at a slightly lower elevation. The location was carefully chosen to obtain a homogenous frame scale: the distance from the installation site to the object varies between 500 m and 650 m. The pose of the camera stations is slightly convergent to ensure maximum overlap. The interior parameters of the optical system and the camera position and orientation have been determined with a calibration procedure. For more accurate details about the system architecture, stability and metrological analysis, see [150].

Figure 4.10: Left the watertight IP67 box housing the system. Right one of the system units and the solar panel installed on a pole

Once an image pair is received, the DSM of the scene is generated by digital image correlation on a processing server computer and the points extracted are triangulated

using the parameters of the reference image orientation. Data are thus available for archiving or analysis. Following an accurate description of the system automatic DSM generation workflow is described and, next, some results will be discussed.

4.2.2 Image processing workflow

A service running on a host remote server waits for incoming images from both stations; once a synchronous image pair is available, the DSM generation sequence is activated, as described below:

1. The original images are resampled to remove the lens distortion;
2. well defined points are extracted on both images with the SURF (Speeded Up Robust Feature) interest operator [14]); preliminary point pairs are established, looking for a similar score of their descriptor (see [11]); these pairs are then filtered by a RANSAC procedure considering as consistent a projective transformation induced by an (unknown) rotation of the support;
3. the transformation is used to resample the input images, removing the effect of unwanted movements of the box;
4. the new images are used in the matching procedure after epipolar resampling (see [141]). The points on the master image are always selected on the same regular grid (commonly with a 2 pixel spacing);
5. to filter vegetation or gross matching errors, every matched point on the master grid is compared with its neighbours: if there is a sudden change in the disparity values (corresponding to an equally sudden change in depth in 3D space) the point is labelled as an outlier and discarded;
6. after the matching step, the points extracted are triangulated using the orientation parameters of the reference image pair;
7. a colour-coded map of the displacements measured by the system at the new epoch is automatically generated; format).

The whole process takes about 15 minutes on a medium performance server with eight processing cores to produce a DSM with on average 600-700 thousand points (excluding those removed by the filtering step). The new SGM algorithm was considered a strategic improvement of the processing pipeline since can provide in much less time (ca. 6 minutes) a more dense surface model (the matching can be performed pixel-wise) and is capable of filtering most of the noise connected to low-texture, presence of vegetation or snow, different luminosity conditions (the shoots are not perfectly synchronized) just implementing the regularity constraints.

4.2.3 Results

In a first stage, as a consistency test to evaluate the level of performance achieved by the system using LSM, it was decided to produce and compare, either automatically and manually, a series of 20 DSM referring to the time interval that goes from July 10th, 2013 to September 10th, 2013.
The DSM were selected considering the best days and time of the day according to two criteria: first, trying to monitor some periods of consecutive days evenly distributed on the two months' period, and then selecting pairs with the higher number of matched points. Comparing DSM's of consecutive days, the measurements repeatability can be evaluated: on such short periods the possible landslide displacements can be considered insignificant for the system accuracy level, and the models should not manifest any change. In other words, the differences shown by the comparison should be considered derived by the measurement noise of the system itself. The comparison is presented in Figure 4.11 in which the colours represent the distances (absolute differences) between two models separated by four hours. Most of the points are coloured in blue (corresponding to a movement of 2 ÷ 4 cm) with peaks reaching, in some cases (green), ca. 10 cm. It should be noted that usually the maximum discrepancies show up in those areas where the complexity of the object or the presence of shadows makes the matching procedure more troublesome.

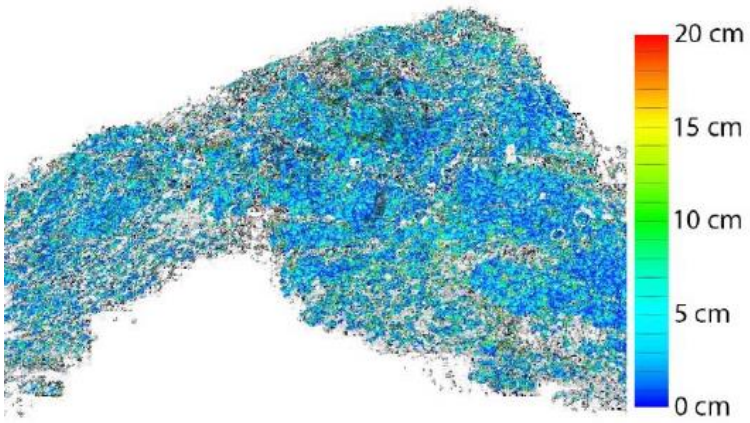


Figure 4.11: Repeatability between two 4-hours separated DSMs.

Evaluating the differences point wise (in this case considering the signed distance of more than 600,000 points), and analysing their distribution, some remarks can be made. The data are approximately normally distributed with a mean distance between the two epochs considered in the different comparisons usually not null. A sort of systematic misalignment of the two DSM (likely due to a residual, not modelled box movement) is

found quite often. However, the mean distance between two epochs is always lower than 2 cm and can be considered negligible if compared to actual system accuracy. In fact, considering a distance from the object Z of 500 m and a base-length B of ca. 100 m using a slightly convergent stereo camera system with a 21 Mpixel resolution (i.e. a 6.5 μm pixel FX sensor), equipped with a 50 mm focal optic, and considering that, in outdoor good lighting condition, current state of the art area based matching algorithms can achieve 0.2 pixel accuracy quite systematically, a final depth restitution precision of ca. 6.5 cm can be obtained. Therefore, the above described results are in agreement with the theoretical accuracy of the system.

This evaluation were performed using the Adaptive Least Squares Matching (see [68]) but, the following and actual development of the SGM algorithm has allowed the landslide investigation with this method as well. In fact, thanks to the intrinsic characteristics of a semi-global method we can achieve important advantages:

- the regularization constraints on adjacent pixels displacement allows, not only to obtain more complete reconstruction, but also to increase the repeatability of the reconstructed digital models since the improved results smoothness decreases the measurement noise;
- the processing pipeline becomes faster thanks to the better performance of the SGM technique, with respect to the LSM method: to have an order of magnitude, the first takes about 2 minutes for obtaining the final DSM, the second 11 minutes approximately. Low computational time might allow an higher frequency of the monitoring activity and, especially, the possibility of elaborating the DSM on-board directly, avoiding to send the data to the remote server.

In order to evaluate the SGM algorithm performance, two image pairs of the landslide, taken at a short time distance (the first was shot the 11th September 2014, the second to the 22 September 2014), were analysed comparing the results obtained with the old implemented LSM strategy and the novel SGM approach.

First, in Figure 4.12, the digital surface models obtained with the two algorithm are presented. the evident noise of the Digital model obtained using LSM technique is not evident, on the contrary, in the Semi-global solution, which reaches a more accurate and complete surface reconstruction. Holes and noise, which characterize the LSM solution (Figure 4.12 (a)) have a much lower magnitude in the SGM results. These noisy areas, together with the easily recognizable lack of completeness, allow to say that the semi-global matching algorithm is able to obtain better results and, as previously said, with much lower computational efforts.

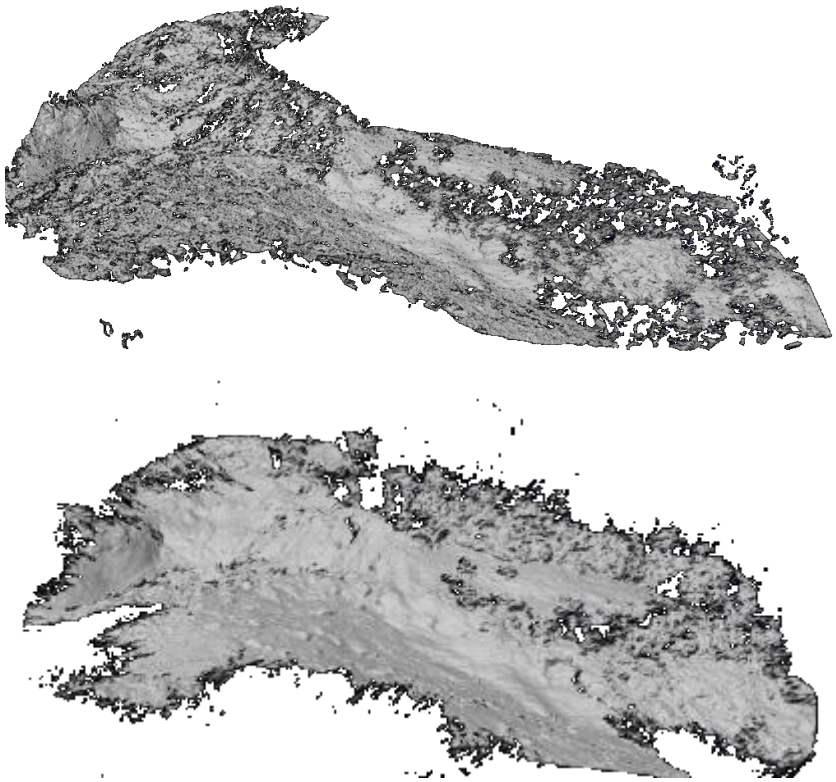


Figure 4.12: a) Mont de la Saxe reconstructed DSM (relative to the 11 September 2014 epoch) using: a) LSM algorithm ; (b) SGM algorithm.

Second, in Figure 4.13, the comparisons results are presented: on top (Figure 4.13 (a)) the reconstructed LSM and SGM digital models relative to the 11 September, on the bottom (Figure 4.13 (b)) those regarding the 22 September 2014 period.
In both cases, a good agreement between the LSM and SGM solutions is shown: most of the surface differences are in the range of ± 5 cm. However, as better visible in Figure 4.14, where a close up of the results is shown, it is possible to see isolated points that considerably deviate from the range of ± 5 cm, since they can reach up to difference values equal to ± 30 cm.

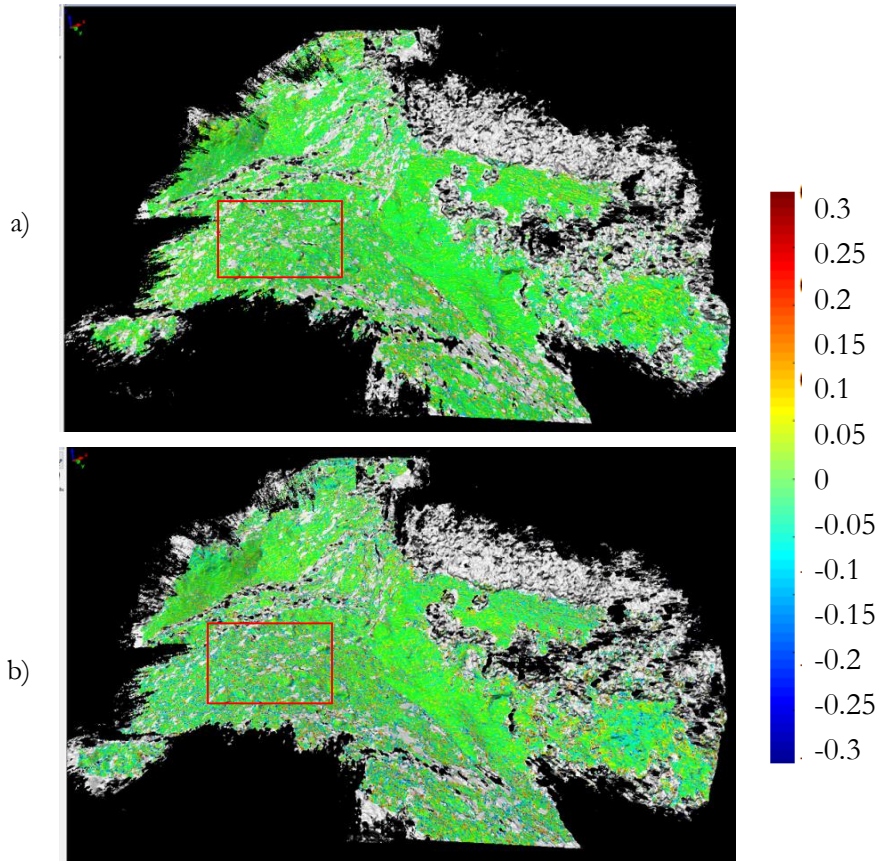
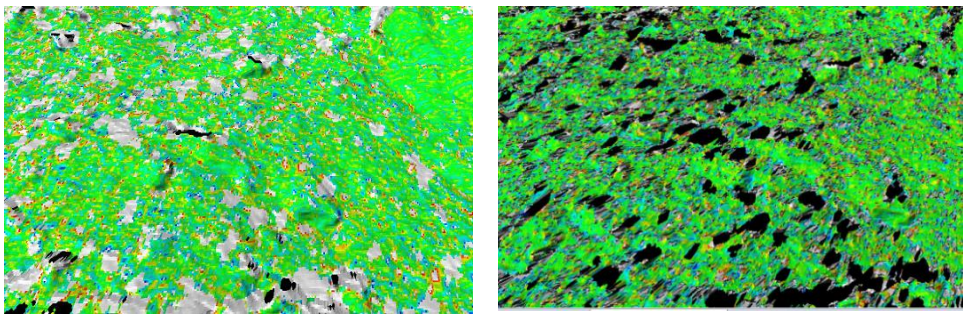


Figure 4.13: Comparison [in m] between the LSM and SGM reconstructed DSMs. a) The distance map of the DSMs of the slope acquired on September 11th at 17 PM.; b) The distance map of the DSMs of the slope acquired on September 22nd at 17 PM.;



a) b)

Figure 4.14: In The 11 September (a) and the 22 September image pairs error maps with a zoom on the red delimited areas shown in Figure X.

After the analysis of the SGM method reliability, the algorithm was finally tested in the condition required by the monitoring application, i.e. comparing the 11th and 22nd September DSMs for evaluating the landslide displacements during such period (see Figure 4.15 (a)). The same map has been obtained using the two LSM digitals surface models for a further visual comparison of the two strategies performances (Figure 4.15 (b)).

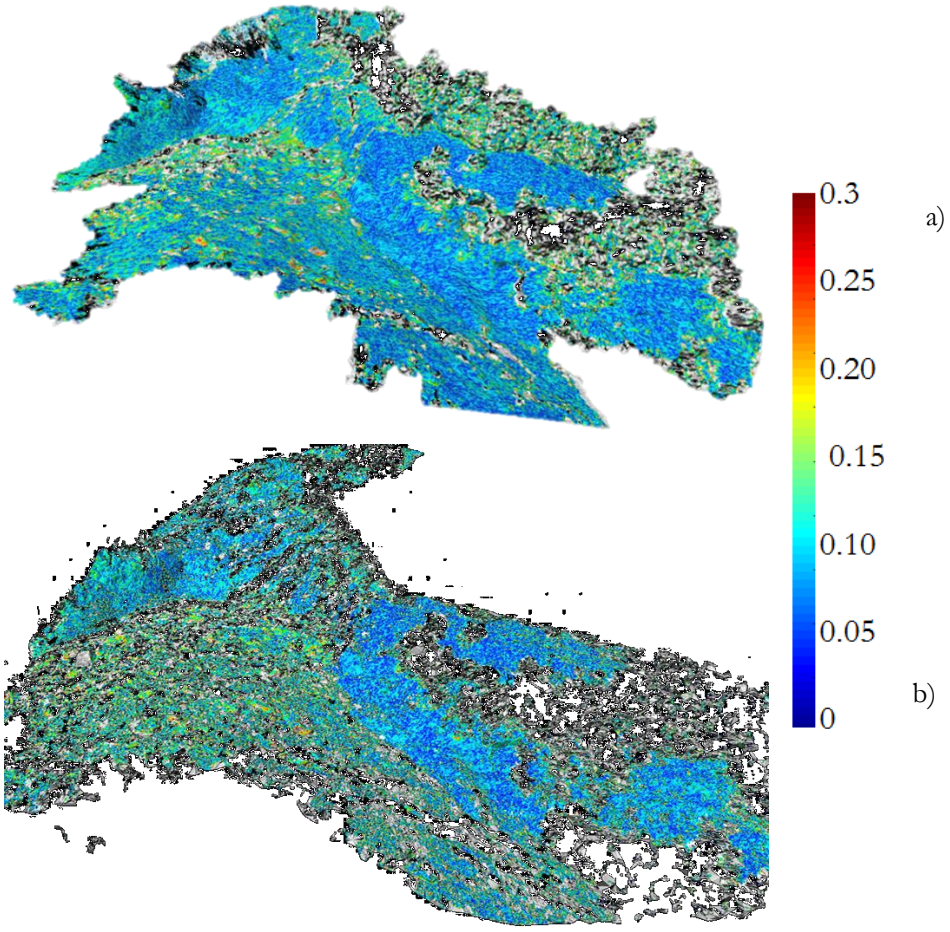


Figure 4.15: Map of slope displacements [in m] between the 11th and 22nd September. a) The results obtained with SGM digital surface models. b) The results obtained with LSM digital surface models

The results confirm the considerations just made: there is a good agreement between the displacements measured by the two method. Summarizing, during the short considered time period, important slope movements were not observed and just some

areas presents medium-high displacements that reach 30 cm. These regions are probably rocks or small stones that collapsed.
Overall, the SGM method (Figure 4.15(a)) appears more accurate and complete, with respect to the LSM: the lower noise level makes the results more complete and detailed. Its employment within the image processing workflow of the photogrammetric monitoring system might allow to reach an higher quality and higher frequency monitoring of the landslide.

4.3 3D surface reconstruction of the Parma Baptistery zoophorous

The zoophorous of Benedetto Antelami is a series of seventy-five sculpted panels in red Verona marble, realized with bas relief technique on the lower level of the external marble façade of the Baptistery of Parma (Figure 4.16). The decoration wants to represent the fantastic in sculpture: human figures, real and imaginary animals are shown, almost following the directions of a medieval bestiary (see Figure 4.17).

Figure 4.16: Some panel of the Baptistery south-west (on the top) and north-west side (on the bottom).

Figure 4.17: The Lion wounded by an arrow shot by a satyr with a cap on his head and the rooster, considered as a symbol of vigilance

After the laser scanning survey (more details on the application and work objectives of are presented in [47]) and the photogrammetric analysis of indoor and outdoor environments of the Parma Baptistery, an accurate close-range photogrammetric survey was computed to generate high-resolution digital surface models of the zoophorous panels. A Nikon D3X camera, with 50 mm and 105 mm optics and image resolution of 6048 x 4032 pixels, was used for the photogrammetric investigation. The baptistery zoophorous was firstly surveyed with the 50 mm focal lens (as shown in Figure 4.18) to reconstruct the whole zoophorus and after, for each panel, sequences of about three-four frames each were taken with the 105 mm optic for having an high-resolution description of the panels details.

The 3D reconstruction of some panels (see for instance the panels in Figure 4.19) extracted from both sequences, was performed using the semi-global strategy both in stereo and multi-image processing conditions.

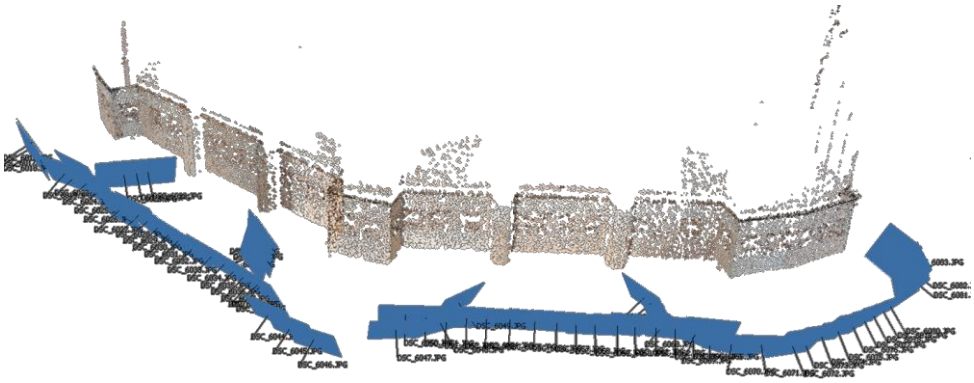


Figure 4.18: The alignment of an images sequence of the zoophorous, taken with the 50 mm focal lens.

In particular, the 50 mm sequence allowed to obtain the overall three-dimensional reconstruction of the panels sequences.

After the image idealization and epipolar rectification processes, the image correlation was performed using the semi-global algorithm in a pairwise mode: i.e. considering an image sequence of five frame, the process was computed considering consecutive pair of stereo images and performing the image correlation of one stereo pair at a time. The four obtained parallax maps were then triangulated and the digital surface models were joined, producing the final three-dimensional reconstruction, which is shown in Figure 4.19 in grey scale colours.

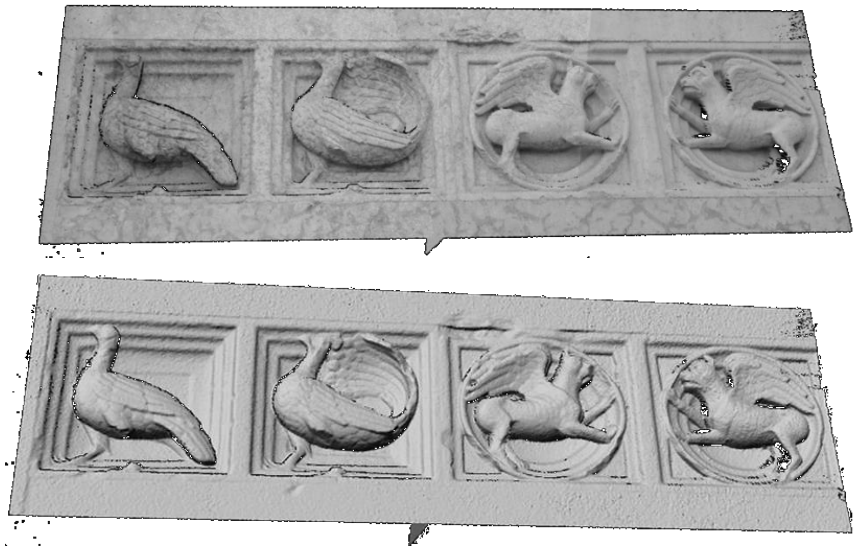


Figure 4.19: The 3D reconstruction of a Baptistery images sequences taken with Nikon D3X with 50 mm focal lens. The image correlation was computed with SGM in stereo mode. a) The point cloud in grey scale colours. b) The mesh of the reconstructed DSM.

In a second steps, the attention was focused on the photogrammetric survey of each individual panel, which was performed using the 105 mm macro optic. In this case the image sequences were composed by only three photographs which capture the object from three very different perspective view (a high-resolution sequence is shown in Figure 4.20). In this case the great perspective changes between the three images has made the employment of the SGM particularly interesting. Indeed, as explained in paragraphs 2.6 and 3.1.10, the use of a multi-image method seems to be a better approach to perform the image correlation when the accuracy, reliability and completeness of the results are weakened by significant changes of the images point views. Moreover, since the objective is the achievement of a high-resolution and complete model the SGM multi-image method has been demonstrated to be a more appropriate choice also than the LSM one.

Figure 4.20: The "lion" images sequence.

The case of the "Lion" panel is following described. For obtaining the DSM the proprietary Semi-Global matching code has been applied in the multi-image pairwise software variation (paragraph 3.1.10.1).
Being the sequence composed three images, the method identifies six possible stereo image pairs and, for each one of these, the correlation process with the SGM algorithm is performed, obtaining as output the six depth maps presented in Figure 4.21.
Once the depth maps have been obtained, the method proceeds to the data combination for the achievement of a unique result. The maps merging process includes also the application of a median filter in order to reduce/remove depth data noise (as evident from some of the depth maps in Figure 4.21). Once the final result is produced, the digital surface model can be easily reconstructed (Figure 4.22 on the left).

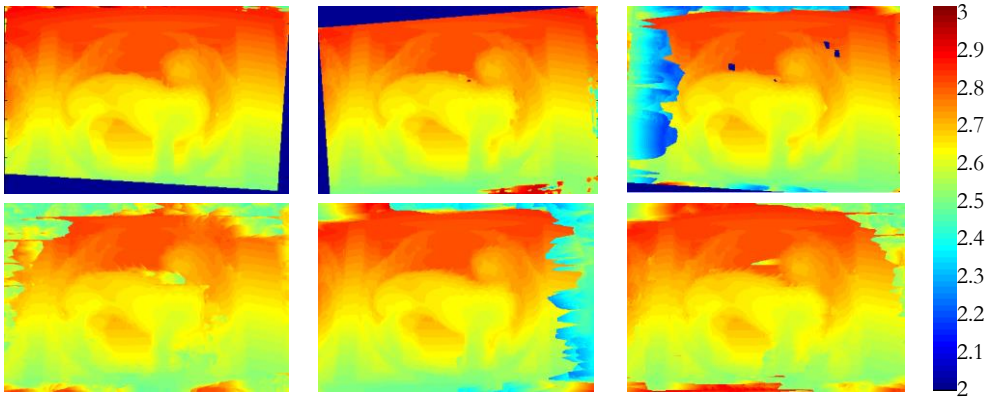


Figure 4.21: The six depth maps [m] obtained with the pairwise SGM proprietary algorithm.

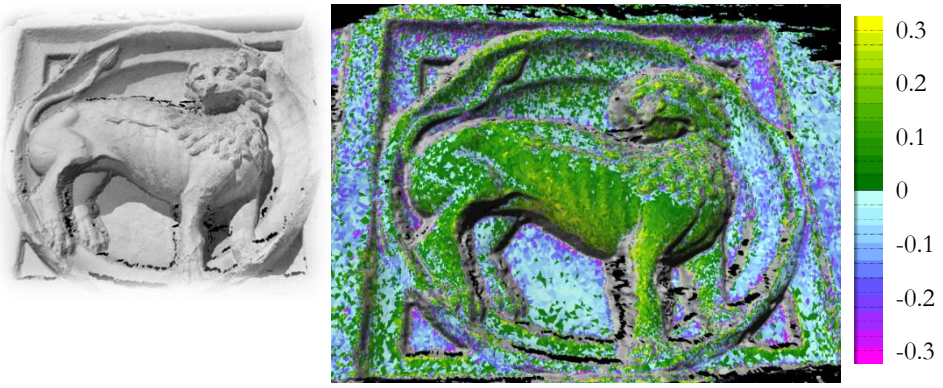
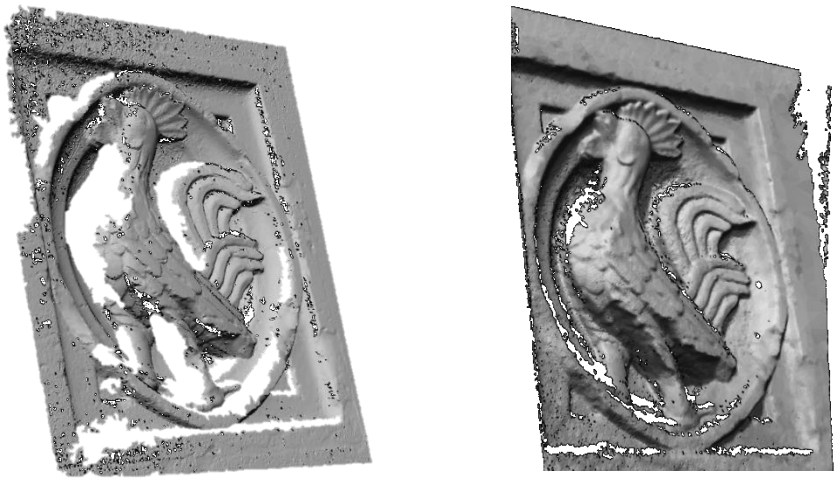


Figure 4.22: On the left the DSM of the "lion" panel and the ccomparison (on the right) between the reconstructed DSM obtained using the multi-image SGM and PhotoScan software [mm].

In order to have an evaluation of the method accuracy and reliability, the lion mesh has been compared with that obtained from the sequence processing in Agisoft PhotoScan [138] software. The comparison result is presented in Figure 4.22 and it shows a good agreement between the two software. It is possible to say that the results are almost identical: the lion body is well reconstructed by both the applications and the majority of the differences regards object discontinuity (such as the panel boundaries and small local details) where a discrepancy of about ±0.1 mm is revealed. Higher differences (that reach up to ± 0.3 mm) characterize the occluded regions where our algorithm, differently from PhotoScan, doesn't provide post-processing tasks capable of filling holes and smooth surfaces.

A final investigation has regarded a different zoophorous panel (shown previously in Figure 4.17). In this case, the SGM algorithm was compared with the Least Square Matching method employed in the same processing conditions (i.e. using only a stereo-pair). This test allowed analysing the different behaviour of the algorithms with regard to the robustness of the stereo matching methods in correlating images characterized by high depth and perspective changes and, moreover, to evaluate the lacks and the qualities of both the Least Squares and Semi Global matching methods.



a) b)

Figure 4.23: The DSMs of the "Rooster" panel from matching of an image stereo pair. a) the LSM solution. b) SGM solution.

The result is presented in Figure 4.23: the LSM model lack of completeness is evident, as well as the noisier reconstructed surface, in particular in the planar region. On the contrary, the SGM strategy shows its abilities in regularizing object shape and detecting points on high discontinuities and depth changes. However, in this kind of applications

where the examined objects present high depth variations and, at the same time, require high completeness level, the multi-image method represents a better strategy than stereo correlation.

4.4 UAS flight on Veio archaeological site

Multi-image matching algorithms are essential component in the restitution process of images sequences taken from UAS flights since the mission (flight and data acquisition) is planned with the aim to have high longitudinal and transversal overlaps between the images; therefore, many of the surveyed areas are captured by three or more frames.
For this reason, the investigation of the proprietary algorithm abilities also in this field of interest has been evaluated. In particular the performances analysis were focused on the UAS flight of an archaeological area (data are provided from the authors of [147]), the ancient town of Veio, near Rome, in Italy. The survey was performed with a Microdrones MD4-200 system by Zenit company: images were taken with a Pentax Optio A40 digital camera, with 8 mm focal length, 1.87 pixel size and image resolution equal to 4242 x 3178 pixel. The whole area of the Veio site includes a big (35 x 30 m^2) and a small (30 x 10 m^2) camp (se Figure 4.24). The image correlation process has been performed with the Object Multi-Image SGM software code (see paragraph 3.1.10.2).
In a first stage, the attention was focused on the algorithm processing performances, in terms of quality of the reconstruction: many several tests have been made to understand the optimum parameter configuration of the Object SGM algorithm.

Figure 4.24: On the left the route taken by the drone on the big and small camp of the Veio Park. On the right the orthophoto of the small archeological camp.

Since that, in a previous work [156], several evaluations and analysis were made on the small camp one, the SGM technique has been also applied on that area. Moreover, the

smaller size of the camp has allowed computing several, which were less onerous in terms of memory and time requirements.

4.4.1 Image acquisition and processing

The mission (flight and data acquisition) was planned with dedicated software, starting from the area of interest (AOI), the required ground sample distance (GSD) or footprint, and knowing the intrinsic parameters of the mounted digital camera. Thus fixing the image scale and camera focal length, the flying height is derived. The camera perspective centres ('waypoints') are computed fixing the longitudinal and transversal overlap of strips, while the presence of GNSS/INS on board is usually exploited to guide the image acquisition. In Table 4.2, the flight planning parameters are summarized.

Focal length [mm]	8.0331
Sensor size [mm]	7.9386 x 5. 9462
Pixel size [mm]	1.87
Sensor format [pixel]	4242 x 3178
Area of interest [m]	30 x 10

Table 4.2: Planning parameters of the UAS flight on the Veio small camp.

The designed flight altitudes is 40 meters that have corresponded operationally in 38 meters for the flight on the small camp. The camera characteristics and the flight altitude has allowed to extract a Ground Sampling Distance (GSD) of the order of about 0.8 cm. In this case a single strip composed of six images was sufficient.
After the data acquisition stage, the interior orientation parameters of the camera have been extracted from a camera calibration process performed in the Photomodeler software [136] and, using the already obtained calibration parameters, the images idealization (for removing distortion components) has been computed. Following, the image exterior orientation has been achieved using the EyeDEA [149] proprietary software which has delivered about 8464 image corresponding points, well-distributed on the images sequence, and a theoretical precision of the computed object Z coordinates of 3 cm.
Finally, starting from the known exterior and interior orientation parameters, the small camp has been reconstructed by means of automated dense image matching techniques.

Different image matching algorithms and solutions have been evaluated and compared for producing the Veio DTM: (i) the digital terrain model generation program DenseMatcher [146], which implements a LSM algorithm and processes consecutive stereo image pair; (ii) the Agisoft Photoscan software [138], (iii) the proprietary Object-based SGM multi-image algorithm.
Hereafter the results description and evaluation are presented.

4.4.2 Results and comparisons

The Object-based SGM method has required the following processing setting:

- definition of the Area of Interest;
- characterizations of the object volume in term of cells size and depth resolution; both the variables were set to 1 cm;
- identification of the elevation search range (i.e. the maximum and the minimum of the object Z-coordinates for defining the equivalent, in the object space, of the images disparity search range).

The other matching parameters remain unchanged with respect to a normal stereo or pairwise multi-image matching process: in this case, for example, a five pixel block size and the NCC similarity function were used. In Figure 4.25 the obtained point cloud and DTM of the Veio small camp are shown.

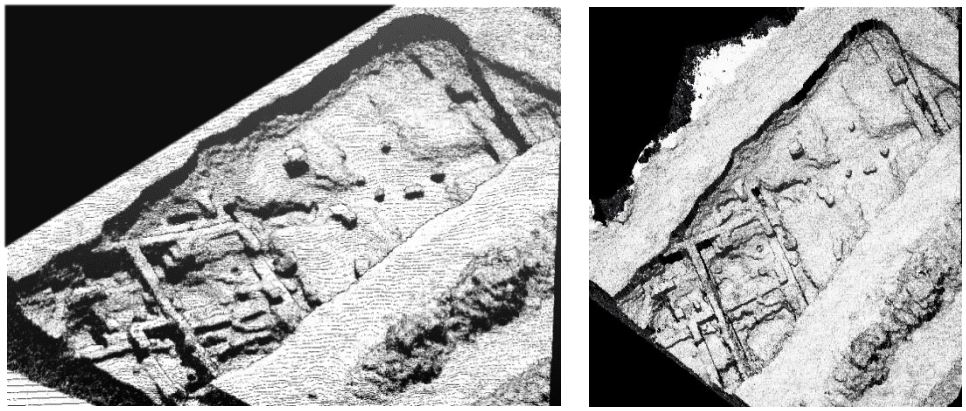


Figure 4.25: On the left the point cloud of the Veio Park small camp and on the right the reconstructed DTM.

It is interesting to notice how the matching parameters intervene in the quality of the final results. Indeed, considering that the images taken from UAS are often lightly blurred and not highly resolute, the choice of the similarity function and optimal

regularization parameters to use in the semi-global optimization step is important and can vary the quality and noise level of the digital model (as shown in the DSM zoom areas depicted in Figure 4.26).

At the same time, with regard to the surface reconstruction noise level, a second issue have to be considered: the pixel locking effect. Therefore, a second digital surface model have been obtained, with the same processing parameters but enabling the sub-pixel symmetric refinement strategy described in paragraph 3.1.9.1

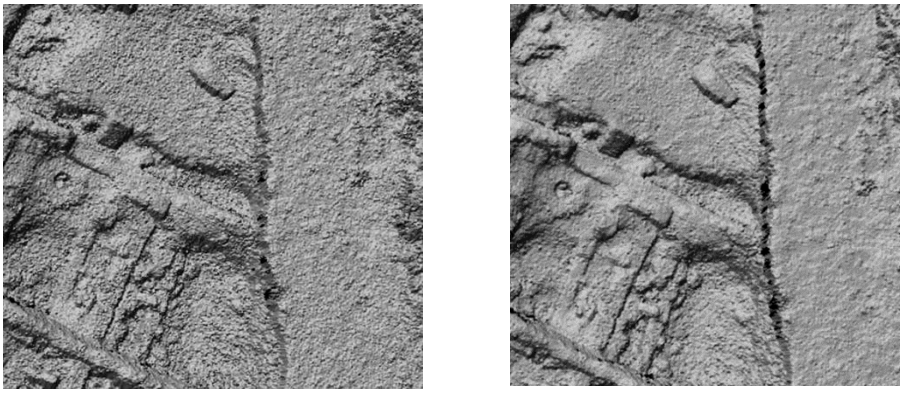


Figure 4.26: A zoom area of the two DSM of the Veio small camp obtained using different semi-global matching parameters combination.

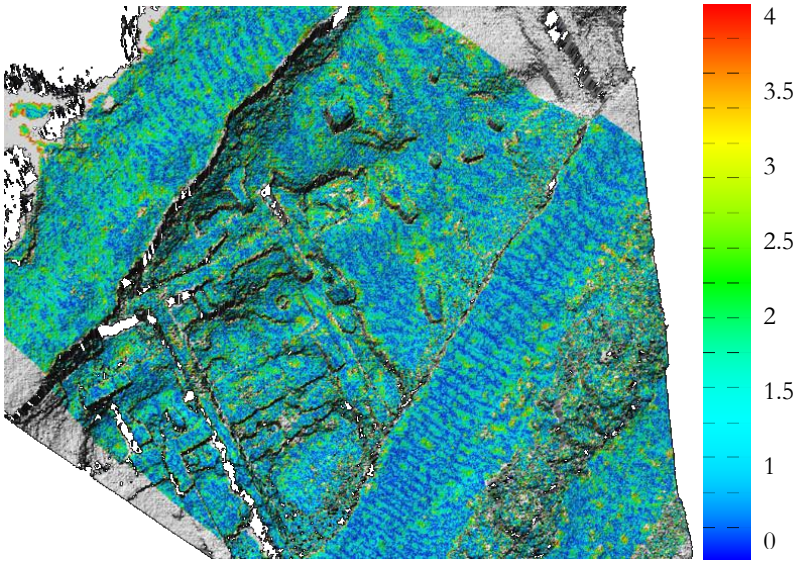


Figure 4.27: Comparison [in cm] between the two digital surface models obtained enabling and disabling the pixel locking decrease strategy.

The comparison between the two results is depicted in Figure 4.27: the staircase pattern, in particular on planar surfaces, of the first model is an evident indication of the pixel locking effect. However, the accuracy of each solution must be also highlight: indeed, despite the just emphasized differences between the results, the two DSM are in agreement showing differences in the range of 1 -3 cm.

A final analysis, as previously said, has regarded the comparison of our result with respect to other digital surface model performed with different matching algorithms and software. The results are show in Figure 4.28 where the proprietary solution has been compared with those obtained by Agisoft Photoscan and DenseMatcher.

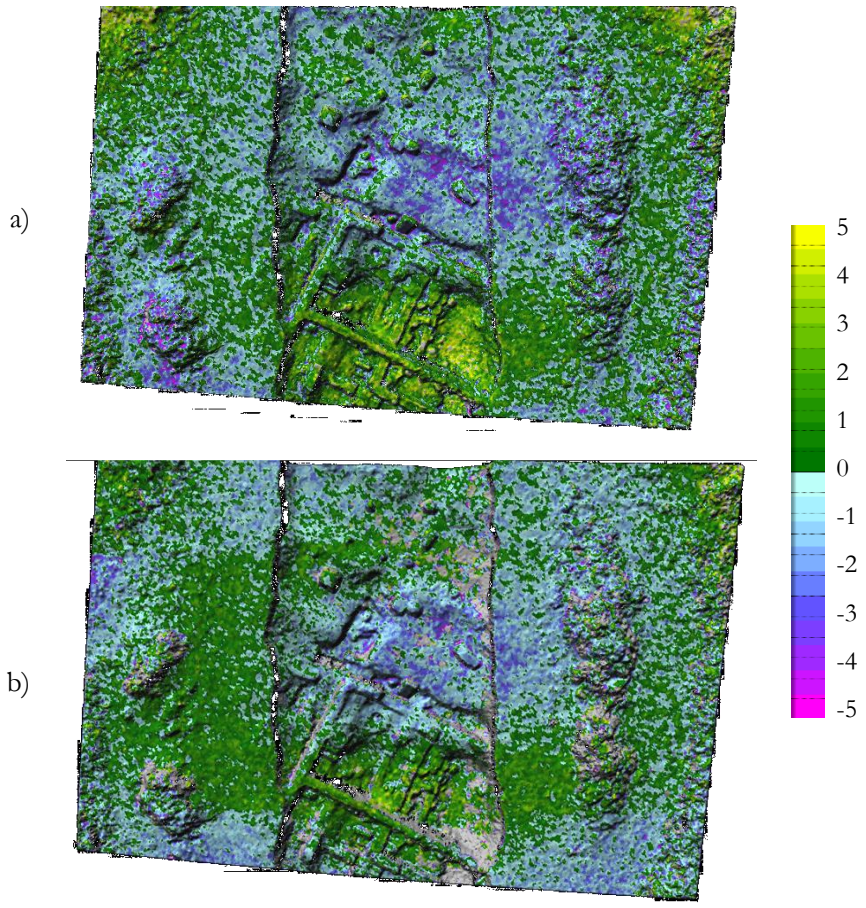


Figure 4.28: Error map of the differences between the DSMs obtained with: a) Object-based SGM algorithm and Photoscan software; b) Object-based SGM algorithm and the DenseMatcher (i.e. LSM algorithm) software.

It is possible to highlight a good agreement between the two different error maps: the higher differences are located, for both the comparison plots, in the same areas that correspondence of the excavation areas. Here, the discrepancies reach also ±4-5 cm. At the same time some differences can be seen in the two error maps and, specifically, in the case presented in Figure 4.28 (a) also the vegetation regions that surround the excavation present higher differences that those shown in the case (b). Therefore, it is possible to assert that a better agreement is shown between the SGM and LSM solutions than respect the Photoscan result even if, it is important noticing that the LSM digital surface model presents some lacks of completeness shown as grey coloured areas. Nevertheless, as shown in Table 4.3, the difference in terms of amount of inlier compared points is low as well as the difference in standard deviation and mean values of the differences.

	Points number	**Mean [cm]**	**Std Dev [cm]**
Object-based SGM – PS	389'337	0.0801	1.85
Object-based SGM – DM	336'555	0.0373	1.67

Table 4.3: Statistics of the DSMs comparisons

CHAPTER 5

TESTS OF STEREO SEMI-GLOBAL MATCHING SOFTWARE CODE FOR 2D DISPLACEMENTS DOMAINS

Introduction

The novel algorithmic extension to a bi-dimensional correspondences research through images was applied and verified in different areas of engineering interest, starting from motion tracking of particles within fluids, to the mechanical behaviour of materials, up to the 2D displacement field reconstruction of rock glaciers surface. Following, a brief description of each application together with the obtained results, will be presented.
What is common to all these applications is the search and the inspection of the continuous displacement field: in fact, the images points tracking through two or more images can give a complete and accurate description of the variations between different temporal epochs. Therefore, the image correlation with the use of the semi-global algorithm can be an appropriate method since that the hypothesis at the base of the algorithm (i.e. the regularity of the displacement field between adjacent pixels) perfectly fits the physical geometry of these problem. The method variation for performing 2D correspondences search has been applied in the applications shown below, and its performance has been also compared with the local LSM method one.

5.1 Strain measurements in material mechanics by image correlation

Knowledge of the microstructure/deformation relationship in materials is a crucial issue in mechanics because microstructure-related strain heterogeneity and damage accumulation control fundamental phenomena of the materials mechanical behaviour, such as fracture, fatigue and creep. Therefore, experimental solid mechanics testing often requires accurate measurements of full-field strain components.
DIC is a non-contact optical technique which allows measurement of displacements and strains in materials. It works by tracking the same points between two consecutive images of the material specimen at different stage of its deformation. The feature tracking is usually achieved using Area Based Matching (ABM), i.e. extracting the image correspondences by evaluating the similarities between grey values. To have more details about the state of the art in DIC applications and methods, an in depth examination was already illustrated in paragraph 1.2.4.
In short, all the different image matching techniques aim at the same result: comparing images that present radiometric and geometric differences due to a relative (three-dimensional) motion between the camera (the observer) and the object (the scene

framed by the camera), and tracking the (two-dimensional) movement of specific elements on the image.

DIC performances depend on the algorithm capability to identify the same feature in different images: well-contrasted and recognizable pattern facilitate the tracking process, achieving good accuracies and high matching reliability. Consequently, specimen usually requires a preliminary surface treatment to ensure a successful imaging acquisition and the subsequent application of the DIC method. The technique involves measurements of the greyscale level at each pixel location of the image thus very well-contrasted images are fundamental for achieving a high measurement accuracy.

In this regard, in order to evaluate the reliability and the accuracy of the novel SGM implementation for 2D displacement solution space, the algorithm was tested for measuring strain fields in some highly deformable materials. Indeed, the implemented SGM code allows constraining the regularity of the displacement field, significantly improving the reliability of the evaluated strains, especially in highly deformable materials.

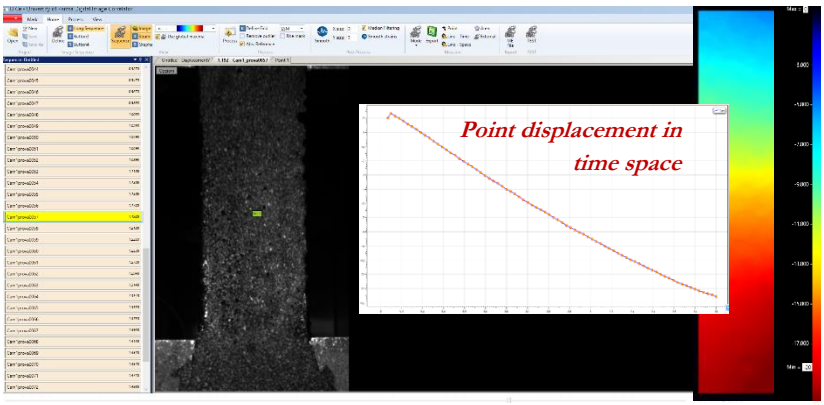


Figure 5.1: Screenshot of DICe, developed at University of Parma, which implements LSM and SGM algorithms to perform the image correlation process.

SGM novel approach is compared with other algorithms used commonly application in DIC applications:

- a proprietary developed Digital Image Correlation System, based on an initial points correspondences research performed using the NCC algorithm, followed by a local Least Squares Matching approach for refining the correlation analysis. It has been developed for providing a dense and accurate full displacement/strain field measurement of composite materials and for detecting materials cracking behaviour. The software (named DICe) allows tracking a dense set of features along the acquired image sequence using an

approximate value of the correlation window position estimated at the previous step (see [22] for more detail). The displacement is computed as the difference of the feature location between each image frame in the sequence and the reference one (which is fixed). The graphic user interface of the software was conceived on purpose to pick the point of interest, choosing which information is requested (horizontal/vertical/shear strains, horizontal/vertical displacements). The strain values are then exported to plot stress-strain responses at the specific chosen point or by means of false colour maps. A screenshot of the DIC in-house implemented program in shown in Figure 5.1.

- the open source code Ncorr [123] that allows a user-friendly, efficient and flexible computation of DIC analysis. The image processing workflow is easy and intuitive and, at the end of the correlation analysis, smooth displacement and strain maps can be obtained. Differently from the in-house developed LSM system, in Ncorr the initial guess obtained by applying the NCC algorithm are used as initial input to the iterative non-linear optimization scheme performed with Inverse Compositional Gauss-Newton (ICGN) method [24].

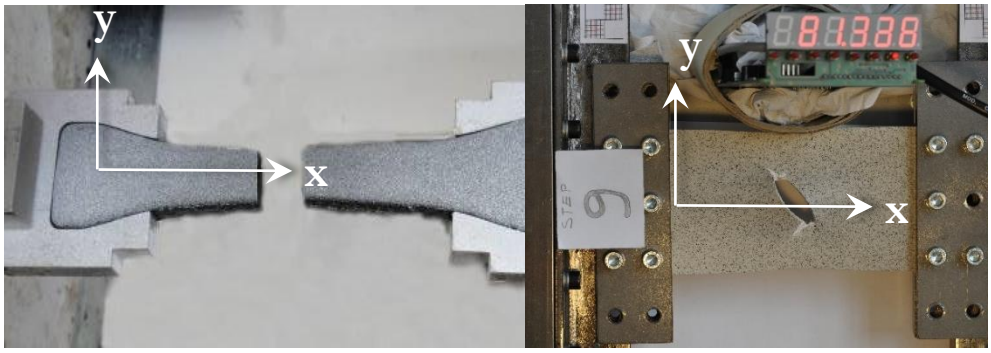
The investigation using the present DIC methods focus on 2D full-field strain maps of plain and notched specimens under tensile loading made of two different highly deformable materials: hot mix asphalt and thermoplastic composites for 3D-printing applications. In the latter specimens, an elliptical hole is introduced to assess the potentiality of the method in capturing high strain gradients in mixed-mode fracture tests.

Since in this tests reference results, for comparing the algorithm performance independently, are not available, the DIC methods accuracy has been evaluated comparing the results of each software (results are presented in the following paragraph 5.1.2) and with FEM (Finite Element Method) simulations.

5.1.1 Materials and specimens

Asphalt mixture is a composite material consisting of interspersed aggregates, asphalt binder and air voids. Its cracking behaviour is affected by its heterogeneity, specifically by the interaction between the aggregates and the mastic. The stiffness of the mastics affects the ability of the mixture to resist permanent deformation and their fracture and fatigue strength. Accurate description of strain evolution and distribution in mastics is essential for revealing significant information on the influence of microstructure properties on asphalt mixture macroscopic behaviour.

Eight asphalt mastics with very different properties were used in this study. The cracking behaviour of the mastics is investigated using a Modified Direct Tension Test (MDTT) performed in a servo-hydraulic load frame to allow the material to stretch up to rupture. A digital Basler piA1600-35gm camera (resolution 1608x1308, focal length 8mm, pixel size 7.4 micrometers, 35 fps@max resolution), directly connected to the testing control system, is located on a support inside the climatic chamber where tests are performed. Since the crack phenomenon is very fast and short-lasting (1-2 seconds), the camera is properly set up to acquire the images in a smaller area of the sensor (1600x500 pixel) reducing the bandwidth required for transmitting each frame and, consequently, allowing a higher frame rate (about 80 fps).
To achieve high accuracies in the strain field measurements, the specimen surface must present a well-contrasted grey scale speckle pattern, easily obtainable by a water paint-based treatment.
Strains are obtained from DIC system, interpolating all the strain values of the grid points defined in the specimen central area. An example of the test configuration at failure is shown in Figure 5.2 (a).



(a) (b)

Figure 5.2: (a) Modified direct tension test for a mastic specimen at failure; (b) Tensile test for a polylactic specimen at failure.

The second tested material is a biodegradable thermoplastic aliphatic polyester derived from renewable resources, called Polylactic acid or polylactide (PLA). PLA is commonly processed by 3D printing. In this work, PLA specimens are produced by a FFF (Fused Filament Fabrication) 3D printing technique, which is based on a layer-by-layer deposition, controlled by a slicing software, of plastic filaments (of 0.3 to 0.5 mm in diameter) supplied by an extrusion nozzle. PLA can be mechanically characterized by a visco-elastic behaviour which can be described for instance by the Maxwell-Wiechert model [188].

Tensile tests are performed on five 3D printed rectangular specimens with a width of 80 mm, a length of 162 mm (clear length 119.2 mm) and a thickness of ca. 2 mm, containing a central hole with maximum size of 30mm and different shape and orientation: (i) a circular hole; (ii) an elliptical hole with aspect ratio 0.2 normal to the load axis; (iii) an elliptical hole, with aspect ratio 0.1, normal to the load axis; (iv) an elliptical hole, with aspect ratio 0.1, 60° inclined with respect to the load axis; (v) an elliptical hole, with aspect ratio 0.1, 45° inclined with respect to the load axis. The load is acting in the direction of the major dimension of the specimens and is applied by displacement control of regular 0.3 mm steps every 30 seconds. Note that in each layer the deposition of plastic filaments follows the 45° direction with respect to the load axis, so as to induce some degree of orthotropy in the 3D printed specimens.
A full-format Nikon D3X (6048x4032 pixels) digital camera, with 105 mm lens and 6 micrometers pixel size, was used for DIC experiment. The camera was opportunely mounted on a tripod, in order to entirely capture the sample, and the images were taken manually at defined temporal instants. Specimens have been treated by applying an irregular painted surface pattern in order to get a non-uniform, well-contrasted coloured surface suitable for performing DIC analyses.
An example of the test configuration at failure is shown in Figure 5.2 (b).

5.1.2 Results

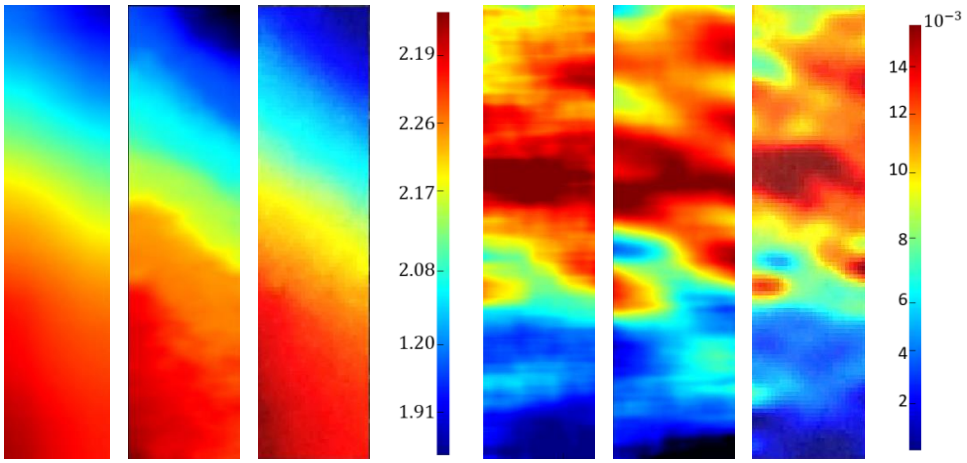
For comparison purposes, the DICe system (using both LSM and SGM methods) and the open source Ncorr code are employed to obtain experimental measures of 2D full-field displacements and strains maps for mastic and PLA specimens. As previously said in the first paragraph, a "ground truth" displacement/strain map for comparison studies was not available, therefore the repeatability evaluation was just performed comparing the three algorithms solutions in the several cases of study.
Deformation components are calculated according to the Green-Lagrange finite strain tensor, which, in the case of small displacement gradient compared to unity, coincides with the small strain tensor. Green-Lagrange strains are obtained by using the four displacement gradients as shown below:

$$E_{xx} = \frac{1}{2}\left(2\frac{\partial u}{\partial x} + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2\right) \tag{5.1}$$

$$E_{xy} = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} + \frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\frac{\partial v}{\partial y}\right) \tag{5.2}$$

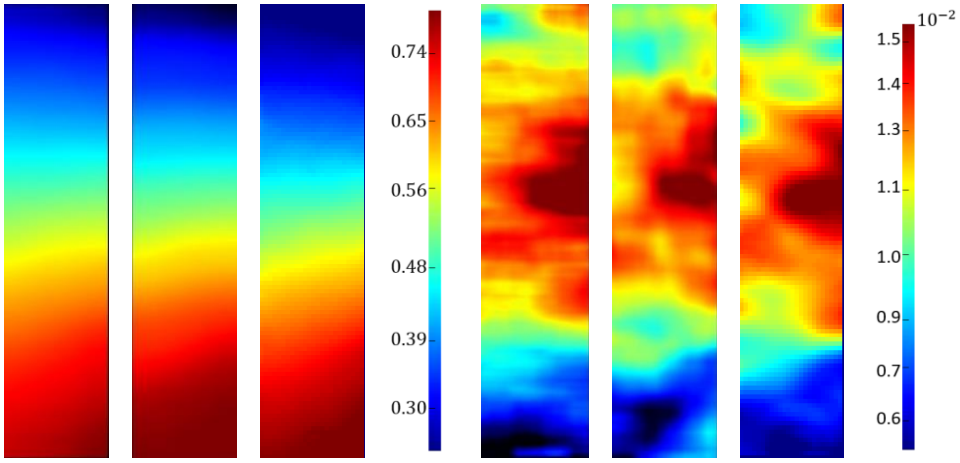
$$E_{xx} = \frac{1}{2}\left(2\frac{\partial v}{\partial y} + \left(\frac{\partial v}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2\right) \tag{5.3}$$

In the following some selected false colour displacements and strains maps are reported (Figure 5.3-5.7 - please note that X direction is aligned with the loading axis). The results correspond to an applied nominal strain of about 30% of the strain at failure.



(a) (b)

Figure 5.3: Asphalt mixture sample 1: results, from left to right, with LSM, SGM, NCORR. (a) Map of the displacement along X axis [mm]; (b) Map of the strain along X axis.



(a) (b)

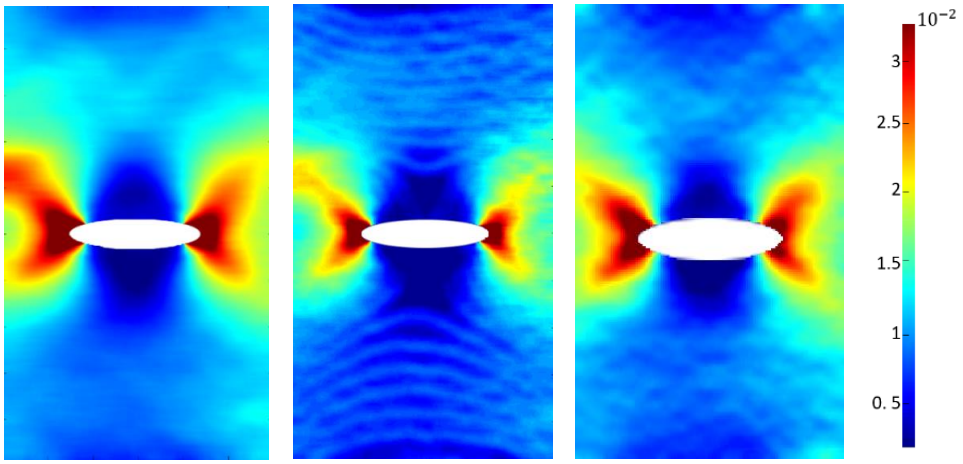
Figure 5.4: Asphalt mixture sample 2: results, from left to right, with LSM, SGM, NCORR. (a) Map of the displacement along X axis [mm]; (b) Map of the strain along X axis.

In the case of PLA specimens, the contour maps for the elliptical hole, with aspect ratio 0.1, normal to the load axis, and for the elliptical hole, with aspect ratio 0.1, 60° inclined

with respect to the load axis are reported in Figures 5.5-5.7, where the longitudinal nominal strain is equal to 2.01% and 3.36%, respectively.

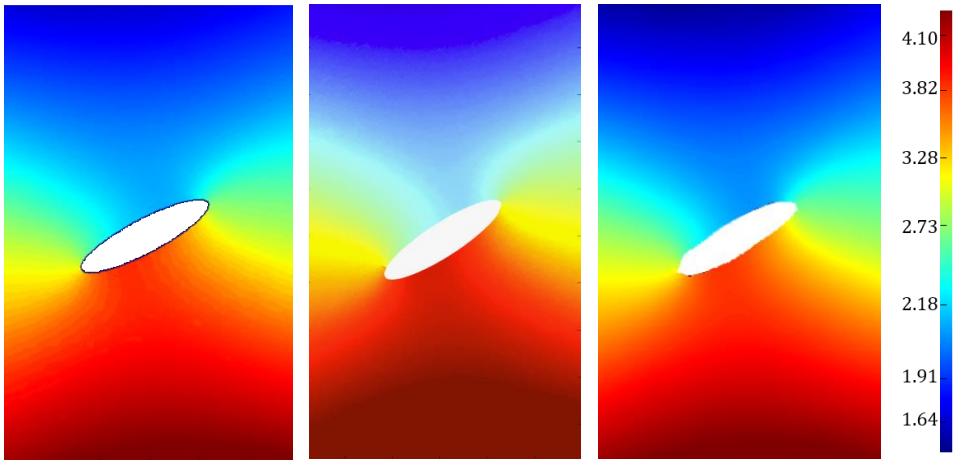
The results presented in Figure 5.3 and 5.4 show that the SGM solutions seem to be in better agreement with the Ncorr results rather than with the LSM maps. The SGM regularity constraint on neighbourhood pixels displacement allows to reduce the streaking effect presented in the LSM results, showing more detailed and accurate results.

As regard the thermoplastic aliphatic polyester specimens results (Figure 5.5, 5.6 and 5.7) a sort of waved/staircase pattern is clearly visible in the SGM strain maps; the same effect is visible, with lesser extent, in the Ncorr results, while it is absent in LSM solutions. This is likely due to the pixel locking effect. If the maximum deformation to be measured is less than 30% the use of simple similarity cost function (like SAD) can be used without any trouble with SGM algorithms, at least if small template sizes are used. However, with these approaches and as far as the displacements are limited, pixel-locking effects, which often occur during the parabolic sub-pixel fitting of the cost function minimization, can produce unwanted biased results.



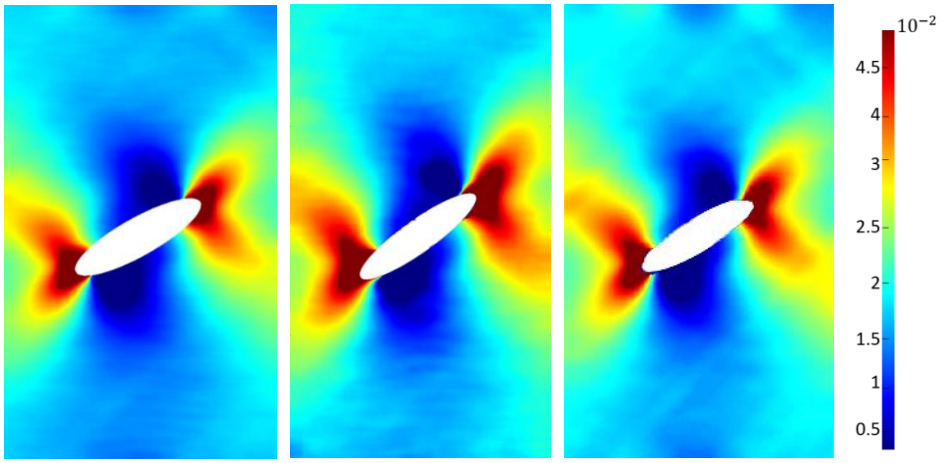
(a) (b) (c)

Figure 5.5: Map of the strain along X axis of the elliptical hole with an aspect ratio 0.1 normal to the load axis. Results computed, from left to right, with (a) LSM, (b) SGM, (c) NCORR.



(a) (b) (c)

Figure 5.6: Map of the displacement along X axis[mm] of the elliptical hole with an aspect ratio 0.1, 60° inclined with respect to the load axis. Results computed, from left to right, with (a) LSM, (b) SGM, (c) NCORR.



(a) (b) (c)

Figure 5.7: Map of the strain along X axis[-] of the elliptical hole with an aspect ratio 0.1, 60° inclined with respect to the load axis. Results computed, from left to right, with (a) LSM, (b) SGM, (c) NCORR.

5.1.3 Comparison with FE analysis and discussion

To provide numerical comparison data, Finite Element (FE) analysis of the PLA plates containing an elliptical hole were performed. Eight node finite elements in plane stress conditions are used. The FE model is loaded in tension by imposing a null axial

displacement at one end and a uniform displacement at the opposite end. Linear and geometrically non-linear analyses were performed.

The elastic equilibrium solution of the plane problem is related to prescribed boundary conditions, with no traction and no body forces being applied. The displacements/strains solution is independent on the Young modulus E (while stresses clearly depends on E), but it does depend on the Poisson ratio ν (a value of $\nu = 0.45$ is adopted, which is typical of plastic materials and it is consistent with the experimental measurements of transversal contraction). This holds true also for the geometrically non-linear analyses with finite strains. The distribution of Green-Lagrange local strain components along the direction of the major axis of the elliptical hole as a function of the distance from the ellipse root is compared in Figure 5.8 for FE analysis and DICe using the SGM algorithm.

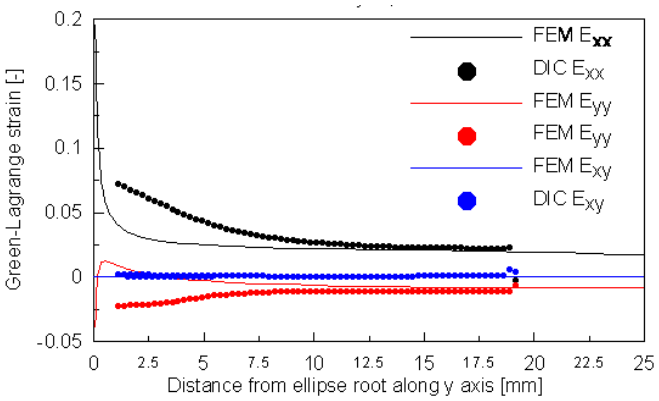


Figure 5.8: Green-Lagrange local strain components along the major axis for the elliptical hole with an aspect ratio 0.1 normal to the load axis (axial normal strain = 2.01%).

Undoubtedly, as far as high deformable materials are concerned, matching algorithms can experience a lot of different issues: the transformation considered between master and slave image can be inadequate to model localized deformations. Numerical problems, such as pixel-locking, can lead to systematic errors: if the maximum deformation to be measured is less than 30% the use of simple similarity cost function (like SAD) can be used without any trouble with SGM algorithms. However, with these approaches and as far as the displacements are limited, pixel-locking effects can produce unwanted biased results. On the other hand, taking into account that the displacement solution should present continuity to some extent (even if fractures occur, the discontinuity is limited to one specific direction), the use of regularization constrain in the matching technique is an efficient approach to improve the final results reliability. In this context the use of the new SGM algorithm developed by the authors seems to

adequately address such requirement. However, if higher deformation rates are expected, SGM algorithms should be discarded: the use of more complex similarity functions that consider also a patch deformation model, even if in principle can be implemented, will make the entire process computational unfeasible. At the same time, the simplistic regularization function in eq. (3.8) can be unsuitable to constrain the displacement field of the solution which, with high deformation, can show more likely localized effects. Providing appropriate regularization functions and penalization terms can be tricky and introduce might bias the final results. In this context we found that the use of other cost function (i.e. Census transform cost function [204]) improves to some extent the results.

A final remark concerns the possibility to help the matching algorithm through the change of the master (reference) image during the DIC processing: in this way, comparing more similar images should improve the matching algorithm performance (no matter if LSM or SGM is used). Even if, in principle, that is true, the user have to consider that, in this way, accumulating the displacements computed from every reference and slave image, also the errors accumulate; in the end, it is likely that the entire process becomes less accurate than the case where the master image is fixed.

5.2 Rock glacier monitoring system

Monitoring the surface creep of mountain permafrost is important to understand the effect of ongoing climate change on slopes dynamics. Rock glaciers are widespread landforms that can show rapid acceleration and destabilization [49]. In heavily anthropic areas like the Alps, the accelerating creep of perennially frozen talus/debris with high ice content will probably become an increasing problem, notably for human infrastructures [74]. However, traditional techniques (e.g. topographical survey) cannot easily be applied in such scenarios: for example, the glacier surface is rough and presents hazards like crevasses. Only an operator with adequate training is able to realize a survey, often with some risks to his safety.

An evaluation of movements and volumetric changes of an Italian rock glacier, obtained by multi-temporal analysis of UAS images over the period 2012-2015, has been investigated. The study area is located in the western Alps at the head of the Valtournenche Valley (Valle d'Aosta, Italia) on the Italian side of Matterhorn (see Figure 5.9). The body of the rock glacier is composed by two lobes, spanning an elevation range between 2600 and 2750 m. It is nearly 400 m long, between 150 and 300 m wide and has an apparent thickness (based on the height of the front) of 20-30 m.

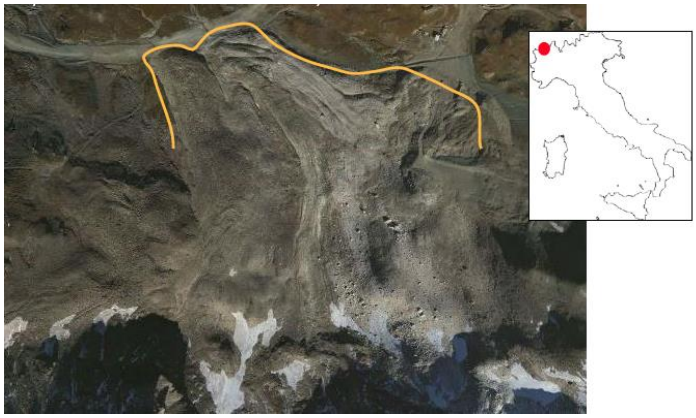


Figure 5.9: The study area location and (Valle d'Aosta, Italia) and the body of the rock glacier.

Since 2012, the surface movements of the glacier are monitored by ARPAVdA (*Agenzia Regionale per la Protezione dell'Ambiente Regione Autonoma Valle d'Aosta* - Environmental Protection Agency of Valle d'Aosta) as a case study for the possible impact of climate change on high-mountain infrastructures: in fact, this glacier juts on a ski slope of the Cervinia resort, causing every year maintenance issues to professionals. For these reasons, a multi-approach monitoring system, based on repeated UAS-photogrammetry and GNSS (Global Navigation Satellite System) survey, has been setup.

The current dataset of observation consists of three UAS flights (October 2012, October 2014 and July 2015) and three GNSS campaigns (mid August 2012, 2013, 2014). The fixed-wing Swinglet CAM produced by SenseFLY has been used for the photogrammetric UAS flight. It was equipped with a 12 Mpixel CANON IXUS 220 HS camera for the 2012 flight, and with a 16 Mpixel CANON IXUS 125HS camera for the 2014 flight. The former flight was performed at 150 m height with a longitudinal overlap of 60% and a sidelap of 70% between adjacent strips, with a resultant GSD (Ground Sampling Distance) of 5 cm/pixel. The number of images acquired and used in the bundle block adjustment was 110. For the 2014 flight, the same GSD was obtained changing slightly the flight altitude. At the same time, to make the image block more rigid, the longitudinal and side overlap were respectively of 80 and 85%. Given the flight characteristics, the images acquired in the photogrammetric block were 239. Table 5.1 summarises the design parameters of the two UAS flights.

The GNSS data can be used as ground truth for validating the displacement obtained by orthoimage analysis and DSM comparison and check the accuracy of the monitoring system.

	2012	2014
Date	October 24th	August 18th
N° images acquired	110	239 (two flights)
N° images used	110	239
Side overlap	70%	80%
Longitudinal overlap	60%	85%
Ground resolution	5 cm/pixel	5 cm/pixel

Table 5.1. Summary of the UAS flights characteristics.

5.2.1 Measurement and analysis of the displacement field

The reconstruction of the rock glacier surface movements is obtained by comparing the orthophotos and the DSM of the three UAS photogrammetric surveys of the investigated area. The photogrammetric workflow has allowed to obtain three raster DSM (with a cell size of 20 cm), and three different sets of orthophotos of the inspected region, with 5 cm pixel size. The orthophotos were analysed to identify the rock glacier displacements using two different methods:

- a manual identification of well-recognizable points (on the glacier surface) on the orthophotos of three epochs (see Figure 5.10);
- an automatic tracking method able to recognize a dense grid of corresponding points between the images (through image correlation analysis).

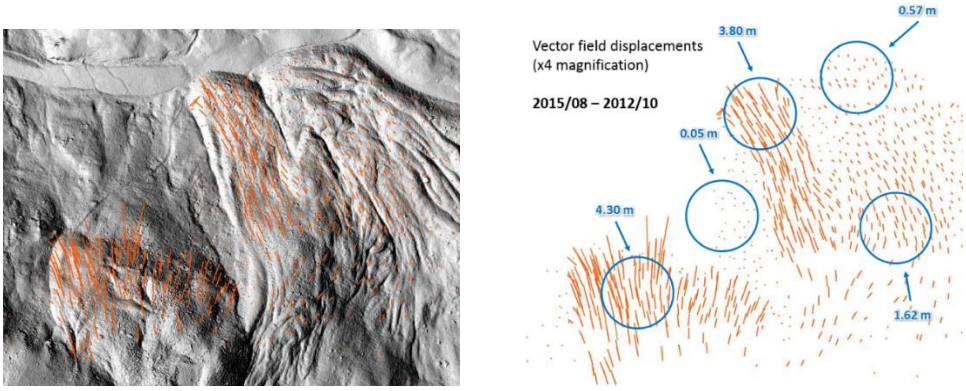


Figure 5.10: Manual Measurement of the Displacements: Vector field displacements (x4 magnification) above the period 2015/08–2012/10.

On one hand, the manual measurements are capable to describe all the survey area and allow to better understand the displacements dynamics; however, this activity is very time consuming: the measurement of ca. 1.000 points to describe 0.25 Km^2 required 6 -8 hours of operator work. On the other hand, displacement and/or deformation measurements of the object surface can be obtained by automatically tracking corresponding points between the images. Features tracking can usually be achieved using Feature (FBM) or Area Based Matching (ABM) algorithms and a much more dense displacements map of the region of interest (with little or no working load for the operator) can be reconstructed.

The automatic procedure can be applied directly to the orthophotos. Usually, image matching software packages work on 8 bit indexed images (if RGB images are used, the software generally converts them in gray scale). However, the application of ABM algorithms on orthophotos can lead to inaccurate and erroneous matching results since image texture changes, low contrast regions, radiometric transformation problems and, especially, slope illumination variations over the monitored period can worsen the matching algorithms performance (as it be described in the following paragraph 5.2.2). Another option is to represent the local shape of the DSM converting its height map to some other representation (e.g. with a shaded relief map) and exporting it to a common image format (see Figure 5.10). Also in this case, however, the raster product has an 8-bit colour depth. Moreover, as far as hill-shading rendering is concerned, being such technique based on the computation of the angle between DSM surface normals and incident light rays direction, height data noise can be amplified by the procedure. On the other hand, smoothing effects, where shape discontinuities occur, should be expected. A better choice would probably be the use of particular image filters such as the Wallis filter [196] to improve the height map local contrast (Figure 5.11).

However, the best way to address the problem would probably be to match the two height maps directly: even if lighting or texture changes occurs in the monitoring period the matching algorithm would not be affected working only on the shape of the glacier surface. Although the vast majority of commercial or free software expect to work on 8 bit image data, both DenseMatcher [146] and the new SGM algorithm are able to perform the correlation process taking in input also 32- bit floating point rasters, so all the previous limitations are overcome. An automatic tracking of the displacements, using both the orthophotos and the DSMs will be presented in the next section.

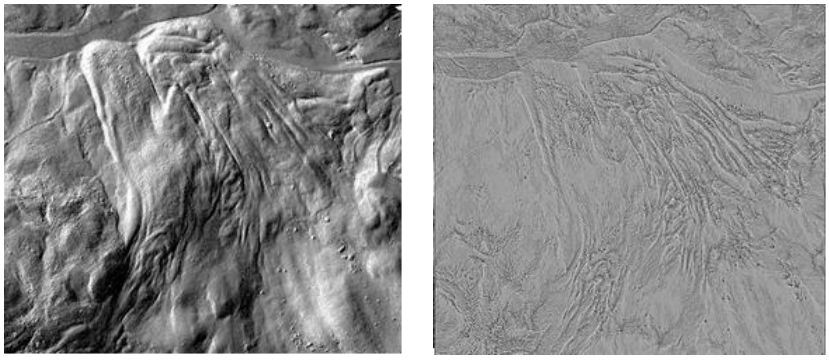


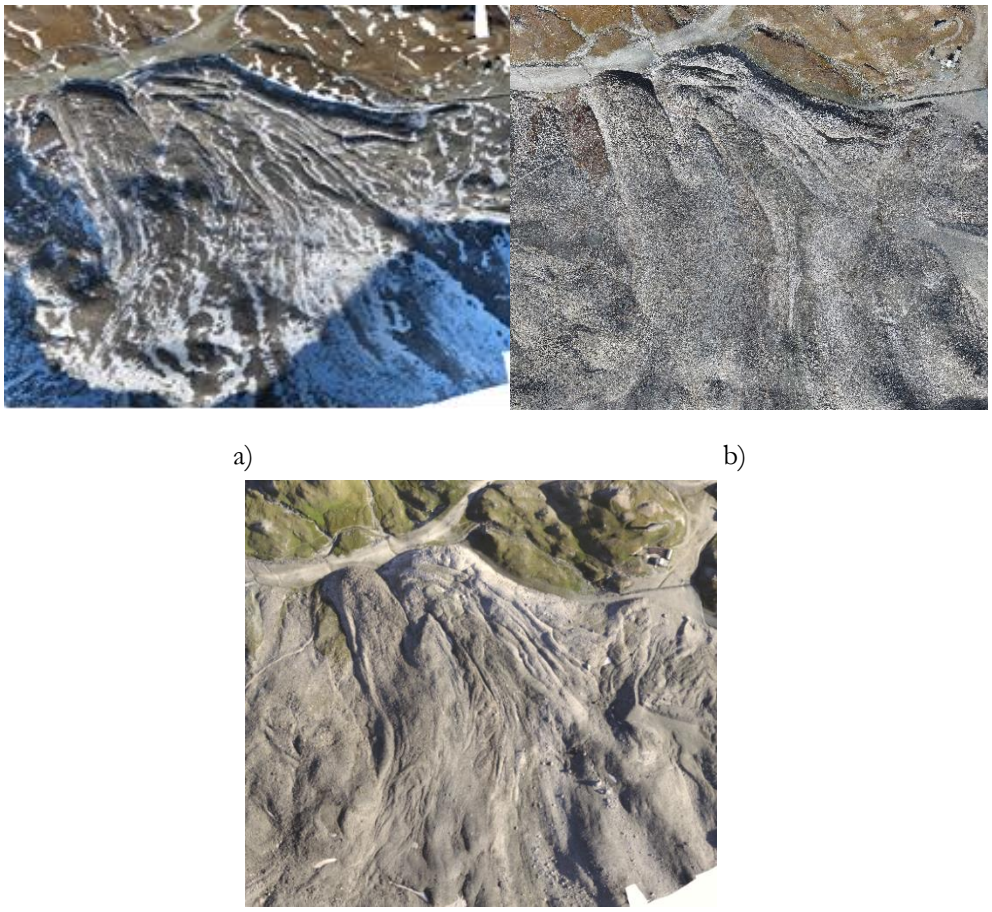
Figure 5.11: On the left: Shaded DSM of the October 2012 flight; on the right: Elevation raster map, after the Wallis filter application, of the October 2012 DSM.

5.2.2 Automatic displacement measurements on orthophotos

Observing the evaluated ortho-image pairs (shown in Figure 5.12), the main issue for a matching procedure is represented by the illumination conditions of the slope, which change drastically over the monitoring period, producing significant contrast and brightness variations between the images. Moreover, the weathered surface of the rock glacier can produce relevant texture changes of the surface. Taking into account this, a first series of tests has been performed for evaluating and verifying the ABM algorithm accuracy and precision on the ortho-images matching process. In order to compare the displacement maps with the subsequent DSMs analysis, the results obtained using a 20 cm pixel size orthophoto was used (instead of the original 5 cm high-resolution orthoimages).

The results, illustrated in Figure 5.13, show a significant number of outliers (out of range red areas), probably due to false positive in the low contrasted regions. Indeed, the evident contrast and illumination changes, together with the presence of snow-clad areas, cannot be identified reliably by the correlation algorithm, producing erroneous matched points. At the same time, it is worth noting that, where the images radiometry shows suitable characteristics for the ABM application, the results accuracy is very good, especially considering that, in these comparisons, a pixel size of 20 cm was used. The results have been indeed validated against a GPS survey on the rock glacier that has consisted, in a first stage, in the points materialization by fixing well-recognizable target on glacier stones and, in a second moment, in the points measuring survey which was carried out several times, i.e. before each photogrammetric flight. The comparison between displacements obtained from the automatic comparison and the GNSS measurements has shown a final RMSE of about 10 cm (see Figure 5.13 right). However, these results are validated against only 28 points of the 48 GNSS total

measurements (just 58% of the control points are in agreement with those measured with the GPS campaign). The remaining 20 check points were discarded in the analysis since they fell in regions affected by outliers.



a) b)

c)

Figure 5.12: a) October 2012 orthophoto; b) October 2014 orthophoto; c) August 2015 orthophoto.

To provide a more reliable comparison of the two epochs (possibly in all the extent of the rock glacier) a different method, described in the next paragraph, which considers the information derived from the raster elevation models, has been investigated.

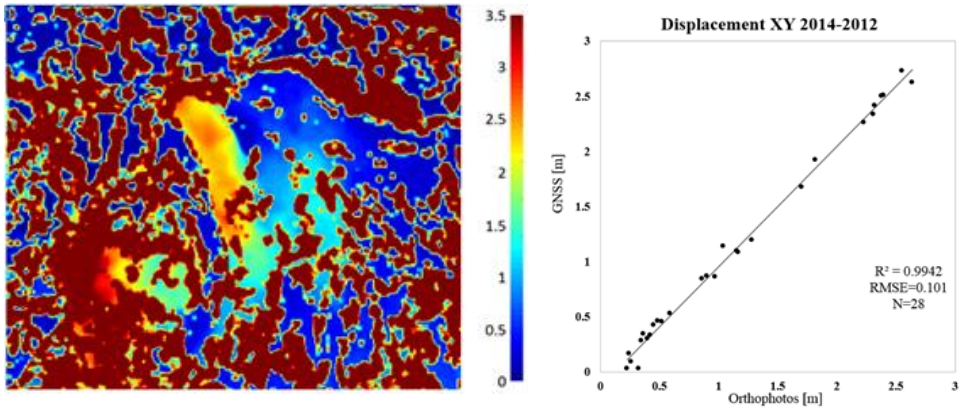


Figure 5.13: On the left: 2D Displacement [m] coloured map obtained by ABM process on the orthophotos; on the right: Scatterplot between displacements obtained from the automatic comparison of UAS orthophotos and GNSS measurements on 28 points.

5.2.3 Automatic displacement measurements on raster elevation models

Forasmuch as the rock glacier is subjected to sliding, the local surface shape does not change so much, while the colour variations of the surface elements can be significant from one epoch to the other. As already pointed out, the proprietary image matching techniques can use floating-point rasters and the identification of homologous features can be performed on the DSM directly. Such solution represents the simplest and most efficient way to overcome the limitation represented by the use of long time-separated data.

It is important to highlight that the application of the matching algorithms to the elevation data has requested the use of a big image block. Indeed, the level of detail of the DSM is not high enough to represent clearly the smaller blocks and, in analogy with traditional image matching techniques, the area inside the matching template should not be "flat", unless high uncertainties are considered acceptable. For this reason, the use of large templates is necessary to perform an accurate and reliable identification of the same areas on the two investigated raster. In all the tests performed (presented in the following section) a template size corresponding to a ground size of ca. 6÷10 m has been used, because it was verified as the template size that produces the best results.

The automatic displacement measurements final validation has been performed by comparing the achieved displacement vector with those acquired with the manual

measurement process: such comparison, being measured on 5 cm orthophotos resolution, can be considered the most suitable and accurate comparison data at the time being.

5.2.3.1 Data analysis and results

The first series of tests have regard the following two monitoring periods: a two-years period between October 2012 and October 2014, and a one-year survey between October 2012 and August 2015). With regard to the requirements of ARPAVdA, a periodic survey on a monthly basis will be also expected, in order to better investigate the seasonal velocity changes, especially during the summer period. During the winter, being the rock glacier body completely covered by the snow, the proposed methodology cannot be applied and different techniques should be implemented (e.g. GNSS survey on large rock blocks). At this time, being such solutions very expensive, and expecting little or no movement in winter, the monitoring activities are planned only during the spring, summer and autumn seasons.

With the proposed automatic methodology, a very detailed description of the creep/displacement behaviour of the glacier can be produced. The results has been obtained using the LSM proprietary algorithm [44], and the novel Semi-Global strategy, with the aim to compare the solutions and thus verify the SGM reliability. Figure 5.14, and 5.15 describe respectively the rock glacier displacements occurred in different time periods, combined with the relative displacements scatterplot in order to validate the automatic measurements against the manual ones.The measurements validation has been performed on *785 points* for the *LSM* solution and *872 points* for the *SGM* one, in the case of two-years monitoring activity and on *881 points* (*LSM* solution) and *996 points* (*SGM* solution), in the one-year monitoring.

The results are very good (Table 5.2): the R^2 correlation index indicates a good correlation between the automatic and the manual measurements in both cases, and the Root Mean Squares Error (RMSE) is also good, showing values lower than 12 cm, which approximately corresponds to half pixel of the analysed raster DSM cell size (of 20 cm). In other words, in these two cases, the methods have demonstrated that dense displacement fields, by tracking automatically homologous areas in raster DSMs, can be obtained with accurate and reliable results and sub-pixel precision. At the same time, it is interesting to notice that the SGM strategy achieves, in all the investigated cases, greater inlier percentage: the introduction of the displacements regularization constraint allows to reduce the presence of mismatches and outlier and thus the statistic can be computed considering bigger data samples.

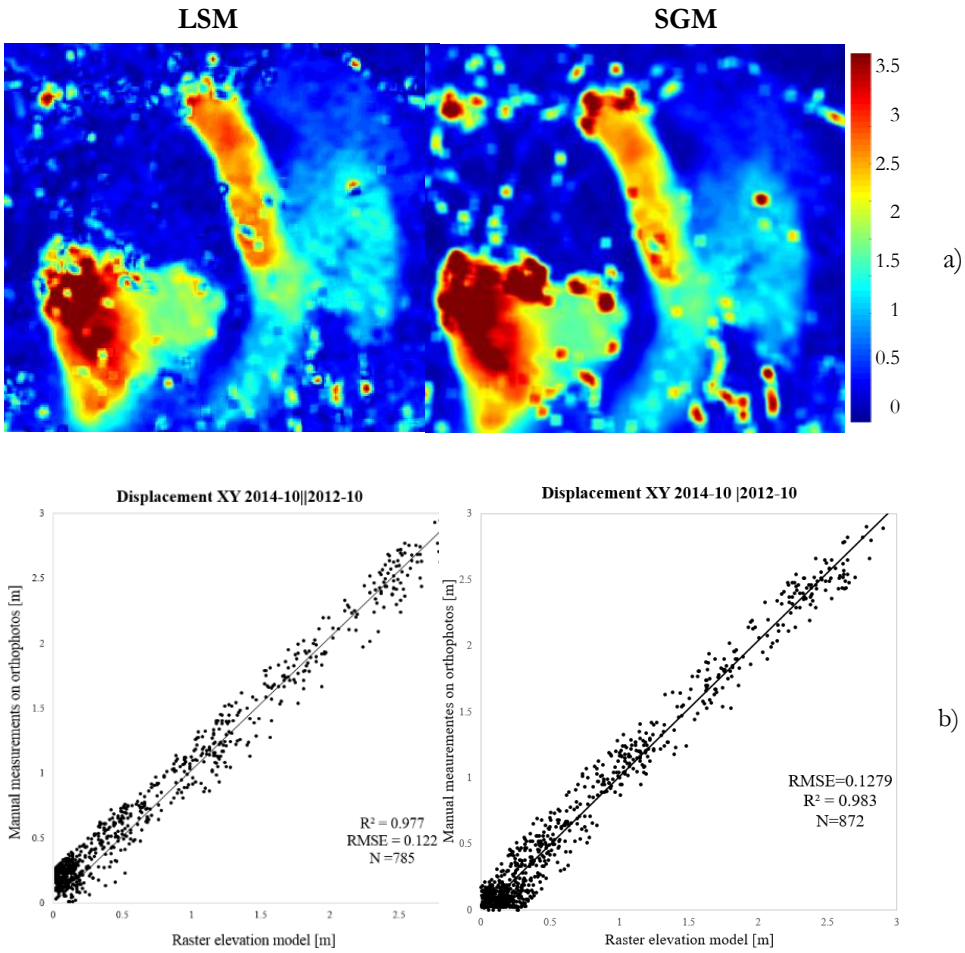


Figure 5.14: Two-years monitoring activity. a) 2D Displacement [m] coloured map calculated automatically with LSM and SGM algorithm; b) Scatterplot of displacements obtained from the automatic comparison of DSM and the manual displacement.

		RMSE	R^2	N
Two-years monitoring	*LSM*	0.122	0.977	785
	SGM	0.127	0.983	872
One-year monitoring	*LSM*	0.097	0.948	881
	SGM	0.128	0.940	996

Table 5.2: Summary of the comparison statistics between the automatic DSM comparison technique and the manual identification of the displacements.

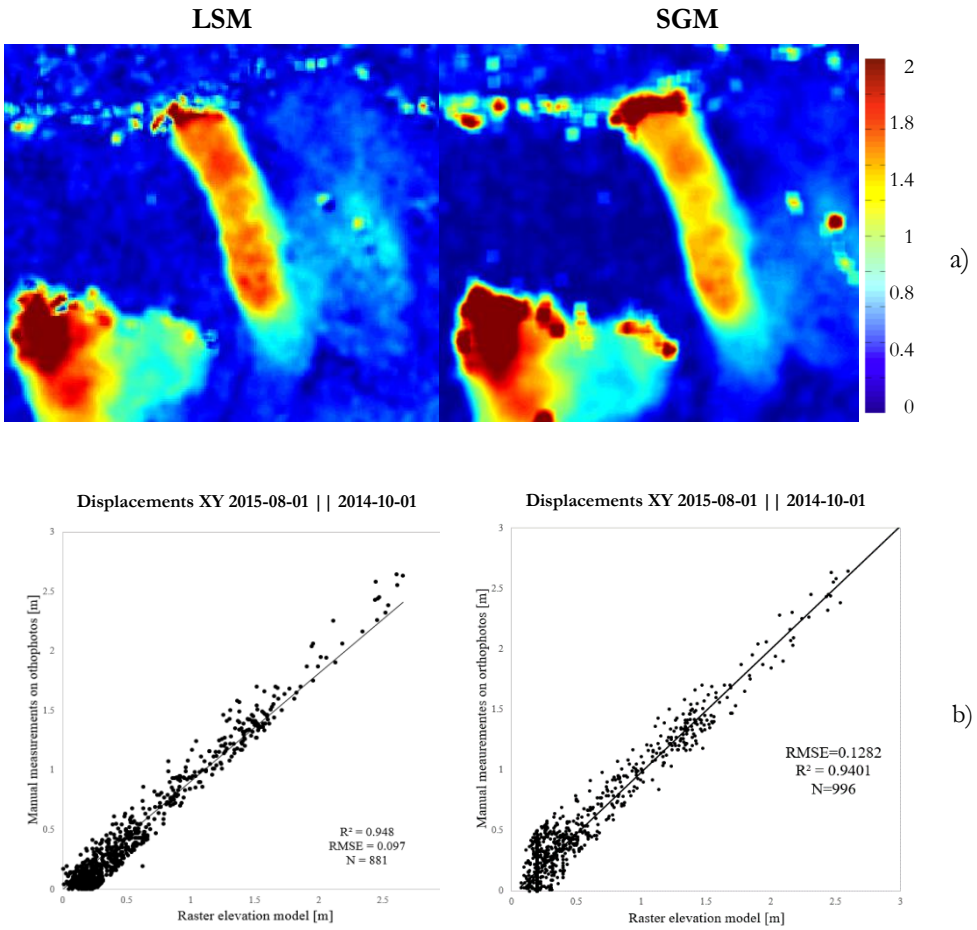


Figure 5.15: One-year monitoring activity**.** a) 2D Displacement [m] coloured map calculated automatically with LSM and SGM algorithm; b) Scatterplot of displacements obtained from the automatic comparison of DSM and the manual displacement measurements.

As far as the monthly/seasonal surveys are concerned, a new DSM restitution, characterized by smaller raster cell size will be provided in the future: the expected displacements range is much lower than the other ones (contained within 20 cm), and the current DSM cell size is too big to allow distinguishing the small movement of the rock glacier.

5.3 Particle Image Velocimetry analysis

Particle Image Velocimetry (PIV) is a non-intrusive whole-flow-field technique that provides instantaneous velocity vector measurements in a cross-section of a flow. Nowadays, PIV is widely used in fluid mechanics applications and problems (diagnostics into flow, turbulence, microfluidics, spray atomization, combustion processes, etc.) where unsteady and separated flows are predominant.

The fluid is seeded with tracer particles which, for sufficiently small particles, are assumed to faithfully follow the flow dynamics; the section to be examined is illuminated with two consecutive and closely spaced laser light pulses (few nanoseconds), converted by means of an optical apparatus in light blades. The particles reflect the light, which is captured by a CCD or CMOS digital camera with the use of a synchronizer: the camera is able to capture each light pulse and the two particle images captured at time t and t' and stored on separated frames of the CCD or CMOS sensor. This results in two images of the particles position at two different, very close time instants (an in depth explanation of the subsequent stages on image acquisition in PIV is presented in [198]).

Comparing the two images, the particle displacement vector field on the plane of the light blade can be obtained. Assuming that a correctly seeding has been chosen, the particles will follow the motion of the fluid. Dividing the displacement by the time delay between the two images, the velocity field of the flow is obtained (it will be much closer to the instantaneous speed as far as the time interval is small). Considering the image matching aspects, the most important issue becomes the choice of the correlation method or function to use, for obtaining a reliable and dense displacement field, also when a low density or a poor quality distribution of the particles within the fluid occurred. Rapid temporal or spatial changes (transition from laminar to turbulent flow, coherent structures, pitching airfoils in transonic flows with shocks, rotors, etc.) can generate sudden flow variations, producing for example air bubbles in the fluid or noisy-blurred image areas.

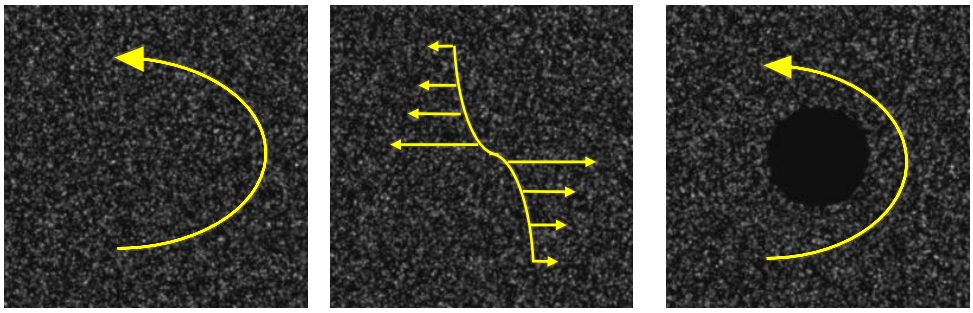
The most commonly used PIV-images correlation algorithm is cross-correlation function, which represents the heart of most of the commercial software. However, the use of correlation method based on Cross Correlation (CC), Normalized Cross Correlation or Fast-Fourier Transform (FFT) algorithms require the use of a search window to perform the correlation between the image pairs. Since usually, in these approaches, the disparity values inside the template are supposed to be constant, the use of big correlation windows, when the fluids investigated present very strong local inhomogeneity in terms of velocity fields, should be discarded. At the same time, in

order to have an high density displacement vector field, small template size should be used. However, in PIV applications, reliable results (in term of points/particles correspondences) are generally achieved by using large window size (32x32 or 16x16 pixels): the template radiometric content does not present optimal characteristic for the matching algorithm performance and this makes essential the use of big block size. In this regard, the use of a Semi-Global method could represent a valid alternative: SGM allows to perform a pixel-wise matching supported by a smoothness constraint. This allows obtaining very dense measurements field (it is possible to obtain one displacement measure per pixel) and, furthermore, thanks to the displacements field regularization function it is successful also in flow fields where the radiometric content is rather poor and rapid temporal and spatial changes intervene.
The novel SGM algorithm has been tested with PIV datasets, in order to obtain a dense field of fluid velocity measurements. The results reliability was investigated by means of direct comparison with some analytical calculated data (synthetic PIV images) and the TSI PIV System commercial software [189](which performs the image correlation process with a classical local method).

5.3.1 PIV tests cases

The presented studies have been regarded two different PIV datasets: in a first stage computer generated flow field has been analytically simulated and the relative synthetic image sequences have been rendered. In Figure 5.16 the three synthetic case of study are shown and, for each case, the simulated flow field is highlighted with a yellow arrow.

SYNTHETIC DATASETS

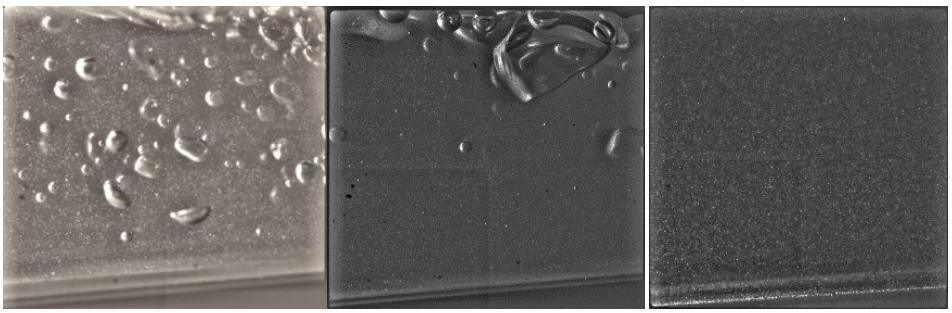


A) B) C)

Figure 5.16: The analyzed PIV syntethic datasets.

The three test cases has been opportunely chosen in order to simulate very different velocity field and thus covering a wide range of different situations.
The use of controlled computer generated dataset has allowed a numerical investigation of the proprietary algorithm performances; after that, a second analysis has regarded the evaluation of real PIV dataset. Three real case of study has been considered: the image sequences have been obtained with the Insight v.3.5 (TSI Inc.) PIV System. TSI was the first company to offer Particle Image Velocimetry technology for flow measurements; the equipment used for the tests is based on a laser imaging technique that combines the accuracy of nonintrusive point measurements with the global flow imaging capability of flow visualization to obtain time-resolved, instantaneous velocity information over an extended (customizable) interest region of the flow. The pulsed laser beam "freeze" the particles location in the planar measurement region; subsequent laser pulses are separated by a defined time (Δt). For each laser pulse one digital image is acquired and transferred to a computer for the image processing: at this time images are divided into many small regions and particles X–Y displacements between images is determined through a cross-correlation algorithm. Finally, the velocities are found by dividing the particles displacements by the time between pulses. The process is repeated for all region until the instantaneous velocity field is reconstructed for the whole image. The software allows, at the end, obtaining and visualizing the planar velocity vector fields and provides a temporal evolution of the flow characteristics.

REAL DATASETS



(D) (E) (F)

Figure 5.17: The real PIV image datasets.

Images were acquired with the 630049 PowerWiev Camera (of the TSI System) which has a 2048 x 2048 format size and a squared 7.4 μm pixel size.
The final results have been obtained with Insight v.3.5 and the related global flows have been calculated considering an image grid size of a 16 × 16 pixel. In Figure 5.17 the

three datasets are shown: the low contrast and brightness that characterize the image radiometric content has made necessary the application of the Wallis filter for improving the visibility of the fluid particles.

5.3.2 Results

In this paragraph, the results that has been obtained applying the proprietary algorithm in the PIV analysis will be presented.
First, the method accuracy has been investigated considering the synthetic flow fields: in this case, being known the analytic vector displacements field between each image pair, the direct comparison with the reconstructed 2D parallax maps, obtained using the SGM algorithm, was possible. The SGM particles flows are highlighted in the maps shown in Figure 5.18 were, for each computer generated dataset, the displacement vectors of the whole area of interest are depicted. The SGM reconstructed particles motions has been overlapped on the synthetic solutions for evaluating the agreement between the two datasets: for each case Figure 5.18 shows also, in a zoom region of the SGM vector maps, the coherence between the achieved displacement data (in blue arrows) and the reference displacement field (in green arrows).
The SGM results seems to be in a good agreement with respect to the mathematical solution, producing high density displacement maps: the achieved flow fields has been computed pixel-wise.
In order to further examine the method accuracy, an evaluation of the mean and standard deviation values of the two solution differences has been computed. The results are shown in Table 5.3. The presented data are satisfying: in all the case of study the SGM method reaches sub-pixel accuracies; at the same time cases (A) and (B) show better results with respect to the (C) one. This is due to the presence of the central hole (see Figure 5.16 (C)): in this area, where there is not particles transit, the SGM strategy has found erroneous solutions which have led to an overall worsening of the reconstruction accuracy of the whole displacement field.

		Mean [pix]	Std. Dev [pix]
DATASET	**A**	0.17	0.087
	B	0.05	0.098
	C	0.31	0.26

Table 5.3: Comparison between the reconstructed and the reference vector flow fields for the synthetic PIV datasets evaluation.

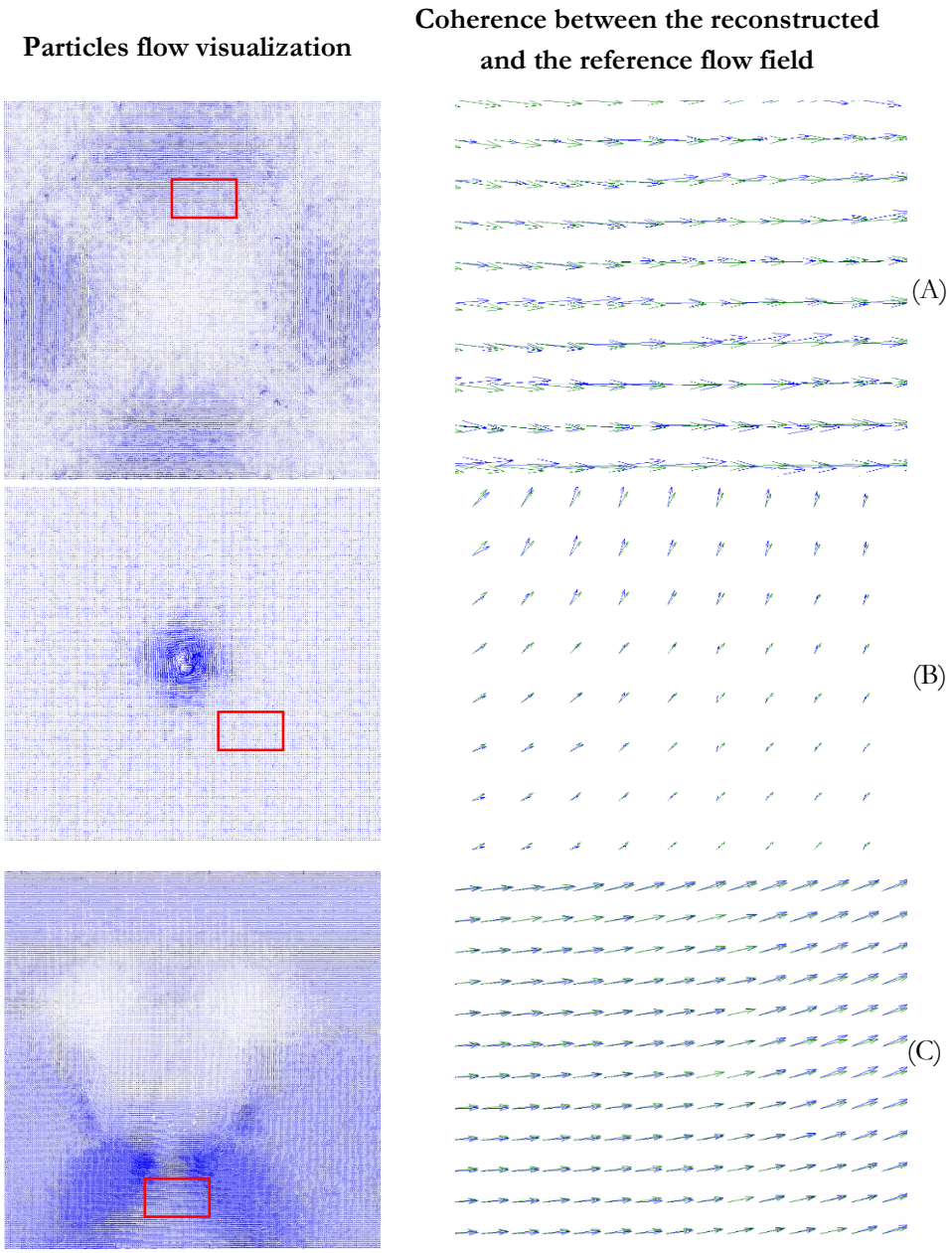
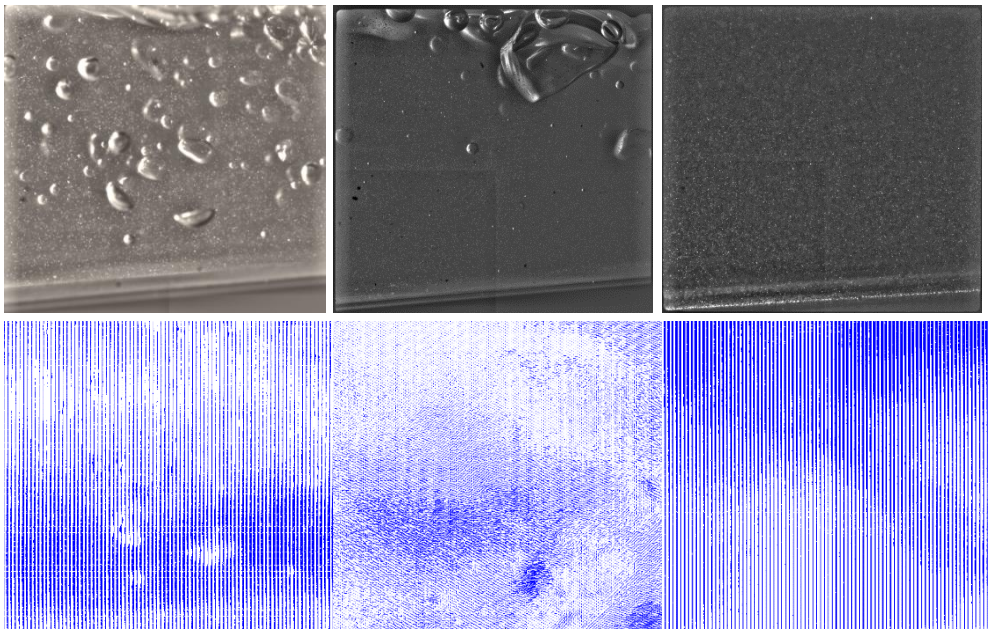


Figure 5.18: For each sythetic dataset the reconstructed flow fields obtained with the 2D SGM strategy and, in a zoom area, the comparison between the reconstructed vectorial displacement map (in blue) and the analytical solution (in green).

Once the algorithm accuracy and reliability was evaluated, a second analysis has regarded the real image data acquired with the PIV TSI System.

The SGM strategy was in this case applied for obtaining high density vector displacement fields (with a pixel-wise resolution), which can allow a better and deepen investigation of the turbulence and the characteristic of the fluid (the vectorial displacement field is presented in Figure 5.19). Differently, the commercial application force the minimum results resolution to a 16 x 16 sampling matching step.



D) E) F)

Figure 5.19: For each real dataset, the reconstructed vectorial displacement field obtained using the 2D SGM algorithm variation.

Finally, the obtained SGM solutions has been compared to the PIV TSI one computing, as in the previous section, the mean and the standard deviation of the differences between the two algorithms . In this case, it is important to highlight that disparities "ground-truth" were not available and the methods comparison just wants to offer an evaluation of the two strategies agreement (or disagreement). Results are shown in Table 5.4.

Looking at the data it is possible to said that the two applications seems to differ not too much; with respect to the syntethic analysys (see Table 5.3), however, the standard deviations are much higher. The higher agreement between the two methods is achieved for case (F): the comparison shows a standard deviation of about 0.5 pixel and a mean value very close to zero. Case (F) considerably differ from the other two cases (D and E): the absence of air bubbles and turbolent phenomena makes more uniform the

particles motion and this results in a better quality of the image radiometric characteristics. This leads to more accurate points correspondences identification for both the correlation algorithms.

		Mean [pix]	Std. Dev [pix]
DATASET	**A**	0.0972	0.9859
	B	-0.0371	0.9706
	C	-0.0133	0.5855

Table 5.4: Statistics of the differences between the results obtained using the proprietary SGM and the commercial TSI PIV System strategies.

Finally, a visual evaluation of the achieved displacement maps, for each case of study, is presented in Figure 5.20. SGM solutions produces less noisy displacements maps, allowing to obtain continuous particles flow motion. On the contrary, the cross-correlation algorithm used in the commercial PIV package presents more irregular and noisy displacement values.

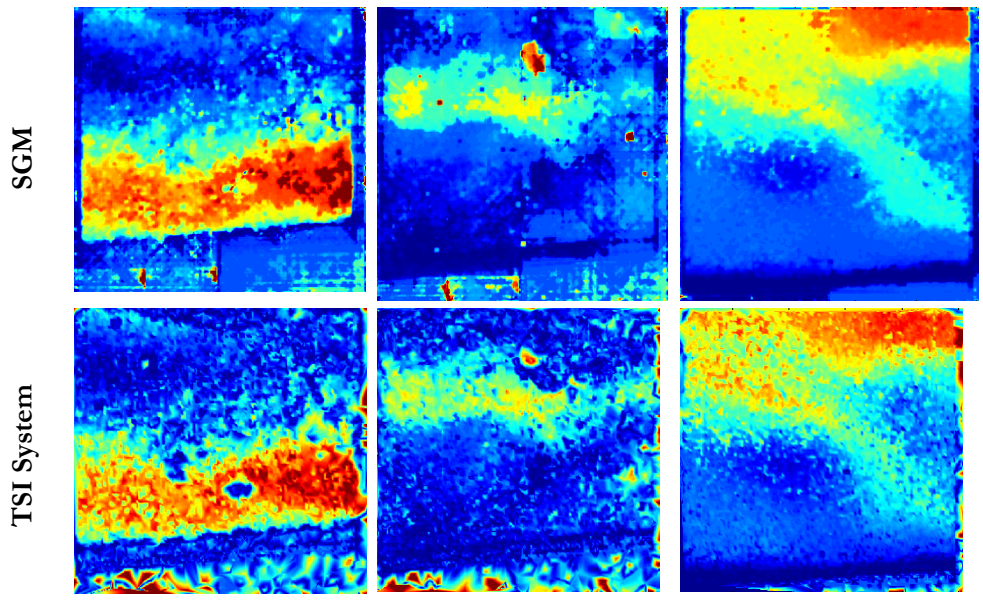


Figure 5.20: For each real dataset, the reconstructed 2D displacement maps obtained with the SGM and the commercial (TSI System) strategies.

The possibility of intervening on the numerical SGM code allows adapting the matching parameters according to the case of study: in this particular application field, the force of parallax regularization can considerably modify the final displacement maps completeness, smoothness level and regularity. In Figure 5.21 is shown an example of the 2D parallax map changes by increasing the intensity of the regularization term in the matching step.

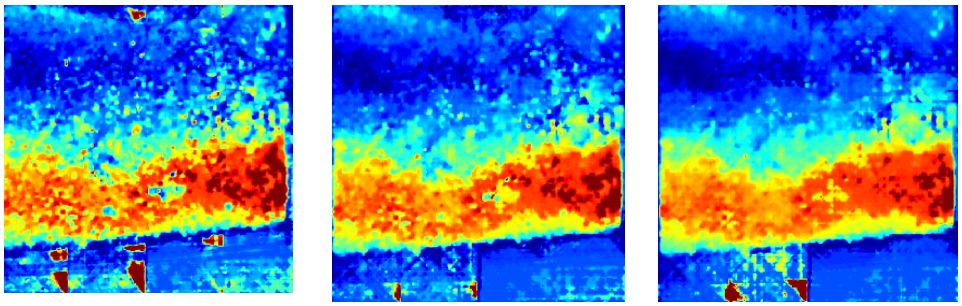


Figure 5.21: Case (D) displacement maps assessed with different combination of regularization values.

It is clearly visible that increasing the parallaxes filed penalization values, the displacements regularity also increase. If such behaviour actually improves the results and the algorithm robustness towards outliers and mismatches it's however hard to assess.

CONCLUSIONS

This work has been essentially directed to verify the actual capability of a proprietary implemented Semi-Global Matching software code, mainly aimed at three-dimensional surfaces reconstruction. In fact, after an initial literary review stage, an in-depth analysis of the wide range of stereo matching algorithms has allowed to study the pros and cons of most local methods (i.e. the adaptive Least Squares Matching technique) and novel, increasingly widespread, Global and Semi-Global techniques. The choice to develop and implement a proprietary algorithm is mainly justified from the desire of maximizing the matching method accuracy and results completeness, paying attention to the process computational efforts as well. The investigation of these three aspects represent the heart of this work of thesis: it is increasingly evident that, both in Computer Vision and Photogrammetry, the achievement of high accuracy and complete results, taking however into consideration the time for obtaining the final result as well, are key factor for a state-of-art matching process.

The algorithm calibration has allowed verifying the method performances in terms of (i) algorithm accuracy in finding corresponding points on image space; (ii) results completeness and level of detail; and (iii) matching core computational time. The processing variables, which decisively intervene in the algorithm metrological characterisation, have also been evaluated with the aim of understand how the method performances change varying the correlation window size, the regularization and similarity cost functions, the penalty values combinations and the use of the multi-resolution and image tiling strategies. An efficient use of the semi-global technique depends extremely on the choice of the best strategies and parametric combinations that are able to well-describe and represent the object typology under evaluation.

The results of the algorithm calibration stage, performed on several image datasets, has allowed to verify all the above described strategy variables. As far as the disparity regularization is concerned, penalty values optimal ranges were identified according to the applications and objects of interest. In particular, using small template size (e.g. three or seven pixels window dimension) it is necessary to pay more attention on the optimal penalty values to use in the correlation process. The influence of the template size on the parallax accuracy has also been evaluated considering the variation of the image similarity functions: it was observed that the enforcement of the disparity continuity allows to obtain comparable results between the analysed cost functions, showing light improvements of the accuracy with the increase of the template size; at the same time

there seems to be an optimal windows size beyond which the cost functions produce less accurate and reliable results.

The accuracy of the second SGM algorithmic implementations has been investigated w.r.t. the first implemented method showing some lightly differences, in terms of accuracy, for small template size: the memory efficient method seems to provide better performances with NCC w.r.t. to the first implementation, but with SAD cost function the same is not true. That's probably due to the fact that SAD with small template sizes relies heavily on a good disparity regularization: the first implementation, although much slower than the new one, performs a more rigorous penalization stage. Overall, the low highlighted differences (in terms of accuracy) are negligible if the significant computational time improvement of the novel method is considered. In other words, the novel SGM implementation seems to be more convenient, thanks to its more efficient computing approach.

Finally, considering the results completeness, a first analysis was computed testing the method on the Middlebury stereo matching datasets: the SAD similarity function seems to be perform better on this type of images; at the same time NCC data tend towards high accuracies if big template size are used. The worst results are achieved by the Rank cost function, while Census shows a good robustness with respect to outliers identification proving, however, to be less accurate than the other similarity measures.

The same datasets were also used for comparing the proprietary software with other SGM strategies: in two of the five cases considered, the best solution was obtained with the proprietary semi-global algorithm; on the contrary, the remaining three cases have presented slightly worse results. Considering that, differently from the compared strategies, the proprietary method does not implement post-processing steps (such as hole-fitting, interpolation, median filtering, etc.) it is possible to say that the results of the comparison are very promising.

After the first method calibration and characterization stages, its performance was tested in real applications fields, where problems related to image radiometric content, points visibility, lacks of completeness and the achievement of established metric precisions, significantly intervene. The method achieved not only complete and appealing digital surface/terrain models, showing its higher robustness w.r.t. Area Based local method, but also highly accurate reconstructions. This concerns both the developed semi-global implementations for surface reconstruction, i.e. the classical stereo approach, and the SGM variation for the 2D displacement domain analysis. In fact, the algorithm metrological characterization stage and the development of computational optimized strategies has allowed a continuous improvement of the algorithm performances and the possibility to extend the method to the 2D search of

image correspondences. Finally, the multi-image algorithmic extensions have also been tested successfully in dense surface reconstruction in different application fields and images acquisition configurations.
In most of the studied application cases, the "pixel locking effect" has produced a random or coherent noise in the final reconstructed scenes, due to costs biasing in the sub-pixel refinement strategy. It represents currently an open-topic since it interests most of the image matching methods: the sub-pixel symmetric refinement strategy has allowed obtaining a light improvement of the disparity refinement stage accuracy. However, further test and research work should be made in the future for investigating the causes and identifying more powerful resolution methods.
Concluding, although several image matching open source and commercial applications are available in the photogrammetric images processing panorama, the choice to completely develop the semi-global software code has originated from the need to better acquire competence and knowledge about the several variables and strategies that are involved and influence the image matching process. The use of a proprietary source code has the advantage of allowing the complete control of the process: the problem solutions can be adapted to the user needs, the data input quality, the application field, the precision and accuracy of the results and the required time of process.
Future developments will regard the creation of a graphical interface for increasing the "user-friendliness" of the matching code (and the settings of all its parameters) making it usable for a wider public and expanding its application fields. On the other hand, an important step for obtaining an accurate solution is the identification of the best algorithm parameters combination: in this context a significant improvement will be the automatic/adaptive assessment of the optimal parametrizations in relation if the application field, the image content and quality, as well as the required accuracy.
Finally, the integration of the processing workflow with post-processing methods (such as images cross-checking, outliers filtering, interpolation techniques, noise reduction methods, etc.) for increasing the method robustness with regard to occlusions, points visibility and data noise problems, will provide further improvement in the final results quality.

Although the most obvious and rapid developments in image matching field are attributed, in recent decades, to the computer vision field, the photogrammetric discipline must try to constantly develop novel solutions that must be mainly pertinent with their own goals and research areas. The importance of developing proprietary systems lies in the choice of be suspicious and curious of the results offered by open-source or commercial applications (of which we do not have full control). In the current

state of the art, photogrammetry, as far as image matching is concerned, no longer seems to be oriented to the design of innovative algorithms or optimization strategies as in the past. Rather, a deeper understanding of each variable and process involved in the data analysis is requested, with the ultimate goals of getting controlled results and achieving a result which is not the first and only that has been obtained, but rather the best one of a series of detailed processes and analyses. The photogrammetric research in the image matching field cannot be said to be completed, but it will continue growing along with the birth and the development of new applications that make the image analysis a key of innovation and originality.

BIBLIOGRAPHY

[1] Abellán, A., Jaboyedoff, M., Oppikofer, T. & Vilaplana, J. (2009). Detection of millimetric deformation using a terrestrial laser scanner: experiment and application to a rockfall event. Nat. Hazards Earth Syst. Sci., 9, 365-372.

[2] Ackermann, F. (1984). Digital image correlation: performance and potential application in photogrammetry. Photogrammetric Record, 11(64): 429–439.

[3] Adhyapak, S., Kehtarnavaz, N. & Nadin, M. (2007). Stereo matching via selective multiple windows", Journ. of Electronic Imaging, 16(1):013012.

[4] Alsadik, B., Gerke, M., Vosselman, G., Daham, A. & Jasim, L. (2014). Minimal Camera Networks for 3D Image Based Modeling of Cultural Heritage Objects. Sensors, 14, 5785–5804.

[5] Baker, H. H. & Binford, T. O. (1981). Depth from edge and intensity based stereo. Proceedings of the 7th International Joint Conferences on Artificial Intelligence. 631–636.

[6] Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J. & Szeliski, R. (2011). A Database and Evaluation Methodology for Optical Flow. International Journal of Computer Vision 92, 1, 1-31.

[7] Baltsavias, E. (1984). Digital image correlation in photogrammetry. M.Sc. technical report, Department of Geodetic Science and Surveying, Ohio State University, Columbus, Ohio.

[8] Baltsavias, E., Li, H., Mason, S., Stefanidis, A. & Sinning, M. (1996). Comparison of two digital photogrammetric systems with emphasis on DTM generation: case study glacier measurement. International Archives of Photogrammetry and Remote Sensing, 31, 104-109.

[9] Banks, J. & Corke, P. (2001). Quantitative evaluation of matching methods and validity measures for stereo vision. Int. J. Robotics Research, vol. 20, pp. 512-532.

[10] Barazzetti, L., Roncella, R., Forlani, G., Remondino, F. & Scaioni, M. (2011). Orientamento Automatico di Blocchi Fotogrammetrici in Ambito Close-range: Stato dell'arte e Prospettive Della Ricerca. In Proceedings of Geomatica: Le Radici del Futuro, Pavia, Italy, 10–11 February 2011; SIFET: Roma, Italy; pp. 131–140.

[11] Barazzetti, L., Forlani, G., Remondino, F., Roncella, R. & Scaioni, M. (2011). Experiences and achievements in automated image sequence orientation for

close-range photogrammetric projects. SPIE Optical Metrology. International Society for Optics and Photonics.

[12] Barazzetti, L. & Scaioni, M. (2010). Development and implementation of image-based algorithms for measurement of deformations in material testing. Sensors, 10, 7469–7495.

[13] Barnard, S. T. & Fischler, M. A. (1982). Computational stereo. ACM Computing Surveys (CSUR), 14(4), 553-572.

[14] Bay, H., Tuytelaars, T. & Van Gool, L. (2006). Surf: Speeded up robust features. Computer Vision–ECCV 2006, 404-417.

[15] Bay B.K., Smith T.S., Fyhrie D.P. & Saad M. (1999). Digital volume correlation: three-dimensional strain mapping using X-ray tomography. Exp Mech 39(3):217–226

[16] Bellman, R. (1956). On a routing problem (No. RAND-P-1000). Rand Corp Santa Monica CA.

[17] Bethmann, F. & Luhmann, T. (2014). Object-based multi-image semi-global matching-concept and first results. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(5), 93-100.

[18] Birchfield, S. & Tomasi, C. (1999). Multiway Cut for Stereo and Motion with Slanted Surfaces, Proc. Int'l Conf. Computer Vision, pp. 489- 495.

[19] Birchfield, S. & Tomasi, C. (1998). A pixel dissimilarity measure that is insensitive to image sampling. IEEE TPAMI, 20(4):401-406.

[20] Birchfield, S. & Tomasi, C. (1999). Depth discontinuities by pixel-to-pixel stereo. International Journal of Computer Vision, 35(3): 269-293.

[21] Birgisson B., Montepara A., Romeo E., Roque R. & Tebaldi G. (2010). Influence of mixture properties on fracture mechanics in asphalt mixtures, Road Mater Pavement Des 11: 68-88. DOI:10.3166/rmpd.11hs.61-88.

[22] Birgisson, B., Montepara, A., Romeo, E., Roncella, R., Roque, R. & Tebaldi, G. (2009). An optical strain measurement system for asphalt mixtures.Materials and Structures, 42(4), 427-441.)

[23] Bitelli, G., Dubbini, M. & Zanutta, A. (2004). Terrestrial Laser Scanning and Digital Photogrammetry Techniques to Monitor Landslide Bodies. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXV, Part B5, pp. 246-251. Istanbul (Turkey).

[24] Blaber, J., Adair, B. & Antoniou, A. (2015). Ncorr: Open-source 2D digital image correlation Matlab software. Experimental Mechanics, 1-18.

[25] Bobick, A. F. & Intille, S.S. (1999). Large occlusion stereo. IJCV, 33(3):181-200.

[26] Boykov, Y. & Kolmogorov, V. (2003). Computing Geodesics and Minimal Surfaces via Graph Cuts, Proc. Int'l Conf. Computer Vision, pp. 26-33.

[27] Boykov, Y. & Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(9), 1124-1137.

[28] Boykov, Y., Veksler, O. & Zabih, R. (2001): Fast Approximate Energy Minimization via Graph Cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (11), pp. 1222- 1239.

[29] Boykov, Y., Veksler, O. & Zabih, R. (1998). A variable window approach to early vision. IEEE TPAMI, 20(12):1283-1294.

[30] Bradsky, G. & Kaehler, A. (2008). Learing OpenCV: Computer Vision with the OpenCV Library. Mike Loukides.

[31] Buttlar, W. G., Hill, B. C., Kim, Y. R., Kutay, M. E., Millien, A., Montepara, A., et. al. (2014). Digital image correlation techniques to investigate strain fields and cracking phenomena in asphalt materials. Materials and Structures, 47(8), 1373-1390.

[32] Cardenal, J., Mata, E., Perez-Garcia, J., Delgado, J., Andez, M., Gonzalez, A. & Diaz-de-Teran, J. (2008). Close Range Digital Photogrammetry Techniques applied to Landslide Monitoring. International Archives of the Photogrammetry, Remote sensing and Spatial Information Sciences, Vol XXXVII. Part B8.

[33] Casson, B., Baratoux, D., Delacourt, C. & Allemand, P. (2003). La Clapière landslide motion observed from aerial differential high resolution DEM. Eng. Geol., 68, 123-139.

[34] Chambon S. & Crouzil A. (2003). Dense matching using correlation: new measures that are robust near occlusions. In: In Proc. British Machine Vision Conference (BMVC 2003), volume 1, pages 143-152.

[35] Chan, S., Wong, Y. & Daniel, J. (2003). "Dense stereo correspondence based on recursive adaptive size multi windowing", in Proc. Image and Vision Computing New Zealand, volume 1, pages 256–260.

[36] Chen, J. & Katz, J. (2005). Elimination of peak-locking error in PIV analysis using the correlation mapping method. Measurement science and technology, 16(8), 1605.

[37] Choi S. & Shah S.P. (1997). Measurement of deformations on concrete subjected to compression using image correlation, Exp Mech 37.3: 307-313

[38] Chu T.C., Ranson W.F., Peters W.H. & Sutton M.A. (1985). Applications of digital-image-correlation techniques to experimental mechanics, Exp Mech 25.3: 232-245

[39] Cochran, S. D. & Medioni, G. (1992). 3-D surface description from binocular stereo. IEEE TPAMI, 14(10):981-994.

[40] Collins, R. T. (1996). A space-sweep approach to true multi-image matching. In Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on. IEEE, p. 358-363.

[41] Cox, I. J. (1994). A maximum likelihood n-camera stereo algorithm. In Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on IEEE, pp. 733-739.

[42] Cox, I. J., Roy, S. & Hingorani, S. L. (1995, October). Dynamic histogram warping of image pairs for constant image brightness. In Image Processing, 1995. Proceedings., International Conference on IEEE, Vol. 2, pp. 366-369.

[43] Crosta, G., Cancelli, P., Tamburini, A., Alberto, W., Broccolato, M., Castellanza, R. et. al. (2012). Chasing a complete understanding of a rapid moving rock slide: the La Saxe landslide. EGU General Assembly 2012. Vienna, Austria.

[44] Dall'Asta, E., Delaloye, R., Diotri, F., Forlani, G., Fornari, M., di Cella, U. M. et. al. (2015). Use of Uas in a High Mountain Landscape: the Case of Gran Sommetta Rock Glacier (AO). The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(3), 391.

[45] Dall'Asta, E. & Roncella, R. (2014). A comparison of semiglobal and local dense matching algorithms for surface reconstruction. International Archives of The Photogrammetry, Remote Sensing And Spatial Information Sciences, 40(5), 187-194

[46] Dall'Asta, E., Ghizzardi, V., Brighenti, R., Romeo, E., Roncella, R. & Spagnoli, A. (2015). New experimental techniques for fracture testing of highly deformable materials. The 5th International Conference on Crack Paths (CP 2015), Ferrara, Italy, 16-18 September, 2015.

[47] De Bellis, A. (2013). Architettura e luce: controlli digitali per il rilievo dello spazio sacro. Il caso del Battistero di Parma (Doctoral dissertation, Università degli Studi di Parma. Dipartimento di Ingegneria Civile ed Architettura).

[48] Deilami, K. & Hashim, M. (2011). Very high resolution optical satellites for DEM generation: a review. European Journal of Scientific Research, 49(4), 542-554.

[49] Delaloye, R., Morard, S., Barboux, C., Abbet, D., Gruber, V., Riedo, M. & Gachet, S. (2013). Rapidly moving rock glaciers in Mattertal. GRAF, C. (Red.) Mattertal–ein Tal in Bewegung. Publikation zur Jahrestagung der Schweizerischen Geomorphologischen Gesellschaft, 29, 21-31.

[50] Demoulin, C. & Van Droogenbroeck, M. (2005). A method based on multiple adaptive windows to improve the determination of disparity maps" In Proc. IEEE Workshop on Circuit, Systems and Signal Processing, pages 615–618.

[51] Ebner, H. & Heipke, C. (1988). Integration of digital image matching and object surface reconstruction. International Archives of Photogrammetry and Remote Sensing, 27(B11): 534–545.

[52] Egnal, G. (2000). Mutual information as a stereo correspondence measure. Technical Reports (CIS), 113.

[53] Eisert, P., Steinbach, E. & Girod, B. (1999). Multi-hypothesis, volumetric reconstruction of 3-D objects from multiple calibrated camera views. In ICASSP 99, pp. 3509-3512.

[54] El-Hakim, S., Remondino, F., Gonzo, L. & Voltolini, F. (2007). Effective High resolution 3D geometric reconstruction of heritage and archeological sites from images. In Proceedings of CAA 2007 Conference, Berlin, Germany, April 2007.

[55] Favalli, M., Fornaciai, A., Isola, I. & Tarquini, S. (2012). Nannipieri, L. Multiview 3D reconstruction in geosciences. Comput. Geosci, 44, 168–176.

[56] Felzenszwalb, P. F. & Huttenlocher, D. P. (2006). Efficient belief propagation for early vision. International journal of computer vision, 70(1), 41-54.

[57] Ford, L. & Fulkerson, D. (1962). Flows in Networks. Princeton Univ. Press.

[58] Förstner, W. (1982). On the geometric precision of digital correlation. International Archives of Photogrammetry, 24(3): 176–189.

[59] Förstner, W. (1986). A feature based correspondence algorithm for image matching. International Archives of Photogrammetry and Remote Sensing, 26(3), 150-166.

[60] Fröba, B. & Ernst, A. (2004). Face detection with the modified census transform. In Proc. Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2004), pages 91-96. IEEE Computer Society.

[61] Fua, P. (1993). A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications, 6:35-49.

[62] Fusiello, A., Roberto, V. & Trucco, E. (1997, June). Efficient stereo with multiple windowing. In cvpr on IEEE, pp. 858.

[63] Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Machine Intell, vol. PAMI-6, pp. 721-741.

[64] Gennert, M., A. (1988). Brightness-based stereo matching. In ICCV, pp. 139–143.

[65] Gong, J., Li, Z., Zhu, Q., Sui, H. & Zhou, Y. (2000). Effects of various factors on the accuracy of DEMs: an intensive experimental investigation. Photogrammetric Engineering & Remote Sensing, 66(9): 1113–1117.

[66] Greig, D., Porteous, B. & Seheult, A. (1989). Exact Maximum A Posteriori Estimation for Binary Images, J. Royal Statistical Soc., Series B, vol. 51, no. 2, pp. 271-279.

[67] Grimson, W. E. L. (1985). Computational experiments with a feature based stereo algorithm. IEEE TPAMI, 7(1):17-34.

[68] Grun A. (1985). Adaptive Least Squares Correlation: A Powerful Image Matching Technique. South African J. Photogramm. Remote Sensing and Cartography, pp. 14(3): 175-187.

[69] Grün, A. & Baltsavias, E.P. (1988). Geometrically constrained multiphoto matching. PE&RS, 54(5), pp. 633-641.

[70] Gruen, A. & Baltsavias, E. (1989, April). Automatic 3-D measurement of human faces with CCD-cameras. In Biostereometrics' 88: Fifth Intl Mtg, pp. 106-117. International Society for Optics and Photonics.

[71] Gruen, A. (1996). Digital photogrammetric stations revisited. International Archives of Photogrammetry and Remote Sensing, 31(B2): 127–134

[72] Gruen, A., Remondino, F. & Zhang, L. (2003). Computer Reconstruction and Modeling of the Great Buddha Statue in Bamiyan, Afghanistan. In ISPRS International Workshop on Visualization and Animation of Reality-based 3D Models,(Tarasp-Vulpera, Switzerland, 24-28 February 2003), International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences (Vol. 34, No. 5/W10).

[73] Gruen, A. (2012). Development and status of image matching in photogrammetry. Photogramm Record 27(137):36–57

[74] Haeberli, W. (2013). Mountain permafrost - research frontiers and a special long-term challenge. Cold Regions Science and Technology, 96, 71-76.

[75] Hartley, R. I. (1997). Lines and points in three views and the trifocal tensor. International Journal of Computer Vision, 22(2), 125-140.

[76] Hartley, R.I. & Zisserman, A. (2000). Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge.

[77] Helava, U. V. (1972). Digital processing and analysis of image data. International Archives of Photogrammetry, 20(1), pp.354.

[78] Helava, U. V. (1976). Digital correlation in photogrammetric instruments. International Archives of Photogrammetry, 21(2), pp.26.

[79] Helava, U. V. (1978). Digital correlation in photogrammetric instruments. Photogrammetria, 34(1): 19–34.

[80] Helava, U. V. (1988). Object-space least squares correlation. Photogrammetric Engineering & Remote Sensing, 54(6): 711–714.

[81] Hirschmuller, H. (2001). Improvements in real-time correlation-based stereo vision. In IEEE workshop on Stereo and MultiBaseline Vision. IJCV (this issue).

[82] Hirschmuller, H., Innocent, P. & Garibaldi, J. (2002). Real-time correlation-based stereo vision with reduced border errors", Int. Journ. of Computer.

[83] Hirschmuller, H. (2005, June). Accurate and efficient stereo processing by semi-global matching and mutual information. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on IEEE, Vol. 2, pp. 807-814.

[84] Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(2), 328-341.

[85] Hirschmuller, H. & Scharstein, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(9), 1582-1599.

[86] Hirschmuller, H. (2011). Semi-Global Matching-Motivation, Developments and Applications.

[87] Hirschmüller, H., Buder, M., & Ernst, I. (2012). Memory efficient semi-global matching. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3, 371-376.

[88] Hobrough, G. L. (1959). Automatic stereo plotting. Photogrammetric Engineering, 25(5), 763-769.

[89] Ishikawa, H. (2003). Exact Optimization for Markov Random Fields with Convex Priors, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 10, pp. 1333-1336.

[90] James, M.R. & Robson, S. (2014). Mitigating systematic error in topographic models derived from UAV and ground-based image networks. Earth Surf. Proc. Landf. 2014, 39, 1413–1420.

[91] Jensen, F. V. (1996). An Introduction to Bayesian Networks. Berlin, Germany: Springer-Verlag.

[92] Jepping, C., Bethmann, F. & Luhmann, T. (2014). Congruence analysis of point clouds from unstable stereo image sequences. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1, 301-306.

[93] Kalarot, R., Morris, J., Berry, D. & Dunning, J. (2011). Analysis of real-time stereo vision algorithms on GPU. International Conference on Image and Vision Computing New Zealand.

[94] Kanade, T. & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 16(9), 920-93.

[95] Kang, S. B., Szeliski, R. & Chai, J. (2001). Handling occlusions in dense multi-view stereo. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on IEEE, Vol. 1, pp. I-103.

[96] Kim, J. & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems.

[97] Kim, J., Kolmogorov, V. & Zabih, R. (2003). Visual Correspondence Using Energy Minimization and Mutual Information, Proc. Int'l Conf. Computer Vision, pp. 1033-1040.

[98] Kim Y.R. & Wen H. (2001). Fracture energy from indirect tension test, J Assoc Asphalt Paving Technol 71: 779-793

[99] Kolmogorov, V. & Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. International Conference for Computer Vision, pp. 508-515.

[100]Kolmogorov, V. & Zabih, R. (2002). Multi-Camera Scene Reconstruction via Graph Cuts, Proc. European Conf. Computer Vision, vol. 3, pp. 82-96.

[101]Konolige, K. (1998). Small vision systems: Hardware and implementation. In Robotics Research, pp. 203-212. Springer London.

[102]Krauß, T., Reinartz, P., Lehner, M., Schroeder, M. & Stilla, U. (2005). DEM generation from very high resolution stereo satellite data in urban areas using dynamic programming. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36, 1.

[103]Kutulakos, K. & Seitz, S. (2000). A theory of shape by space carving. IJCV, 38(3):199-218.

[104]Kutulakos, K. (2000). Approximate N-view stereo. In ECCV, vol.I, pp. 67-83.

[105]Kwatra,V., Scho¨dl, A.,Essa, I. ,Turk, G. & Bobick, A. (2003). Graph cut Textures: Image and Video Synthesis Using Graph Cuts, ACM Trans. Graphics, Proc. SIGGRAPH 2003.

[106]Lewis, J. P. (1995, May). Fast normalized cross-correlation. In Vision interface, Vol. 10, No. 1, pp. 120-123.

[107]Lourakis M. (2004). levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++, URL http://www.ics.forth.gr/lourakis/levmar.

[108]Lucas, B.D. & Kanade, T. (1981). An iterative image registration technique with an application in stereo vision. In Seventh International Joint Conference on Artificial Intelligence (IJCAI-81), Vancouver, pp. 674-679.

[109]Martin, J. & Crowley, J. L. (1995, March). Comparison of correlation techniques. In International Conference on Intelligent Autonmous Systems, Karlsruhe (Germany), pp. 86-93.

[110]Mantovani, F., Soeters, R. & Van Westen, C. J. (1996). Remote sensing techniques for landslide studies and hazard zonation in Europe. Geomorphology, 15(3), 213-225.

[111]Makarovic, B. (1980). Image correlation algorithms. International Archives of Photogrammetry, 23(B2): 139–158.

[112]Makki, M. M. & Chokri, B. (2015). Determination of Stress Concentration for Orthotropic and Isotropic Materials Using Digital Image Correlation (DIC). In Multiphysics Modelling and Simulation for Systems Design and Monitoring, pp. 517-530. Springer International Publishing.

[113]Marr, D. & Poggio, T. (1976). Cooperative computation of stereo disparity. Science, 194: 283-287.

[114]Masad, E., Muhunthan, B., Shashidhar, N. & Harman, T. (1998). Aggregate orientation and segregation in asphalt concrete. In Application of Geotechnical Principles in Pavement Engineering, Proceedings of Sessions of Geo-Congress 98.

[115]Middlebury Stereo Evaluation Dataset. (last accessed 2106). http://vision.middlebury.edu/stereo/eval3/ (2015).

[116]Mikolajczyk, K., Tuytelaars, T., Schmid C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. & Van Gool, L. (2005). A comparison of affine region detectors. IJCV 65(1/2):43-72, 2005.

[117]Mora, P., Baldi, P., Casula, G., Fabris, M., Ghirotti, M., Mazzini, E. & Pesci, A. (2003). Global Positioning Systems and digital photogrammetry for the monitoring of mass movements: application to the Ca' di Malta landslide (northern Apennines, Italy). Eng. Geol., 68, 103-121.

[118]Moré, J. J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. In Numerical analysis, pp. 105-116. Springer Berlin Heidelberg

[119]Muszynski L., Wang F. & Shaler S.M. (2002). Short term creep tests on phenol resorcinol formaldehyde (PRF) resin undergoing moisture cement, Wood Fiber Sci 34.4:612-624.

[120]Muszynski L., Lopez-Anido R. & Shaler S.M. (2000). Image correlation analysis applied to measurement of shear strains in laminated composites, Paper presented at the SEM XI International Congress on Experimental Mechanics. Orlando, FL,

[121]Nahshon, K., Hoffman, W. A. & Ullagaddi, C. B. (2015). In Springer International Publishing, Characterization of Structural Scale Ductile Fracture of Aluminum Panels Using Digital Image Correlation, Fracture, Fatigue, Failure, and Damage Evolution, (2015), vol.5 81-87

[122]Narayanan, P.,Rander, P. & Kanade,T. (1998). Constructing virtual worlds using dense stereo. In ICCV, pp. 3-10.

[123]Ncorr. (last accessed 07.10.15). Algorithms used in Ncorr, http://www.ncorr.com/index.php/dic-algorithms,

[124]Nehab, D., Rusinkiewiez, S. & Davis, J. (2005, October). Improved sub-pixel stereo correspondences through symmetric refinement. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on IEEE, Vol. 1, pp. 557-563.

[125]Nex, F. & Rinaudo, F. (2008). Multi-image matching: An "Old And New" photogrammetric answer to lidar techniques. Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 37, 621-626.

[126]Okutomi, M. & Kanade, T. (1992). A locally adaptive window for signal matching. IJCV, 7(2):143-162.

[127]Okutomi, M. & Kanade, T. (1993). A multiple-baseline stereo. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 15(4), 353-363.

[128]OpenCV 2.4.8 StereoSGBM method, single-pass variant. (2006). Reimplementation and modification of H. Hirschmüller's SGM method In CVPR 2006; PAMI 2008.

[129]Open Source Computer Vision Library (last accessed 21.01.2016). StereoSGBM C++ code. https://github.com/Itseez/opencv.

[130]Overmars, E. F. J., et al. (2010). Bias errors in PIV: the pixel locking effect revisited. In: 15th international symposium on applications of laser techniques to fluid mechanics, Lisbon, Portugal, pp. 5-8.

[131]Paparoditis, N., Maillet, G., Taillandier, F., Jibrini, H., Jung, F., Guigues, L. & Boldo, D. (2001). Multi-image 3D feature and DSM extraction for change detection and building reconstruction. Automatic Extraction of Man-Made objects from aerial and space images (III), 217-230.

[132]Pateraki, M. N. (2003). Analysis and performance of the Adaptive Multi-Image matching algorithm for airborne digital sensor ADS40. ETH, Swiss Federal Institute of Technology Zurich, Institute of Geodesy and Photogrammetry.

[133]Pearl, J. (1982, August). Reverend Bayes on inference engines: A distributed hierarchical approach. In AAAI (pp. 133-136).

[134]Pearl, J. (1988). Probabilistic reasoning in intelligent systems: networks of plausible inference. San Francisco, CA: Morgan Kaufmann.

[135]Pertl, A. (1984). Digital image correlation with the analytical plotter Planicomp C100. International Archives of Photogrammetry, 25(3b): 874–882.

[136]Petrikova, I., Marvalova, B., Samal, S. & Cadek, M. (2015). Digital Image Correlation as a Measurement Tool for Large Deformations of a Conveyor Belt. In Applied Mechanics and Materials, vol.732: 77-80.

[137]Photomodeler. (2015). (last accessed 07.09.2015). http://www.photomodeler.com/index.html

[138]PhotoScan. (2015). Algorithms used in Photoscan (last accessed 05.10.15). http://www.agisoft.ru/forum/index.php?topic=89.0

[139]Pierrot-Deseilligny, M. & Paparoditis, N. (2006). A Multiresolution and Optimization-Based Image Matching Approach: An Application To Surface Reconstruction From Spot5-Hrs Stereo Imagery. In IAPRS vol XXXVI-1/W41 in ISPRS Workshop on Topographic Mapping FromSpace (With Special Emphasis on Small Satellites). Ankara, Turquie.

[140]Pierrot Deseilligny, M. & Clery, I. (2011). Apero, An Open Source Bundle Adjusment Software For Automatic Calibration And Orientation Of Set Of Images. In IAPRS Vol. XXXVIII-5/W16, Trento.

[141]Pollefeys, M., Kock, R. & Van Gool, L. (1999). A simple and efficient rectification method for general motion. International Conference on Computer Vision, (p. 496-501). Corfù.

[142]Poon, J., Fraser, C. S., Chunsun, Z., Li, Z. & Gruen, A. (2005). Quality assessment of digital surface models generated from IKONOS imagery. The Photogrammetric Record, 20(110), 162-171.

[143]Psarakis, E. Z. & Evangelidis, G. D. (2005, October). An enhanced correlation-based method for stereo correspondence with subpixel accuracy. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on Vol. 1, pp. 907-912).

[144]Prazdny, K. (1985). Detection of binocular disparities. Biological Cybernetics, 52(2):93-99.

[145]Re, C. (2014). Three-dimensional reconstruction of planetary surfaces from stereo satellite images. PhD Thesis at Centro di Ateneo di Studi e Attività Spaziali CISAS "Giuseppe Colombo" – University of Padova, Italy.

[146]Re, C., Cremonese, G., Dall'Asta, E., Forlani, G., Naletto, G. & Roncella, R. (2012, November). Performance evaluation of DTM area-based matching reconstruction of Moon and Mars. In SPIE Remote Sensing (pp. 85370V-85370V). International Society for Optics and Photonics.

[147]Remondino, F., Barazzetti, L., Nex, F., Scaioni, M. & Sarazzi, D. (2011). UAV photogrammetry for mapping and 3D modeling-current status and future perspectives. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 38(1), C22.

[148]Remondino, F., Spera, M. G., Nocerino, E., Menna, F. & Nex, F. (2014). State of the art in high density image matching. The Photogrammetric Record, 29(146), 144-166.

[149]Roncella, R., Re, C. & Forlani, G. (2011). Comparison of two structure and motion strategies. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3816, 343-350.

[150]Roncella, R. & Forlani, G. (2015). A fixed terrestrial photogrammetric system for landslide monitoring. In Modern Technologies for Landslide Monitoring and Prediction,pp. 43-67. Springer Berlin Heidelberg.

[151]Rothermel, M., Wenzel, K., Fritsch D. & Haala, N. (2012). SURE: Photogrammetric surface reconstruction from imagery. Proceedings LowCost3D Workshop 2012, 04th – 05th December 2012, Berlin.

[152]Roy, S. & Cox, I. (1998). A Maximum-Flow Formulation of the n-Camera Stereo Correspondence Problem, Proc. Int'l Conf. Computer Vision.

[153]Sadeghian, S., Zoej, M. J. V., Delavar, M. R. & Abootalebi, A. (2001). Precision rectification of high resolution satellite imagery without ephemeris data. International Journal of Applied Earth Observation and Geoinformation, 3(4), 366-371.

[154]Saint-Marc, P., Chen, J. S. & Medioni, G. (1989, June). Adaptive smoothing: A general tool for early vision. In Computer Vision and Pattern Recognition, 1989. Proceedings CVPR'89., IEEE Computer Society Conference on (pp. 618-624). IEEE.

[155]Sakai, S., Ito, K., Aoki, T., Masuda, T. & Unten, H. (2013). An efficient image matching method for multi-view stereo. In Computer Vision–ACCV 2012 (pp. 283-296). Springer Berlin Heidelberg.

[156]Santise, M. (2012). Verifiche sperimentali nella produzione di modelli digitali del terreno tramite sistemi UAV: applicazioni nel campo del rilievo archeologico e del telerilevamento. Master thesis at DICATeA, University of Parma (PR), Italy.

[157]Scharstein, D. & Szeliski, R. (2002). A taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision, 47 (1/2/3), pp. 7-42.

[158]Scharstein, D. (1994). Matching images by comparing their gradient fields. In ICPR, volume 1, pages 572-575.

[159]Scharstein, D. (1999). View Synthesis Using Stereo Vision, volume 1583 of Lecture Notes in Computer Science (LNCS). Springer-Verlag.

[160]Schenk, T. (1999). Digital Photogrammetry: Volume I. TerraScience, Laurelville, Ohio, USA. 421 pages.

[161]Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X. & Westling, P. (2014). High-resolution stereo datasets with subpixel-accurate ground truth. In Pattern Recognition, pp. 31-42. Springer International Publishing.

[162]Schlüter, M. (1998). Multi-image matching in object space on the basis of a general 3-D surface model instead of common 2. 5-D surface models and its application for urban scenes. International Archives of Photogrammetry and Remote Sensing, 32, 545-552.

[163]Scholten, F., Gwinner, K., Roatsch, T., Matz, K.-D., Wählisch, M., Giese, B., et. al. (2005). Mars Express HRSC data processing - Methods and operational Aspects, Photogrammetric Engineering & Remote Sensing, 71(10):1143-1152.,

[164]Seitz, S. & Dyer, C. (1999). Photorealistic scene reconstruction by voxel coloring. IJCV, 35(2):151-173.

[165]Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. & Szeliski, R. (2006, June). A comparison and evaluation of multi-view stereo reconstruction algorithms. In Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on IEEE, Vol. 1, pp. 519-528.

[166]Seo Y., Kim Y.R., Schapery R., Witczac M. & Bonaquist R. (2004). A study of crack-tip deformation and crack growth in asphalt concrete using fracture mechanics, J Assoc Asphalt Paving Technol 73: 697-730

[167]Shashua, A. (1994). Trilinearity in visual recognition by alignment. In Computer Vision-ECCV'94 (pp. 479-484). Springer Berlin Heidelberg.

[168]Slabaugh, G., Malzbender, T., Culbertson, B. & Schafer, R. (2000). Improved voxel coloring via volumetric optimization. TR 3, Center for Signal and Image Processing.

[169]Slabaugh, G., Culbertson, B., Malzbender, T. & Stevens, M. (2004). Methods for volumetric reconstruction of visual scenes. IJCV, 57(3):179-199.

[170]Snow, D., Viola, P. & Zabih, R. (2000). Exact Voxel Occupancy with Graph Cuts,Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 345-352.

[171]Stein, F. (2004). Efficient computation of optical flow using the census transform. In C. E. Rasmussen, H. H. Bülthoff, B. Schölkopf, and M. A. Giese, editors, Pattern Recognition, volume 3175 of Lecture Notes in Computer Science, pp. 79-86. Springer, 2004.

[172]Stein, A. N., Huertas, A. & Matthies, L. (2006, May). Attenuating stereo pixel-locking via affine window adaptation. In Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on IEEE, pp. 914-921.

[173]Strecha, C., von Hansen, W., Gool, L. V., Fua, P. & Thoennessen, U. (2008, June). On benchmarking camera calibration and multi-view stereo for high resolution imagery. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on IEEE, pp. 1-8.

[174]Sun, C. C., Ruan, S. J., Shie, M. C. & Pai, T. W. (2005). Dynamic contrast enhancement based on histogram specification. Consumer Electronics, IEEE Transactions on, 51(4), 1300-1305.

[175]Sun, J., Zheng, N. N. & Shum, H. Y. (2003). Stereo matching using belief propagation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 25(7), 787-800.

[176]Sutton, M. A. & Hild, F. (2015). Recent Advances and Perspectives in Digital Image Correlation. Experimental Mechanics, 55.1: 1-8

[177]Szeliski, R. (1999). A multi-view approach to motion and stereo. In CVPR, vol. 1, pp. 157-163.

[178]Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov,V., Agarwala,A., Tappen, M. & Rother, C. (2008). A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30, 6, pp. 1068-1080

[179] Taylor, C. (2003). J. Surface reconstruction from feature based stereo. In ICCV, pp. 184-190.

[180]Thoeni, K., Giacomini, A., Murtagh, R. & Kniest, E. (2014). A Comparison of Multi-view 3D Reconstruction of a Rock Wall using Several Cameras and a Laser Scanner. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2014, 45, 573–580.

[181]Thomos, I., Malasiotis,S. & Strintzis, M.G. (1998). Optimized Block Based Disparity Estimation in Stereo Systems Using a MaximumFlow Approach," Proc. SIBGRAPI '98 Conf..

[182]Tian, Q. & Huhns, M.N. (1986). Algorithms for subpixel registration. Computer Vision, Graphics, and Image Processing, 35:220- 233.

[183]Tomasi, C. & Kanade, T. (1991). Detection and tracking of point features. Pittsburgh: School of Computer Science, Carnegie Mellon Univ.

[184]Toutin, T. (2004). Comparison of stereo-extracted DTM from different high-resolution sensors: SPOT-5, EROS-A, IKONOS-II, and QuickBird. Geoscience and Remote Sensing, IEEE Transactions on, 42(10), 2121-2129.

[185]Treuille, A., Hertzmann, A. & Seitz, S. Example-based stereo with general BRDFs. In CCV, vol. II, pp. 457-469, 2004.

[186]Triggs, B. (1995, June). Matching constraints and the joint image. In Computer Vision, 1995. Proceedings., Fifth International Conference on IEEE, pp. 338-343.

[187]Triggs, B., McLauchlan, P., Hartley, R. & Fitzgibbon, A. (2000). Bundle Adjustment -- A Modern Synthesis. Lecture Notes in Computer Science, Vol 1883, pp 298-372.

[188]Tschoegl, N. W. (2012). The phenomenological theory of linear viscoelastic behavior: an introduction. Springer Science & Business Media, pp. 119-126.

[189] TSI PIV systems (2015). Particle Image Velocimetry (PIV) systems:image acquisition and analysis (last accessed 25.01.2016). http://www.tsi.com/stereo-piv-system.

[190]Van Meerbergen, G., Vergauwen, M., Pollefeys, M. & Van Gool, L. (Apr.-June 2002). A Hierarchical Symmetric Stereo Algorithm Using Dinamic Programming. Int'l J. Computer Vision, vol. 47, nos. 1/2/3, (p. pp. 275-285).

[191]Vasuki, Y., Holden, E.J., Kovesi, P. & Micklethwaite, S. (2014). Semi-automatic mapping of geological Structures using UAV-based photogrammetric data: An image analysis approach. Comput. Geosci. p. 69, 22–32.

[192]Veksler, O. (1999) Efficient Graph-based Energy Minimization Methods in Computer Vision. PhD thesis, Cornell University.

[193]Veksler, O. (2001). Stereo matching by compact windows via minimum ratio cycle. In Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on IEEE, Vol. 1, pp. 540-547.

[194]Veksler, O. (2005). Stereo correspondence by dynamic programming on a tree. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on IEEE, Vol. 2, pp. 384-390.

[195]Viola, P. & Wells III, W. M. (1997). Alignment by maximization of mutual information. International journal of computer vision, 24(2), 137-154.

[196]Wallis, R. (1976, November). An approach to the space variant restoration and enhancement of images. In Proc. of symp. On current mathematical problems in image science, naval postgraduate school, Monterey CA, USA, November, pp. 329-340.

[197]Weiss, Y. & Freeman, W. T. (2001). On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. Information Theory, IEEE Transactions on, 47(2), 736-744.

[198]Westerweel, J. (1993). Digital Particle Image Velocimetry — Theory and Application, (PDF). Delft University Press. ISBN 90-6275-881-9.

[199]Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J. & Reynolds, J.M. (2012). "Structure-from-Motion" photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology 2012, 179, 300–314.

[200]Wrobel, B. (1987). Facets stereo vision (FAST vision)—a new approach to computer stereo vision and to digital photogrammetry. Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland. 437 pp. 231–258.

[201]Yang, R., Pollefeys, M. & Welch, G.(2003). Dealing with textureless regions and specular highlights - a progressive space-carving scheme using a novel photo-consistency measure. In ICCV, pp. 576-584.

[202]Yedidia, J. S., Freeman, W. T. & Weiss, Y. (2003). Understanding belief propagation and its generalizations. Exploring artificial intelligence in the new millennium, 8, 236-239.

[203]Yoon, K. J., & Kweon, I. S. (2005, June). Locally adaptive support-weight approach for visual correspondence search. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on IEEE, Vol. 2, pp. 924-931).

[204]Zabih, R., & Woodfill, J. (1994). Non-parametric Local Transforms for omputing Visual Correspondence. Proceedings of the European Conference of Computer Vision, (p. pp. 151-158). Stockholm, Sweeden

[205]Žbontar, J. & LeCun, Y. (2015). Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. arXiv preprint arXiv:1510.05970.

[206]Zhang, L. & Seitz, S. (2001). Image-based multiresolution shape recovery by surface deformation. In SPIE: Videometrics and Optical Methods for 3D Shape Measurement, pp. 51-61,

[207]Zhang, L. & Gruen, A. (2004). Automatic DSM extraction from linear array imagery data. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 35, Part B3, pp. 128-133, XXth ISPRS Congress, Istanbul, Turkey,12-23 July 2004.

[208] Zhang, L. (2005). Automatic digital surface model (DSM) generation from linear array images. Mitteilungen- Institut fur Geodasie und Photogrammetrie an der Eidgenossischen Technischen Hochschule Zurich.

Ringraziamenti

É difficile in poche righe ricordare tutte le persone che, a vario titolo, hanno contribuito allo sviluppo di questo lavoro e alla mia crescita durante questi tre anni.

Un ringraziamento particolare va alla "famiglia" ICAR/06:
al *Professore Riccardo Roncella*, per gli innumerevoli aiuti, il costante supporto e la tenace ironia con cui mi ha sempre sostenuta, infondendomi fiducia e consapevolezza;
al *Professore Gianfranco Forlani*, guida sapiente e generosa del nostro gruppo di ricerca, per le critiche, le osservazioni ed i preziosi suggerimenti, e per avermi insegnato l'importanza della curiosità;
a *Marina*, amica e collega fidata, per essermi stata sempre vicina, avere condiviso con me lavori, pensieri ed emozioni, ed avermi donato ogni giorno la sua luce e il suo calore;
a *Matteo*, amico e collega paziente, per avermi infuso forza e stabilità, e per avermi aiutato, con la sua allegria e serenità, a superare i momenti più difficili;
a *Nazarena,* amica e collega affettuosa, per l'aiuto scientifico e morale, il supporto e l'incoraggiamento, e per aver stimolato in me grinta e decisione.

Desidero inoltre ringraziare l'amichevole collaborazione di tutti coloro che hanno contribuito allo sviluppo di questa tesi e alla mia crescita personale: l'Ing. *Fabrizio Diotri* e il Dott. *Umberto Morra di Cella*, l'Ing. *Klaus Thoeni* e la Prof.ssa *Anna Giacomini* e l'Arch. *Andrea Zerbi*.

Un affettuoso ringraziamento va anche a *Camilla, Roberta, Giorgia* e *Cristina*, premurose colleghe di dottorato, per la loro amicizia ed il grande incoraggiamento.

Non avrei mai potuto concludere questo lavoro se non avessi avuto il supporto della *mia famiglia* e di *Marco,* che mi hanno sostenuta con affetto e pazienza insegnandomi ad avere fiducia nelle mie capacità.