State of the art in high density image matching

Article *in* The Photogrammetric Record · June 2014

DOI: 10.1111/phor.12063

5 authors, including:

Fabio Remondino
Fondazione Bruno Kessler
282 PUBLICATIONS **3,890** CITATIONS
SEE PROFILE

Erica Nocerino
Fondazione Bruno Kessler
45 PUBLICATIONS **283** CITATIONS
SEE PROFILE

Fabio Menna
Fondazione Bruno Kessler
55 PUBLICATIONS **481** CITATIONS
SEE PROFILE

Francesco Nex
University of Twente
72 PUBLICATIONS **759** CITATIONS
SEE PROFILE

THE PHOTOGRAMMETRIC RECORD

The Photogrammetric Record 29(146): 144–166 (June 2014)
DOI: 10.1111/phor.12063

STATE OF THE ART IN HIGH DENSITY IMAGE MATCHING

Fabio Remondino (remondino@fbk.eu)
Maria Grazia Spera (spera@fbk.eu)
Erica Nocerino (nocerino@fbk.eu)
Fabio Menna (fmenna@fbk.eu)
Francesco Nex (franex@fbk.eu)
Bruno Kessler Foundation (FBK), Trento, Italy

Abstract

Image matching has a history of more than 50 years, with the first experiments performed with analogue procedures for cartographic and mapping purposes. The recent integration of computer vision algorithms and photogrammetric methods is leading to interesting procedures which have increasingly automated the entire image-based 3D modelling process. Image matching is one of the key steps in 3D modelling and mapping. This paper presents a critical review and analysis of four dense image-matching algorithms, available as open-source and commercial software, for the generation of dense point clouds. The eight datasets employed include scenes recorded from terrestrial and aerial blocks, acquired with convergent and normal (parallel axes) images, and with different scales. Geometric analyses are reported in which the point clouds produced with each of the different algorithms are compared with one another and also to ground-truth data.

Keywords: DSM, evaluation, image matching, photogrammetry, point cloud

It is quite evident that even with our past progress, we have only scratched the surface of the possibilities in the use of photogrammetry.
(*George D. Hardy in 1973*)

Introduction

The 3D reconstruction of scenes and objects at different scales is generally performed today using range or image data. Range or active sensors (Vosselman and Mass, 2010), such as laser scanners or structured light systems, are a common source of dense point clouds due to their ease of use, speed and ability to capture millions of points in a very short time. The associated data-processing procedure is also quite straightforward and is based on reliable and powerful commercial software. On the other hand, the image-based approach (Remondino and El-Hakim, 2006), advanced by recent developments in computer vision (Hartley and Zisserman, 2003; Goesele et al., 2007; Pollefeys et al., 2008; Snavely

et al., 2008; Furukawa et al., 2010; Shan et al., 2013), is presently offering additional automated procedures for both image orientation (Remondino et al., 2012) and 3D reconstruction (Lafarge and Mallet, 2012; Haala, 2013) at different scales. Complex scenes and objects can be surveyed and reconstructed using a large set of images with very satisfactory results (Fig. 1). In particular, methods for dense point-cloud generation (dense image matching) are increasingly available for professional and amateur applications such as 3D modelling and mapping, robotics, medical imaging, surveillance, tracking and navigation.

Due to the availability of a number of different low-cost and open-source software systems, automated 3D reconstruction methods are becoming very popular. Nevertheless, the metrological and reliability aspects of the resulting 3D measurements and modelling should not be ignored, particularly if the community wishes to adopt such solutions not only for quick 3D modelling and visualisation but also for accurate measurement purposes. To this end, clear accuracy statements, benchmarking and evaluations must be carried out.

This paper presents a critical review and analysis of selected dense image-matching algorithms. The algorithms considered are from both the commercial and open-source domains. The datasets adopted for the testing (Table I and Fig. 3) include terrestrial and aerial image blocks, acquired with convergent and normal (parallel axes) images at different scales and resolution. With respect to other reported benchmarking datasets, the imagery considered here is of higher resolution and it covers more complex scenes. Moreover, the evaluations presented are performed on the raw output of the matching (that is, on the point cloud) and not at the mesh level. The algorithms are evaluated according to their ability to produce dense and high-quality 3D point clouds, as well as according to computation time. Geometric analyses are reported, in which the point clouds produced with each of the different algorithms are compared with one another and also to ground-truth data.

Laser Scanning or Photogrammetry?

Since 2000, range sensors, both airborne and terrestrial, have been employed for various applications; continuous improvements in both hardware and software have been made as a result of technological advancements. Thus, for more than a decade, range sensors have been growing in popularity as a fundamental source of dense point clouds for 3D documentation, mapping and visualisation purposes at various scales. Over the same period, until quite recently, photogrammetry was not able to efficiently deliver dense and detailed 3D point clouds similar to those produced by ranging instruments. Consequently,

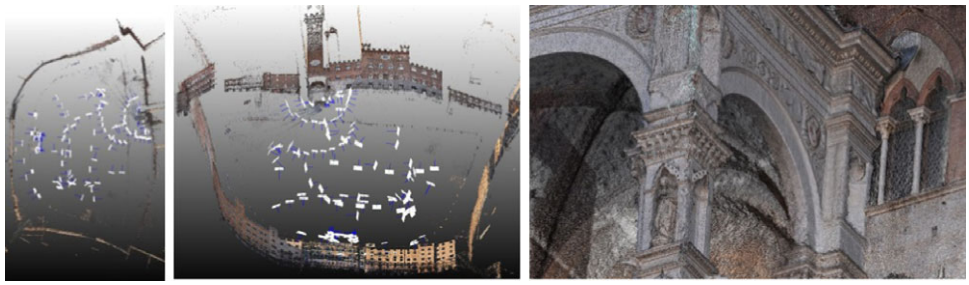

FIG. 1. Example of a very complex image network (more than 150 convergent high-resolution images) and the derived dense point cloud of the surveyed scene.

range sensors became the dominant technology for dense 3D recording, replacing photogrammetry in many application areas. Furthermore, many photogrammetric scientists shifted their research interests to laser scanning, slowing advances and new developments in automated procedures using photogrammetric technology. Thanks to recent significant improvements in hardware (such as better dynamics and radiometry) and algorithms (for example, structure from motion (SfM) or innovative image-matching algorithms), photogrammetry has re-emerged as a competitive technology and a resurgence in automated photogrammetric methods is now evident. Image-based surveying and 3D modelling can now deliver, in reasonable time, results of comparable geometric characteristics to those of laser scanning for many terrestrial and aerial applications. Therefore, the market, which was previously primarily dominated by airborne and terrestrial range sensors, nowadays offers more image-based measurement tools for 3D recording and modelling.

Range sensors, although smaller and lighter than they were some years ago, are still relatively cumbersome and expensive compared to terrestrial digital cameras (off-the-shelf or SLR types); their bulkiness can be problematic in some field campaigns and research projects. The point clouds recorded with range instruments may be metrically correct at the outset, but they are not based on redundant measurements, which can be problematic for projects concerned with absolute accuracy. Typical photogrammetric quality indicators derived in adjustment procedures (variance estimations or other statistical measures) are not available for quality evaluation of point clouds produced with range sensors. Moreover, few statistics (normally provided by vendors) are provided to describe errors for the entire dataset. Range sensors are normally treated as "black boxes" as they lack well-defined procedures to assess quality on an individual project basis.

On the other hand, photogrammetric processing, although more labour intensive, can be carried out so that calibration procedures, systematic error corrections and error metrics are explicitly stated. This is mainly valid for pure photogrammetric processes, while SfM tools also fit into the black-box category where bundle adjustment divergence or geometric deformations may be experienced (Remondino et al., 2012). Photogrammetric processing algorithms can suffer problems due to either the initial image quality (noise, low radiometric quality, shadows and so on) or to certain surface materials (shiny or textureless objects). This can result in noisy point clouds and/or difficulties in feature extraction. Furthermore, in order to derive metric 3D results from images, a known distance or ground control points (GCPs) are required in an image-based measurement project. In the case of aerial acquisitions, the typical point density of laser scanning datasets is 1 to 25 points/m^2 while an aerial photogrammetric image typically has a ground sampling distance (GSD) of the order of 10 cm, which could theoretically be used to produce a dense point cloud with 100 points/m^2.

Several recent publications have compared ranging and imaging techniques based on factors such as accuracy, resolution and dense 3D reconstructions (Kersten and Lindstaedt, 2012; Opitz et al., 2012; Koutsoudis et al., 2014). The choice between the two techniques nowadays depends primarily upon project constraints and requirements, budget and experience, and rather less on the geometric properties of point density and accuracy.

DENSE IMAGE-MATCHING ALGORITHMS

Concepts and History

Matching can be defined as the establishment of correspondences between various datasets (for example, images, maps and 3D shapes). In particular, image matching is

concerned with the establishment of correspondences between two or more images (Schenk, 1999). In computer vision, image matching is often called the *stereo correspondences problem* (Szeliski, 2011; Sonka et al., 2014). Image matching requires the establishment of correspondences between primitives extracted from two or more images, along with the determination of the 3D coordinates of matched feature points via a collinearity or projective model. In image space, this process produces a *depth map* that assigns relative depths to each pixel of an image. The corresponding outcome in object space is the *3D point cloud*. Considering an image pair, the disparity (or parallax, that is, horizontal discrepancy) is inversely proportional to the camera-to-object distance. Even in situations where the visual understanding and basic geometry relating to the disparity and scene structure are well understood, the automated measurement of such disparities by establishing dense and accurate image correspondences remains a challenging task.

For historical reasons, photogrammetric developments in the field of image matching were mainly related to aerial images and topographic mapping problems. The earliest matching algorithms were developed in the photogrammetry community in the 1950s (Hobrough, 1959; Williams, 1959). In the 1970s, the concepts of epipolar geometry and cross-correlation for image matching were introduced (Helava, 1978). With the advent of digital imaging, research was focused more on automated procedures to both replace manual operator intervention and achieve more powerful matching performance for single points (Förstner, 1982; Ackermann, 1984; Gruen, 1985). In Gruen (1985) and Gruen and Baltsavias (1988), the *multi-photo geometrical constraints* (MPGC) concept was introduced. Subsequently, the matching procedure was also generalised to object space through the introduction of the concept of the "groundel" or "surfel" (Wrobel, 1987; Ebner and Heipke, 1988; Helava, 1988).

The 1990s was a time of consolidation for image matching and a large number of commercial photogrammetric systems appeared for digital surface model (DSM) and digital terrain model (DTM) generation from large blocks of near-nadir aerial images. Thus, from sparse (single point) matching, algorithms moved to determination of dense point correspondences. Zhang et al. (1992) presented innovative global matching with probabilistic relaxation, while Maas (1996) introduced a MPGC approach where features were searched for and matched along epipolar lines. However, in spite of this success, the photogrammetric matching of convergent images in non-traditional reference systems (namely, close-range image blocks) was proving to be problematic and thus new developments in this area were largely confined to the computer vision community where accuracy was not a high priority. In this field, stereomatching was investigated as early as the mid-1970s (Marr and Poggio, 1976) and developments continued in the 1980s mainly for terrestrial applications (Baker and Binford, 1981; Marr, 1982; Ohta and Kanade, 1985; Dhond and Aggarwal, 1989). Then, in the 1990s, the focus moved to multi-view approaches (Okutomi and Kanade, 1993; Fua and Leclerc, 1995; Narayanan et al., 1998), and then more recently to field programmable gate array (FPGA) and graphics processing unit (GPU) developments in computer architecture (Kalarot et al., 2011), per-pixel measurement (Birchfield and Tomasi, 1999), global energy minimisation algorithms (Roy and Cox, 1998; Hirschmüller, 2008) and dynamic programming approaches (Kolmogorov et al., 2006).

Algorithms and Classifications

The large number of image-matching algorithms developed to date in the scientific community precludes the presentation of a summary here. The reader is instead referred to surveys and comparisons of matching algorithms presented in Makarovic (1992), Scharstein

and Szeliski (2002), Brown et al. (2003), Seitz et al. (2006), Hirschmüller and Scharstein (2009) and Gruen (2012).

The first intuitive classification of image-matching algorithms is based on the utilised primitives, namely, image intensity patterns (windows composed of grey values around a point of interest) or features (for example, edges and regions), leading to *area-based matching* (ABM) or *feature-based matching* (FBM) algorithms (Remondino et al., 2008).

The image-matching problem is nowadays solved using stereopairs (*stereomatching*) (Hirschmüller, 2008; Gehrig et al., 2009; Tola et al., 2010; Haala and Rothermel, 2012; Hermann and Klette, 2013) or via identification of correspondences in multiple images (*multi-view stereo* – MVS) (Collins, 1996; Zhang, 2005; Pierrot-Deseilligny and Paparoditis, 2006; Goesele et al., 2007; Remondino et al., 2008; Furukawa and Ponce, 2010; Vu et al., 2012; Toldo et al., 2013).

According to Szeliski (2011), stereo methods can be local or global. Local (or window-based) methods compute the disparity at a given point using the intensity values within a finite region, with implicit smoothing assumptions and a local "winner-take-all" optimisation at each pixel. On the other hand, global methods make explicit smoothness assumptions and then solve for a global optimisation problem using an energy minimisation approach, based on regularised (variational) Markov random fields (MRFs), graph-cut, dynamic programming or max-flow methods. Most of these procedures apply consistency measures only to single stereopairs, while geometric constraints are applied only during the fusion of the point clouds derived by the stereopairs or via some volumetric approaches.

The most common multi-view algorithms rely on silhouette and fusion (Hernández Esteban and Schmitt, 2004), volumetric graph-cut (Vogiatzis et al., 2007), patch-based methods (Furukawa and Ponce, 2010) and global optimisation (Vu et al., 2012).

Most of the proposed matching methods are based on *similarity* or *photo-consistency measures*; in other words, they compare pixel values between the images. These measures can be defined in either image or object space, according to the algorithms (stereo or multi-view). The most common measures (or matching costs) include squared and absolute intensity differences, normalised cross-correlation (NCC), dense feature descriptors, census transform, mutual information, gradient-based algorithms and bidirectional reflectance distribution functions (BRDFs). Dense multi-view reconstructions apply multi-image radiometric consistency measures and geometric constraints (Gruen and Baltsavias, 1988; Zhang, 2005). Some algorithms use normalised and distortion-free images, whose adoption both simplifies and speeds up the search for correspondences. In Campbell et al. (2008) and Toldo et al. (2013), each pixel of an image is associated with a number of candidate depths. Then, using MRF optimisation, a final and unique depth map is created.

Many approaches require a rough surface model of the scene in order to initialise the matching procedure (Lhuillier and Quan, 2002; Furukawa and Ponce, 2010). Such models can be derived in different ways, for example, by using a point cloud interpolated on the basis of tie points obtained from the orientation stage or using already existing 3D models from low-resolution range data. Other methods are organised in a hierarchical framework which first generates a rough surface reconstruction, which is refined and made denser at a later stage (Sinha and Pollefeys, 2005; Vogiatzis et al., 2007).

Other algorithms are still based on local (window-based) correlation models where a constant disparity within a correlation window is adopted, hence implicitly assuming a smooth surface. The larger the size of this window, the greater is the robustness of the matching. However, the implicit assumption of constant disparity inside the window is violated for elements with geometric discontinuities which leads to blurred object boundaries and over-smoothing. Matching algorithms based on intensity differences are very

sensitive to recording and illumination differences and are not reliable in poorly textured or homogeneous regions. On the other hand, dense pixel-wise matching based on global algorithms assumes an explicit formulation for the smoothing assumption and solves it with a global optimisation approach (Szeliski, 2011). Such methods, however, only operate on stereopairs, thus reducing the potential to recover statistical information intrinsic in multi-view methods.

Accuracy Analyses

Algorithm developments have always been accompanied by accuracy assessment studies. Early analyses are reported in Schewe and Förstner (1986) for industrial applications, while Rosenholm (1986) investigated the potential of matching algorithms in the case of aerial images. His tests were limited to quantifying single point accuracy. Lately, accuracy investigations have been carried out on surface model generation (Smith and Smith, 1996; Gong et al., 2000; Remondino and Menna, 2008; Bartelsen et al., 2012; Kersten and Lindstaedt, 2012) by comparing the results achieved to ground truth. Benchmarking datasets for close-range and aerial image configurations was proposed in Seitz et al. (2006), Strecha et al. (2008) and Haala (2013). The establishment of a good reference dataset (ground truth) with an accuracy which is two to three times better than the expected matching results is not a straightforward task. It is also complicated to define just what should be compared and evaluated (the entire surface, small patches or single points?) and with which procedure (for example, the Euclidean distance). Seitz et al. (2006) evaluated completeness (how much of the scene has been reconstructed with respect to the ground truth) and accuracy (how close the result is to the ground truth), based on Euclidean distances and nearest points. Haala (2013) correctly reported that pixel-wise matching provides for surface representations at a geometric resolution which is frequently higher than that available from standard lidar flights. Therefore, he adopted a median DSM from all the surface models being evaluated, though such a measure still falls short of constituting independent ground truth at a higher order of accuracy. However, the approach is very suitable as a means of highlighting differences between the different matching algorithms. In cases where known geometric shapes (such as spherical or cuboid reference objects) are employed for the algorithm evaluation (Hosseininaveh Ahmadabadian et al., 2013), the accuracy analysis can be performed using evaluation parameters defined in the German engineering VDI/VDE (2002; 2008) guidelines.

Recent Innovations

The real innovation that has been introduced in several dense image-matching methods during recent years, besides the pixel-wise approach, is the integration of different basic correlation algorithms, consistency measures, visibility models, shape knowledge, constraints and minimisation approaches into a multi-step procedure. In many cases this works through a multi-resolution approach. Commercial packages are also moving in this direction, although the ability to process both close-range convergent images (or oblique aerial imagery) as well as vertical aerial images is still very limited.

As dense image matching is a task involving a large computing effort, researchers have started to use advanced techniques such as parallel computing and implementation at the GPU/FPGA level in order to reduce the computational time and allow real-time depth map production (Kalarot et al., 2011; Hirschmüller et al., 2012).

EVALUATED DENSE IMAGE-MATCHING ALGORITHMS

Each of the four tested and analysed dense image-matching algorithms will now be briefly summarised. Further details can be found in the referenced publications. A qualitative example of image-matching results achieved with the evaluated software systems is shown in Fig. 2.

SURE

SURE (photogrammetric SUrface REconstruction from imagery) (http://www.ifp.uni-stuttgart.de/publications/software/sure/index.en.html) is a MVS method (Haala and Rothermel, 2012; Rothermel et al., 2012) where a reference image is matched to a set of adjacent images using a semi-global matching (SGM) type of stereo algorithm (Hirschmüller, 2008). For each pair, a disparity map is computed; then all disparity maps sharing the same reference view are merged into a unique final point cloud capitalising on the redundancy across the stereopairs. Within a premodule, a network analysis and selection of suitable image pairs for the reconstruction process is performed. Epipolar images are then generated and a time- and memory-efficient SGM algorithm is applied to produce depth maps. All these maps are then converted in 3D coordinates using a fusion method based on geometric constraints that both help in reducing the number of outliers and increase precision. With respect to the classical SGM approach, SURE searches pixel correspondences using dynamic disparity search ranges and a tube-shape structure is employed to store costs of potential correspondences. Moreover, SURE implements a blunder removal approach and it is computationally very efficient. Although SURE is a commercial product, a free version is available for research purposes.

MicMac

MicMac (http://www.MicMac.ign.fr) is a multi-resolution and multi-image method (Pierrot-Deseilligny and Paparoditis, 2006) which implements a coarse-to-fine extension of the maximum-flow image-matching algorithm presented in Roy and Cox (1998). The surface measurement and reconstruction is formulated as an energy function minimisation problem, that is, a minimal cut is found on a graph. The global energy function accounts for both the correlation and the smoothing term. Thus, the problem is solved in polynomial time (implying it is reasonable to compute) with classical minimal cut and maximal flow graph theory algorithms. The procedure was originally developed to deal with large, high-resolution satellite images, but it can now also process large and complex terrestrial sequences or aerial blocks. MicMac uses pyramidal processing: starting from a lower resolution, the matching results achieved in each pyramid level guide the matching at the next, higher resolution, level in order to improve the quality of the matching, up to the full resolution.

MicMac operates according to two different strategies, called "*image*" and "*ground*" geometries. In the former approach, the user selects a set of master images for the correlation procedure; then for each candidate 3D point a patch in the master image is identified and projected to all the neighbouring images, and a global similarity is derived. In the *ground* geometry strategy, a voxel is defined according to the block size and camera-to-object distance; then every candidate 3D point is back-projected onto the images and a global similarity is derived. Finally, for both strategies, an energy minimisation approach is applied to enforce surface regularities and avoid undesirable jumps. MicMac is open source

and provides for detailed and accurate 3D reconstructions preserving surface discontinuities, thanks to its optimisation process. MicMac can be linked to the Institut Géographique National (IGN) orientation module *Apero* and to the ortho-image generator *Porto*. In the following tables and figures, MicMac is referred to by MM.

PMVS

PMVS (patch-based multi-view stereo, http://www.di.ens.fr/pmvs/) is a matching method (Furukawa and Ponce, 2010) which follows a multi-step approach that does not need any initial approximation of the surface. A "patch" p is a local tangent rectangle approximating a surface whose geometry is fully defined by the position of its centre $c(p)$ and the unit normal vector oriented towards a reference image $R(p)$ where it is viewed. After the initial matching step, a propagation of reconstructed semi-dense patches is performed with a final filtering to remove possible local outliers. In its original implementation, the surface-growing method simultaneously used all the images of the processed dataset, which can lead to a very large demand on computer memory. This issue was solved via a clustering of the input images, with the reconstruction of sub-spaces of the scene. PMVS software is open source and uses oriented and distortion-free images.

Photoscan

Agisoft Photoscan (http://www.agisoft.ru) is a commercial package able to automatically orient and match large datasets of images. Due to commercial considerations, little information is available concerning the internal matching algorithms employed. Nevertheless, from the authors' experience and from the achievable 3D measurement results, the implemented image-matching algorithm seems to be a stereo SGM-like method (for this study, version 0.9.0 of Photoscan was used). Normally the software delivers results that are already meshed, but for the evaluations in this study the "raw" point clouds were computed and exported (not the vertices of the mesh). Photoscan is referred to in the following tables and figures by PS.

FIG. 2. An image dataset processed with all the tested algorithms. Original image (left) and derived dense point clouds generated with SURE, MicMac (MM), PMVS and Photoscan (PS). In the lower images, the enlargements show notable differences between the point clouds.

TABLE I. Main characteristics of the employed datasets for the dense image-matching evaluation.

Dataset	*A* *Cube*	*B* *Augusto*	*C* *Stele*	*D* *Fountain*	*E* *Buergerhaus*	*F* *Ventimiglia*	*G* *Marseille*	*H* *Primiero*
Width (m)	0·1	0·3	0·7	5	9·5	20	2000	800
Length (m)	0·1	0·2	2·0	6	9·5	30	1300	500
Height (m)	0·1	0·2	0·35	2·5	1·5	5·5	80	20
Number of images	9	6	8	12	12	10	6	9
Camera	Nikon D3X	Nikon D3X	Canon EOS 5D Mark II	Canon EOS D60	Nikon D90	Nikon D3X	Z/I DMC	Nikon D3X
Image size (pixel)	6048 × 4032	6048 × 4032	5616 × 3744	3072 × 2048	2144 × 1424	6048 × 4032	13 824 × 7680	6048 × 4032
Pixel size (μm)	5·9	5·9	6·4	7·4	11	5·9	12	5·9
Focal length (mm)	50	50	50	20	20	50	120	50
Approx. image scale	1:10	1:15	1:65	1:450	1:450	1:500	1:10 000	1:16 000
Min–max GSD (mm)	0·06–0·07	0·07–0·08	0·3–0·5	2·8–3·9	3·8–6	2·6–3	120	95

Datasets and Evaluation Methodology

In order to evaluate the performance and potential of the four dense image-matching algorithms under various conditions, eight test datasets were selected. Specifically, five terrestrial photogrammetric datasets were considered: A – Cube, B – Augusto, C – Stele, D – Fountain (Strecha et al., 2008) and E – Buergerhaus. Additionally, there were three aerial cases: F– Ventimiglia, G – Marseille and H – Primiero (Fig. 3). The main characteristics of the employed datasets are summarised in Tables I and II. The datasets are characterised by different image scales (ranging from 1:16 000 for the Marseille aerial case to 1:10 for the Cube dataset), image resolution, number of images, camera network, baseline length, object texture and size.

With respect to other benchmarking datasets available in the photogrammetric and computer vision communities (for example, Middlebury – www.vision.middlebury.edu), the employed images have a higher resolution and feature more complex real scenes and objects, with no black background or controlled image acquisition. Moreover, the evaluations presented are performed on the raw output of the matching (that is, on the 3D point cloud) and not at mesh level. In order to have a common starting point for testing the image-matching algorithms, all the datasets were first oriented via a bundle adjustment and the images were undistorted. The orientation parameters achieved and the idealised images were then used to run all the image-matching tests. The matching algorithms were run trying to ensure similar settings between them.

All the tests were carried out using the second-level image pyramid, that is, at a quarter of the original image resolution. There were two reasons for this choice: (a) to keep the time and processing effort low (a reasonable trade-off between processing effort and resolution); and (b) to allow a comparison with the ground truth where available. Thus, the derived dense point clouds have a sampling resolution of twice the original image GSD. The image-matching results were evaluated using the following procedures:

(1) For the Augusto, Fountain and Buergerhaus case studies, the photogrammetric point clouds (PH) were compared against a meshed model obtained with a terrestrial laser scanner (TLS).
(2) For the Cube dataset, a flatness measurement was performed on both the top and a side face.
(3) For the Stele and aerial cases, selected cross sections were extracted and the profiles obtained were compared.

Table II. Main characteristics of the ground truth employed for each dataset.

Dataset	*Ground truth*	*Evaluation*	*Spatial resolution (mm)*	*Accuracy*
A – Cube	Best-fit plane	Flatness error	–	–
B – Augusto	Triangulation TLS mesh	Euclidean distances	0·5	0·2 mm
C – Stele	–	Profiles	–	–
D – Fountain	ToF TLS mesh	Euclidean distances	3	<5 mm @ 10 m
E – Buergerhaus	ToF TLS mesh	Euclidean distances	10	<5 mm @ 10 m
F – Ventimiglia	ToF TLS mesh	Profiles	3	<5 mm @ 10 m
G – Marseille	–	Profiles	–	–
H – Primiero	–	Profiles	–	–

TLS = terrestrial laser scanner; ToF = time of flight using phase shift.

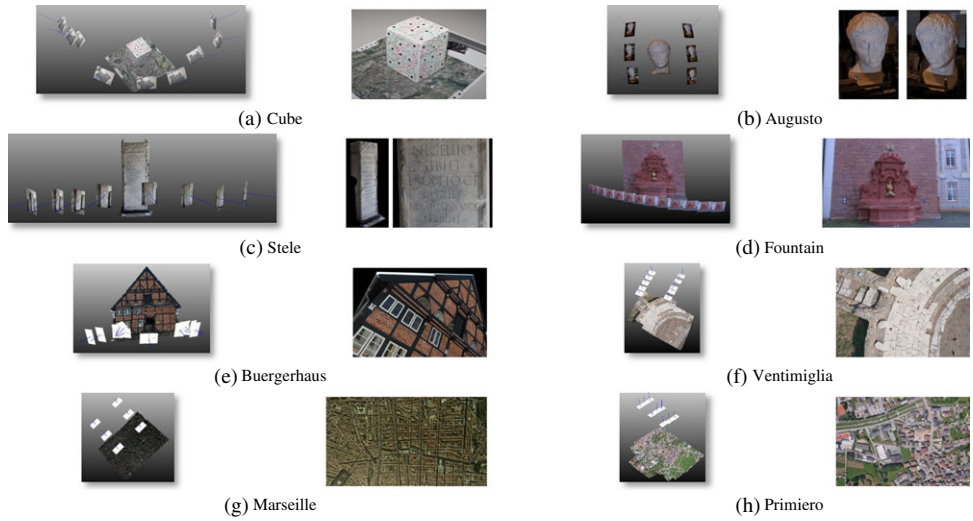

FIG. 3. The image datasets used for the evaluation and analysis of dense matching algorithms. For each dataset, the image network (left) and one to two images are shown.

EVALUATION RESULTS

A – The Cube Dataset

The Cube is a porous object made of building materials (breeze block). The dataset features a very small GSD (smaller than 0·07 mm) and convergent images with a maximum angle of about 80°. The planar top face and one side face of the cube, highlighted in red and blue, respectively, in Fig. 4(a), were selected from the measured dense point clouds and a best-fit plane was computed for each face. The photogrammetric targets placed on the object (Fig. 4(b)) were excluded from the computation. The deviations from the best-fitting planes are shown in Fig. 5. For both faces, all the matching algorithms delivered an average flatness error smaller than the GSD (equal to 0·13 mm on the object) employed in the second-level pyramid matching.

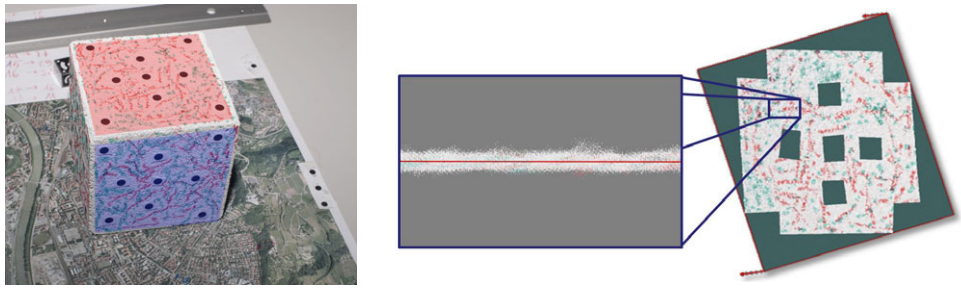

FIG. 4. (a) The top and a side face of the cube, highlighted in red and blue, respectively. (b) Best-fitting plane computed on one side.

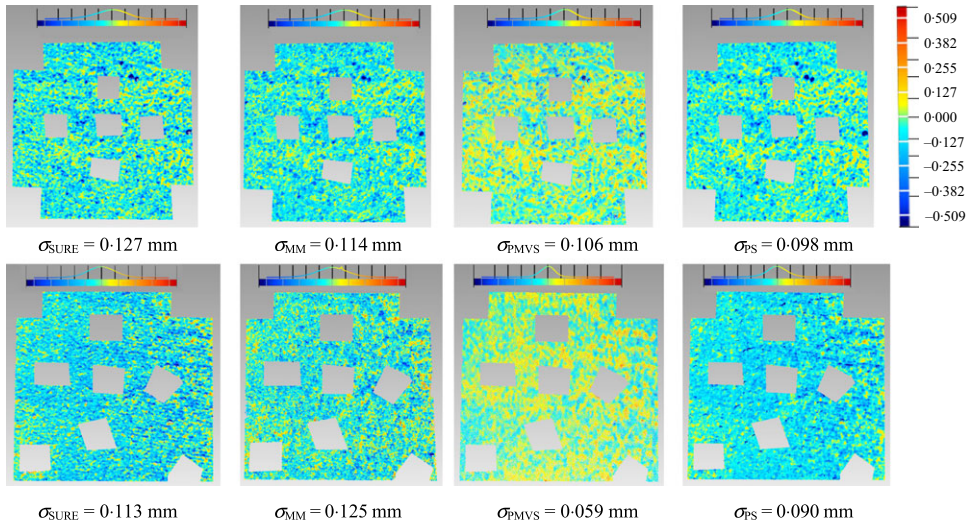

FIG. 5. Evaluation results for the Cube. Upper: top face (red in Fig. 4(a)). Lower: frontal side face (blue in Fig. 4(b)). Small deviations from ground truth shown in light blue and yellow; large deviations in dark blue and red.

The four matching algorithms showed similar behaviour, except for PMVS which was characterised by an opposite systematic deviation trend. Nevertheless, for the top face the highest deviations (of the order of four times the average flatness error) were localised in the same position for all the matching results, corresponding to the true roughness of the object surface. This was not the case on the side face as the images featured a somewhat inclined view, and the algorithms tested did not always correctly match the images and reconstruct the true roughness of the object.

B – The Augusto Dataset

Augusto is a small statue made of white matte plastic, approximately 35 cm in height and 20 cm in width. A mesh model from a triangulation-based laser scanner was used as ground truth for the analyses.

The photogrammetric survey of the statue was realised with a Nikon D3X camera with a random grey pattern projected onto the object. The deviations from the ground truth for all four matching solutions were of the order of the matching sampling resolution (about 0·15 mm on the object), except for MM which produced higher deviations (Fig. 6).

C – The Stele Dataset

The Stele, an inscribed marble slab, has quite a homogenous texture, so it constitutes a challenging dataset for matching algorithms. To evaluate the matching results, two profiles, across two pairs of letters, were extracted from the photogrammetric point cloud, as indicated in Fig. 7.

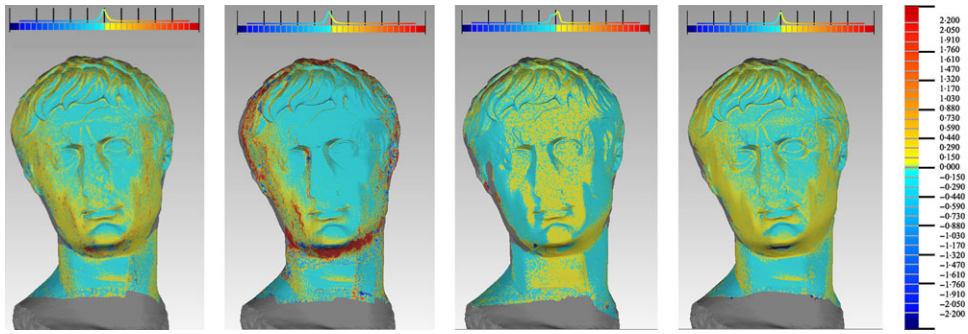

FIG. 6. Evaluation results for the Augusto statue dataset. Small deviations from ground truth shown in light blue and yellow; large deviations in dark blue and red.

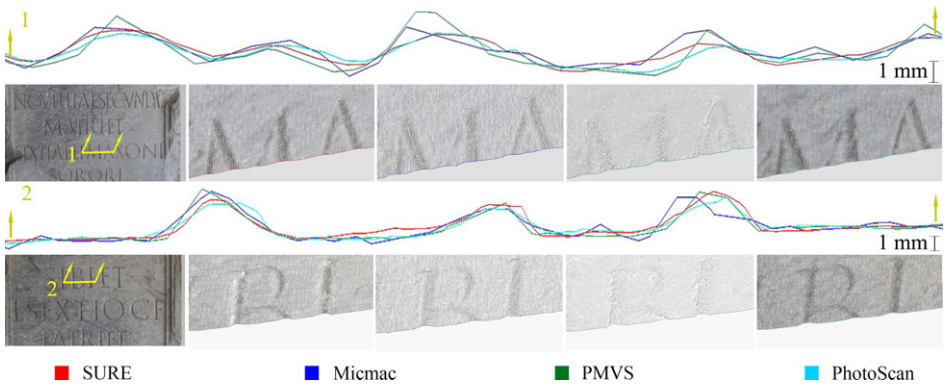

FIG. 7. Evaluation results for the Stele dataset along the two profiles. Upper images for the letter pair "MA". Lower images for letter pair "BI".

The profiles refer to the pairs of letters "MA" and "BI". In the first pair of letters, the profiles from SURE and Photoscan were smoother and more regular than those from MicMac and PMVS. In the second pair, the differences between the profiles from each of the algorithms were less evident. The maximum difference was of the order of 1 mm, which corresponds to the mean GSD of the image pyramid level used for the matching.

D – The Fountain Dataset

The Fountain dataset was selected for its interesting shape, with undercuts, complex relief detail and a quite uniform texture. The image-matching results were compared to a reference TLS dataset and the surface deviations, indicated by Euclidean distances, are shown in Fig. 8. On the wall beside the fountain and on flat fountain elements, the deviations for three of the four matching algorithms were of the order of the image-matching sampling resolution (a GSD of about 6 mm on the object); the exception was MicMac which produced slightly noisier results than the other matching solutions. For all

the matching algorithms, the highest deviations were concentrated in edge areas. Overall, the surfaces produced by MicMac and Photoscan were nosier than those from PMVS and SURE.

E – The Buergerhaus Dataset

This terrestrial dataset of a historic German building features convergent imagery with highly variable baselines and camera-to-object distances. Possibly due to the non-conventional camera network, the following behaviour of the image-matching algorithms was observed: (a) SURE was not able to reconstruct the bottom left part of the façade; (b) a similar problem was observed with MicMac as the software requires a reference image while the dataset lacked an image depicting the whole façade; (c) PMVS delivered the whole façade but the surface area at the bottom left contained excessive noise. The surface differences between the photogrammetric point clouds and the reference TLS data are indicated in Fig. 9.

The largest standard deviations, of the order of 10 mm, were observed in the MicMac and PMVS point clouds, whereas the smallest, equal to 7 mm, was obtained with SURE. It is noteworthy that these values are all of the order of the point-cloud spatial resolution (the GSD of the image pyramid level used for the matching was close to 10 mm). As expected, the largest deviations of ± 40 mm were observed in textureless and dark regions. A small curvature in the shape of the photogrammetric point clouds obtained with MicMac and Photoscan (especially) was also observed.

F – The Ventimiglia Dataset

The flight over the Roman amphitheatre at Ventimiglia in Liguria, Italy, was realised with an unmanned aerial vehicle (UAV) (remotely piloted aircraft system – RPAS) platform (a model helicopter) equipped with a Nikon D3X camera. A time of flight (ToF) phase-shift TLS was also used to survey the amphitheatre and the derived mesh model was used as the reference surface. The image-matching sampling resolution was about 6 mm on the object, which is twice the original image GSD. For the geometric evaluation, the photogrammetric point clouds were segmented into steps, as indicated in Fig. 10. The matching algorithms were able to correctly reconstruct the step tread surfaces, but not the vertical surfaces of the

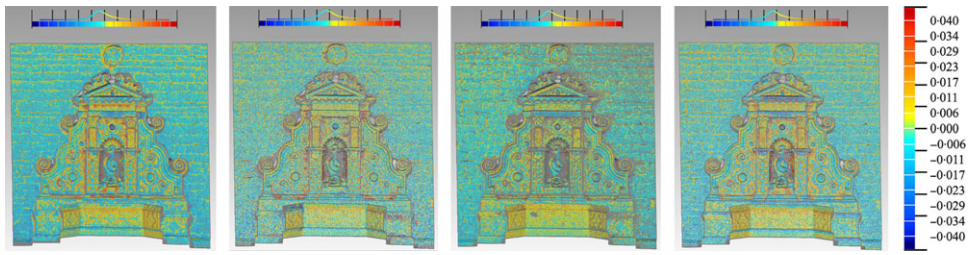

FIG. 8. Evaluation results for the Fountain dataset. Comparison and standard deviations between the TLS ground truth and the photogrammetric point clouds. Small deviations from ground truth shown in light blue and yellow; large deviations in dark blue and red.

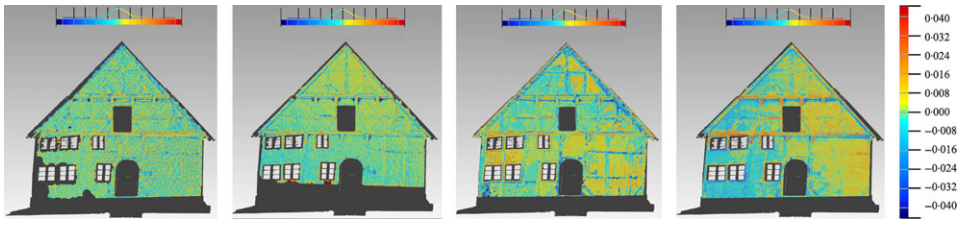

FIG. 9. Evaluation results for the Buergerhaus dataset. Differences between the photogrammetric point clouds and the ground truth (TLS) are reported. Small deviations from ground truth shown in light blue and yellow; large deviations in dark blue and red. Black areas represent no data.

risers. This was largely anticipated because of the low incidence angles on these surfaces from the vertical photography. A smoothing effect can be observed in the extracted surfaces produced by all four matching algorithms.

G – The Marseille Dataset

This dataset in southern France is composed of aerial images acquired with a DMC large-format camera. It is characterised by a small image scale, but a high image resolution (large number of pixels). For the comparative analysis, as no ground truth was available, a reference building was selected and two cross sections were extracted from the photogrammetric point clouds (yellow P1 and green P2 planes in Fig. 11).

In this case, the image-matching sampling resolution was about 24 cm on the object, which is again double the original image GSD. The extracted profiles show that the ground level was equally well reconstructed by three of the matching algorithms, but less well produced by PMVS. The bell tower was reconstructed by MicMac and PMVS, while points on the vertical walls were provided only by MicMac and Photoscan.

H – The Primiero Dataset

These images with a 9·5 cm GSD over the Trentino region of Italy, were acquired with a Nikon D3X camera from a helicopter flying over an urban area. Shown in Fig. 12 are height profiles 1 and 2 taken from the point clouds. These traverse three buildings and a large roof, respectively. The SURE profiles reveal that the software was able to correctly reconstruct the ground surface and the shape of the roofs, with little noise and few outliers. Similar results were obtained with MicMac, while PMVS and Photoscan produced noisier and over-smoothed profiles, respectively. Photoscan was able to reconstruct the vertical walls of the buildings, illustrating that its matching algorithm uses stereopairs whose point clouds are then merged into a unique final point cloud (the other algorithms might use higher thresholds for a minimum number of rays for each 3D point).

CONCLUSIONS

This paper has presented a review of the performance of four state-of-the-art dense image-matching software systems, two each from the commercial and open-source domains. From the accuracy analyses of point clouds derived via image matching conducted in this

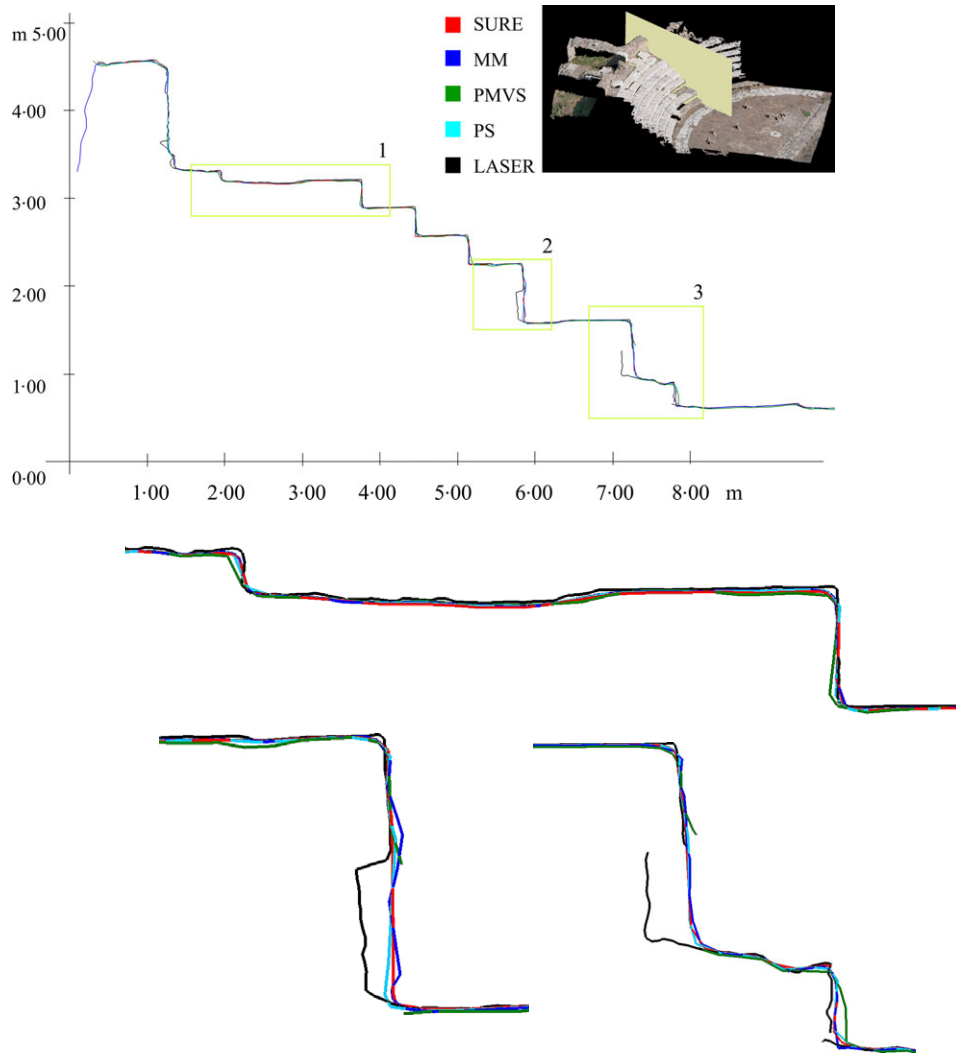

FIG. 10. Evaluation results for the Ventimiglia dataset. The profiles along the theatre steps (1, 2 and 3 in the upper image) are enlarged and highlighted in the central (1) and lower (2 and 3) images.

investigation, which covered eight varying objects and scenes, it can be clearly seen that photogrammetry continues to be a very viable technique for dense 3D reconstruction purposes. Photogrammetry could be said to have re-emerged from lidar's shadow into a viable highly-automated technology, able to provide precise and dense 3D surface measurements of complex and detailed objects at various scales. Stereo or multi-image approaches are available, based on local or global minimisation methods. The results from this investigation highlight both the significant advantages and some disadvantages of the photogrammetric approach utilising dense matching. Many of these are functions of input parameter selection. On one hand this is good because the user can control and adjust the

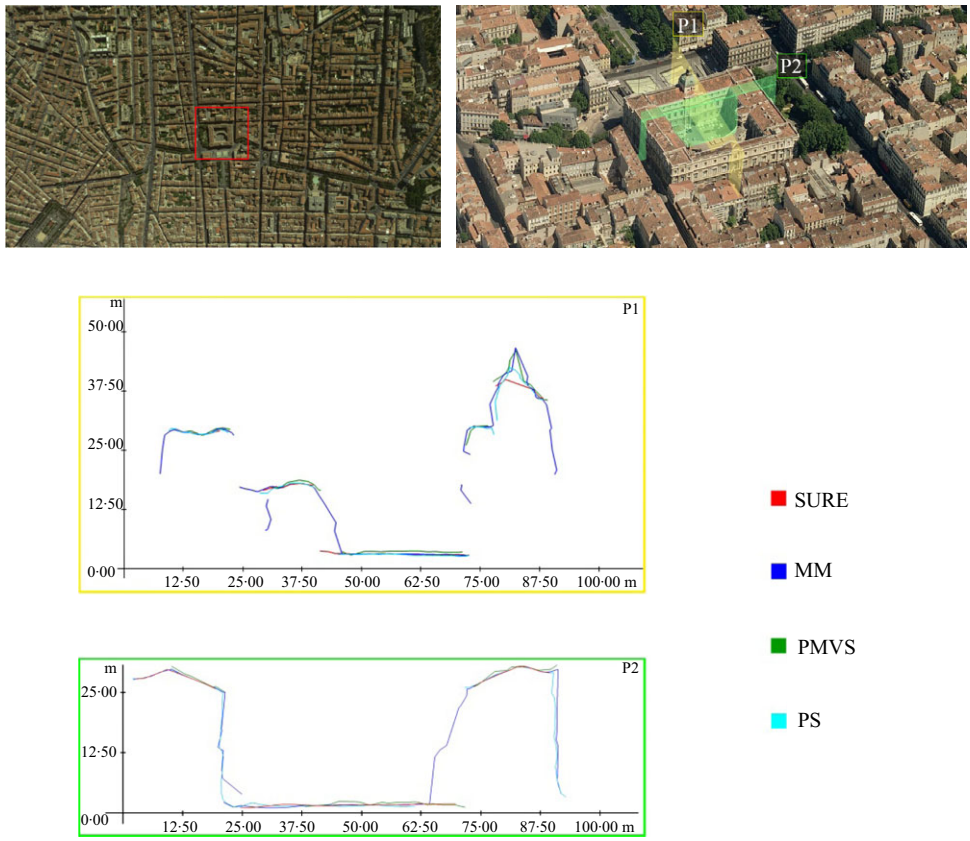

FIG. 11. Evaluation results for the Marseille dataset. The two profiles P1 and P2, orthogonal to each other and to the building quadrangle highlighted in the upper right image, are shown in the two lower plots.

object reconstruction performance according to the characteristics of the dataset. On the other hand, it is not optimal to have too many, and often unclear, parameters as they can affect the results achieved by non-experts who prefer fully automated "black-box" tools.

Comparative assessments of the accuracy and performance of the four dense matching algorithms have not been an easy task. Different evaluations have been reported using eight varying datasets and imaging configurations which cover a broad range of real applications and provisions of ground truth. No ranking has been reported, primarily due to the fact that dense image matching is a topic of ongoing research and new developments are continuously being implemented, notwithstanding the difficulties in defining appropriate settings to run the matching. While the reported results illustrate that all four matching algorithms have considerable potential, it should be recalled that due care must be taken in the image acquisition and parameter selection phases, otherwise even a well-proven and robust matching method may be found to have produced an unsatisfactory 3D reconstruction. In the experiments reported, the parameters for the four matching algorithms

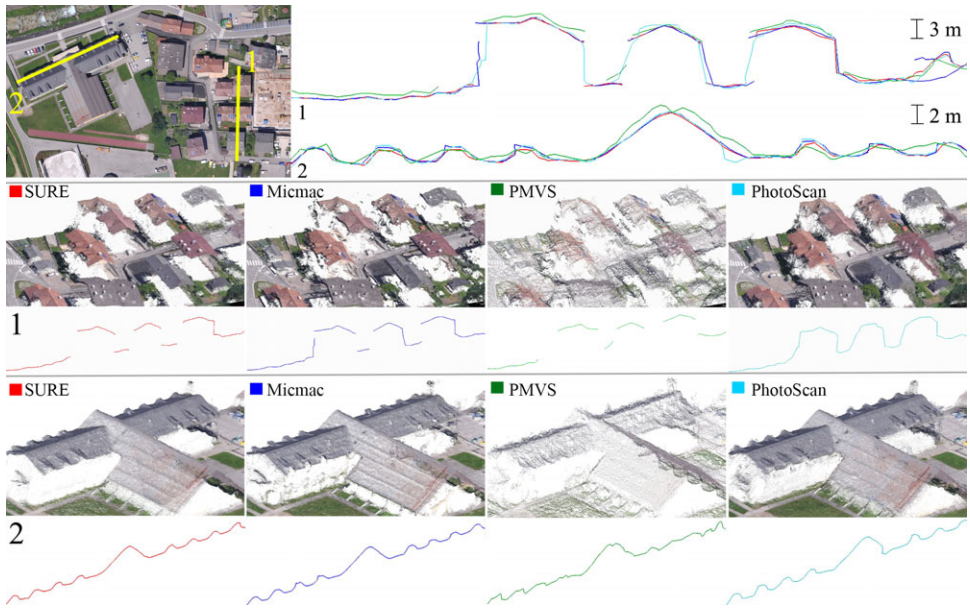

FIG. 12. Evaluation results for the Primiero dataset along profiles 1 and 2 in the upper left image. The profiles show certain noise in the data as well as some smoothing effects.

were optimally tuned on the basis of the operator's experience, gained after familiarisation with the software. A deep analysis of the reported results has highlighted that some imaging configurations and scene/object characteristics – such as the presence of shadows, sharp discontinuities and small structures – still cause problems for dense matching algorithms.

An evaluation of the computational time necessary to perform the dense 3D reconstructions (Fig. 13) shows that the processing time is, as expected, directly proportional to the dimensions of the dataset (total number of pixels) and also to the number of reconstructed 3D points.

It should be noted that large differences in the total number of points generated by the four matching algorithms (see Fig. 13) do not necessarily indicate significant quality differences in the final 3D reconstruction. For example, PMVS employs a true multi-image matching approach, meaning that for each object point visible in multiple images only one unique 3D point (which satisfies certain geometric conditions) is computed. On the other hand, a 3D point is computed for each pixel in the overlapping area of each stereopair in the Photoscan and SURE methods. In such cases, for n stereopairs, n 3D points corresponding to the same object point can be computed. This is particularly true in the case of large GSD and sub-pixel matching, leading to clusters of 3D points grouped near each other in the object space (but representing the same 3D point). This large number of points then can be successively averaged or statistically reduced to a cloud of unique points, but the user needs to consider a proper workflow that takes into account the point-cloud processing requirements for point averaging, de-noising and filtering.

Doubtless, new algorithms that incorporate the best features of existing methods, as exemplified by the four algorithms considered in this paper, are presently being developed. The requirements for such an optimal image-based 3D reconstruction system providing successful, powerful and reliable image matching should be:

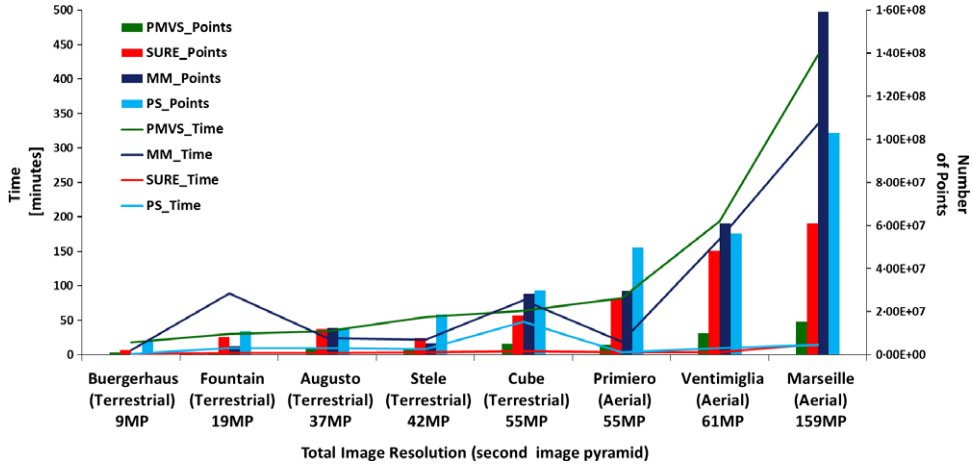

FIG. 13. Evaluation of the computational time for the image datasets employed. The time is compared with the total number of pixels of the datasets and with respect to the number of retrieved 3D points. An exponential behaviour of the time with respect to the number of derived points is, as expected, clear, but not overly strong for all the evaluated solutions.

(1) utilise accurately calibrated cameras and oriented images;
(2) employ image networks with strong geometric configurations;
(3) use local and global image information;
(4) combine different matching strategies to overcome textureless areas and preserve geometric discontinuities;
(5) operate a global consistency check amongst the matching candidates;
(6) exercise constraints to restrict the search space;
(7) consider an estimated shape of the object as a priori information;
(8) use multiple images simultaneously for reliability reasons;
(9) employ strategies to monitor and report matching results and quality;
(10) produce dense point clouds with a sufficient resolution to describe the object's surface and its discontinuities;
(11) adaptively tune the parameters in order to preserve edges and to avoid too many points in near-planar areas;
(12) guarantee 3D points in regions with poor textures or illumination and scale variations; and
(13) provide for quality indicators, including matching statistics and accuracy and reliability measures.

Some of these points provide interesting future research topics.

ACKNOWLEDGEMENTS

The authors are sincerely thankful to Konrad Wenzel and Mathias Rothermel (IFP, Stuttgart University, Germany) for useful discussions and support for the SURE matching, and also to Thomas Kersten (HCU Hamburg, Germany) for providing the Buergerhaus dataset.

REFERENCES

ACKERMANN, F., 1984. Digital image correlation: performance and potential application in photogrammetry. *Photogrammetric Record*, 11(64): 429–439.

BAKER, H. H. and BINFORD, T. O., 1981. Depth from edge and intensity based stereo. *Proceedings of 7th International Joint Conference on Artificial Intelligence*, 2: 631–636.

BARTELSEN, J., MAYER, H., HIRSCHMÜLLER, H., KUHN, A. and MICHELINI, M., 2012. Orientation and dense reconstruction from unordered wide baseline image sets. *Photogrammetrie, Fernerkundung. Geoinformation (PFG)*, 4: 421–432.

BIRCHFIELD, S. and TOMASI, C., 1999. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3): 269–293.

BROWN, M. Z., BURSCHKA, D. and HAGER, G. D., 2003. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8): 993–1008.

CAMPBELL, N. D. F., VOGIATZIS, G., HERNANDEZ, C. and CIPOLLA, R., 2008. Using multiple hypotheses to improve depth-maps for multi-view stereo. *Lecture Notes in Computer Science*, 5302: 766–779.

COLLINS, R. T., 1996. A space-sweep approach to true multi-image matching. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 35: 8–363.

DHOND, U. R. and AGGARWAL, J. K., 1989. Structure from stereo – a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6): 1489–1510.

EBNER, H. and HEIPKE, C., 1988. Integration of digital image matching and object surface reconstruction. *International Archives of Photogrammetry and Remote Sensing*, 27(B11): 534–545.

FÖRSTNER, W., 1982. On the geometric precision of digital correlation. *International Archives of Photogrammetry*, 24(3): 176–189.

FUA, P. and LECLERC, Y. G., 1995. Object-centered surface reconstruction: combining multi-image stereo and shading. *International Journal of Computer Vision*, 16(1): 35–56.

FURUKAWA, Y., CURLESS, B., SEITZ, S. M. and SZELISKI, R., 2010. Towards internet-scale multi-view stereo. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 143: 4–1441.

FURUKAWA, Y. and PONCE, J., 2010. Accurate, dense and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8): 1362–1376.

GEHRIG, S. K., EBERLI, F. and MEYER, T., 2009. A real-time low-power stereo vision engine using semi-global matching. *Lecture Notes in Computer Science*, 5815: 134–143.

GOESELE, M., SNAVELY, N., CURLESS, B., HOPPE, H. and SEITZ, S. M., 2007. Multi-view stereo for community photo collections. *11th International Conference on Computer Vision*, 2: 265–270.

GONG, J., LI, Z., ZHU, Q., SUI, H. and ZHOU, Y., 2000. Effects of various factors on the accuracy of DEMs: an intensive experimental investigation. *Photogrammetric Engineering & Remote Sensing*, 66(9): 1113–1117.

GRUEN, A., 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3): 175–187.

GRUEN, A., 2012. Development and status of image matching in photogrammetry. *Photogrammetric Record*, 27(137): 36–57.

GRUEN, A. and BALTSAVIAS, E. P., 1988. Geometrically constrained multiphoto matching. *Photogrammetric Engineering & Remote Sensing*, 54(5): 633–641.

HAALA, N., 2013. The landscape of dense image matching algorithms. *Photogrammetric Week '13* (Ed. D. FRITSCH). Wichmann, Berlin/Offenbach, Germany. 271–284.

HAALA, N. and ROTHERMEL, M., 2012. Dense multi-stereo matching for high quality digital elevation models. *Photogrammetrie, Fernerkundung. Geoinformation (PFG)*, 4: 331–343.

HARTLEY, R. and ZISSERMAN, A., 2003. *Multiple View Geometry in Computer Vision*. Second edition. Cambridge University Press, Cambridge, UK. 655 pages.

HELAVA, U. V., 1978. Digital correlation in photogrammetric instruments. *Photogrammetria*, 34(1): 19–34.

HELAVA, U. V., 1988. Object-space least squares correlation. *Photogrammetric Engineering & Remote Sensing*, 54(6): 711–714.

HERMANN, S. and KLETTE, R., 2013. Iterative semi-global matching for robust driver assistance systems. *Lecture Notes in Computer Science*, 7726: 465–478.

HERNÁNDEZ ESTEBAN, C. and SCHMITT, F., 2004. Silhouette and stereo fusion for 3D object modeling. *Computer Vision and Image Understanding*, 96(3): 367–392.

HIRSCHMÜLLER, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2): 328–342.

HIRSCHMÜLLER, H. and SCHARSTEIN, D., 2009. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9): 1582–1599.

HIRSCHMÜLLER, H., BUDER, M. and ERNST, I., 2012. Memory efficient semi-global matching. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(3): 371–376.

HOBROUGH, G. L., 1959. Automatic stereo plotting. *Photogrammetric Engineering*, 25(5): 763–769.

HOSSEININAVEH AHMADABADIAN, A., ROBSON, S., BOEHM, J., SHORTIS, M., WENZEL, K. and FRITSCH, D., 2013. A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. *ISPRS Journal of Photogrammetry and Remote Sensing*, 78: 157–167.

KALAROT, R., MORRIS, J., BERRY, D. and DUNNING, J., 2011. Analysis of real-time stereo vision algorithms on GPU. *International Conference on Image and Vision Computing New Zealand*.

KERSTEN, T. P. and LINDSTAEDT, M., 2012. Automatic 3D object reconstruction from multiple images for architectural, cultural heritage and archaeological applications using open-source software and web services. *Photogrammetrie, Fernerkundung Geoinformation (PFG)*, 6: 727–740.

KOLMOGOROV, V., CRIMINISI, A., BLAKE, A., CROSS, G. and ROTHER, C., 2006. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9): 1480–1492.

KOUTSOUDIS, A., VIDMAR, B., IOANNAKIS, G., ARNAOUTOGLOU, F., PAVLIDIS, G. and CHAMZAS, C., 2014. Multi-image 3D reconstruction data evaluation. *Journal of Cultural Heritage*, 15(1): 73–79.

LAFARGE, F. and MALLET, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision*, 99(1): 69–85.

LHUILLIER, M. and QUAN, L., 2002. Match propagation for image-based modeling and rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8): 1140–1146.

MAAS, H.-G., 1996. Automatic DEM generation by multi-image feature based matching. *International Archives of Photogrammetry and Remote Sensing*, 31(3): 484–489.

MAKAROVIC, B., 1992. Considerations on image matching – an engineering perspective. *International Archives of Photogrammetry and Remote Sensing*, 29(B2): 613–622.

MARR, D., 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco, California, USA. 397 pages.

MARR, D. and POGGIO, T., 1976. Cooperative computation of stereo disparity. *Science*, 194: 283–287.

NARAYANAN, P. J., RANDER, P. W. and KANADE, T., 1998. Constructing virtual worlds using dense stereo. *6th International Conference on Computer Vision*. 3–10.

OHTA, Y. and KANADE, T., 1985. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2): 139–154.

OKUTOMI, M. and KANADE, T., 1993. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4): 353–363.

OPITZ, R., SIMON, K., BARNES, A., FISHER, K. and LIPPIELLO, L., 2012. *Close-range photogrammetry vs 3D scanning: comparing data capture, processing and model generation in the field and the lab*. Computer Applications and Quantitative Methods in Archaeology, Southampton, UK.

PIERROT-DESEILLIGNY, M. and PAPARODITIS, N., 2006. A multiresolution and optimization-based image matching approach: an application to surface reconstruction from SPOT5-HRS stereo imagery. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(1/W41). 5 pages (on CD-ROM).

POLLEFEYS, M., NISTÉR, D., FRAHM, J.-M., AKBARZADEH, A., MORDOHAI, P., CLIPP, B., ENGELS, C., GALLUP, D., KIM, S.-J., MERRELL, P., SALMI, C., SINHA, S., TALTON, B., WANG, L., YANG, Q., STEWENIUS, H., YANG, R., WELCH, G. and TOWLES, H., 2008. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision*, 78(2–3): 143–167.

REMONDINO, F. and EL-HAKIM, S., 2006. Image-based 3D modelling: a review. *Photogrammetric Record*, 21(115): 269–291.

REMONDINO, F., EL-HAKIM, S. F., GRUEN, A. and ZHANG, L., 2008. Turning images into 3D models – development and performance analysis of image matching for detailed surface reconstruction of heritage objects. *IEEE Signal Processing Magazine*, 25(4): 55–65.

REMONDINO, F. and MENNA, F., 2008. Image-based surface measurement for close-range heritage documentation. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B5–1): 199–206.

REMONDINO, F., DEL PIZZO, S., KERSTEN, T. P. and TROISI, S., 2012. Low-cost and open-source solutions for automated image orientation – a critical overview. *Lecture Notes in Computer Science*, 7616: 40–54.

ROSENHOLM, D., 1986. Accuracy improvement of digital matching for evaluation of digital terrain models. *International Archives of Photogrammetry and Remote Sensing*, 26(3/2): 573–587.

ROTHERMEL, M., WENZEL, K., FRITSCH, D. and HAALA, N., 2012. SURE: photogrammetric surface reconstruction from imagery. *LC3D Workshop*, Berlin, Germany. 9 pages.

ROY, S. and COX, I. J., 1998. A maximum-flow formulation of the N-camera stereo correspondence problem. *6th International Conference on Computer Vision*. 492–499.

SCHARSTEIN, D. and SZELISKI, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1–3): 7–42.

SCHENK, T., 1999, *Digital Photogrammetry: Volume I*. TerraScience, Laurelville, Ohio, USA. 421 pages.

SCHEWE, H. and FÖRSTNER, W., 1986. The program PALM for automatic line and surface measurement using image matching techniques. *International Archives of Photogrammetry and Remote Sensing*, 26(3/2): 608–622.

SEITZ, S. M., CURLESS, B., DIEBEL, J., SCHARSTEIN, D. and SZELISKI, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1: 519–528.

SHAN, Q., ADAMS, R., CURLESS, B., FURUKAWA, Y. and SIETZ, S. M., 2013. The visual Turing test for scene reconstruction. *Vision Conference.*, 3D: 25–32.

SINHA, S. N. and POLLEFEYS, M., 2005. *Multi-view reconstruction using photo-consistency and exact silhouette constraints: a maximum-flow formulation. 10th International Conference on Computer Vision*, 1: 349–356.

SMITH, M. J. and SMITH, D. G., 1996. Operational experiences of digital photogrammetric systems. *International Archives of Photogrammetry and Remote Sensing*, 31(B2): 357–362.

SNAVELY, N., SEITZ, S. M. and SZELISKI, R., 2008. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80(2): 189–210.

SONKA, M., HLAVAC, V. and BOYLE, R., 2014. *Image Processing, Analysis, and Machine Vision*. Fourth edition. Cengage Learning, Stamford, Connecticut, USA. 870 pages.

STRECHA, C., VON HANSEN, W., VAN GOOL, L., FUA, P. and THOENNESSEN, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1–8.

SZELISKI, R., 2011. *Computer Vision – Algorithms and Applications*. Springer, Heidelberg, Germany. 812 pages.

TOLA, E., LEPETIT, V. and FUA, P., 2010. DAISY: an efficient dense descriptor applied to wide-baseline matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5): 815–830.

TOLDO, R., FANTINI, F., GIONA, L., FANTONI, S. and FUSIELLO, A., 2013. Accurate multiview stereo reconstruction with fast visibility integration and tight disparity bounding. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(5/W1): 243–249.

VDI/VDE 2634/Part2, 2002. *Optical 3-D Measuring Systems – Optical Systems based on Area Scanning*. http://www.vdi.eu/guidelines/vdivde_2634_blatt_2-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/ [accessed: 21st May 2014].

VDI/VDE 2634/Part3, 2008. *Optical 3-D Measuring Systems – Multiple View Systems based on Area Scanning*. http://www.vdi.eu/guidelines/vdivde_2634_blatt_3-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/ [accessed: 21st May 2014].

VOGIATZIS, G., HERNANDEZ, C., TORR, P. H. S. and CIPOLLA, R., 2007. Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12): 2241–2246.

VOSSELMAN, G. and MAAS, H.-G. (Eds.), 2010, *Airborne and Terrestrial Laser Scanning*. Whittles, Caithness, UK. 318 pages

VU, H.-H., LABATUT, P., PONS, J.-P. and KERIVEN, R., 2012. High accuracy and visibility-consistent dense multiview stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5): 889–901.

WILLIAMS, R. E., 1959. The automatic map compilation system. *Photogrammetric Engineering*, 25(1): 103–110.

WROBEL, B., 1987. Facets stereo vision (FAST Vision) – a new approach to computer stereo vision and to digital photogrammetry. *ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland. 473 pages: 231–258.

ZHANG, L., 2005. *Automatic digital surface model (DSM) generation from linear array images*. Doctoral thesis. Mitteilungen 88. Dissertation No. 16078. Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland. 218 pages. http://e-collection.library.ethz.ch/view/eth:28126 [accessed: 21st May 2014].

ZHANG, Z., ZHANG, J., WU, X. and ZHANG, H., 1992. *Global image matching with relaxation method*. International Colloquium on Photogrammetry, Remote Sensing and GIS, Wuhan, China, 175–188.

Résumé

L'histoire de l'appariement d'images remonte à plus de cinquante ans, lorsque les premières expériences ont été réalisées avec des procédures analogiques pour des applications cartographiques. L'intégration récente d'algorithmes de vision par ordinateur et de méthodes photogrammétriques a rendu possibles des procédures très intéressantes dans lesquelles la modélisation 3D à base d'images est de plus en plus automatisée. L'appariement d'images est une des étapes essentielles de la modélisation et de la cartographie tridimensionnelles. Cet article passe en revue et analyse de manière critique quatre algorithmes d'appariement dense d'images, disponibles dans des logiciels libres et commerciaux, pour la production de nuages denses de points. Les huit jeux de données utilisés sont des scènes issues de blocs d'images terrestres et aériennes, acquises avec des axes de prise de vue convergents et parallèles, et à différentes échelles. L'analyse géométrique présentée consiste à comparer les nuages de points obtenus par les différents algorithmes, entre eux ainsi qu'avec des données de terrain.

Zusammenfassung

Die digitale Bildzuordnung hat seit den ersten analogen Ansätzen für die automatisierte Kartierung eine über 50-jährige Geschichte. Die Integration von Computer Vision Algorithmen und photogrammetrischen Methoden hat zu Verfahren geführt, die den gesamten Prozess der bildbasierten 3D-Modellierung zunehmend automatisieren. Die Bildzuordnung ist einer der zentralen Schritte in der 3D-Modellierung und der Kartierung. Dieser Beitrag gibt einen kritischen Überblick und eine Analyse von vier Bildzuordnungsverfahren, die als Open-Source oder kommerziell erhältlich sind und durch dichte Zuordnung dichte Punktwolken erzeugen. Die Basis für die Evaluation bilden acht Datensätze aus terrestrischen und Luftbildblöcken, die mit konvergenten und mit Normalaufnahmen in verschiedenen Maßstäben erzeugt wurden. Die vorgestellten geometrischen Analysen umfassen Vergleiche der erzeugten Punktwolken untereinander aber auch zu Solldaten.

Resumen

La correspondencia de imágenes tiene una historia de más de 50 años, desde los primeros experimentos realizados con procesos analógicos y fines cartográficos. La reciente integración de algoritmos de visión por computador y métodos fotogramétricos está dando lugar a procedimientos interesantes que automatizan cada vez más todo el proceso de modelado 3D basado en imágenes. La correspondencia de imágenes es uno de los pasos clave en el modelado 3D y la cartografía. Este artículo presenta una revisión crítica y un análisis de cuatro algoritmos de correspondencia de imágenes, disponibles como software libre y comercial, para la generación de nubes densas de puntos. Los ocho conjuntos de datos usados incluyen escenas pertenecientes a bloques terrestres y aéreos, adquiridos con imágenes convergentes y de ejes paralelos, y a diferentes escalas. Se analiza geométricamente las nubes de puntos producidas con cada algoritmo comparándolas con la de los otros y también con la referencia.

摘要

影像匹配技术在模拟摄影测量中首次应用开始,已经有*50*年的发展历史。近几年,计算机视觉算法和摄影测量方法融合进行影像匹配促进了基于影像进行三维建模自动化程度。影像匹配是三维建模和制图的关键步骤,本文详细阐述和分析了四种开源软件和商业软件中为制作点云的密集匹配算法,采用通过中心投影或平行光投影方式获取的不同分辨率的地面摄影和航空摄影的八组影像进行测试分析,比较了不同算法制作的点云并和地面真值进行比较。