

Accurate 3D shape measurement of multiple separate objects with stereo vision

Hien Kieu^{1,2}, Tongyan Pan³, Zhaoyang Wang¹, Minh Le², Hieu Nguyen² and Minh Vo⁴

¹ Department of Mechanical Engineering, The Catholic University of America, Washington, DC 20064, USA

² Department of Electrical Engineering, The Catholic University of America, Washington, DC 20064, USA

³ Department of Civil Engineering, The Catholic University of America, Washington, DC 20064, USA

⁴ Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

E-mail: wangz@cua.edu

Received 17 September 2013, revised 27 November 2013

Accepted for publication 17 December 2013

Published 30 January 2014

Abstract

3D shape measurement has emerged as a very useful tool in numerous fields because of its wide and ever-increasing applications. In this paper, we present a passive, fast and accurate 3D shape measurement technique using stereo vision approach. The technique first employs a scale-invariant feature transform algorithm to detect point matches at a number of discrete locations despite the discontinuities in the images. Then an automated image registration algorithm is applied to find full-field point matches with subpixel accuracy. After that, the 3D shapes of the objects can be reconstructed according to the obtained point matching and the camera information. The proposed technique is capable of performing a full-field 3D shape measurement with high accuracy even in the presence of discontinuities and multiple separate regions. The validity is verified by experiments.

Keywords: three-dimensional shape measurement, three-dimensional sensing, three-dimensional image acquisition

(Some figures may appear in colour only in the online journal)

1. Introduction

3D shape measurement has become one of the most prevalent topics in numerous fields because of its wide and ever-increasing applications. Some notable examples include constructing 3D images of human body parts for cosmetic surgery, building 3D images of world famous pieces of art to archive and replicate, digitizing real objects to make animations for 3D games or cartoons, and creating 3D images for 3D printer to produce product prototypes. At present, there are many techniques capable of creating 3D images from real objects such as the fringe projection [1], laser scanning [2] and stereo vision [3] techniques. In this paper, the technique being employed is the stereo vision one that uses two cameras to capture objects from two different views and builds their 3D images based on the matches of the two images and the camera information.

The key component of stereo-vision-based 3D shape measurement is the matching of the same physical points in two different images. The existing image matching algorithms mainly fall into two categories: area-based matching and feature-based matching. The area-based matching algorithm uses a subset area around the point to be matched in one image (reference image) and finds the corresponding subset area in another image (target image) according to the correlation of the two subsets [4, 5]. This algorithm assumes that there is a shape transformation model for the two subsets around each point, and the task is to find the best parameters of the model through optimizing the matching or correlation coefficients with an iterative approach such as the Newton–Raphson method or the Levenberg–Marquardt method. These parameters provide information to locate the corresponding point in the target image for each interrogated point in the reference image, which indicates point matching. After the completion of the

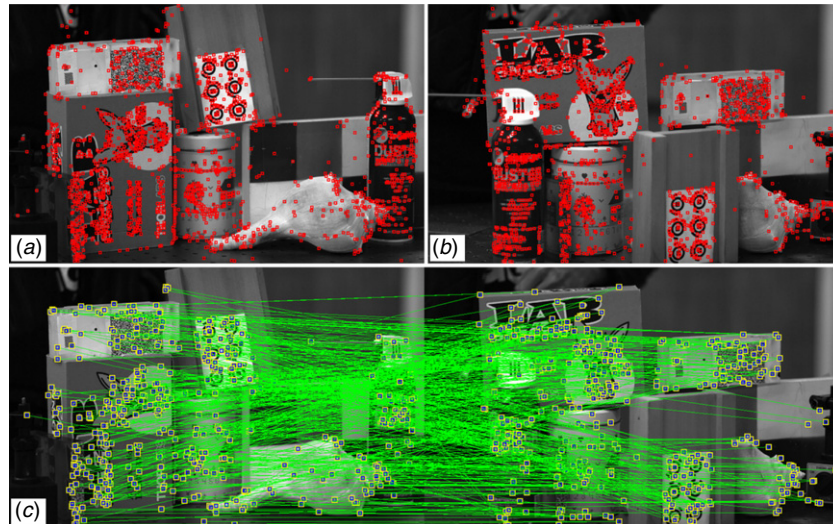


Figure 1. Scale-invariant feature transform (SIFT) matching result of a few objects placed in various positions. (a) and (b) are two images with detected SIFT feature points, and (c) shows the detected fast library for approximate nearest neighbors (FLANN) correspondences.

matching process at a point, the process will propagate to the neighbor points and find their correspondences. This matching technique is capable of yielding point matching with subpixel accuracy, but the disadvantage is that it will not propagate through the discontinuities in the images caused by occlusions and shadows as well as the regions that lack sufficient intensity variations. Another matching algorithm, the feature-based one, extracts the features in the images and builds a unique descriptor for each feature point. The feature points in the reference and target images can then be matched by comparing their descriptors. This algorithm is able to detect good matches in the images without any difficulty of coping with discontinuities. However, it only gives matches at a number of discrete points and cannot produce dense and full-field matching. It should be noted though, that the term *full-field* does not necessarily mean every pixel in the images because there are always unmatchable points, such as the points in the homogeneous regions or the points in one image but not in another image.

The technique proposed in this paper will exploit the aforementioned two matching algorithms to achieve automated full-field matching for 3D shape measurement in the presence of discontinuities or multiple separate regions, which are common in practice. The basic idea is: first, the feature-based matching is adopted to detect a number of matching pairs in the images. Then, the matching is expanded to full-field by performing the area-based matching. After that, the matching result is employed to construct dense-point or full-field 3D images of objects. The technique is elaborated below.

2. Algorithm

2.1. Scale-invariant feature transform (SIFT) for discrete point matching

The feature-based matching task of the proposed work uses the scale-invariant feature transform (SIFT) [6] and the fast library for approximate nearest neighbors (FLANN) algorithms [7].

The SIFT algorithm aims to extract local features in the images and describe each feature point with a 128-dimensional descriptor, whereas the FLANN algorithm helps detect the matching of the feature points between two images according to the SIFT descriptors. For simplicity, the matching method is called SIFT matching technique hereafter. It is very robust and can cope well with the scaling, translation, rotation, distortion (affine transformation), occlusion and illumination-variation issues encountered in real images. Figure 1 illustrates an example of the SIFT matching result, where a number of matching pairs are detected from the two images containing objects placed in various positions.

The SIFT algorithm has a few parameters that should be set prior to the matching process, but there is often no need to specify them because the default parameters, supported by numerous tests under various circumstances, work well for general matching applications [6]. Therefore, the SIFT matching is typically an automated process. It is noteworthy that the SIFT algorithm has become the most widely used technique in computer vision for feature detection, so the details of the algorithm are not introduced in this paper.

Because the SIFT matching technique detects the matches at only a number of discrete locations and there usually exist some false matching pairs, the technique alone is insufficient for the subsequent reconstruction of 3D images.

2.2. Automated image registration (AIR) for full-field point matching

The area-based matching can provide full-field point-to-point matching between two images. It is normally called image alignment or image registration in the computer vision field [8] and digital image correlation in the optics and mechanics fields [5]. The former usually uses natural or passive patterns, while the latter commonly requires speckle patterns to be fabricated on object surfaces. Another primary difference in practice is that the computer vision applications often put emphasis on speed rather than accuracy [9], whereas high accuracy is

usually essential in optics and mechanics applications [10]. The proposed approach in this paper tends to use natural or passive patterns to obtain 3D shape measurement with high accuracy, and the area-based matching technique is named automated image registration (AIR) matching technique.

The AIR matching technique aims to match every pixel, if possible, in the reference image with the corresponding point at subpixel locations in the target image. For a representative pixel (x_0, y_0) in the reference image to be matched, a square region called a subset with the size of $(2M + 1) \times (2M + 1)$ pixels with its center located at (x_0, y_0) is usually selected. The corresponding subset in the target image, i.e., the target subset, is often of irregular shape. Denoting the shift amount between the centers of the two matching subset patterns as (ξ, η) , the shape mapping or transformation function for the entire reference and target subsets can be expressed as:

$$\begin{aligned} x'_i &= x_i + \xi + \xi_x \Delta_x + \xi_y \Delta_y + \xi_{xx} \Delta_x^2 + \xi_{yy} \Delta_y^2 + \xi_{xy} \Delta_x \Delta_y \\ y'_j &= y_j + \eta + \eta_x \Delta_x + \eta_y \Delta_y + \eta_{xx} \Delta_x^2 + \eta_{yy} \Delta_y^2 + \eta_{xy} \Delta_x \Delta_y, \end{aligned} \quad (1)$$

where i and j range from $-M$ to M , $\Delta_x = x_i - x_0$, $\Delta_y = y_j - y_0$ and ξ , ξ_x , ξ_y , ξ_{xx} , ξ_{yy} , ξ_{xy} , η , η_x , η_y , η_{xx} , η_{yy} and η_{xy} are the transformation parameters. The determination of these parameters can be achieved by minimizing the least-squares-based correlation coefficient defined as [5, 11]:

$$C = \sum_{i=-M}^M \sum_{j=-M}^M [af(x_i, y_j) + b - g(x'_i, y'_j)]^2, \quad (2)$$

where a is a scale factor, b is an offset of intensity and $f(x_i, y_j)$ and $g(x'_i, y'_j)$ indicate the intensity values at a representative pixel in the reference subset and the potential matching pixel in the target subset, respectively. M is usually in the range of 7 to 15.

The establishment of matching for pixel (x_0, y_0) requires only ξ and η , but the optimization process must involve determining all the transformation parameters $(\xi, \xi_x, \dots, \eta_{yy}, \eta_{xy})$ and the scale and offset parameters (a, b) associated with equation (2). The AIR technique typically employs the Newton–Raphson or the Levenberg–Marquardt method to carry out the registration optimization by iterative calculation, which is capable of providing fast and accurate analysis [8, 12]. Nevertheless, the AIR matching is an overall time-consuming process because the area-based matching is required for each interrogated point. To increase the processing speed in practice, the AIR matching will not run on every pixel but on pixels at a step size of N , and the full-field matching results will be obtained by data interpolation from those grid pixels. The step size N is usually set to 3 to 5, and can be set to larger for a faster processing speed with reduced analysis accuracy. Another important drawback of the existing AIR technique is that a reasonably good initial guess for the transformation parameters is required, and this highly impedes the application of the technique to real-world problems where multiple separate regions usually exist.

2.3. Proposed approach

The proposed technique involves a novel approach taking the advantages of both the SIFT and the AIR matching techniques. The procedure is as follows.

- (i) For the two captured images, the SIFT matching process is employed to detect the matching of a number of discrete points. This includes extracting the SIFT feature points with descriptors for both images and running the FLANN algorithm to determine point correspondences.
- (ii) In the reference image, every SIFT matching point and its two closest SIFT neighbor points that can form a regular-shape triangle (every angle is no less than 15°) are selected. Together with the matching points in the target image, the six points can be used to solve the transformation parameters $(\xi, \xi_x, \xi_y, \eta, \eta_x, \eta_y)$ in equation (1). The other parameters are set to zero for the initial guess task, except for the scale factor a , which is set to one. The integer centroid of the triangle in the reference image becomes a seed point for the AIR analysis.
- (iii) If the AIR analysis of the seed point is successful based on the convergence criterion (e.g., 30 iteration cycles) and disparity limitation criterion (e.g., 100%), the AIR process will propagate its matching results to its nearest neighbor point located on the calculation grids, and this point is the starting point. The subsequent matching computation will then follow a path guided by the best matching: for all the grid points that have been processed, the one that has the smallest correlation coefficient is chosen as the next point to be propagated. This means that its matching results are used as the initial guess for its non-processed grid neighbors. A processed point will not be propagated if its analysis fails the convergence or disparity limitation criteria, or its correlation coefficient is above the pre-defined threshold, or all its neighbors have already been processed. By doing this, the analysis will stop when it reaches any bad region and prevent the analysis from producing incorrect results. This step is illustrated in figure 2.
- (iv) If the AIR analysis of the seed point fails (this happens if a SIFT matching result is false) or the AIR propagation stops, the process will go back to step 2 and move to the next SIFT point in no particular order. If a grid point has been processed successfully in a previous computation, it will not be analyzed again in any subsequent AIR processing to avoid redundant processing. Otherwise, it will be processed again in the subsequent analysis that originates from a different seed point.
- (v) Data interpolation is applied to obtain the full-field matching at unmatched grid points and non-grid points that are closely surrounded by matched grid points. As previously mentioned, although the matching is called full-field, there are always unmatchable points in the images.

After the full-field point matching, the classic triangulation algorithm [13] is used to calculate the 3D coordinates of the matched points upon a prior camera calibration [14, 15].

3. Experiments

A few experiments have been conducted to demonstrate the performance of the proposed technique. In the experiments,

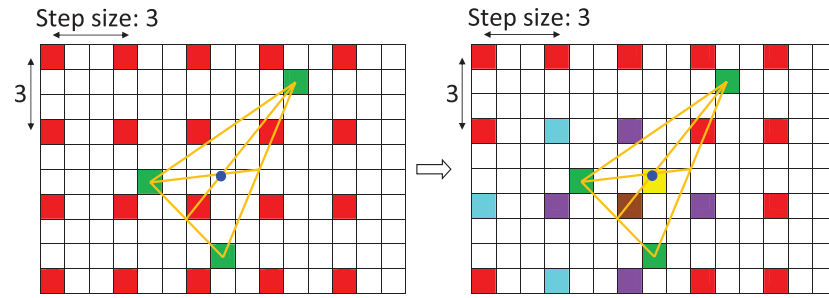


Figure 2. Illustration of the initial guess and automated image registration (AIR) propagation process. The image on the left shows the AIR computation grids in red and a scale-invariant feature transform (SIFT) point with its two closest SIFT points in green. The SIFT points form a triangle in orange. The small blue circle indicates the centroid of the triangle in both images. The image on the right shows the seed point in yellow and starting point in brown. The four points in purple are analyzed following the starting point, and the three points in cyan are analyzed subsequently assuming the left-most purple point has the best matching. After that, one point among the other three purple and the three cyan points that has the best matching will be the next point to be propagated.

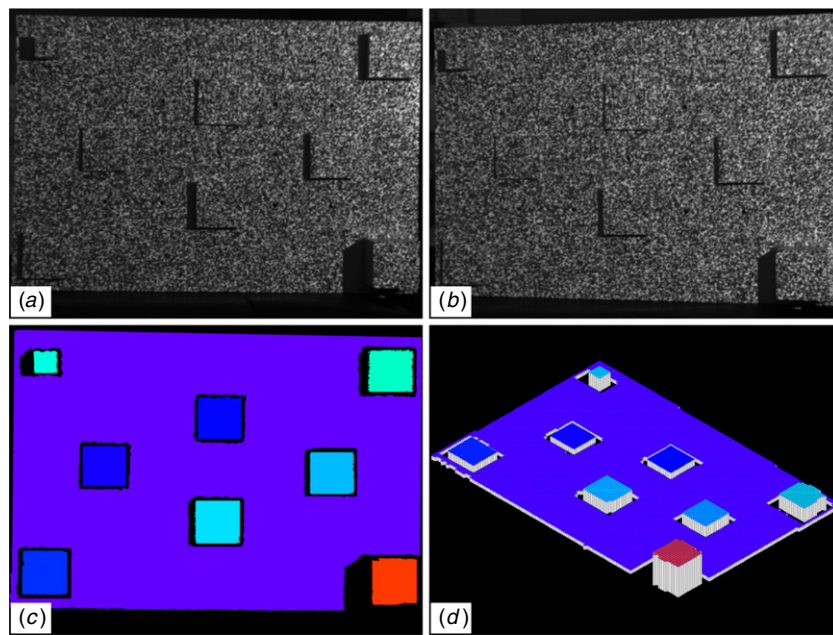


Figure 3. A plate with eight gage blocks. (a) and (b) are captured images, and (c) and (d) are the schematic 2D and 3D height plots, respectively.

the camera parameters are obtained by using a calibration technique that utilizes the frontal image concept and hyper-precise control point detection scheme [16, 17]. In addition, all the images are captured with a resolution of 2048×1536 pixels. The viewpoint angle between the two cameras is approximately 15° and the distances from the objects to the cameras are around 1.3 m. It takes less than two minutes to accomplish the entire analysis for each experiment with a regular laptop computer.

The first experiment aims at testing the 3D measurement accuracy on multiple separate regions, where a flat plate of $457.2 \text{ mm} \times 304.8 \text{ mm}$ with eight high-precision gage blocks is selected as the specimen. A digital random speckle pattern is projected on the white specimen to create the required features for the point matching purpose, as shown in figures 3(a) and (b). Because of occlusions and shadows, the gage blocks are separate from each other in terms of the captured patterns. In the analysis, the subset size and step size are chosen to be

Table 1. Actual and measured heights of the blocks.

Block	Actual	Mean measured	Error
1	25.400	25.387	-0.013
2	19.050	19.030	-0.020
3	6.350	6.351	0.001
4	6.350	6.354	0.004
5	12.700	12.685	-0.015
6	15.875	15.887	0.012
7	9.525	9.548	0.023
8	50.80	50.727	-0.073

Unit: mm

21 and 3, respectively. The processing result shows that the proposed algorithm can automatically detect discontinuities and handle the multiple separate regions. Figures 3(c) and (d) illustrate the obtained 2D and 3D height plots, and the detailed height information of the blocks relative to the plate is summarized in table 1. The result indicates that the

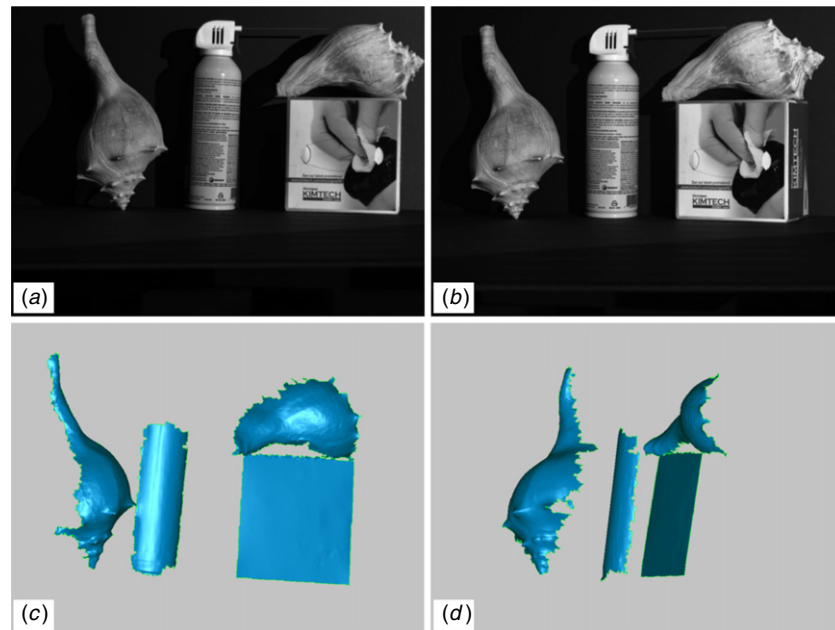


Figure 4. Reconstruction of the 3D images for a few objects. (a) and (b) are the captured images, and (c) and (d) are two representative 3D plots.

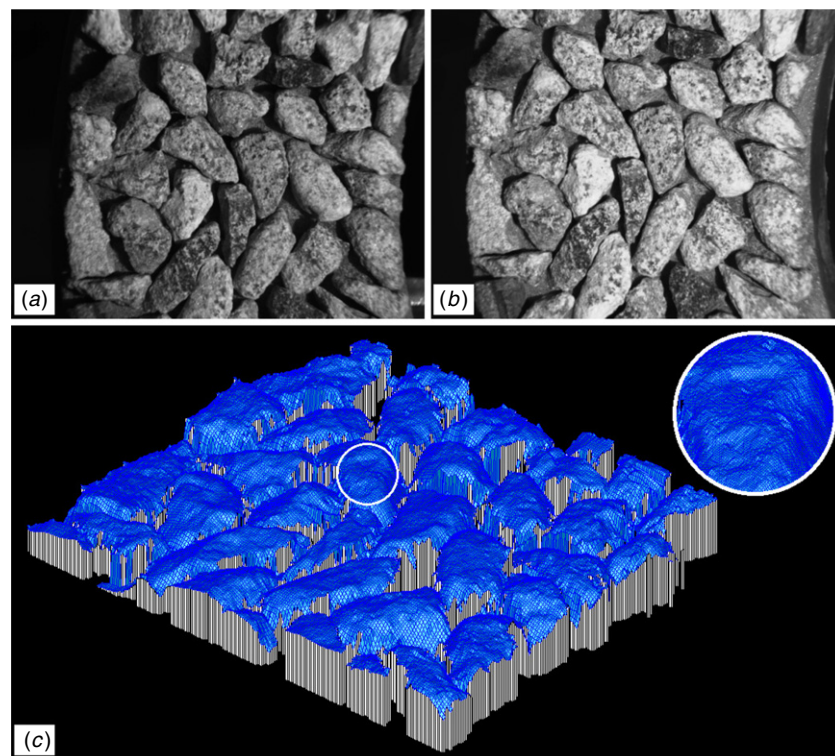


Figure 5. 3D surface microtexture measurement of a road pavement sample. (a) and (b) are captured images and (c) is the 3D plot.

maximum error of the mean measured heights over the entire field is 0.073 mm, which yields a relative accuracy (defined as the ratio of the out-of-plane measurement accuracy to the in-plane dimension) of 0.016%. This confirms the validity of the proposed 3D shape measurement approach.

The other two experiments are intended to demonstrate the flexibility of the proposed technique for real applications. Figure 4 shows an example of building the 3D images of a

few objects including two conch shells, an air duster can and a tissue box from their two 2D images. In this experiment, the natural patterns on the object surfaces provide the required intensity variations for the SIFT and AIR matching analysis.

The last experiment is an application to the 3D surface microtexture measurement of a road pavement sample for friction analysis. A zoom lens is used to magnify the measurement region of 72 mm wide to fit into the captured

images, which are shown in figures 5(a) and (b). Figure 5(c) shows the plot of the 3D surface profile acquired by the proposed technique. In this experiment, the specimen surface has natural microtexture patterns available for the imaging matching task. It is noted that the different illuminations presented in the captured images are automatically handled by the matching algorithms.

4. Discussion and conclusion

The scale-invariant feature transform (SIFT) matching scheme provides robust matches at a number of feature points by coping well with scaling, translation, rotation (up to 50°), lighting condition, partially occlusion and affine distortion. The automated image registration (AIR) matching scheme yields matches at dense pixel locations with very high accuracy thanks to the iterative and fast Newton–Raphson or Levenberg–Marquardt algorithm. Compared with the SIFT technique alone, the proposed technique can obtain matching for many more points with higher density in broader regions, which helps produce more complete and smoother 3D shape reconstructions. Compared with the conventional AIR approach for 3D shape measurement, the proposed technique does not need manual operation for the initial guess task and does not perform image matching from a single starting point but from all of the SIFT points, which allows automated and reliable processing in one or multiple separate regions. Furthermore, the proposed technique follows reliable paths to perform the AIR matching and automatically stops the matching process at unmatchable points to avoid false matching and prevent error propagation. In addition to the experiments presented in this paper, numerous other experiments conducted by the authors also verified that the proposed technique is superior to the conventional AIR technique in terms of the covered regions and the measurement accuracy.

The proposed technique does not need extra pre- or post-processing steps, and it can normally accept the default or commonly used SIFT and AIR parameters. Consequently, it allows automated processing. In addition, similar relative accuracy can typically be achieved in different general applications. The actual measurement accuracy of an application depends on many factors, such as surface patterns, camera resolution, lighting condition, etc. Depending on the cases, the processing parameters may be adjusted to achieve optimized analysis results. During the grid point matching process, there can be an option for requiring each grid point to be successfully analyzed more than once, say twice, to achieve possibly slightly better analysis by selecting the results associated with better correlations. This minor improvement is in practice insufficient to compensate for the extra processing time induced.

It should also be pointed out that the random sample consensus (RANSAC) technique is commonly adopted in the 3D computer vision and shape measurement field to remove incorrect matches from the SIFT matching result [18, 19]. In the proposed technique, the wrong matches can be automatically detected by the subsequent AIR matching, so

there is no need to integrate the RANSAC algorithm into the SIFT matching. This is an additional notable advantage of the proposed technique.

As technologies evolve, the passive 3D shape measurement technique has gained considerable progress in recent years. The technique can now provide accuracy close to that offered by active 3D shape measurement and vision techniques (e.g., fringe projection [20] and laser scanning [21]), as demonstrated in this paper and other recent work [22]. It is expected that the proposed technique can help broaden the applications of passive 3D shape measurement.

Acknowledgments

This work was supported by the US Army Research Office (ARO) under grant no. W911NF-10-1-0502.

References

- [1] Wang Z, Du H, Park S and Xie H 2009 Three-dimensional shape measurement with a fast and accurate approach *Appl. Opt.* **48** 1052–61
- [2] Liu Z, Zhu J, Yang L, Liu H, Wu J and Xue B 2013 A single-station multi-tasking 3D coordinate measurement method for large-scale metrology based on rotary-laser scanning *Meas. Sci. Technol.* **24** 105004
- [3] Hasler M, Haist T and Osten W 2012 Stereo vision in spatial-light-modulatorbased microscopy *Opt. Lett.* **37** 2238–40
- [4] Pan B, Qian K, Xie H and Asundi A 2009 Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review *Meas. Sci. Technol.* **20** 062001
- [5] Sutton M A, Orteu J J and Schreier H W 2009 *Image Correlation for Shape, Motion, and Deformation Measurements* (New York: Springer)
- [6] Lowe D 2004 Distinctive image features from scale-invariant keypoints *Int. J. Comput. Vis.* **60** 91–110
- [7] Muja M and Lowe D 2009 Fast approximate nearest neighbors with automatic algorithm configuration *Int. Conf. Comput. Vis. Theory Appl.* **24** 331–40
- [8] Baker S and Matthews I 2004 Lucas-Kanade 20 years on: a unifying framework *Int. J. Comput. Vis.* **56** 221–55
- [9] Scharstein D and Szeliski R 2002 A taxonomy and evaluation of dense two-frame stereo correspondence algorithms *Int. J. Comput. Vis.* **47** 7–42
- [10] Hu Z, Xie H, Lu J, Hua T and Zhu J 2010 Study of the performance of different subpixel image correlation methods in 3D digital image correlation *Appl. Opt.* **49** 4044–51
- [11] Pan B, Xie H and Wang Z 2010 Equivalence of digital image correlation criteria for pattern matching *Appl. Opt.* **49** 5501–9
- [12] Luu L, Wang Z, Vo M, Hoang T and Ma J 2011 Accuracy enhancement of digital image correlation with B-spline interpolation *Opt. Lett.* **36** 3070–2
- [13] Hartley R and Zisserman A 2003 *Multiple View Geometry in Computer Vision* 2nd edn (Cambridge: Cambridge University Press)
- [14] Zhang Z 2000 A flexible new technique for camera calibration *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 1330–4
- [15] Huang J, Wang Z, Xue Q and Gao J 2012 Calibration of a camera projector measurement system and error impact analysis *Meas. Sci. Technol.* **23** 125402
- [16] Douchamps D and Chihara K 2009 High-accuracy and robust localization of large control markers for geometric camera calibration *IEEE Trans. Pattern Anal. Mach. Intell.* **31** 376–83

- [17] Vo M, Wang Z, Luu L and Ma J 2011 Advanced geometric camera calibration for machine vision *Opt. Eng.* **50** 110503
- [18] Fischler M and Bolles R 1981 Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography *Commun. ACM* **24** 381–95
- [19] Chum O and Matas J 2005 Matching with PROSAC—progressive sample consensus *IEEE Conf. on Computer Vision and Pattern Recognition* pp 220–6
- [20] Vo M, Wang Z, Pan B and Pan T 2012 Hyper-accurate flexible calibration technique for fringe-projection-based three-dimensional imaging *Opt. Exp.* **20** 16926–41
- [21] Ouyang W and Xu B 2013 Pavement cracking measurements using 3D laser-scan images *Meas. Sci. Technol.* **24** 105204
- [22] Beeler T, Bickel B, Noris G, Marschner S, Beardsley P, Sumner R and Gross M 2012 Coupled 3D reconstruction of sparse facial hair and skin *ACM Trans. Graph.* **31** 117