Registration and integration algorithm in structured light three-dimensional scanning based on scale-invariant feature matching of multi-source images

Haihua Cui (崔海华)*, Wenhe Liao (廖文和), Ning Dai (戴 宁), and Xiaosheng Cheng (程筱胜)

Jiangsu Key Laboratory of Digital Medical Equipment Technology, Nanjing University of Aeronautic and Astronautic, Nanjing 210016, China

**Corresponding author: cuihh@nuaa.edu.cn*

Received February 28, 2012; accepted March 14, 2012; posted online May 16, 2012

Based on the homography between a multi-source image and three-dimensional (3D) measurement points, this letter proposes a novel 3D registration and integration method based on scale-invariant feature matching. The matching relationships of two-dimensional (2D) texture gray images and two-and-a-half-dimensional (2.5D) range images are constructed using the scale-invariant feature transform algorithms. Then, at least three non-collinear 3D measurement points corresponding to image feature points are used to achieve a registration relationship accurately. According to the index of overlapping images and the local 3D border search method, multi-view registration data are rapidly and accurately integrated. Experimental results on real models demonstrate that the algorithm is robust and effective.

OCIS codes: 100.0100, 120.0120, 120.2650, 150.0150.

doi: 10.3788/COL201210.091001.

Three-dimensional (3D) registration and integration are difficult but necessary steps in performing structured light 3D measurements[1]. A multiple-view measurement is needed to fully cover an object because of object occlusion, shadows or limited view depth, and so on. The overlapping layers of the same object must be integrated before the whole object is completely measured. The goal of registration is to determine the transform function of a view from one coordinate system to another[2−8].

Different methods of registration have been developed[2−8], and can be divided into three categories. The first method involves the use of supporting hardware, such as turntable methods[2], gyroscopic self-positioning devices[3], etc. This method relies on precision supporting devices, and has a limited application range. The second method adds some additional features to the surface of the object and identifies the designed features to achieve registration. This method includes the paste landmark technique[4] and fixing the calibration plate on the object[5], among others. However, one main drawback of this method is that the object may not be completely measured because of the obstruction caused by the additional feature. The third method is completely based on the 3D measurement data. This technique starts from an approximate registration, and iteratively reaches the final registration by minimizing an error functional, such as the iterative closest point algorithm[6] and various improved algorithms[7−8]. This method has the smallest measurement constraint and is most widely applied. However, its biggest drawback is the requirement of a good initial position; otherwise, convergence is not guaranteed[7−8].

To maximize the convenience of application, 3D measurement data and the related measurement process are used to simplify the complex problem. The matching scale-invariant features from two-dimensional (2D) gray images and two-and-a-half-dimensional (2.5D) range images are accurately recognized. The homographic relationship is then used to solve 3D registration and integration. The novelty of this approach is characterized by the following features: the 3D registration and integration are realized only with the captured 2D image, which is feasible; a scale-invariant feature point detection based on phase shifting 2.5D range image is proposed; the registration based on feature matching can be used for large view change; and it has the smallest measurement constraint and is most widely applied. This registration method is flexible, and the only main constraint is that two view images must contain the overlapping region.

The 2D gray and 2.5D range images are the sources of data of the structured light 3D measurement. Therefore, the complicated 3D registration problem can be simplified by achieving the transform relation of these measurement images.

Previous studies[9,10] have presented the scale-invariant feature transform (SIFT) algorithm for extracting distinctive features from images that can be invariant to image scale, rotation, and zoom. The algorithm has a good effect on image affine, view, and brightness change. The fundamental principle of SIFT is the detection idea of the Laplace of Gaussian (LoG) method[9−11].

To detect the invariant feature using a simple and efficient image operation, previous studies[9] have employed the difference of Gaussian (DoG) operator instead of LoG. The locations invariant to the scale change of the image can be determined by searching for stable features across all possible scales using a continuous scale function known as a scale space[11].

Two images of the dental plaster model are captured at view change of about 45°, as shown in Fig. 1. The blue cross markers of the image are the detected scale-invariant feature points using SIFT algorithm. The red circle markers are the matching points. The algorithm is found to be fast; however, its effective view change range is limited. As shown in Fig. 1, there are only two matching points. The current study also found that the

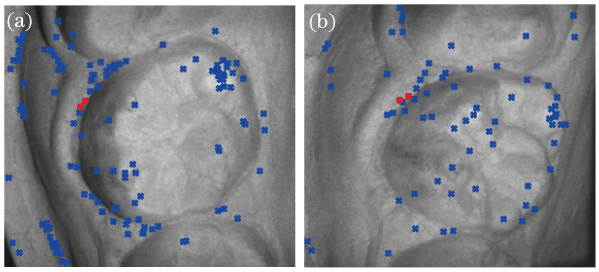

Fig. 1. Scale-invariant feature points. Detected feature points of (a) the initial-view image and (b) the image view change.

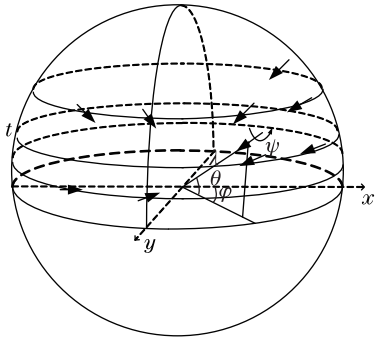

Fig. 2. Transition tilt simulation figure of different camera views.

algorithm cannot detect feature points, which have poor texture gray information despite obvious 3D characteristics of the region. The middle area of the tooth 3D surface is shown to be complex. The gray image is also insufficient, which indicates that the algorithm cannot achieve feature detection (Fig. 1).

To determine the invariant features at the large view change, previous studies[12] have developed a fully affine SIFT method, called affine-SIFT (ASIFT). This method simulates a set of sample views of the initial images obtained by varying the axis orientations (latitude θ and the longitude φ angles) of two cameras, as shown in Fig. 2. The SIFT method is then applied to all generated images. The method is mathematically proven to be fully affine-invariant. The basic principle of the method is that any planar smooth deformation can be approximated around each point by an affine map using the first order Taylor formula. The total simulated area is about 13.5 times the area of the original image[12]. Therefore, the complexity of the ASIFT feature computation is 13.5 times the complexity of computing SIFT features.

The ASIFT method is claimed to achieve full affine transform feature detection and a greater number of detected feature points. Two identical images are used to examine this method, as shown in Fig. 3. More feature points are detected using ASIFT than using SIFT. However, feature points that have poor-texture gray information still cannot be detected, as shown in the middle part of the tooth. There are also almost no matching feature point on the surface of the tooth.

The structured light measurement system is based on sinusoidal phase-shifting stripes[13]. At least three phase-shifting stripes can solve the primary phase. The primary phase can be revealed with the unwrapping algorithm to acquire the continuous phase[14]. The unwrapped phase value expresses the object height, and is also known as the range image. The range image is a 2D image showing the distance points in a scene from a specific point. It has pixel values that correspond to distance (i.e., brighter values mean shorter distance, and vice versa), as shown in Figs. 4(b) and 5(e). It is also called a 2.5D image. This type of imaging modality provides direct and explicit geometric information that is considered useful in many applications. This modality is also potentially unchanged by variations in lighting and viewpoint[15]. The shape of an image surface should remain constant under viewpoint changes. Hence, the local distribution of curvatures observed on the surface should also remain constant.

The computation of curvatures based on the range image is simpler than 3D data, because the pixel coordinate of the range image is a regular grid. The neighborhood topology information is also known and does not need complex topology triangular construction. The range image can be considered as a discrete quadratic surface. The first and second partial surface derivatives can be computed with the first image differences $[f_x(i,j),\ f_y(i,j)]$ and the second-order image differences $[f_{xx}(i,j),\ f_{yy}(i,j),$ and $f_{xy}(i,j)]$[16]. The surface curvature can be used to divide the principle, mean, and Gaussian curvatures[15]. The mean and Gaussian curvatures are computed as

$$H(i,j) = \{[1+f_y^2(i,j)]f_{xx}(i,j) + [1+f_x^2(i,j)]f_{yy}(i,j) - 2f_x(i,j)f_y(i,j)f_{xy}(i,j)\} / \{2[\sqrt{1+f_x^2(i,j)+f_y^2(i,j)}]^3\}, \quad (1)$$

$$K(i,j) = \frac{f_{xx}(i,j)f_{yy}(i,j) - f_{xy}^2(i,j)}{[1+f_x^2(i,j)+f_y^2(i,j)]^2}. \quad (2)$$

The degree of surface curving based on estimates of H and K can be computed and utilized as a weight to indicate the strength of the curvature at a particular point[15]:

$$CL(i,j) = \sqrt{2H^2 - K}. \quad (3)$$

Figures 4(a)–(c) are the texture gray image, range image based on phase shifting, and curvature image of one view, respectively. Figures 4(d)–(f) are the respective images of another view. Figures 4(c) and (f) are the curvature images based on range image; these are called feature image in this letter. The value of each pixel is computed with the Eq. (3). The computed values are normalized as $[0, 255)$, and are shown with gray value as Figs. 4(c) and (f), respectively. The newly developed feature image based on phase shifting is detected using

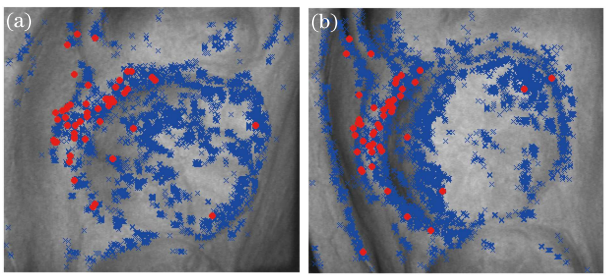

Fig. 3. Affine scale-invariant feature points. Detected feature points of (a) the initial-view and (b) the image view change.

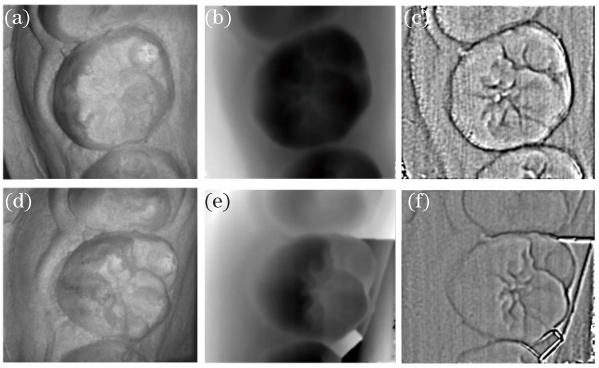

Fig. 4. Range image of an object and its feature image. (a) Texture gray image at initial view. (b) Range image at initial view; (c) feature image at initial view; (d) texture gray image at view change; (e) range image at view change; (f) feature image at view change.

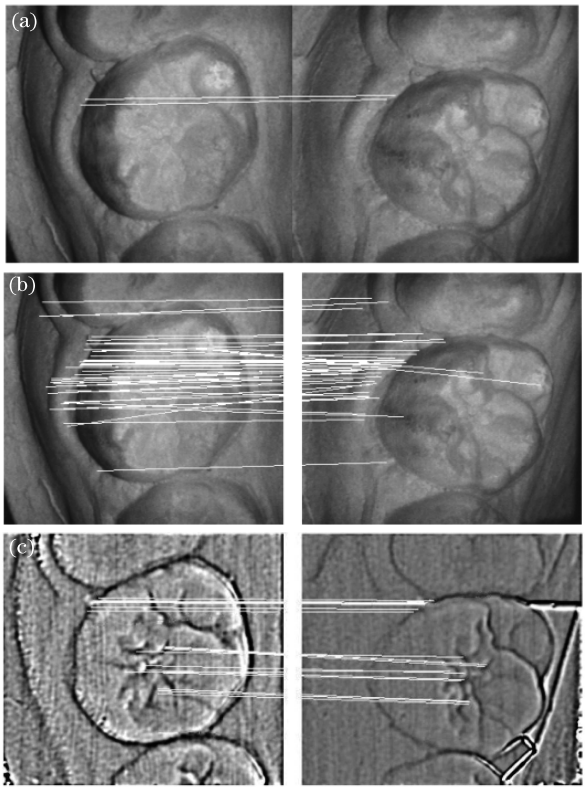

Fig. 5. Correct matches of multi-source images after purification. (a) Correct matches of a gray image using (a) the SIFT algorithm, (b) the ASIFT algorithm, and (c) the SIFT and ASIFT algorithms.

the ASIFT and SIFT algorithms. The number of the detected feature points is large. Hence, some points based on their strengths are shown in Fig. 6. The matching feature points are similar with the two algorithms, and are marked with red circle points in Fig. 6. The detected features are distributed in the feature regions of 3D surface. Compared with Figs. 1 and 3, the detected scale-invariant feature points of a 2.5D range image can reflect 3D object feature information, and have less impact on view changes. These characteristics are complementary to the method of feature detection based on a texture gray image.

The previous operations have assigned an image location, scale, and orientation to each feature point. This step is performed to compute a descriptor for the local image region. The method is the same as the SIFT algorithm[9].

The acquired matching points are not completely correct based on multi-source images. Thus, the purification operation is needed to filter out the incorrect matching points and ensure the accuracy of registered data. An improved random sample consensus (RANSAC) algorithm[17] is developed for this purification operation.

Different view images of the same scene have homographic relationships. In a homogeneous coordinate, images $I_X(x_i, y_i, 1)^{\mathrm{T}}$ and $I_{x'}(x'_i, y'_i, 1)^{\mathrm{T}}$ undergo a perspective transformation, as shown below:

$$k \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix}, \tag{4}$$

where $\mathbf{H} = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8)$ is a transformation matrix with eight degrees. The matching points are purified using the transformation relationship.

The RANSAC algorithm is an iterative method for estimating the parameters of a mathematical model from a set of observed data containing outliers. This algorithm is improved in this letter based on our application. In the first step of the improved RANSAC algorithm, the subset of the sample data is alternatively extracted from matching data characteristics instead of performing complete random extraction by the original algorithm. The squared ratio of the distance between the nearest and second nearest matching points can basically express the accuracy of the matching points. Hence, the first sort is based on the ratio value, and the sample matching point subset is then extracted according to the ratio order. This operation increases the accuracy of the initialization model, reduces the computation procedure, and improves the efficiency of purification. The error tolerance of the fitting model is 0.1 in the improved algorithm.

The matching results after purification are shown in Fig. 5. The correct matching result using the SIFT algorithm is shown in Fig. 6(a). There are only two correct matches on the gray image at a large view change. Figure 5(b) shows the correct matching result using the ASIFT algorithm. The correct matches are almost entirely distributed in the tooth root, and the complex top tooth surface has almost no correct match. Figure 5(c) shows the correct matching result based on the proposed range feature image. The correct matches are complementary

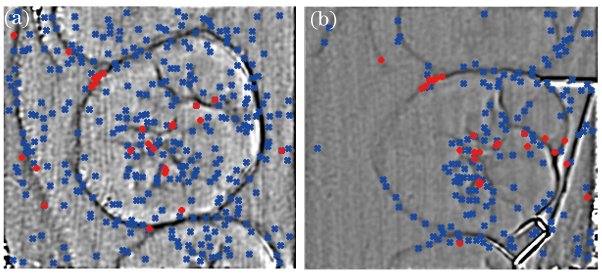

Fig. 6. Scale-invariant feature points of a range image. Detected feature points of (a) an initial-view image and (b) an image view change.

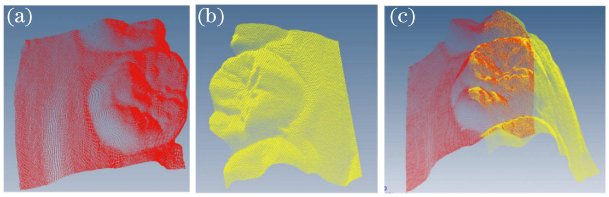

Fig. 7. Registration result of two image views. 3D measurement result of (a) one view and (b) another view; (c) registered data of the two views.

to those based on the texture gray image. There are many correct matches on the top tooth surface. All of the correct matches will be used in the 3D registration.

The goal of registration is to find the fitting transformation matrix that can be acquired with a unit quaternion[18]. The 3D data corresponding to 2D matches are set to $\mathbf{Q} = \{q_i\}$ and $\mathbf{P} = \{p_i\}$. The computed rotation matrix $\mathbf{R}$ and the translation matrices $\mathbf{T}$ make up the registration relationship $\mathbf{H}$ of the two-view 3D data. The two-view 3D measurement data are then registered, as shown in Fig. 7. The degree of the two views is about 90°. Multi-view measurements are registered according to the above method.

The integration of registered data is divided into two steps. A view image containing feature points Q is set to $I_Q(i,j)$ and $(0 \leqslant i \leqslant i_w, 0 \leqslant j \leqslant j_h)$, as shown in Fig. 8(a). Another view image containing feature point P is set to $I_P(i',j')(0 \leqslant i' \leqslant i_w, 0 \leqslant j' \leqslant j_h)$, as shown in Fig. 8(b). (i,j) and (i',j') are the captured image pixel coordinates, while i_w and j_h are the width and height of the captured image, respectively. With the registration transformation $\mathbf{H}$:

$$I_{P'}(i'',j'') = I_P(i',j') * \mathbf{H}, \tag{5}$$

where $I_{P'}(i'',j'')$ is the transformed image. The pixel coordinate point (i',j') is the overlapping image pixel if its corresponding transformation coordinate meets $0 \leqslant i'' \leqslant i_w, 0 \leqslant j'' \leqslant j_h$, and vice versa. Therefore, the overlapping regions and boundaries are found. According to the index, only one-view image data are computed to achieve integration, as shown in Fig. 8(g).

After the first integration based on 2D image registration, some gaps are observed at the boundary of the overlapping regions. These gaps are caused by perspective transformation, as shown in Fig. 8(g). This problem is solved using a local 3D search method according to the index of the overlapping image boundary. A belt region is expanded along the boundary line of the overlapping region, as shown in the blue line in Fig. 8(c). The 3D points corresponding to the pixels of the region represent the local overlapping data. The width of the belt is set to 10 pixels. The distance between the 3D points from one view to another view is used to judge the same overlap point. If the distance is less than a threshold (d_t =0.04 mm), the two points of the two views only retain one point. Hence, boundary integration is accurately performed, as shown in Fig. 8(h).

To verify the proposed method, an experiment was conducted on a plaster tooth model, as shown in Fig. 9. Tooth surfaces are highly complex and have various surface morphologies. Theoretically, their gray information is also abundant. However, the captured gray images have poor information (such as images presented in paper sometimes), which may be caused by structured light, camera angle, image zooming, and so on. Therefore, a range image can be a very good complement. We contracted a miniaturized 3D measurement system to scan the tooth model. Figures 9(a) to (d) show the four tooth model measurement data from different views, respectively. Figures 9(e) and (f) show four registered and integrated data based on the above method, respectively. Figure 9(g) is the surface model of the four-view integration result. The time cost of four view registra-

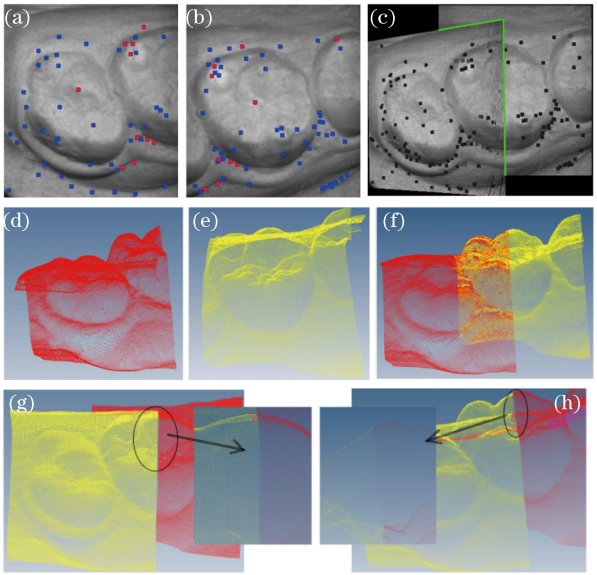

Fig. 8. 3D data integration based on a registered image. (a) One-view gray image. (b) Another-view gray image. (c) Registration result of two-view images. (d) 3D measurement data of a one-view image. (e) 3D measurement data of an another-view image. (f) Registration result of two-view 3D measurement data. (g) Integration result based on image registration. (h) Precise integration result.

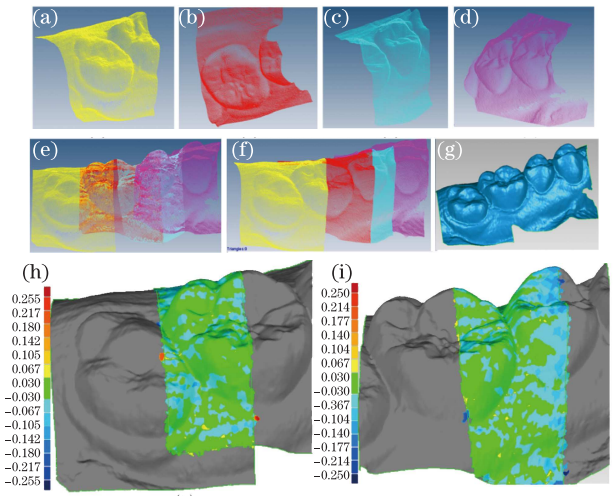

Fig. 9. Registration and integration results of a plaster tooth model. (a–d) Four views measurement data. (e) Registration results of multi-view 3D measurement data. (f) Integration results of multi-view 3D measurement data. (g) 3D geometric surface model of multi-views. (h) Registration accuracy of the first and second view. (i) Registration accuracy of the second and third view.

tion is about 8 s. The whole procedure contains image capture, data computing, and the result shown. The registration and integration of two views only took less than 500 ms. In order to verify the registration accuracy, we used Geomagic Studio software to analyze the registration result. The statistical results of the first and second view are as follows: average distance of 0.0291 mm and Std. Dev of 0.0272 mm, as shown in Figs. 9(h) and (i) is the registration accuracy result of the second and third views. The registration accuracies of all views are less than 0.035 mm. The estimation value roughly represents the accuracy of our developed registration algorithm.

In conclusion, this letter presents a new 3D registration and integration method for measurement data. The method completed 3D registration successfully. Firstly, the 2D gray image and 2.5D range image of the measurement process are considered as multi-source images. Scale-invariant feature algorithms are integrated and used for accurate detection and matching. These algorithms also overcome incorrect matching due to shortcomings in a large view change, view depth change, and certain poor light reflections on an image. Multi-view 3D measurement data are then registered and integrated based on the matched feature points. The algorithms used resolve a difficult problem, and have a wide range of potential applications.

This work was supported by the National "863" Program of China (No. 2009BAI81B02) and the Doctoral Foundation of the Ministry of Education (No. 20070287055). This research shared partial source code of the SIFT algorithm (Rob Hess) and ASIFT algorithm (Jean-Michel Morel, Guoshen Yu). The authors express their gratitude to them.

References

1. F. Chen, G. M. Brown, and M. Song, Opt. Eng. **39,** 10 (2000).
2. X. Long, Y. Zhong, R. Li, and Z. You, J. Tsinghua University (Sci. and Tech) (in Chinese) **42,** 477 (2002).
3. X. Wang, H. Wei, S. Liu, C. Shi, and L. Zhang, "Multi-view auto-registration algorithm based on attitude sensor," (in Chinese) Chinese Patent: CN101539405 (2009).
4. X. Luo, Y. Zhong, and R. Li, J. Tsinghua University (Sci. and Tech) (in Chinese) **44,** 1104 (2004).
5. T. Ren, J. Zhu, Y. Li, and S. Ye, Chin. J. Mechanical Engineering (in Chinese) **44,** 137 (2008).
6. P. J. Besl and N. D. McKay, IEEE T. Pattern. Anal. **14,** 239 (1992).
7. C. Yang and G. Medioni, Image Vision. Comput. **10,** 145 (1992).
8. N. J. Mitra, N. Gelfand, H. Pottmann, and L. Guibas, in *Proceedings of Eurographics Symposium on Geometry* 23 (2004).
9. D. G. Lowe, Int. J. Comput. Vision. **60,** 91 (2004).
10. X. Liu, Z. Tian, W. Yan, and X. Duan, Chin. Opt. Lett. **9,** 061001 (2011).
11. T. Lindeberg, Int. J. Comput. Vision. **30,** 79 (1998).
12. J. M. Morel and G. Yu, SIAM J. Imaging Sci. **2,** 438 (2009).
13. H. Cui, W. Liao, X. Cheng, N. Dai, and T. Yuan, Chin. Opt. Lett. **8,** 33 (2010).
14. H. Cui, W. Liao, N. Dai, and X. Cheng, Opt. Eng. **50,** 063602 (2011).
15. R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Porcessing Using MATLAB* (Publishing House of Electronics Industry, Beijing, 2005).
16. T. W. R. Lo, J. P. Siebert, and A. F. Ayoub, in *Proceedings of MICCAI 2006 Workshop on Craniofacial Image Analysis for Biology, Clinical Genetics, Diagnostics and Treatment* 75 (2006).
17. M. A. Fischler and R. C. Bolles, Commun. Acm. **24,** 381 (1981).
18. D. W. Eggert, A. Lorusso, and R. B. Fisher, Mach. Vision. Appl. **9,** 272 (1997).