Registration algorithm based on image matching for outdoor AR system with fixed viewing position

L. Lin, Y. Liu, W. Zheng and Y. Wang

Abstract: An image registration method based on the Fourier–Mellin transform is introduced for an outdoor augmented reality (AR) system. For this type of AR system, the observation position is fixed, and a complex 3-D registration problem can be reduced to a 2-D image registration for this fixed viewing position system. An observation globe model for this method is proposed. Under this supposition, a Fourier–Mellin transform is used in image registration, and the architecture of this system is illustrated. Experimental results show that this image registration algorithm is accurate and robust. It is effective for an outdoor AR system with a fixed viewing position.

1 Introduction

Augmented reality (AR) is a new computer technology, which joints virtual objects with realistic environments. More and more applications have been found recently by applying this technique, such as applications in visual medical surgery, visual computer-aided education, military, industry, entertainment etc. [1–6]. The basic problem in an AR system is how to perfectly 'cling' a virtual computer-generated 3-D graphic model to a real scene, i.e. a 3-D image registration problem. It means proper alignment of virtual objects with the real environment in 3-D space, and the virtual objects must behave as their real counterparts would do.

Generally, the registration process can be realised in the following three steps: positioning, rendering, and merging [7]. Positioning is to locate the observer's position and orientation, by which the virtual objects can be affine transformed and then added to the real scene correctly. Rendering is to obtain a 2-D projected image from the 3-D model, which is the real image of the 3-D model seen by the observer. Merging is an image processing procedure to merge the virtual objects with the real environments in order to make it look like a real part of the scene. As virtual objects can be added in a real scene, this technique is useful in many situations, such as the digital reconstruction of historic sites, visualised surgery etc.

As indicated above, the key problem of an AR system is image registration. Recently, there are mainly three kinds of registration methods being used. First, registration-based on direction-tracking equipment, such as GPS and gyrometer, to obtain the 6 degrees of freedom (DOF) of an observer [8, 9]. Second, registration-based on computer vision [10–12]. Third, hybrid registration, which combines the virtue of GPS, gyrometer and computer vision [8, 9, 13, 14]. All these registration methods have their own advantages and shortcomings. For example, registration by GPS and gyrometer has faster speed but much lower accuracy.

IEE Proceedings online no. 20045181

doi:10.1049/ip-vis:20045181

Paper first received 22nd September 2004 and in revised form 17th March 2005

The authors are with the School of Information Science and Technology, Beijing Institute of Technology, Beijing 100081, People's Republic of China

E-mail: linliang@bit.edu.cn

Many computer vision registration methods used in AR systems require the placement of markers in the scene beforehand, which are captured by the calibrated cameras mounted on the observer's helmet. And then, the 6DOF parameters can be worked out from the cameras' extrinsic parameters. Such approaches have been applied successfully in indoor AR systems [15]. However, it is hard to apply them in outdoor AR systems. First, marker-based approaches are very sensitive to outdoor lighting [15] and they are not robust. Secondly, such approaches become feeble when some of the markers are randomly hidden from view by pedestrians etc. Finally, and the most importantly, it is not feasible to place markers in many outdoor applications.

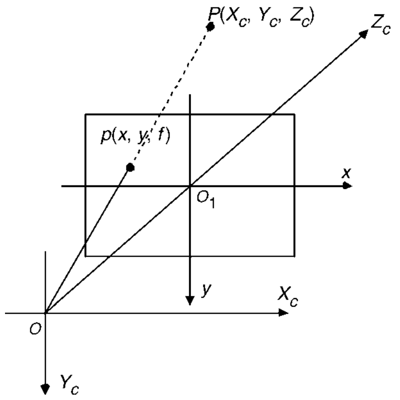
Nowadays, research on markerless registration has become a hot issue in outdoor AR systems [16–19]. In order to obtain accurate and robust image registration results in an outdoor environment, a novel method to be applied in a fixed-viewing-position outdoor AR system is proposed in this paper, which is a Fourier–Mellin-based image registration algorithm. The proposed approach utilises the overall vision information in the scene, and thus can obtain stable registration results. Though the proposed method requires the support of a reference image database, which is a set of prestored reference images of the surrounding scene with the different orientations of the camera, it can achieve more accurate and robust registration results and avoid the accumulation of tracking errors. Experimental results show the correctness of theoretical analysis and the feasibility of its application in an outdoor AR system.

2 Principle

2.1 *Projection model in a fixed-viewing-position outdoor AR system*

Image registration in an outdoor AR system is a difficult problem. However, for an AR system with fixed-viewing-position as described in Section 3 of this paper, it can be simplified significantly by reducing a 3-D image registration problem to a 2-D problem.

A pin-hole model can be used to approximate the projection of a camera. Figure 1 shows the pin-hole model of a camera.



$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

Fig. 1 *Pin-hole camera model*

In an AR system, the optical centre of the camera is fixed, and the 3-D objects are projected onto the 2-D image plane of the camera. This optical system can be described as an observation globe, which is illustrated in Fig. 2. Its centre is the camera's optical centre; its radius is the normalised focal length of the camera; θ is the camera's field angle. Objects in the scene are projected onto the globe and string AB denotes a projected image of the current scene. When the camera is yawing and pitching, its orientation changes accordingly. A new 2-D image A′B′ will be obtained from the observation globe.

For a camera mounted on a revolving table, when the change of yawing and pitching angle is not significant, such motion can be approximated as the horizontal and vertical translation of the captured 2-D image, its skew angle to a rotation angle of the image, and its zooming to the scale. Then, a registration problem on a fixed-viewing-position AR system is equivalent to a 2-D image registration with translation, rotation, and scale changes between the captured and the reference images.

2.2 Principle of Fourier–Mellin transform based 2-D image registration algorithm

In a 2-D image registration, translation, rotation and scale are the most common factors and automatic registration can be performed by using a phase correlation technique.

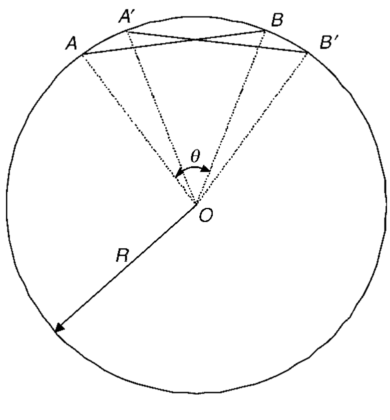


Fig. 2 *Observation globe model*

2.2.1 Phase correlation technique for detection of translation: Let f_s and f_r denote two images that differ only by a displacement (x_0, y_0), where f_r is a reference image and f_s is an input one to be registered, i.e.

$$f_s(x, y) = f_r(x - x_0, y - y_0) \tag{1}$$

F_s, F_r are the corresponding Fourier transforms with the following Fourier shift theorem

$$F_s(u, v) = e^{-j2\pi(ux_0+vy_0)} F_r(u, v) \tag{2}$$

The cross-spectrum of F_s and F_r is defined as

$$R = \frac{F_s(u, v)F_r^*(u, v)}{|F_s(u, v)F_r(u, v)|} = e^{j2(ux_0+vy_0)} \tag{3}$$

where F^* is the complex conjugate of F. By taking the inverse Fourier transform of R, an impulse function, as shown in (4), will be obtained

$$F^{-1}(R) = \delta(x - x_0, y - y_0) \tag{4}$$

By detecting the peak of $\delta(x - x_0, y - y_0)$, the translation parameters (x_0, y_0) can be acquired.

2.2.2 Phase correlation technique for detection of translation and rotation [20]: If $f_s(x, y)$ is a rotated replica of $f_r(x, y)$ with rotation θ_0 and translation (x_0, y_0), then

$$f_s(x, y) = f_r(x\cos\theta_0 - y\sin\theta_0 - x_0, x\sin\theta_0 + y\cos\theta_0 - y_0) \tag{5}$$

According to the property of a Fourier transformation, the Fourier transforms F_s and F_r have the following relationships:

$$F_s(u, v) = e^{-j2\pi(ux_0+vy_0)} \times F_r(u\cos\theta_0 - v\sin\theta_0, u\sin\theta_0 + v\cos\theta_0) \tag{6}$$

$$|F_s(u, v)| = |F_r(u\cos\theta_0 - v\sin\theta_0, u\sin\theta_0 + v\cos\theta_0)| \tag{7}$$

This means $|F_s|$ is also a rotated replica of $|F_r|$ with rotation θ_0 and without translation. The rotation can be obtained by representing the $|F_s|$ and $|F_r|$ with polar coordinates as shown in (8)

$$|F_s(\rho, \theta)| = |F_r(\rho, \theta - \theta_0)| \tag{8}$$

Then angle θ_0 can be obtained by using phase correlation.

2.2.3 Fourier–Mellin transform and detection of scale, rotation, and translation (R, S, T): If f_s is a replica of f_r with scale k, rotation θ_0, and translation (x_0, y_0), then,

$$f_s(x, y) = f_r[k(x\cos\theta_0 - y\sin\theta_0) - x_0, k(x\sin\theta_0 + y\cos\theta_0) - y_0] \tag{9}$$

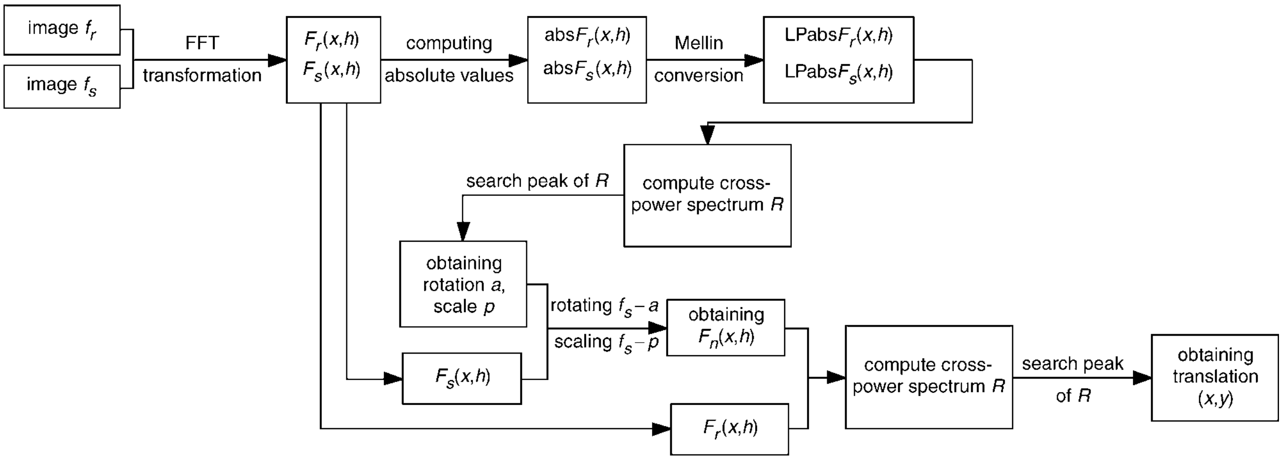


Fig. 3 *Flow chart of Fourier–Mellin-based image registration algorithm*

The Fourier transforms F_s and F_r are related as

$$F_s(u,v) = \frac{1}{k^2} e^{-j2\pi((ux_0/k)+(vy_0/k))} \times F_r\left(\frac{u\cos\theta_0 - v\sin\theta_0}{k}, \frac{u\sin\theta_0 + v\cos\theta_0}{k}\right) \tag{10}$$

$$|F_s(u,v)| = \frac{1}{k^2}\left|F_r\left(\frac{u\cos\theta_0 - v\sin\theta_0}{k}, \frac{u\sin\theta_0 + v\cos\theta_0}{k}\right)\right| \tag{11}$$

Let G_s and G_r denote the magnitude spectra, i.e.

$$G_s(u,v) = \frac{1}{k^2} G_r\left(\frac{u\cos\theta_0 - v\sin\theta_0}{k}, \frac{u\sin\theta_0 + v\cos\theta_0}{k}\right) \tag{12}$$

Let H_s and H_r denote the transforms of G_s and G_r converted from Cartesian coordinates to polar coordinates, i.e.

$$\begin{cases} \rho = (u^2 + v^2)^{1/2} \\ \theta = \tan^{-1}(u/v) \end{cases} \tag{13}$$

$$H_s(\rho,\theta) = \frac{1}{k^2} H_r(\rho/k, \theta + \theta_0) \tag{14}$$

After introducing a log transform for ρ in the polar coordinates, which is derived from the Mellin transform, the scale factor k can be resolved by the phase correlation technique in the log-polar coordinate

$$\begin{cases} r = \log\rho \\ \varphi = \theta \end{cases} \tag{15}$$

Let Q_s and Q_r denote the corresponding transform of H_s and H_r in the log-polar coordinate,

$$Q_s(r,\varphi) = Q_r(r - \log k, \varphi + \theta_0) \tag{16}$$

Then $\log k$ and θ_0 can be obtained by using the phase correlation technique.

Taking the Fourier transform of a log-polar map is equivalent to the computing of the Fourier–Mellin transform [21, 22]

$$F_M(k_1,k_2) = \int_{-\infty}^{\infty}\int_0^{2\pi} f(e^r\cos\varphi, e^r\sin\varphi) e^{j(k_1 r + k_2\varphi)} dr\, d\varphi \tag{17}$$

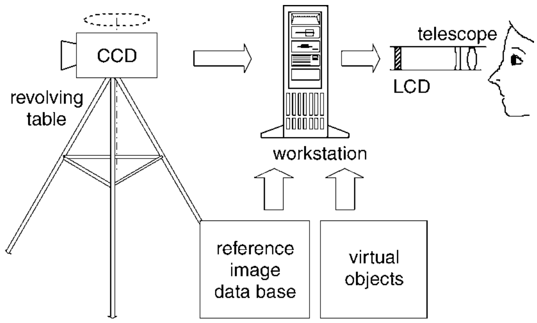


Fig. 4 *Architecture of the fixed-viewing-position outdoor AR system*

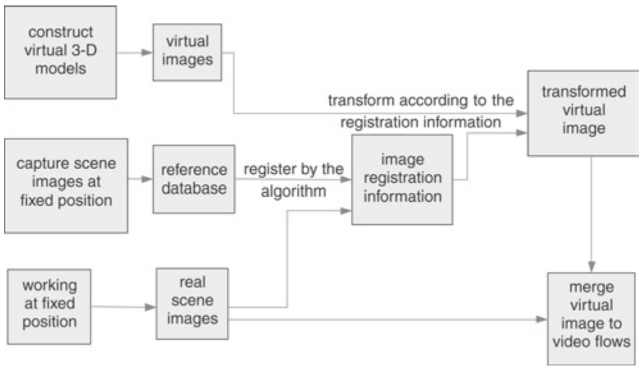


Fig. 5 *Working flow chart of the fixed-viewing-position AR system*

The modulus of the Fourier–Mellin transform is rotation and scale invariant.

After obtaining rotation θ_0 and scale k, a new replica f_3 can be created which is affined from f_1 according to these two factors. The new image f_3 is the same as f_2 except for translation. Then, by using a phase correlation technique again, translation (x_0, y_0) can be obtained, which means all parameters of R, S, T have been obtained and a 2-D image registration has been performed.

The working flow chart of the image registration algorithm based on the Fourier–Mellin transform is illustrated in Fig. 3.

3 Architecture of outdoor AR system

Our fixed-viewing-position outdoor AR system is composed of the following parts as shown in Fig. 4: revolving table,

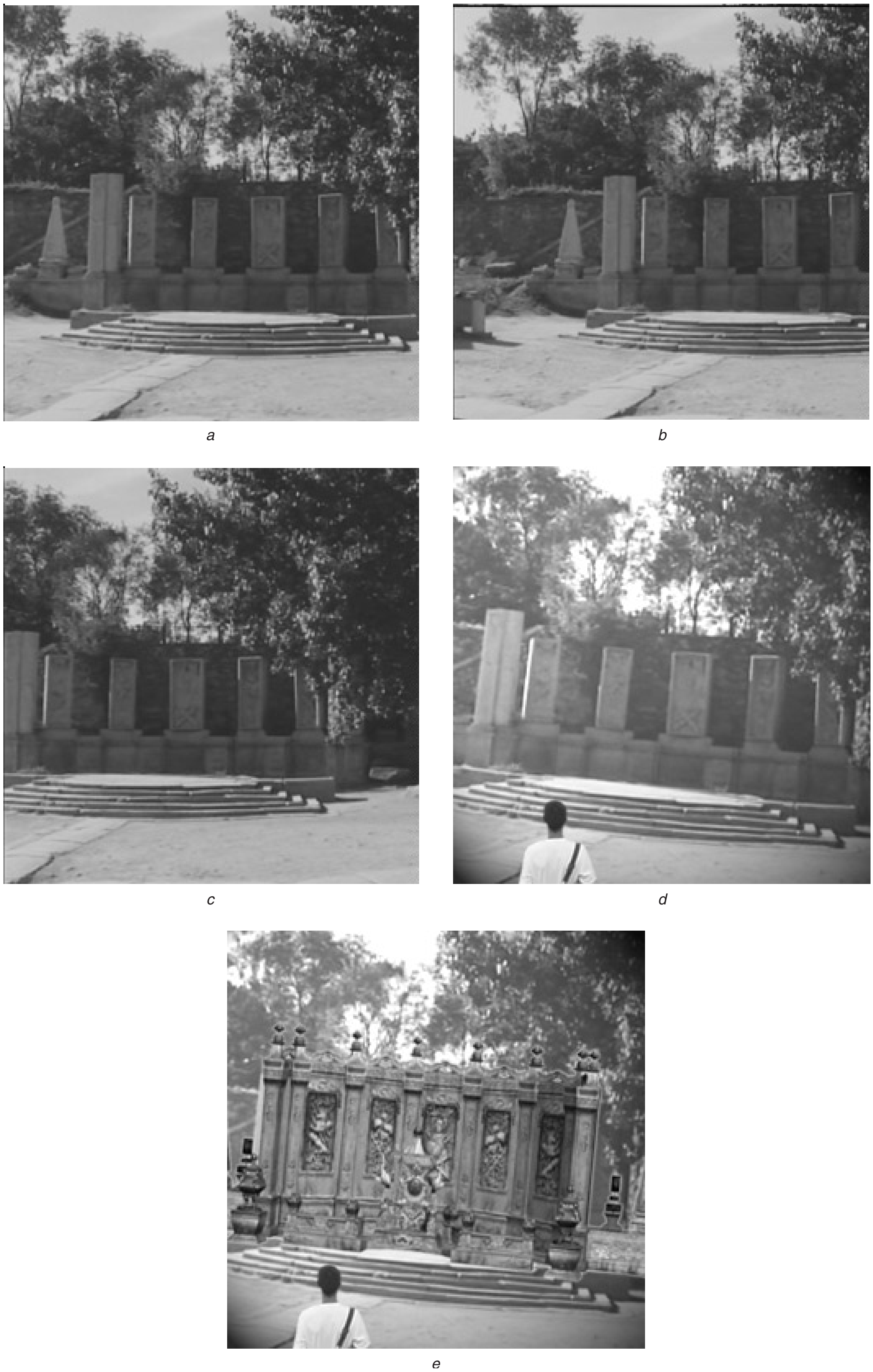


Fig. 6 *Experimental images*

a, *b*, *c* Reference images prestored in database
d Acquired image of the real scene
e Final augmented image

CCD, reference image and virtual objects database, computer workstation, LCD and viewing optics.

The CCD camera is mounted on a revolving table, which can perform the function of panning, tilting and zooming. The real scene image captured by the camera is sent to the workstation, where it is compared with the reference images in the database. A magnetic compass is used in this system to provide a rough observation direction to reduce the searching scope in the reference images database. With the Fourier–Mellin transform algorithm, the accurate orientation of the camera can be obtained. As the position of the camera is fixed, the 6DOF of the camera can then be obtained. The virtual objects stored in the computer can be affine transformed and added to the current scene. After the process of rendering and merging, the augmented image is displayed on the LCD and can be viewed through an eyepiece.

The working flow chart of the proposed fixed-position outdoor AR system is shown in Fig. 5. The whole working procedure is described as follows:

(1) Preparation

(*a*) Establish the computer generated 3-D models;

(*b*) Build up the reference image database, which can be realised by capturing images at different angles when revolving the camera and recording their orientation information.

(2) Combine virtual objects with the real scene

(*a*) A piece of current scene image is captured when a user is revolving the camera;

(*b*) The captured image is compared with the reference images, and the closest image is found as the reference image to be used. Then the accurate orientation of the camera with respect to the reference image can be obtained by using the Fourier–Mellin transform;

(*c*) The virtual 3-D images are transformed and rendered according to the 6DOF information of the camera, which are then combined with the captured image of the scene.

(3) Display

The combined images are sent to the two pieces of LCDs, which could be seen by an observer through a binocular eyepiece.

4 Experiments

The system proposed in this paper is part of a project for the digital reconstruction of Yuanmingyuan Garden, a magnificent ancient Chinese royal garden set up during the Qing Dynasty, which had been completely destroyed mainly in the 1860s; what is left there are only the huge stone columns that were part of the ancient palaces.

After the reference image database has been established for our AR system by panning and rotating the camera, the system captures the current scene, acquires the registration data, transforms the virtual archaic architectures perspectively, and then adds them to the real scene image.

Experimental images are shown in Fig. 6. Figures 6*a*, 6*b*, and 6c are reference images prestored in our database. The resolution of the reference images is 256 × 256. Figure 6*d* is one of the acquired images of the scene, which is going to be registered. This image is obtained when the camera is panned and rotated randomly and the environment lighting condition changed. In the meantime, a visitor intrudes in the scene. According to the rough observation direction provided by magnetic compass, Fig. 6*b* is considered as the reference image corresponding with the real scene in Fig. 6*d*. Then, the proposed algorithm is applied to Figs. 6*b* and 6*d*, and the result of registration is shown in Table 1. Finally, the virtual image is merged with the captured scene accordingly, which is shown in Fig. 6*e*.

Table 1: Result of registration

Rotation (degree)	Translation (Δx)	Translation (Δy)	Scale (times)
5.625	26.000	6.000	0.843

5 Conclusions

In this paper, an image registration approach based on the Fourier–Mellin transform for a fixed-viewing-position AR system has been proposed. As the optical centre is fixed, the difficult 3-D image registration problem can be simplified to a 2-D image registration problem. An observation globe model is introduced for the fixed-viewing-position AR system and under such supposition, the Fourier–Mellin transform can be used to realise image registration. As the overall information of the image is used, when there are intrusive objects in the captured scene during the process of registration, the registration result is also correct as long as the percentage of the area occupied by the intrusive object is not significant in the whole image, which means that it is insensitive to the obtrusive objects and the performance of registration is robust to some extent.

However, because of the complexity of the Fourier–Mellin transform, the algorithm cannot be performed in real time on a computer. The period of time for registration on images with resolution 256 × 256 is about 3 seconds on our workstation. The proposed system is still in the experimental phase and efforts are being undertaken to realise the proposed algorithm by DSP and FPGA in order to perform the system in real time.

6 Acknowledgments

This project is supported by the National Basic Research Program of China (National 973 Project, Grant No. 2002CB312104) and National Natural Science Foundation of China (Grant No. 60025513).

7 References

1 Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B.: 'Recent advances in augmented reality', *IEEE Comput. Graph. Appl.*, 2001, **21**, (6), pp. 34–47

2 Clarke, J., Vines, J., and Mark, E.: 'An augmented virtuality scientific data center'. Proc. User Group Conf., 9–13 June 2003, pp. 354–357

3 Hicks, J.D., Flanagan, R.A., Petrov, P.V., and Stoyen, A.D.: 'Eyekon: augmented reality for battlefield soldiers'. Proc. 27th Annual NASA Goddard/IEEE Software Engineering Workshop, 5–6 December 2002, pp. 156–163

4 Tang, S.-L., Kwoh, C.-K., Teo, M.-Y., Sing, N.W., and Ling, K.-V.: 'Augmented reality systems for medical applications', *IEEE Eng. Med. Biol. Mag.*, 1998, **17**, (3), pp. 49–58

5 Zagoranski, S., and Divjak, S.: 'Use of augmented reality in education'. 'Computer as a tool', IEEE Region 8 EUROCON 2003, vol. 2, pp. 339–342

6 Sims, D.: 'New realities in aircraft design and manufacture', *IEEE Comput. Graph. Appl.*, 1994, **14**, (2), p. 91

7 Bajura, M., Henry, F., and Ohbuchi, R.: 'Merging virtual reality with the real world: seeing ultrasound imagery within the patient', *Comput. Graph.*, 1992, **2**, pp. 203–210

8 Azuma, R., Hoff, B., and Neely, H.: 'A motion-stabilized outdoor augmented reality system'. Proc. IEEE Virtual Reality 1999, 13–17 March 1999, pp. 252–259

9 Thomas, B., Demczuk, V., Piekarski, W., Hepworth, D., and Gunther, B.: 'A wearable computer system with augmented reality to support terrestrial navigation'. Digest Int. Symp. on Wearable Computers, 1998, pp. 168–171

10 Bajura, M., and Neumann, U.: 'Dynamic registration correction in video-based augmented reality systems', *IEEE Comput. Graph. Appl.*, 1995, **15**, (5), pp. 52–60
11 Koch, R., and Van Gool, L.: '3D structure from multiple images of large-scale environments', *Lect. Notes Comput. Sci.*, 1998, **1506**, pp. 275–289
12 Kutulakos, K.N., and Vallino, J.R.: 'Calibration-free augmented reality', *IEEE Trans. Vis. Comput. Graph.*, 1998, **4**, (1), pp. 1–20
13 Behringer, R.: 'Registration for outdoor augmented reality applications using computer vision techniques and hybrid sensors'. Proc. IEEE Virtual Reality 1999, 13–17 March 1999, pp. 244–251
14 You, S., and Neumann, U.: 'Fusion of vision and gyro tracking for robust augmented reality registration', Proc. IEEE Virtual Reality 2001, 13–17 March 2001, pp. 71–78
15 Chen, J., Shi, Q., and Wang, Y.: 'AR technology and applications', *IEEE Comput. Eng. Appl.*, 2001, **37**, (21), pp. 55–57
16 Simon, G., and Berger, M.-O.: 'Pose estimation for planar structures', *IEEE Comput. Graph. Appl.*, 2002, **22**, (6), pp. 46–53
17 Simon, G., Fitzgibbon, A.W., and Zisserman, A.: 'Markerless tracking using planar structures in the scene'. Proc. 13th Int. Conf. on Pattern Recognition, Vol. 1, 25–29 August 1996, pp. 90–94
18 Genc, Y., Riedel, S., Souvannavong, F., Akinlar, C., and Navab, N.: 'Markerless tracking for AR: a learning-based approach'. Proc. Int. Symp. on Mixed and Augmented Reality 2002 (ISMAR 2002), 30 September–1 October 2002, pp. 295–304
19 Ferrari, V., Tuytelaars, T., and Van Gool, L.: 'Markerless augmented reality with a real-time affine region tracker'. Proc. IEEE ACM Int. Symp. on Augmented Reality 2001, 29–30 Oct. 2001, pp. 87–96
20 Srinivasa Reddy, B., and Chatterji, B.N.: 'An FFT-based technique for translation, rotation, and scale-invariant image registration', *IEEE Trans. Image Process.*, 1996, **5**, (8), pp. 1266–1271
21 Yang, Z.H., and Chen, S.Y.: 'Optical signal processing computing and neural networks' (John Wiley & Sons, Inc., 1992)
22 Ruanaidh, J.J.K., and Pun, T.: 'Rotation, scale and translation invariant digital image watermarking', *Signal Process.*, 1998, **66**, (3), pp. 303–317