Illumination-robust and Anti-blur Feature Descriptors for Image Matching in Abdomen Reconstruction

Huan Liu[1]   Ying Xiao[1]   Wei-Dong Tang[1]   Yan-Hui Zhou[2]

[1]College of Electronic and Information Engineering, Jinggangshan University, Ji′an 343009, China

[2]College of Mechanical and Electronic Engineering, Jinggangshan University, Ji′an 343009, China

Abstract: This paper puts forward a method for abdomen panorama reconstruction based on a stereo vision system. For the purpose of recovering the abdomen completely and accurately under the condition of actual photographing with illumination variance and blur noise, some innovative combined feature descriptors are presented on the basis of Hu-moment invariants. Furthermore, considering the study on the abdomen surface reconstruction, a circle template which is divided into 6 sectors is designed. It is noted that a descriptor merely using gray intensity is not able to provide sufficient information for feature description. Consequently, the sector entropy which denotes the structure characteristics is drawn into the feature descriptor. By means of the combined effect of the gray intensity and the sector entropy, the similarity measurement is conducted for the final abdomen reconstruction. The experimental results reveal that the proposed method can acquire a high precision of abdomen reconstruction similar to the 3D scanner. This stereo vision system has wide practicability in the field of clothing.

Keywords: Stereo vision system, illumination-robust and anti-blur combined invariant, sector entropy, circle template, feature matching, 3D reconstruction.

1 Introduction

The technology of 3D abdomen model reconstruction is crucial to personalized garment design, health care, man-machine engineering, etc. The common problem we confront is how to recover the abdomen model realistically and efficiently.

Mostly, the laser or structured light scanner is used in the body surface imaging, such as America artec 3D scanner, 3D CaMega DCS made in Beijing Boweihengxin Company, etc. Baek and Lee[1] captured the body data by 3D scanner and extracted semantic parameters, then established a correlation between body model and body size by statistical analysis, finally they created a realistic 3D human body model. In this method, a 3D scanner is utilized. This method has a high data acquisition speed and precision, while its disadvantages also exist (e.g., the extremely high cost, the strict constraints to the camera′s positions and the subject, and the additional operations like filling role, removing noise point, data splicing, etc.), which limit its wide extension. In order to overcome these problems, Li et al.[2] proposed a template-based method with parameters gathered from the 4 images in orthogonal views. This method is of low cost, but depends on image processing at early stage. Amstutz et al.[3] introduced a multiple camera-based method to reconstruct the 3D human foot. The standard foot model in database is adjusted for fitting the real foot according to multiple foot images.

Recently, a new stereo vision method based on multiple-camera system has been widely applied to 3D reconstruction. Yu and Xu[4] developed a compact and economical configuration for imaging and completing the whole body based on stereo vision technique. The foot recovery is achieved by stereo vision method[5, 6]. Stereo vision plays an important role in body model reconstruction in virtue of its simple, safe and reliable equipment, low cost, and the fast reconstruction speed.

Feature matching is an extremely crucial technique in 3D reconstruction and its accuracy is directly related to the precision of the final reconstruction model. The feature matching algorithms mainly fall into two categories: global matching based on region gray[7] and local matching based on image feature[8, 9]. On account of its great computational load, the global matching algorithm is unsuitable to real-time system[10]. Therefore, in this paper, the local matching method is implemented. Local matching comprises the feature detection and description. Harris and Stephens[11] proposed the Harris detection algorithm based on Moravec. The Harris is extremely effective, which is robust to the image noise, image translation and rotation, and variances of gray level and illumination[12]. The corners extracted are stable, well-distributed and of high location accuracy. D. Lowe first proposed scale-invariant feature transform (SIFT) in 1999, and perfected it in 2004, which is one of the most widely used feature detection algorithms[13, 14]. SIFT is invariant to the changes of image scale, rotation, illumination and sight angle. But when the angle exceeds 35°, the viewpoint invariance is difficult to be guaranteed. Furthermore, it is inapplicable to the nonlinear gray change.

Speed-up robust features (SURF) is an improved version on SIFT. In most cases, its performance is similar to SIFT, but the running speed is 3−7 times faster than that of SIFT. Maximally stable extremal region (MSER)[15] is an affine invariant algorithm, which was put forward by Matas based on watershed image segmentation method. It has nice anti-noise and affine invariance, however, it is sensitive to fuzzy image. With the boundary becoming blurred, MSER is incapable of obtaining accurate segmentation, hence its repeatability deteriorates rapidly.

Float descriptors include SIFT, SURT, moment invariants and dense wide-baseline stereo matching effectively (DAISY), etc.[16] DAISY is an efficient local image descriptor for dense feature extraction, the ideology of which is the same as SIFT. The difference is that DAISY conducts block convergence of gradient direction histogram by Gauss convolution. Generally speaking, there are shortcomings, such as slow calculation speed and the usage of large memory, though the matching performance is quite nice. In recent years, scholars have brought forward binary descriptors. Binary robust invariant scalable keypoints (BRISK)[17] extracts feature points with the adaptive and generic corner detection based on the accelerated segment test (AGAST)[18], which employs sampling mode for feature detection, and then conducts gray comparison test. However, the sampling points selected are highly correlated without uniqueness. After that, Alahi et al.[19] proposed the fast retina keypoint (FREAK) algorithm. FREAK constructs a sampling mode which is similar to the neural retina, i.e., the nearer the center of the image area, the more intensive the sampling is, while sparse sampling is in peripheral areas. Although this strategy reduces the training complexity, the fixed sampling does not necessarily contain the optimal point. The identifications of BRISK and FREAK are affected by the number and the locations of the sampling points, which are difficult to achieve satisfactory matching results. Binary descriptors have high efficiency, but under the consideration of the same resolution for each descriptor, it results in distance error. Its robustness is inferior to float descriptors.

In the course of capturing image, the captured image is often influenced by bright variance and fuzzy noise. The former is due to the illumination variance and/or the camera's uneven exposure, and different optical properties from multiple cameras. All these result in the different brightness in the images of the same object when imaging at the same time. The latter is because of imaging system aggregation and/or atmospheric turbulence, which make image blurred. In order to overcome the difficulties from the illumination variance and blur noise, a new way is designed to achieve image matching through a 32-dimensional feature descriptor generated from context histogram[20]. The new descriptor is invariant to illumination and blurred change, but its stability decreases. In a word, it is extremely crucial to explore an improved method with higher matching effect as well as being invariant to illumination change and fuzzy noise.

Image moment invariants are highly concentrated features and have been widely applied to image matching. The classical moment invariants include Hu invariant moments, Legendre moment, Zernike moments, composite moments, etc. A 6-dimensional moment invariant which is insensitive to any kind of compound transformations is developed in [21]. Miao et al.[22] combined Radon and polar harmonic transforms (PHTS) so as to produce three new moment invariants which are highly robust to noise. Fuzzy moment invariant was first presented, which was successfully utilized in satellite image detection[23]. However, all of the existing moment invariants have no consideration of the factors including both illumination and blur, which limit their applications in practice. The proposed combined invariant is invariant to the changes of image scale, rotation, and sight angle owing to the construction based on the Hu moments. Furthermore, it is more important that the conjunction of illumination and blur factors is involved in the new invariant creation. Therefore, it has the illumination invariance and fuzzy robustness. In addition, the new combined invariant is a 3-dimensional vector, which is superior to other features in time and computation complexities.

The rest of the paper is organized as follows. The detailed construction of the new illumination-robust and anti-blur combined invariants is presented in Section 2. The improved strategies concerning the circle template and a creative combined feature vector involving sector entropy are expounded in Section 3. Numerical experiments and tests are conducted and the results are discussed in Section 4. Finally, conclusions are drawn in Section 5.

2 New illumination-robust and anti-blur combined invariants

Hu[24] first presented the definition of continuous function moments and pioneered the research concerning moment invariant. Many scholars have done a lot of improvements on it, which have been applied to image recognition, matching, etc. Hu combined the second-order central moment with the third-order one as the result of producing 7 expressions for invariant moments (termed Hu-moment), which involve the invariance about rotation, scaling, contrast and translation. Notably, in the discrete case, Hu-moment has no invariant performance mentioned above. So it is indispensable to explore a set of new and applicable invariant descriptors.

2.1 New combined invariants on Hu-moment

We denote (x, y) and (x', y') as the coordinates of a pixel before and after the transformation. Similarly, $f(x, y)$ and $f'(x', y')$ mean the pixel gray level intensities before and after the transformation. Let ε be a scale factor, and λ be a contrast factor. There exist relations that may be of the 3 basic types as

$$\left[\begin{array}{c} x' \\ y' \end{array} \right] = \varepsilon \left[\begin{array}{c} x \\ y \end{array} \right], \quad \varepsilon > 0 \tag{1}$$

$$f'(x', y') = \lambda f(x, y), \quad \lambda > 0. \tag{2}$$

We assume that μ_{pq} and μ'_{pq} mean the central moments before and after transformation. Similarly, $(\overline{x}, \overline{y})$ and $(\overline{x}', \overline{y}')$ mean the barycentric coordinates before and after transformation. The relation between them can be expressed as

$$\mu'_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} (x' - \overline{x}')^p (y' - \overline{y}')^q f'(x', y') =$$
$$\sum_{i=1}^{M} \sum_{j=1}^{N} \lambda \varepsilon^{p+q} (x - \overline{x})^p (y - \overline{y})^q f(x, y) =$$
$$\lambda \varepsilon^{p+q} \mu_{pq}. \tag{3}$$

The relation between the normalized central moments can be defined by

$$\eta'_{pq} = \frac{\mu'_{pq}}{\mu_{00}^{r'}} = \frac{\lambda \varepsilon^{p+q} \mu_{p+q}}{(\lambda \mu_{00})^r} = \frac{\varepsilon^{p+q}}{\lambda^{\frac{p+q}{2}}} \times \frac{\mu_{p+q}}{\mu_{00}^r} =$$
$$\left(\frac{\varepsilon^2}{\lambda}\right)^{\frac{p+q}{2}} \eta_{pq} = Q^{\frac{p+q}{2}} \eta_{pq} \tag{4}$$

where η_{pq} and η'_{pq} mean the central moments before and after transformation. Set $r = \frac{p+q+2}{2}$ and $Q = \frac{\varepsilon^2}{\lambda}$.

Equality (4) is substituted into Hu-moment formulas, and the derivatives are obtained as

$$\phi'_1 = Q\phi_1, \quad \phi'_2 = Q^2\phi_2, \quad \phi'_3 = Q^3\phi_3, \quad \phi'_4 = Q^3\phi_4,$$
$$\phi'_5 = Q^6\phi_5, \quad \phi'_6 = Q^4\phi_6, \quad \phi'_7 = Q^6\phi_7$$

where $\phi_1, \cdots, \phi_7$ and $\phi'_1, \cdots, \phi'_7$ indicate Hu-moment invariants before and after transformation, respectively.

We combine some new formulas among all of the above equations related by

$$\psi_1 = \frac{\sqrt{\phi_2}}{\phi_1}, \quad \psi_2 = \frac{\phi_4}{\phi_3}, \quad \psi_3 = \frac{\phi_7}{\phi_5},$$
$$\psi_4 = \frac{\phi_3}{\phi_2\phi_1}, \quad \psi_5 = \frac{\phi_5}{\phi_3\phi_4}, \quad \psi_6 = \frac{\phi_6}{\phi_4\phi_1} \tag{5}$$

where $\psi'_1, \cdots, \psi'_6$ are the transformed invariants. For example, the relation between ψ_1 and ψ'_1 is denoted by

$$\psi'_1 = \frac{\sqrt{\phi'_2}}{\phi'_1} = \frac{\sqrt{Q^2\phi_2}}{Q\phi_1} = \frac{\sqrt{\phi_2}}{\phi_1} = \psi_1. \tag{6}$$

From (6), we can recognize that ψ_1 is equivalent to ψ'_1, and $\psi'_2, \cdots, \psi'_6$ are equivalent to $\psi_2, \cdots, \psi_6$, respectively.

2.2 Combine anti-blur with the new invariants

Due to the influence of some factors, such as the camera defocus, atmospheric turbulence and sensor and/or object motion, the target image is often blurred[25], and the process is defined as a convolution of the image with the Gaussian function, i.e.,

$$f'(x, y) = f(x, y) \otimes g(x, y) \tag{7}$$

where $g(x, y)$ means the 2-dimensional Gaussian function, $g(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$.

A pioneer work on this field was done by Flusser and Suk who derived invariants to the convolution with an arbitrary centrosymmetric point-spread function (PSF)[26]. From the geometric point of view, their descriptors are invariant to translation only. Despite of this, the invariants have found successful applications in normalizing blurred images into canonical forms[27], in template to scene matching of satellite images, in blurred digit and character recognition[28], and in registration of images obtained by digital subtraction angiography[29]. Flusser et al.[30] proposed a group-theoretic approach to extend the combined invariant into 3D case. Zhang et al.[31] exploited the combined blur-rotation invariants. In their studies, the blur invariants were calculated on the normalized image which was obtained by transforming from the initial image. It is difficult to ensure the standard form without any blur influence which causes much complexity.

In the paper, the combined invariants with transformation and convolution with Gaussian function are introduced. This ideology is simpler without necessity to perform geometric normalization and/or image deblurring.

Theorem 1. The central moment blurred by Gaussian is

$$\mu^*_{pq} = \sum_{i=0}^{p} \sum_{j=0}^{q} C_p^i C_q^j g_i g_j \mu_{p-i, q-j} \tag{8}$$

where

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

$$g_i = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x_\tau^i e^{-\frac{x_\tau}{2\sigma^2}} dx_\tau =$$
$$\begin{cases} 0, & i \text{ is odd} \\ 1, 3, 5, \cdots, (i-1)\sigma^i, & i \text{ is even.} \end{cases} \tag{9}$$

Some equations can be deduced from (9), such as $g_1 = g_3 = g_5 = g_7 = \cdots = 0$, $g_0 = 1$, $g_2 = \sigma^2$, $g_4 = 3\sigma^4$, $g_6 = 15\sigma^6$, $\cdots$.

There is a relation between the first Hu-moments derived from the original image and fuzzy image respectively, which can be described by

$$\phi^*_1 = \eta^*_{20} + \eta^*_{02} = \frac{\mu^*_{20} + \mu^*_{02}}{\mu_{00}^2} =$$
$$\frac{\mu_{20} + \mu_{02} + 2\sigma^2\mu_{00}}{\mu_{00}^2} = \phi_1 + \frac{2\sigma^2}{\mu_{00}}. \tag{10}$$

From (10), we find that the values of Hu-moment change with the growth of $\frac{2\sigma^2}{\mu_{00}}$.

Similarly, we find that the later 6 moments $\phi_2, \cdots, \phi_7$ can be obtained without any change, i.e., the two values from the pro-change and post-change are equal for the same

moment. Each mathematic derivation is listed as

$$\phi_2 = \phi_2^* = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = \phi_3^* = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = \phi_4^* = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = \phi_5^* = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \times$$
$$[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] +$$
$$(3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = \phi_6^* = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] +$$
$$4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21})$$

$$\phi_7 = \phi_7^* = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \times$$
$$[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] -$$
$$(\eta_{30} - 3\eta_{12})(\eta_{03} + \eta_{21})[3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2].$$

When we substitute all the above equations into (5), only three of those new illumination-robust combined moments are anti-blur simultaneously, which are

$$\psi_2^* = \frac{\phi_4^*}{\phi_3^*} = \frac{\phi_4}{\phi_3} = \psi_2$$

$$\psi_3^* = \frac{\phi_7^*}{\phi_5^*} = \frac{\phi_7}{\phi_5} = \psi_3$$

$$\psi_5^* = \frac{\phi_5^*}{\phi_3^* \phi_4^*} = \frac{\phi_5}{\phi_3 \phi_4} = \psi_5.$$

In conclusions, ψ_2, ψ_3 and ψ_5 are the illumination-robust and anti-blur combined moment invariants.

For the sake of the large variation range, we use logarithm to compress the variation range of feature moment. Since we also take the negative value into account, we take the absolute value before logarithm. The actual invariant moments can be defined by

$$\psi_k' = \lg|\psi_k|, \quad k = 2, 3, 5. \tag{11}$$

3 An improved combined feature description

3.1 Circle template

It is noted that producing a description merely according to intensity is not able to provide sufficient information for feature depiction, and may cause a great deal of error matches. In order to improve the accuracy of identification, a more distinctive description is demanded. Consequently, to take the image structure into account, a unit entropy is proposed in this paper. Meanwhile, to allow for the engineering application of abdomen reconstruction, the template is designed as a circle window, as displayed in Fig. 1. The main direction is decided by the distribution of the peak gradient direction within the adjacent pixels around the key point.

$$m(x, y) =$$
$$\sqrt{(f(x+1, y) - f(x-1, y))^2 + (f(x, y+1) - f(x, y-1))^2}$$

$$\theta(x, y) = a\tan\left(\frac{f(x, y+1) - f(x, y-1)}{f(x+1, y) - f(x-1, y)}\right) \tag{12}$$

where $m(x, y)$ and $\theta(x, y)$ are the gradient and direction of the point at coordinate (x, y). The peak of the gradient histogram is defined as the main direction.
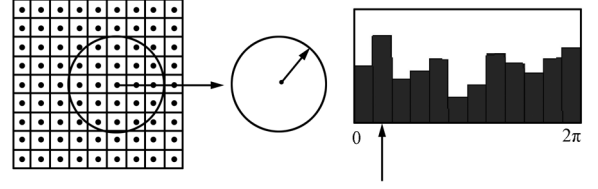

Fig. 1   Main direction of the circle

Beginning with the main direction, the circle area is divided clockwise every $60°$. And the total 6 sector regions can be seen in Fig. 2.
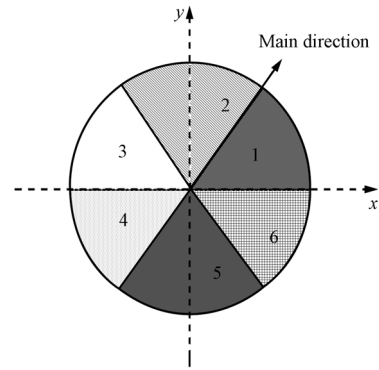

Fig. 2   The distribution of 6 sector regions

3.2 Combined sector entropy by CCA

The canonical correlation analysis (CCA) is that multiple original sets of feature vectors are converted to a few comprehensive variables.

For two random vectors, $x \in \mathbf{R}^m$ and $y \in \mathbf{R}^n$, we find a couple of directional projections α and β. If x' and y' are equivalent to $\alpha^\mathrm{T} x$ and $\beta^\mathrm{T} y$, for a maximal correlation coefficient,

$$\rho = \frac{\alpha^\mathrm{T} C_{xy} \beta}{\sqrt{\alpha^\mathrm{T} C_{xx} \alpha \times \beta^\mathrm{T} C_{yy} \beta}} \tag{13}$$

where C_{xx} and C_{yy} are the covariance matrixes of x and y, C_{xy} means the cross-covariance matrix between x and y. The obtained canonical correlation variables are regarded as two groups of feature vectors, i.e.,

$$X = (\alpha_1^\mathrm{T} x_1, \alpha_2^\mathrm{T} x_2, \cdots, \alpha_m^\mathrm{T} x_m)^\mathrm{T} =$$
$$(\alpha_1, \alpha_2, \cdots, \alpha_m)^\mathrm{T} x = U_x^\mathrm{T} x$$
$$Y = (\beta_1^\mathrm{T} y_1, \beta_2^\mathrm{T} y_2, \cdots, \beta_n^\mathrm{T} y_n)^\mathrm{T} =$$
$$(\beta_1, \beta_2, \cdots, \beta_n)^\mathrm{T} y = U_y^\mathrm{T} y$$

where $U_x = (\alpha_1, \alpha_2, \cdots, \alpha_m)$ and $U_y = (\beta_1, \beta_2, \cdots, \beta_n)$. Then, another vector is assigned as

$$J = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} U_x^\mathrm{T} x \\ U_y^\mathrm{T} y \end{pmatrix} \tag{14}$$

where J is termed combined canonical feature vector.

In this article, the 3 new combined moments from pixel gray level intensity are recombined as a vector, i.e. $x = (\psi_2, \psi_3, \psi_5)$. The 6 feature parameters from sector entropy are recombined as the other vector, i.e., $y = (e_1, e_2, e_3, e_4, e_5, e_6)$. By working out parameters α and β when parameter ρ gets the maximum, the feature vectors with CCA can be gained. We define $(\alpha_1, \alpha_2, \alpha_3)^{\mathrm{T}} x$ as IM_F, and $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^{\mathrm{T}} y$ as E. J can be also expressed as

$$J = \begin{pmatrix} X \\ Y \end{pmatrix} = $$
$$\begin{pmatrix} (\alpha_1, \alpha_2, \alpha_3)^{\mathrm{T}} x \\ (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^{\mathrm{T}} y \end{pmatrix} = \begin{pmatrix} IM\_F \\ E \end{pmatrix}. \quad (15)$$

Making use of CCA, the intensity vector IM_F and entropy vector E of each unit sector construct a new feature vector IM_FE shown in Fig. 3. It is significant to explore feature description by means of synthesizing the above two individual properties.
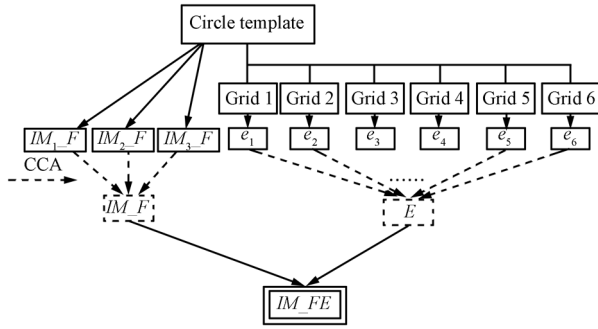

Fig. 3   The IM_FE, IM_F and E

3.3   Similarity measurement

We take the distance between two features as the similarity measurement of the key points as

$$S_{i,j} = \mathrm{e}^{-|D_l(i) - D_r(j)|} \quad (16)$$

where $|D_l(i) - D_r(j)|$ is termed as a distance measure which is determined by

$$|D_l(i) - D_r(j)| = $$
$$|IM\_F_l(i) - IM\_F_r(j)| \times |E_l(i) - E_r(j)|. \quad (17)$$

A suitable threshold for the similarity is set according to the requirement of precision in real application. If the similarity is greater than the threshold, the pair of matching points can be accepted. If the ratio of threshold is reduced, the number of matching points will decrease, but the system will be more stable.

4   Experiment and analysis

4.1   Verification experiment

4.1.1   Verification of the new combined invariants

For verifying the illumination robustness and the anti-blur of the new combined invariants, the initial image is added with levels of blur coefficients defined by (7). Moreover, the illumination darkening coefficients, which made the initial image become greyer, were also added. In the test, both blur and illumination darkening coefficients should ensure the adjusted images to be recognizable shown in Fig. 4. The left picture in Fig. 4 is the initial one, the images with different levels of blur noise coefficients are displayed in the first line, those with different illumination darkening coefficients are shown in the second line, and images added with both different blur and illumination darkening coefficients are exhibited in the third line. The experiment conditions are Celeron(R) dual-core CPU, 2 GHZ, 2 GB memory, Windows XP 2002 operating system, and Matlab7.0 programming environment.
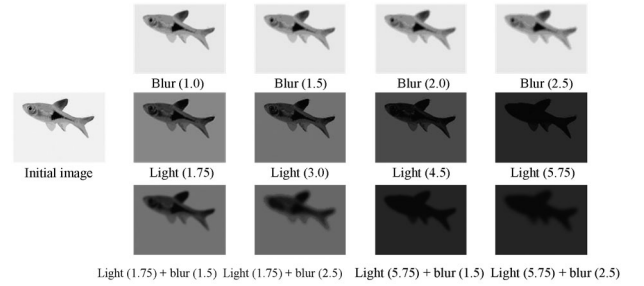

Fig. 4   Fish images with blur and illumination darkening coefficients

Table 1 lists the values of the 3 combined invariants for each figure, and there are a total of 13 figures (deduced in Section 2.2, ψ_2, ψ_3, ψ_5). As can be seen, whether the image is affected by any factor or by two factors together, the 3 invariants are barely changed. Thus, it is proved that the new combined invariants are strongly robust to illumination variance and anti blur noise.

Table 1   3 combined invariants from different fish images

Images	ψ_2	ψ_3	ψ_5
Initial image	0.953261	0.0678106	0.999750
Blur (1.0)	0.953261	0.0678106	0.999750
Blur (1.5)	0.953261	0.0678106	0.999751
Blur (2.0)	0.953261	0.0678105	0.999751
Blur (2.5)	0.953262	0.0678105	0.999751
Blur (1.75)	0.953261	0.0678106	0.999750
Blur (3.0)	0.953261	0.0678106	0.999750
Blur (4.5)	0.953261	0.0678105	0.999750
Blur (5.75)	0.953260	0.0678105	0.999751
Light (1.75) + blur (1.5)	0.953261	0.678106	0.999750
Light (1.75) + blur (2.5)	0.953261	0.678106	0.999750
Light (5.75) + blur (1.5)	0.953260	0.678105	0.999751
Light (5.75) + blur (2.5)	0.953260	0.678104	0.999752

4.1.2   Comparison with other features

To further testify the performance of the proposed descriptors, comparison experiments are conducted among SIFT, SURT, MSER, DAISY, FREAK, BRISK and the

proposed combined invariants. Due to our main discussion on the invariance of illumination and blur, the standard images of the illumination changes of Leuven figures (Leuven1, Leuven2, Leuven3, Leuven4, Leuven5 and Leuven6) and the blur degrees of Bike figures (Bike1, Bike2, Bike3, Bike4, Bike5 and Bike6) from the photo gallery of Oxford are selected. To be fair, Harris is used uniformly to extract features for all the images in advance. Comparisons are shown by calculating the matching rate which is got from the mapping between Leuven1 and Leuven2 (Leuven1 and Leuven3, Leuven1 and Leuven4, Leuven1 and Leuven5, Leuven1 and Leuven6) in Fig. 5 (a), and the matching rates from Bike figures (Bike1 and Bike2, Bike1 and Bike3, Bike1 and Bike4, Bike1 and Bike5, Bike1 and Bike6) in Fig. 5 (b). The higher the matching rate is, the better the performance is. As can be seen from Figs. 5 (a) and (b), when the illumination changes and fuzzy interferes in different degrees, the proposed descriptor can still get good matching rates. In a word, the illumination robustness and anti-blur of the proposed descriptor are better than those of the others.

4.2 Application to abdomen reconstruction

The outline of the way to recover a virtual human abdomen model is as

1) Extract the feature points between the left and right images respectively through Harris algorithm.

2) Obtain the correct matches with the new combined feature descriptor IM_FE.

3) Retrieve 3D cloud data for reconstruction in order to measure the abdominal depth, breadth, and circumference of the subject.

4.2.1 Experimental set-up

In the section, experiments on the abdomen reconstruction are conducted by using the proposed methods. The specific set-up is shown in Fig. 6. A stereo vision system composed of four complementary metal oxide semiconductor (CMOS) cameras (Cyber-shot DSC-W30) is arranged around the model, with a resolution of 2816×2112 pixels and about 6 003 000 effective pixels. The focal length is 18.9 mm (Converted to the 135 mm camera, the value is 114 mm). The image has 624×363 pixels. Two of the cameras (No.1 and No.2) are located in front of the abdomen within a pair for an angle of 60° for obtaining the abdomen images, and the other two cameras (No.3 and No.4) are arranged in the rear of the abdomen with the same angle to the frontal pair for getting the back abdomen images.

620 markers are separately sewn both in front and at the back pieces of the tight dress at the abdominal part. In this section, 3 correlative subsections are arranged corresponding to the three procedures mentioned above in sequence. In Section 4.2.2, the feature extractions with Harris algorithm
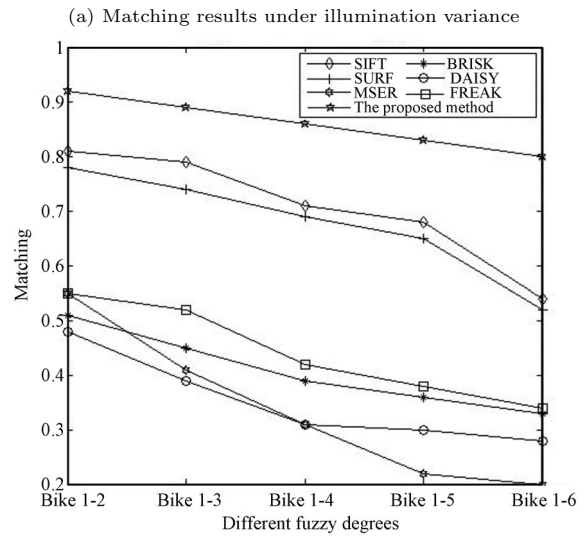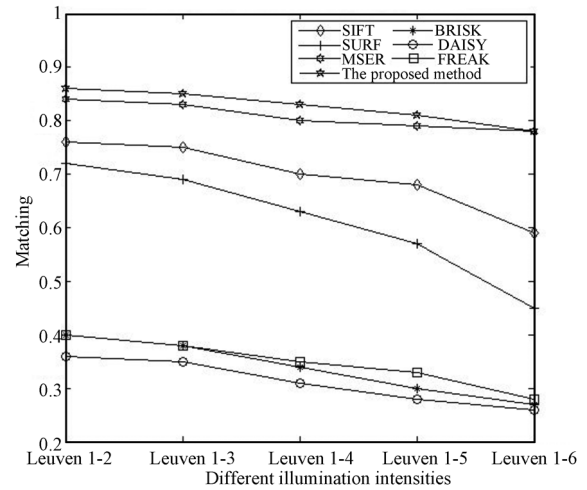

(a) Matching results under illumination variance

(b) Matching results under the fuzzy degree

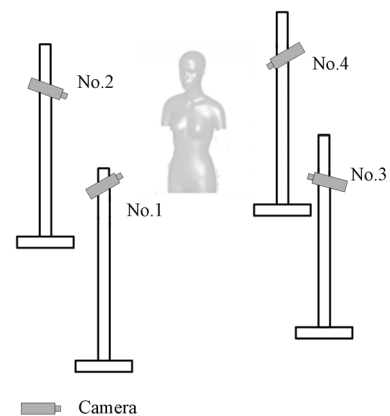Fig. 5 Matching effect for illumination and fuzzy images

Fig. 6 Stereo vision system

are carried out. In Section 4.2.3, the stereo matching with IM_FE description is performed and a thorough analysis about the results is made. Finally, in Section 4.2.4, the 3D cloud data of the human abdomen is utilized for reconstruction. In each subsection, we devote ourselves to enough

tests in 4 typical cases depending on different choices of blur and illumination darkening coefficients. The detailed instructions are given as follows. Case 1 (Figs. 7 (a) and (b)) indicates the left and right belly images in the standard situation. Taking right belly image for example, case 2 (Fig. 7 (c)) is the condition with blur noise of 4.5 in standard right belly image, case 3 (Fig. 7 (d)) is the state combined with blur noise of 4.5 and light darkening of 6.0 in the standard right belly image, and case 4 (Fig. 6 (e)) is the state with blur noise of 5.0 and light darkening of 7.0 in the standard right belly image. With the increase of the blur and darkening coefficients, the images become dimmer and darker. The above 4 situations are representative. It is not necessary to study those with too large or too small coefficient on account of no realistic significance.
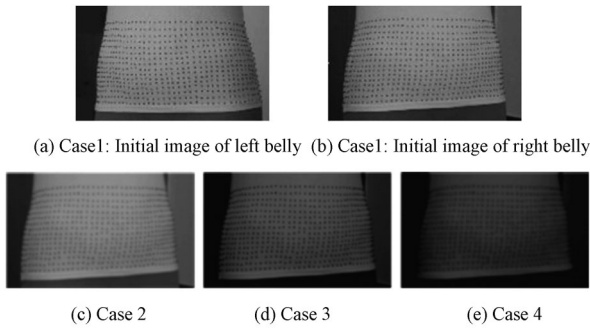
(a) Case1: Initial image of left belly   (b) Case1: Initial image of right belly

(c) Case 2   (d) Case 3   (e) Case 4

Fig. 7   Noise added and brightness weakening on right belly image

4.2.2   Experiments on feature points extraction

One of the critical matters we need to solve is how to extract the accurate markers from the image of the tight dress, which is directly correlated to the accuracy of the later stereo matching and reconstruction. Harris is a classic algorithm commonly used in feature extraction. Its advantages include:

1) Simple calculation: Harris only uses the first order difference and filter wave.

2) The uniform and reasonable results of feature point extraction: It only needs to calculate the interested value for each point, and then select the optimum within the neighborhood.

3) Excellent reliability: Thanks to a differential operation, second-order moment and Gauss scale computations, it is not sensitive to illumination variation, image rotation, gray variation, noise influence, or perspective change.

Depending on the practical engineering background and considering all comprehensive factors, Harris algorithm is adopted in our work. In case 1, the experiment is performed in the standard belly images (Figs. 7 (a) and (b)), altogether 629 left and 638 right feature points are successfully tracked in Figs. 8 (a) and (b).

Figs. 9 (a) and (b) show the results produced by Harris in case 2 (with blur noise of 4.5). Figs. 10 (a) and (b) are the results in case 3 (with blur noise of 4.5 and light darkening of 6.0). Figs. 11 (a) and (b) are yielded from case 4 (with blur noise of 5.0 and light darkening of 7.0).
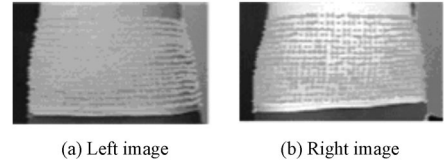
(a) Left image   (b) Right image

Fig. 8   Case 1: feature points detection with Harris algorithm

(a) Left image   (b) Right image
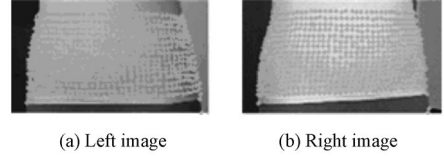
Fig. 9   Case 2: feature points detection with Harris algorithm

(a) Left image   (b) Right image
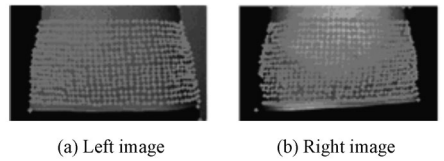
Fig. 10   Case 3: feature points detection with Harris algorithm
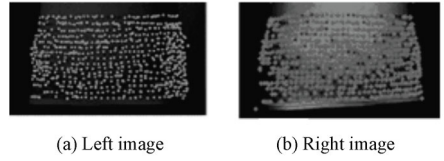
(a) Left image   (b) Right image

Fig. 11   Case 4: feature points detection with Harris algorithm

With the same principle in the above 4 cases, we take the right abdomen images for example. The results of extracting feature points of the right abdomen images for all of the above experiments are shown in Table 2. p_num indicates the number of the extracting points, ρ is the detection rate, D_x and D_y are the deviations between the detection value and the real value in the x and y coordinates. Min_x, Max_x, $Mean\_x$ and Av_x denote the minimum deviation, the maximum deviation, the average deviation and the mean square deviation in the x coordinate, respectively. Min_y, Max_y, $Mean\_y$ and Av_y separately denote the same meanings in the y coordinate.

From Table 2, we can find that in the former 3 cases, the detection performance maintains stable with ρ of more than 90 % in spite of the slight decline, and yet in the 4th case, ρ drops to less than 70 %. Admittedly, when the impact factors increase to an extreme level, it will substantially worsen the extractive correctness. Actually, to that extreme degree, the image quality is seriously poor without any practical value. Besides, regardless of the deviations of x or y coordinate, the errors from Harris are tiny. All in all, within an acceptable and reasonable choice of blur and illumination darkening coefficients, Harris is capable of maintaining good stability and achieving satisfactory results. In the paper, it is suggested that the blur and darkening coefficients should not be more than 5.0 and 7.0 for ensuring the accuracy of later feature matching.

Table 2 Point detection of the right belly image in 4 cases by Harris

Case (right belly image)		Case 1	Case 2	Case 3	Case 4
ρ_num		638	644	633	557
ρ (%)		98.55	91.94	91.13	66.29
D_x (pixel)	Min_x	0	0	0	2
	Max_x	1	1	1	4
	$Mean\_x$	0.0034	0.0068	0.0102	0.1458
	Av_x	0.0034	0.0067	0.0101	0.1358
D_y (pixel)	Min_y	0	0	0	2
	Max_y	1	1	1	3
	$Mean\_y$	0.0022	0.0053	0.0092	0.1291
	Av_y	0.0018	0.0051	0.0090	0.1225

4.2.3 Experiments on feature matching

In the matching test, reliable matching pairs are obtained according to the value of similarity S. The specific procedure can be described as: computing the ratio of the nearest neighbor distance min_d to the second nearest neighbor distance sec_min_d on the basis of feature vectors, and then taking the logarithm of base e of the ratio, signed as $\mathrm{e}^{\frac{min\_d}{secmin\_d}}$. If $\mathrm{e}^{\frac{min\_d}{secmin\_d}} > \varepsilon$ (ε is called similarity threshold, and is 0.5 in this test), then these two points are considered as a correct matching pair. Too small a value of ε may affect the number of the correct matches. The general value is at the range of [0.45 0.55].

The tests on feature points matching are carried out with the new descriptor vector IM_FE. The results of cases 1, 2, 3 and 4 are displayed in Figs. 12 (a), 1,(b), 12 (c) and 12 (d), respectively. The second experiment is performed for case 2. Similarly, the whole procedures are also realized again by IM_FE. The result is displayed in Fig. 12 (b). The third test is conducted on images in case 3, and the results yielded by IM_FE are displayed in Fig. 12 (c). The last test is employed for case 4. The results generated by IM_FE are shown in Fig. 12 (d).

(a) The matching result in case 1

(b) The matching result in case 2

(c) The matching result in case 3

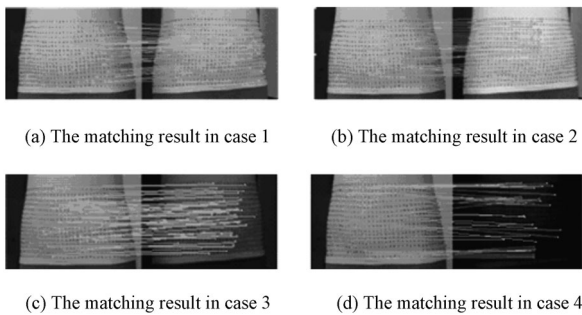(d) The matching result in case 4

Fig. 12 The matching results in 4 cases

Table 3 lists some parameters of feature matching. In the list, n_f signs the number of the initial matching pairs, n_c indicates the correct number, r means the correct matching ratio, and the time spent in constructing descriptor and mapping is marked with T. As shown in Table 3, it can be easily seen that the time is much shorter in virtue of a low

dimensional descriptor derived from the new anti-blur and illumination-robust combined invariants of gray and sector entropy with CCA, which can reduce the time for constructing descriptor greatly. Moreover, under the ideal situation (i.e., case 1), the correct ratio of IM_FE can reach more than 90 %. When the image quality is affected by blur noise and/or illumination darkening, on account of both the blur resistance and the illumination variation insensitiveness of its descriptor, the correct ratio of IM_FE still maintains over 85 % in the former 3 cases within the acceptable error range in practical application in spite of its merely slight descending. The little drop is attributed to the decline of the extracted rate of Harris. But in case 4, the matching correctness of IM_FE is seriously affected with less than 70 % for the failure of the feature point detection by Harris. In the paper, it is better to control the blur and darkening coefficients within 5.0 and 7.0 so as to ensure the effectiveness of IM_FE. In conclusion, IM_FE is attested to be greatly stable and robust as long as the blur and darkening coefficients are within a reasonable range which depends on the practical application.

Table 3 Parameters about feature matching in 4 cases by IM_FE

Parameters	Case 1	Case 2	Case 3	Case 4
n_f	67	58	51	42
n_e	61	52	45	27
r (%)	91.04	89.66	88.24	64.29
T (s)	0.32	0.28	0.25	0.21

4.2.4 Abdomen reconstruction and analysis of shape size

Three dimensional cloud data are obtained by the stereo matching introduced in Section 4.2.3. Table 4 demonstrates the computational time consumed for the complete implementation process of abdomen reconstruction by our method. Here, the computational time is offered at a total of the 3 effective cases verified in Subsections 4.2.2 and 4.2.3. T_D, T_M and T_R denote the time spent on feature detection, feature matching and abdomen reconstruction, respectively. $Total\_T = T\_D + T\_M + T\_R$ means the total time exhausted in the whole abdomen reconstruction process, and the mean value of the 3 different cases is offered at the end of the table. As can be seen from the list, the entire abdomen reconstruction can be completed for a little more than one minute. Consequently, it is indeed a high efficient and strong real-time method. Figs. 13 (a) and (b) illustrate the three-dimensional graphs from different views about anterior and back belly. Figs. 13 (c) and (d) are the 3 dimensional belly figures.

In order to evaluate the accuracy of our proposed method, it is tested on a female model by acquiring the abdomen size parameters, including the belly circumference, breadth and depth. We complete the tests on the same model for 3 independent modeling and measurement procedures. Each process is from first putting on the tight

dress to the final measurement of the abdomen shape. Test points are shown in Fig. 14. With the proposed method IM_FE, the 3D cloud data can be obtained, and then are imported into PPS which is a piece of parameter measurement software, finally those 3 abdomen size parameters can be computed automatically. Comparisons of parameters are made among the above 3 schemes (IM_FE, 3D scanner and manual measurement). As shown in Table 5, all of the data are the average values from the results in 3 independent modeling and measurement procedures. $Error\_M\&FE$ is the mean of errors between manual and IM_FE in the 3 valid cases of different blur and darkening coefficients based on the stereo vision system. Similarly, $Error\_3D\&FE$ is the mean of errors between 3D scanner and IM_FE. In comparison with the manual measurement, the differences between the proposed method and the traditional manual measurement are small. The largest error is the abdomen circumference of 0.423 cm, and the smallest error is the abdomen depth merely of 0.117 cm. According to the acceptable error range 0.5–2 cm in clothing, our approach can

completely meet the requirement. Besides, from the analysis of $Error\_3D\&FE$, the values with the proposed method are extremely close to those with 3D scanner, whose error is 10^{-2} cm. It shows that the method with practicability can entirely reach the same precision as 3D scanner.
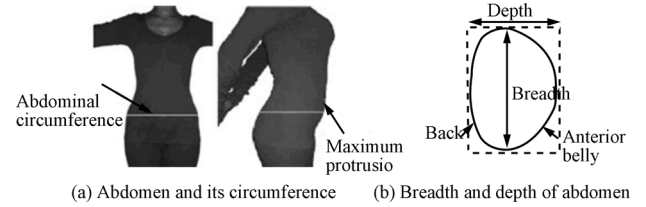

(a) Abdomen and its circumference   (b) Breadth and depth of abdomen

Fig. 14   Definitions of abdomen size parameters on a female model

5   Conclusions

In this paper, we propose a human belly modeling method on the basis of stereo vision system comprised of only 4 cameras fixed up around the subject. We develop the new combined invariants as well as involve the sector entropy that signifies image structure characteristics. Experiments demonstrate that a highly complete abdomen shape and accurate abdomen measurement can be obtained successfully with the help of the feature matching with the innovative feature descriptor vector, even in the condition of blur influence and illumination variance. Briefly, it can be concluded that the distinct advantages of this method are fast image-acquisition, highly accurate, reliable, portable and practicable but low-cost.

Still, it is necessary to notice that the deficiencies are the demand of wearing dress with manual markers and the dependence on the feature extraction algorithm, which limit its scope of applicability and affect its accuracy to a certain extent. For this reason, our research group now begins to conduct the research work about feature extraction algorithm in view of weak texture or no texture. Anyway, all its preferable and popularized performances make it promising for application in a large scale anthropometric survey to collect body dimensions for apparel design. In addition, many applications can be further extended involving human engineering, routine use in clinical medicine settings, rehabilitation training of prostheses, etc.
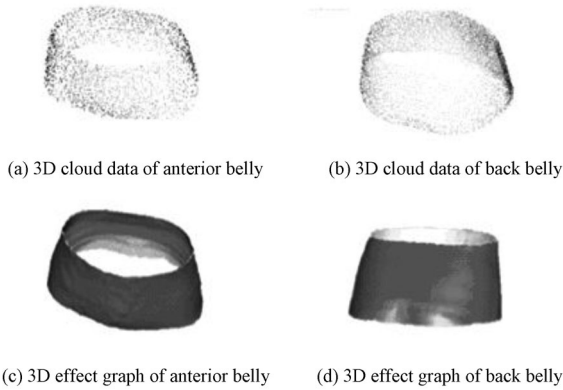

(a) 3D cloud data of anterior belly   (b) 3D cloud data of back belly

(c) 3D effect graph of anterior belly   (d) 3D effect graph of back belly

Fig. 13   Three-dimensional figures of belly from different views

Table 4   Computational costs of the proposed methods in 3 cases

Time (s)	Case 1	Case 2	Case 3
r_D	60.649	55.726	52.485
r_M	1.483	0.889	0.872
r_R	8.565	7.158	7.210
$Total\_T$	70.697	63.773	60.567
Mean of time		65.012	

Table 5   Dimensions of the subject measured by 3 ways

Parameters (cm)	Manual	3D scanner	IM_FE			$Error\_M\&FE$ (cm)	$Error3D\&FE$ (cm)
			Case 1	Case 2	Case 3		
Circumference	78.2	77.78	77.80	77.77	77.76	0.423	0.020
Depth	16.5	16.39	16.39	16.37	16.37	0.117	0.013
Breadth	35.1	34.97	34.98	34.96	0.120	0.010	0.010

References

[1] S. Y. Baek, K. W. Lee. Parametric human body shape modeling framework for human-centered product design. *Computer-aided Design*, vol. 44, no. 1, pp. 56–67, 2012.

[2] J. T. Li, Y. S. Wang, X. Zhou. Personal virtual human modeling from images. *Journal of Computer-Aided Design & Computer Graphics*, vol. 20, no. 5, pp. 554–559, 2008. (in Chinese)

[3] E. Amstutz, T. Teshima, K. Kimura, M. Mochimaru, H. Saito. PCA-based 3D shape reconstruction of human foot using multiple viewpoint cameras. *International Journal of Automation and Computing*, vol. 5, no. 3, pp. 217–225, 2008.

[4] W. R. Yu, B. G. Xu. A portable stereo vision system for whole body surface imaging. *Image and Vision Computing*, vol. 28, no. 4, pp. 605–613, 2010.

[5] J. L. Zhou, S. Y. Zhang. A modeling and comparison method for foot based on passive stereo vision. *Journal of Computer-aided Design & Computer Graphics*, vol. 21, no. 6, pp. 782–788, 2009. (in Chinese)

[6] F. Gao, Q. Wang, W. D. Geng, Y. H. Pan. Three-dimensional dynamic foot shape measuring method. *Journal of Science China*, vol. 6, no. 41, pp. 659–674, 2011. (in Chinese)

[7] Y. B. Zou, F. M. Dong, B. J. Lei, L. L. Fang, S. F. Sun. Image thresholding based on template matching with arctangent Hausdorff distance measure. *Optics and Lasers in Engineering*, vol. 51, no. 5, pp. 600–609, 2013.

[8] F. C. Wu, Z. H. Wang, X. G. Wang. Feature vector field and feature matching. *Pattern Recognition*, vol. 43, no. 10, pp. 3237–3281, 2010.

[9] D. D. Yang, A. Sluzek. A low-dimensional local descriptor incorporating TPS warping for image matching. *Image and Vision Computing*, vol. 28, no. 8, pp. 1184–1195, 2010.

[10] P. R. Sanz, B. R. Mezcua, J. M. S. Pena, J. P. Thiran. Segment-based real time stereo vision matching using characteristic vectors. *Journal of Imaging Science and Technology*, vol. 55, no. 5, pp. 050201-1–050201-7, 2011.

[11] C. Harris, M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, Manchester, UK, pp. 147–151, 1988.

[12] C. Schimid, R. Mohr, C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, vol. 37, no. 2, pp. 151–172, 2000.

[13] D. G. Lone. Object recognition from local scale-invariant fratures. In *Proceedings of IEEE International Conference on Computer Vision*, IEEE, Greece, pp. 1150−1157, 1999.

[14] D. G. Lone. Distinctive image features from scale-invariant Keypoints. *International Joural of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[15] J. Matas, O. Chum, M. Urban, T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[16] E. Tola, V. Lepetit, P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, 2010.

[17] S. Leutenegger, M. Chli, R. Siegwart. BRISK: Binary robust invariant scalable keypoints. In *Proceedings of the 2011 IEEE International Conference on Computer Vision*, IEEE, Switzerland, pp. 2548–2555, 2011.

[18] E. Mair, G. D. Hager, D. Burschka, M. Suppa, G. Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *Proceedings of the 11th European Conference on Computer Vision*, Springer, Berlin Heidelberg pp. 183–196, 2010.

[19] A. Alahi, R. Ortiz, P. Vandergheynst. FREAK: Fast retina keypoint. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, USA, pp. 510–517, 2012.

[20] Y. H. Tang, H. Z. Lu, M. F. Hu. An image matching algorithm based on SCCH feature descriptor. *Journal of Electronics and Information Technology*, vol. 33, no. 9, pp. 2152–2157, 2011. (in Chinese)

[21] G. Wu, C. F. Wu, Q. Y. Hou. Target location method based on invariable moment feature matching. *Optics and Precision Engineering*, vol. 17, no. 2, pp. 460–468, 2009.

[22] Q. G. Miao, J. Liu, W. S. Li, J. J. Shi, Y. D. Wang. Three novel invariant moments based on radon and polar harmonic transform. *Optics Communications*, vol. 285, no. 6, pp. 1044–1048, 2012.

[23] J. Flusser, T. Suk, S. Saic. Recognition of blurred images by the method of moments. *IEEE Transactions on Image Processing*, vol. 5, no. 3, pp. 533–538, 1996.

[24] M. K. Hu. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[25] T. Suk, J. Flusser. Combined blur and affine moment invariants and their use in pattern recognition. *Pattern Recognition*, vol. 36, no. 12, pp. 2895–2907, 2003.

[26] J. Flusser, T. Suk. Degraded image analysis: An invariant approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 590–603, 1998.

[27] Y. N. Zhang, C. Y. Wen, Y. Zhang, Y. C. Soh. Determination of blur and affine combined invariants by normalization. *Pattern Recognition*, vol. 35, no. 1, pp. 211–221, 2002.

[28] X. Wang, R. C. Zhao. Pattern recognition by combined invariants. *Chinese Journal of Electronics*, vol. 10, no. 4, pp. 480–483, 2001. (in Chinese)

[29] Y. Bentoutou, N. Taleb, M. Chikr EI Mezouar, M. Taleb, L. Jetto. An invariant approach for image registration in digital subtraction angiography. *Pattern Recognition*, vol. 35, no. 12, pp. 2853–2865, 2002.

[30] J. Flusser, J. Boldyš, B. Zitová. Moment forms invariant to rotation and blur in arbitrary number of dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 234–246, 2003.

[31] Y. N. Zhang, C. Y. Wen, Y. Zhang, Y. C. Soh. Determination of blur and affine combined invariants by normalization. *Pattern Recognition*, vol. 35, no. 1, pp. 211–221, 2002.
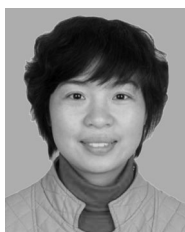
Huan Liu graduated from Nanjing Institute of Technology, China in 2004. She received the M. Sc. degree from Jiangxi Normal University, China in 2008, and the Ph. D. degree from Donghua University, China in 2014. She is currently a lecturer in Jinggangshan University, China.

Her research interests include machine vision, image processing and intelligent algorithm.

E-mail: liuhuan816618@163.com (Corresponding author)

Ying Xiao graduated from Jiangxi Institute of Metallurgy, China in 1985. She received the M. Sc. degree from Yunnan University, China in 2006. She is currently a professor in Jinggangshan University, China.

Her research interests include image processing, visual reality and automatic control.

E-mail: mengya11@126.com

Wei-Dong Tang graduated from Jiangxi Agricultural University, China in 1997. He received the M. Sc. degree and the Ph. D. degree from Jiangsu University, China in 2004 and 2007, respectively. He is currently an associate professor in Jinggangshan University.

His research interests include virtual reality and graphic design, especially the modeling of plant growth (such as virtual plant and intelligent control).

E-mail: metwd@163.com
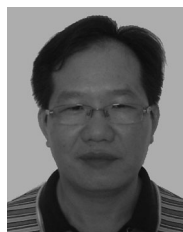
Yan-Hui Zhou graduated from Nanjing Institute of Technology, China in 1994. He received the M. Sc. degree from Nanchang University, China in 2005. He is currently an associate professor in Jinggangshan University, China.

His research interest is electromechanical integration, especially in hydraulic engineering.

E-mail: yhzhou_99@163.com