JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 30, 853-873 (2014)

DOI:10.1688/JISE.2014.30.3.18

Global Image Registration Based on Invariant Representation of Polar Complex Exponential Transform*

MENG YI[1,2,+], BAO-LONG GUO[1] AND CHUN-MAN YAN[1]
[1]*School of Electronic and Control Engineering*
Chang'an University
Xi'an 710071, P.R. China
[2]*Institute of Intelligent Control and Image Engineering*
Xidian University
Xi'an 710064, P.R. China

Robust point matching is a critical and challenging process in feature-based image registration. In this paper, an invariant feature point matching algorithm is presented by introducing the Polar Complex Exponential Transform (PCET), a new kind of orthogonal moment reported recently. Similar to orthogonal complex moments, PCETs is defined on a circular domain. The magnitudes of PCETs are invariant to image rotation and scale. Furthermore, the PCETs are free of numerical instability, so they are more suitable for building shape descriptor. In this paper, the invariant properties of PCETs are investigated, and the accurate moments are selected elaborately. During similarity measurement，the cross correlation function is reconstructed by invariant PCET moments (IPCETs) combining both the magnitude and phase coefficients and maximized to match the control-point pairs. Then the most "useful" matching points that belong to the background are used to find the global transformation parameters between the frames using the projective invariant. The discriminative power of the new IPCETs descriptor is compared with major existing region descriptors (complex moments, SIFT and GLOH). The experimental results, involving more than 10 million region pairs, indicate the proposed IPCETs descriptor has, generally speaking, produce a more robust registration under photometric and geometric performances.

Keywords: image registration, feature matching, polar complex exponential transform (PCET), cross correlation, magnitude and phase components

1. INTRODUCTION

Invariant image representation is crucial for many pattern recognition tasks, such as visual tracking [1], image retrieval [2], image recognition [3], camera calibration [4] and so forth. Image registration [5, 6] resembles these applications in that geometric and photometric invariance is desired. A good registration scheme should be able to estimate the projective model under both photometric transformations (blur, illumination, noise, and JPEG compression) and geometric transformations (rotation, scaling, translation, and viewpoint). Invariant descriptor based registration is a kind of method that can achieve this goal [7]. Descriptor-based registration methods mainly fall back on local techniques, which usually comprise three steps. First, many control points (CPs) are selected or extracted from the reference and sensed images; Second, feature descriptors are used to

Received August 29, 2012; revised November 22, 2012; accepted March 3, 2013.
Communicated by Chung-Lin Huang.
* This work was supproted by the National Nature Science Foundation of China (Nos. 60802077 and 60872136).
+ Corresponding author: yimeng0120@gmail.com.

identify the feature correspondences between image pairs; Third, the parameters of the global transformation are estimated and the sensed image is transformed and resampled by means of the transform model.

The popular invariant descriptors used in the literatures include, but not limited to, filter-based descriptors (steerable filters [8] and Gabor filters [9]), distribution-based descriptors (SIFT [10, 11], GLOH [12], PCA-SIFT [13] and SURF [14]), and moment-based descriptors (geometric moments, Zernike moment (ZM)/pseudo-Zernike moment (PZM)) [7, 15, 16], *et al*.

The steerable filter descriptor uses quadrature pairs of derivatives of Gaussian and their Hilbert transforms to synthesize any filter of a given frequency with arbitrary phase. On the other hand, the Gabor transform uses a number of Gabor filters tuned to various frequencies and orientations to represent the image patterns. However, these filter-based methods are not totally orthogonal and have low dimensions, so their discriminative powers are limited. The idea of SIFT schemes is based on the difference of Gaussians (DOG), utilize a circular window to search for a possible location of a keypoint. However, although the gradient histogram provides satisfactory results against image deformations, the grid partition of the measurement region has high boundary effect [17]. This problem also arises in PCA-SIFT based methods. The moment-based descriptors also provide useful representation for object shapes. The first class of the moment-based descriptors is the geometric moments. Based on the geometric moments, a set of moment invariants can be derived from the nonlinear combinations of the geometric moments to achieve affine invariance [18]. However, geometric moments do not have any of the desired invariance to describe complex shapes. Therefore, the geometric moment invariants are usually only for describing simple images. The ZMs/PZMs based methods are extensively investigated because rotating an image would not change the ZMs/PZMs magnitude [19]. Yet, ZMs/PZMs based methods do not include the phase information, which contains more information than the magnitude part. Furthermore, the higher-order moments are not accurate, thus are not suitable for image matching. Recently, several studies have focused on the phase information of Zernike moments in the comparison process to improve the similarity measure. Although, in that case, the phase information is not invariant to rotation, the moment phase can be used to estimate the rotation angle between two images. One method proposed by Shan and Moon-Chu [20] uses two Euclidean distance measures – the first set consists of ZMs magnitude and the second uses the phase angle. The rotation angle between the overlapping images is estimated by combining phase coefficients of different orders and repetitions to form the rotation invariant. But this method implemented in the discrete space and is sensitive to noise. Another method was presented by Kim and Kim [21] and it proved to be very robust with respect to noise even for circular symmetric patterns. Nevertheless, the probabilistic model used to recover the rotation angle faces multiple peaks of histogram bin and the number of his togram bin values is rather large.

The Polar Complex Exponential Transform (PCET) is a new kind of orthogonal moment defined on the circular domain [22]. Compared to ZMs/PZMs, the computation cost of PCETs is extremely low. Besides, the PCETs are free of numerical instability issues so that high order moments can be obtained accurately. As a result, we believe PCET is more suitable for image matching. In this paper, PCET is introduced into image processing, aiming to achieve more robust under photometric transformations and geo-

metric transformations. The properties of PCETs are investigated and the accurate moments are selected. Then we develop a new rigorously founded approach for comparing two invariant PCETs descriptors (IPCETs) that takes use of both magnitude and phase information. Finally, we develop a projective invariant method that can distinguish more accurate matching points from the less accurate ones and then register the images as accurately as possible. Simulation results show that the proposed scheme shows highly robust both under photometric transformations and geometric transformations. Compared to other major region descriptors, the proposed scheme has the best overall performance.

2. POLAR COMPLEX EXPONENTIAL TRANSFORM

2.1 Mathematical Formulation

Polar Complex Exponential Transform (PCET) is a special name of Polar Harmonic Transform (PHT) [22]. The PCET is defined on a circular domain. For an image $f(x, y)$, it is first transformed into polar coordinates, $f(r, \theta)$, as follows

$$r = \sqrt{x^2 + y^2}, \theta = \arctan\frac{y}{x}. \tag{1}$$

For $f(r, \theta)$, the PCET with order n and repetition l is defined as

$$M_{nl} = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 [V_{nl}(r,\theta)]^* f(r,\theta) r dr d\theta \tag{2}$$

where $|n|, |l| = 0, 1, \ldots, \infty$, $*$ denotes the complex conjugate. $V_{nl}(r, \theta)$ is the kernel of PCET, which consists of a radial component $R_n(r)$ and a circular component $e^{il\theta}$

$$V_{nl}(r, \theta) = R_n(r) \cdot e^{il\theta} \tag{3}$$

with $R_n(r) = e^{i2\pi nr^2}$.

For PCET, the kernel $V_{nl}(r, \theta)$ and its radial component $R_n(r)$ satisfy the following orthogonality conditions.

$$\int_0^{2\pi}\int_0^1 [V_{nl}(r,\theta)]^* f(r,\theta) r dr d\theta = \pi\delta_{nn'}\delta_{ll'} \tag{4}$$

$$\int_0^1 R_n(r)[R_{n'}(r)]^* r dr = \frac{1}{2}\delta_{nn'} \tag{5}$$

Where δ is the Kronecker Delta

$$\delta_{kk'} = \begin{cases} 1, & k = k' \\ 0, & k \neq k' \end{cases} \tag{6}$$

From Eq. (2), we derive the phase relationship of the moments as

$$\phi'_{nl} = \phi_{nl} - l\theta_0. \tag{7}$$

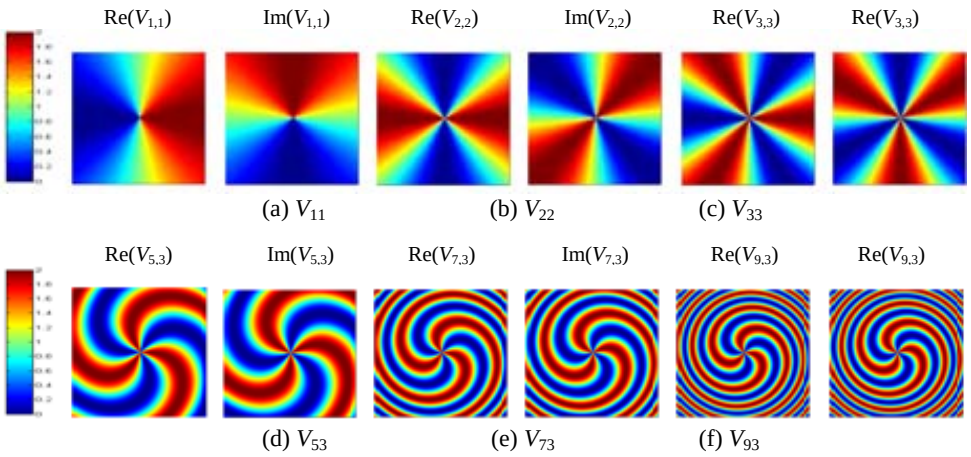
And the magnitude relationship as

$$|M'_{nl}| = |M_{nl}e^{-jl\theta_0}| = |M_{nl}|. \tag{8}$$

The reconstruction of the pattern can be expressed as the sum of every PCET basis function weighted by the corresponding moments:

$$\tilde{f}(x, y) = \sum_{(n,l)\in D} \sum M_{nl} V_{nl}(x, y). \tag{9}$$

Fig. 1 despicts some examples of $V_{nl}(\rho, \theta)$. Notice that the real and imaginary functions of each basis function $V_{nl}(\rho, \theta)$are out of phase $\pi/2$; namely, they form quadrature pairs of filters. In addition, repetition q indicates q sector cycles of the function values along the azimuth angle θ, while n and l jointly specify a different number of annular patterns of the function.



(a) V_{11} (b) V_{22} (c) V_{33}

(d) V_{53} (e) V_{73} (f) V_{93}

Fig. 1. Plots of the part and imaginary part of $V_{nl}(\rho, \theta)$.

2.2 Accurate Moment Selection

In this section, we shall use the PCET magnitude and phase information to design a novel region descriptor, which are invariant to rotation, translation, and scaling in the image. The total number of PCETs is equal to $(n+1)(n+2)/2$. In fact, we do not need all the PCET moments in image registration. Because $V_{n,m}(x, y) = V_{n,-m}(x, y)$, only $M_{nl}(l \geq 0)$ is considered. Table 1 shows samples of PCET features for different max orders.

The sorted PCET moments form a feature vector F as follows:

$$F_k^{PCET} = \left[|M_{31}| e^{j\varphi_{11}}, |M_{51}| e^{j\varphi_{31}}, \ \ldots, \ |M_{nl}| e^{j\varphi_{nl}} \right]^T.$$

Where $|M_{nl}|$ is the PCET magnitude, and φ_{nm} is the PCET phase. Here, the values of M_{00} and M_{11} are not included as the image feature, since they are constant regarding all the normalized images.

Table 1. List of PCETs different max orders.

Order	Moments	No. of Moments	Accumulative
2	$M_{2,0}$; $M_{2,2}$;	2	2
3	$M_{3,1}$; $M_{3,3}$;	2	4
	…		
9	$M_{9,1}$; $M_{9,3}$; $M_{9,5}$; $M_{9,7}$; $M_{9,9}$;	5	28
10	$M_{10,0}$; $M_{10,2}$; $M_{10,4}$; $M_{10,6}$; $M_{10,8}$; $M_{10,10}$;	6	34
	…		

Due to the discrete nature of digital image results in computation error of moments, the PCETs can only be obtained approximately. The inaccuracy moments are not suitable for image registration. Consequently, the initially computed moments need to be carefully selected.

In implementation, we have found that the moments with repetitions $l = 4i$, $i = 0, 1, \ldots, \infty$ are not accurate, thus they cannot be used for construct region descriptor. As a result, the accurate moments used for image registration can be denoted by $S = \{M_{nl}, l \neq 4i, i \in Z\}$.

In order to illustrate this point, we compute the PCETs on a 256×256 image with constant gray value. The magnitudes of some PCETs are listed in Table 2. In this table, we restrict the orders and repetitions by $|n| + |l| \leq 9$ for PCET.

Table 2. Magnitudes of the PCETs for image with constant gray.

	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$	$l = 7$	$l = 8$	$l = 9$
$n = 0$	42.629	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.018	0.000
$n = 1$	0.026	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.018	
$n = 2$	0.026	0.000	0.000	0.000	0.031	0.000	0.000	0.000		
$n = 3$	0.026	0.000	0.000	0.000	0.031	0.000	0.000			
$n = 4$	0.026	0.000	0.000	0.000	0.031	0.000				
$n = 5$	0.026	0.000	0.000	0.000	0.031					
$n = 6$	0.026	0.000	0.000	0.000						
$n = 7$	0.026	0.000	0.000							
$n = 8$	0.026	0.000								
$n = 9$	0.026									

It is easily observed from the tables that the magnitudes of PCETs with repetition $l = 4i$, $i \in Z$, are not zero, and some of them have significant values.

3. PROPOSED SCHEME

3.1 Feature Points Matching Based on IPCETs

It can be noted from Eqs. (2) and (3) that the magnitude remains unchanged, whereas the rotation of an image has an impact on the phase coefficients of the image, so existing moment-based feature descriptors use the magnitude-only PCETs as the image feature [23]. However, losing the phase information cannot effectively describe the original image, and two symmetrical patterns will be classified as identical since their moment magnitudes are the same. In this paper, we form invariant PCETs by combining the magnitude and phase information. As demonstrated in our experiments, our method could represent images more accurately and is also more robust to image noise.

Due to the property of the PCETs, the reconstruction of the pattern $f(x, y)$ can be simply expressed as:

$$f(x,y) \approx \hat{f}(x,y) = \sum_{n}\sum_{l} M_{nl} V_{nl}(x,y). \tag{10}$$

Where, n is a non-negative integer and m is an integer satisfying the conditions: $n - |m|$ is even and $|m| \le n$.

Let I and J be two different images and J_θ be the J image rotated by θ. The cross-correlation function between I and J_θ can be expressed as:

$$\begin{aligned} Corr(I, J_\theta) &= \frac{\sum_{x}\sum_{y} I(x,y) J_\theta(x,y)}{\sqrt{\sum_{x}\sum_{y} (I(x,y))^2 \sum_{x}\sum_{y} (J_\theta(x,y))^2}} \\ &= \frac{\sum_{x}\sum_{y} [\sum_{n}\sum_{m} Z^{I}_{n,m} V_{n,m}(x,y) (\sum_{p}\sum_{q} Z^{J\theta}_{p,q} V_{p,q}(x,y))^*]}{\sqrt{\sum_{x}\sum_{y} \left|\sum_{n}\sum_{m} Z^{I}_{n,m} V_{n,m}(x,y)\right|^2 \sum_{x}\sum_{y} \left|\sum_{p}\sum_{q} Z^{J\theta}_{p,q} V_{p,q}(x,y)\right|^2}} \end{aligned} \tag{11}$$

Due to the orthogonality of the $V_{n,m}$, the scalar product of two PCETs basis functions can be expressed as:

$$\left\langle V_{pq}, V^*_{nm} \right\rangle = \begin{cases} \dfrac{\pi}{n+1} & \text{if } (p,q) = (n,m) \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

Here, the V^*_{nm} denotes the complex conjugate of the V_{nm}. Then we get:

$$\begin{aligned} \sum_{x}\sum_{y} \left|\sum_{n}\sum_{m} Z^{I}_{n,m} V_{n,m}(x,y)\right|^2 &= \sum_{x}\sum_{y} [(\sum_{n}\sum_{m} Z^{I}_{n,m} V_{n,m}(x,y)) (\sum_{n}\sum_{m} Z^{I}_{n,m} V_{n,m}(x,y))^*] \\ &= \sum_{x}\sum_{y} (\sum_{n}\sum_{m} Z^{I}_{n,m} Z^{I*}_{n,m} \left\langle V_{p,q}, V^*_{n,m} \right\rangle) \end{aligned}$$

$$\begin{aligned} &= \sum_n \sum_m (Z^I_{n,m} Z^{I*}_{n,m} \sum_x \sum_y \langle V_{p,q}, V^*_{n,m} \rangle) \\ &= \sum_n \sum_m |Z^I_{n,m}|^2 \frac{\pi}{n+1} \end{aligned} \tag{13}$$

$$\sum_x \sum_y \left| \sum_n \sum_m Z^{J\theta}_{n,m} V_{n,m}(x,y) \right|^2 = \sum_n \sum_m |Z^J_{n,m}|^2 \frac{\pi}{n+1} \tag{14}$$

$$\begin{aligned} &\sum_x \sum_y [\sum_n \sum_m Z^I_{n,m} V_{n,m}(x,y) (\sum_p \sum_q Z^J_{p,q} e^{jq\theta} V_{p,q}(x,y))^*] \\ &= \sum_x \sum_y [\sum_n \sum_m Z^I_{n,m} (Z^I_{n,m} e^{jm\theta})^* \langle V_{n,m}(x,y), V^*_{n,m}(x,y) \rangle] \\ &= \sum_n \sum_m [Z^I_{n,m} (Z^I_{n,m})^* e^{-jm\theta} \sum_x \sum_y \langle V_{n,m}(x,y), V^*_{n,m}(x,y) \rangle] \\ &= \sum_n \sum_m (Z^I_{n,m} (Z^I_{n,m})^* e^{-jm\theta} \frac{\pi}{n+1}) \end{aligned} \tag{15}$$

From the foregoing, we can obtain affine invariant PCETs coefficients. However, we do not need all the PCETs coefficients in image registration. Because the feature descriptors can normally be captured by just a few low-frequency coefficients, the number of coefficients does not need to be large. We use real part of PCET moments to reconstruct cross-correlation function approximately. For our application, the cross-correlation algorithm does not need to reach a high precision since this simple approximation is precise enough for our purpose.

$$Corr(I, J_\theta) \approx \frac{real(\sum_n \sum_m (Z^I_{n,m} (Z^I_{n,m})^* e^{-jm\theta} \frac{\pi}{n+1}))}{(\sum_n \sum_m |Z^I_{n,m}|^2 \frac{\pi}{n+1})(\sum_n \sum_m |Z^J_{n,m}|^2 \frac{\pi}{n+1})} \tag{16}$$

Where, $Z^I_{n,m}$, $Z^J_{n,m}$ represent PCET moments of images I and J, respectively. *real*() means real part. By replacing I and J_θ in Eq. (16) by their exact Zernike reconstruction (9), we obtain Eq. (16). We call it "cross correlation matching reconstruction using PCETs, CCMR-PCET". When two images are registration, the search for optimal distance will result in maximizing.

$$h(\theta) = \sum_m^n Q(m) \cos(m\theta + \varphi(m)), \theta \in [0, 2\pi) \tag{17}$$

It can be seen from Eq. (13), the maximal frequency of $h(\theta)$ is $0.5N/\pi$. In order to restrict the search of the global minimum, $[0, 2\pi]$ can been equally cut into $4N$ intervals. We then find the maxima with the gradient descent method by following the function (17) from $\theta = 0.5\pi i/N (i = 0, 1, \ldots, 4N - 1)$.

To reduce the influence of outliers in the accuracy of the solution, we use RANSAC (M. Fischler, 1981) which uses a distance threshold to find the best transformation matrix by using the greatest number of point pairs between two images instead of using the

entire set. The transformation parameters estimated from the subset which gives the least sum of squared error is then taken as the best fit.

3.2 Projective Invariant

The feature points detected by the Harris detector based on optimal derivative filters are determined up to sub-pixel accuracy. Due to noise, 3-D structures or moving objects in image sequences, some feature points displace when their positions are detected by optimal derivative filters, As a result, there are certain corresponding points remain more invariant than others.

There exist some image properties that remain invariant under projective transformation. For projective transformation, the most fundamental invariant is called the cross-ratio invariant. The cross-ratio can be defined for four collinear points or four concurrent lines, the four concurrent lines is most suitable to our problem as we already have concurrent lines of corresponding points in the image.

For the Delaunay Triangulation of Two-dimensional plane, triangles are mutually disjoint and not contain each other, thus, there exist at most three triangles that share an edge with one triangle, that is, four lines that intersect at one point. As shown in Fig. 2. We will use the cross-ratio invariant of four concurrent lines to choose the accurate correspondences.

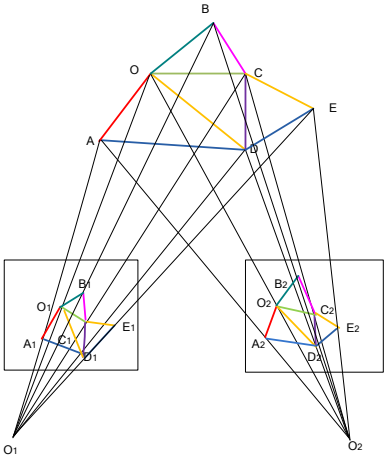


Fig. 2. The cross-ratio invariance of the four lines.

The cross-ratio invariance of the four lines in Fig. 2 is defined as:

$$I = \frac{\sin(OA, OC)\sin(OB, OD)}{\sin(OA, OD)\sin(OB, OC)}. \tag{18}$$

However, this is quite a complex way to compute a cross-ratio. Here we use a more elegant method by determinants. Let L_i, i = 1, …, 4 be any four lines intersecting in pointO, and A_i, i = 1, …, 4 be any four points respectively on these lines, then

$$I(x,y)=\frac{|OA_1A_3||OA_2A_4|}{|OA_1A_4||OA_2A_3|}. \tag{19}$$

Where $|OA_iA_j|$ denotes the determinant of the 3 × 3 matrix, whose columns are the homogeneous coordinate vectors of points O, A_i and A_j. To prove it, let (a, b, 1), (x, y, 1) and (u, v, 1) be the normalized affine coordinate vectors of O, A_i and A_j, respectively, then

$$\begin{aligned}|OA_iA_j| &= \begin{vmatrix} a & x & u \\ b & y & v \\ 1 & 1 & 1 \end{vmatrix} = \begin{vmatrix} a & x-a & u-a \\ b & y-b & v-b \\ 1 & 0 & 0 \end{vmatrix} \\ &= \overrightarrow{OA_i}\times\overrightarrow{OA_j} = |OA_i||OA_j|\cdot\sin(OA_i, OA_j).\end{aligned} \tag{20}$$

If the feature points (x, y) in one frame and the coordinating points (X, Y) are related by the projective transformation, then by replacing (x_i, y_i) with (X_i, Y_i) in Eqs. (18)-(20), we expect $I(x, y) = I(X, Y)$. If $I(x, y)$ and $I(X, Y)$ are not the same, then the smaller their distance

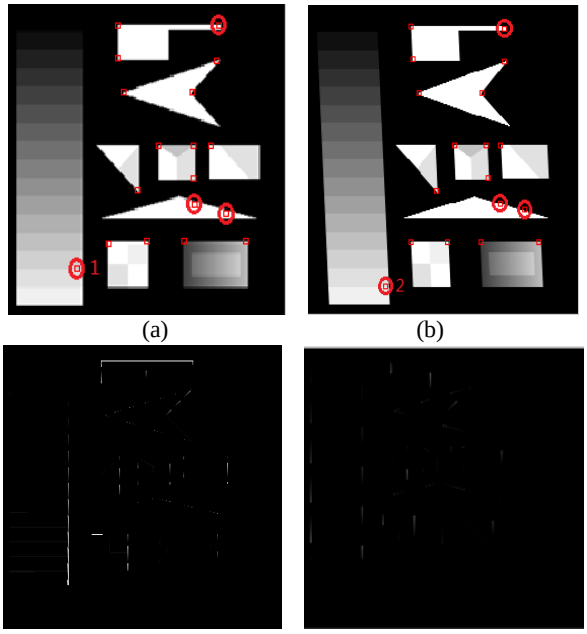
$$D=\sqrt{\left[I_1(x,y)-I_1(X,Y)\right]^2} \tag{21}$$

is the higher the accuracy of the four matching points will be. Then we can select the best combination if the combination giving the smallest distance, and we will select the best 4 corresponding feature points out of n using the projective Invariant in this method.

An example using the projective constraint in image registration is given in Fig. 3. As shown in Figs. 3 (a)-(b), we can see that the correspondence between some points are not accurate. For example, although the white points with label "1" and "2" approximately correspond to each other, the correspondence is obviously inaccurate matching points, and the local feature points will probably result in inaccurate transformation model estimation. The distance D in Eq. (21) is calculated for combination of 4 most accurate points that belong to the background, and the combination can produce the smallest distance. Figs. 3 (c) and (d) show absolute intensity difference of images registered using all the correspondences and using the best four correspondences obtained by the projective invariant, respectively. The difference between the two is significant. The registration result using projective constraint is more accurate than the registration result without the projective invariant.

4. EXPERIMENTAL RESULTS

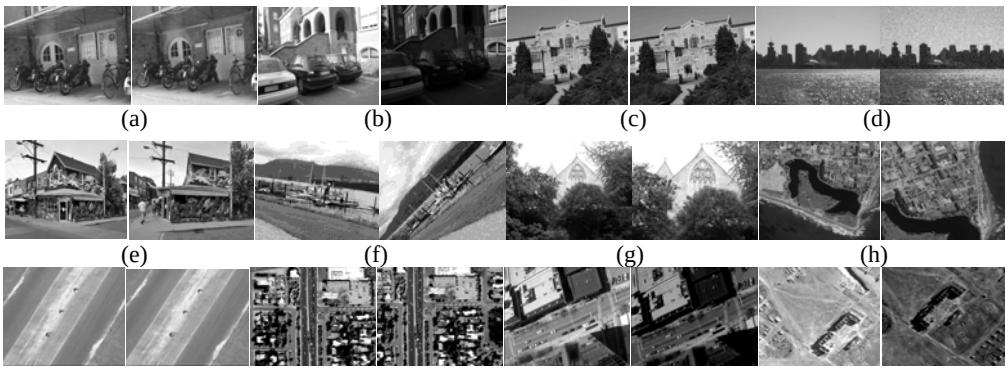
The algorithm has been implemented in C++ and all experiments have been carried out on an IBM Core 2 duo 2.4-GHz desktop computer with 2GB of RAM, Windows XP Professional Edition. Fig. 4 shows 12 sets of scene images that come from the website [24], and our aerial video data, with size 320 × 240, including rural roads, fields and urban buildings, *etc*.



(a) (b)

(c) (d)

Fig. 3. (a)-(b) Two images showing the corresponding feature points; (c) Registration result using all feature points; (d) Registration result using best five matching points.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Fig. 4. Test image pairs taken from the textured and structured scenes under photometric or geometric transformation; (a) Bikes(blur); (b) Leuven (lighting); (c) UBC (JPEG); (d) Sea (noise); (e) Graffito (viewpoint); (f) Boat (rotation); (g) Church (scale); (h)-(l) Aerial photography.

4.1 Feature Matching Performance

This section presents some examples and quantitative results of the proposed matching algorithm. Figs. 4 (a)-(g) show the representative test image pairs taken for the textured and structured scenes. The transformation types contain the common photometric

transformations and geometric transformations.

In regard to the feature descriptor types we evaluate the proposed IPCETs and four descriptors: PCET, SIFT, GLOH and Complex moments. In the beginning of the experiment, we need to choose a feature detector in order to extract the invariant regions of interest point from the given image. Here, we decide to choose Harris_laplace detector. Table 3 lists the descriptor dimensions of the five feature vectors in the experiments.

Table 3. The dimension of the five feature descriptor.

Feature descriptor	SIFT	GLOH	Complex moments	PCET	IPCET
Dimension	128	128	25	25	25

4.1.1 Evaluation criterion

We evaluate the performance of the feature matching using *recall vs.1-precision* graphs. *Recall* is the number of correctly matched regions with respect to the total number of corresponding regions. *Precision* is the number of correct matches to the total number of corresponding regions. The *correct-positive* is the match if the region pairs have the similar region overlap and correspond to the same location. A match is said to be *false-positives*, if the pairs come from different locations. The number of *correct-positives* and the *false-positives* is determined with overlap error [12], which is represented by the overlap ratio between the region intersection area and the union area of region $\varepsilon_e = 1 - (A \cap H^T BH)/(A \cup H^T BH)$, where A and B are two region pair and H is projective transformation between two regions. A match is correct if the error in the image area covered by region pair satisfied $\varepsilon_e < D_t$ for a given overlap error threshold D_t. We can determine the recall and 1-precision as follow:

$$recall = \frac{\text{number of correct - positives}}{\text{total number of positives}},$$

$$1 - precision = \frac{\text{number of false - positives}}{\text{total nmber of matches}}.$$

4.1.2 Performance evaluation

Fig. 5 is the feature region detection results using harris_laplace detector. The correct matches and recall values with different values of overlap error are shown in Figs. 5 (c) and (d), respectively.

The number of the *correct-positive* and total number of positives is computed for a range of overlap errors. For example, the score for 20% is computed for an overlap error from 10% to 20%.

The bold line shows the number of regions correspondences extracted with harris_laplace. We observed that most of the correct match regions are concentrate in the range of 10% and 40% overlap errors. We found that the bold line has a rebound at 30% overlap error. This is because that a descriptor may have several matches and several of them

may be correct. There will obtain more match pairs as the threshold increase. Usually, these new matching pairs are less similar to those at a smaller threshold.

As expected, the recall decreases with increasing overlap error. The recall for proposed method is slightly above the others for the region overlap error in the interval [0.1 0.5]. When the overlap error gets larger, the corresponding regions are less similar. This will result in error to estimate the homography. Therefore, we set the overlap error to 0.3 [12].

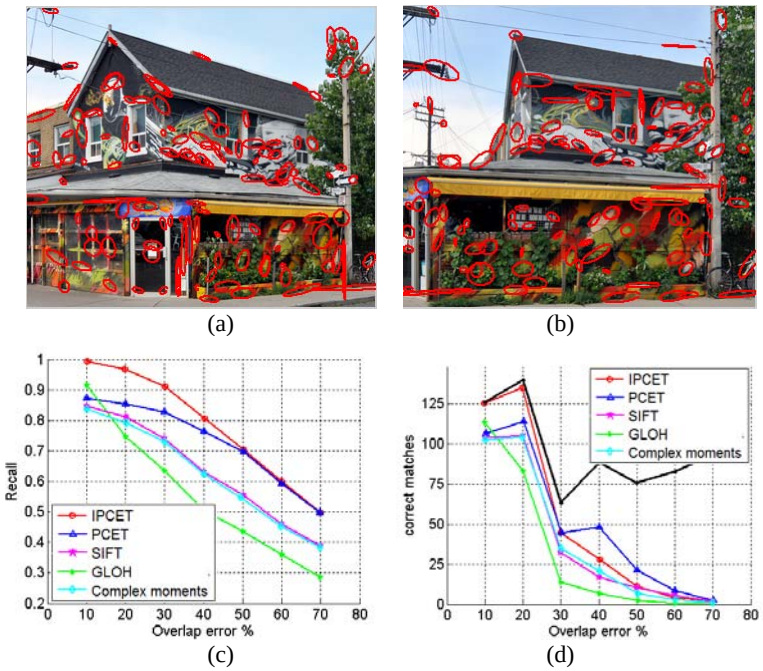


Fig. 5. Curves graph for different overlap errors; (a)-(b) Hessian-affine feature of structured graffti scene under viewpoint change; (c) Correct matches under different; (d) Recall under different overlap error.

After the feature detection, we present and discuss the experimental results of the feature matching evaluation. The performance is compared for image blur, JPEG compression, noise, viewpoints, scale changes, rotation, and illumination changes. We set the overlap error threshold to 40% and the normalized region size to a radius of 30 pixels. The results of these tests are shown in Figs. 6 and 7. A detailed discussion is given below.

(1) *Robustness under Photometric Transformations:*

(a) *Image blur*: The performance is measured with image blur introduced by taking at different lens focus settings. Fig. 6 (a) shows the results for the bike structured scene. The images are displayed in Fig. 4 (a). All the descriptors are computed on normalized image patches. The results show that all descriptors are affected by this type of image blur, IPCETs give the highest scores.

(b) *Illumination*: Fig. 6 (b) shows the results in the presence of illumination changes which are occurred due to imperfect camera calibration or variations of the camera position and direction. The image pair is displayed in Fig. 4 (b). The IPCETs performs best, followed by PCETs and SIFT.

(c) *JPEG*: In Fig. 6 (c), we evaluate the influence of JPEG compression for the UBC structured scene. The quality of the transformed image ranges from 5% to15% of the original one. The performance of IPCETs descriptor is better than the case of blur, and similar to that illumination. The performance increases when the 1-precision of all descriptors decrease.

(d) *Noise*: the performances are evaluated by adding different amount of Gaussian noise from Fig. 4 (d) is displayed in Fig. 6 (d). The best result is obtained with the IPCETs. PCETs obtains slightly better scores than for SIFT and GLOH based method. This is due to that the noise is localized in the high frequencies and has a small effect on the low-order PCETs.

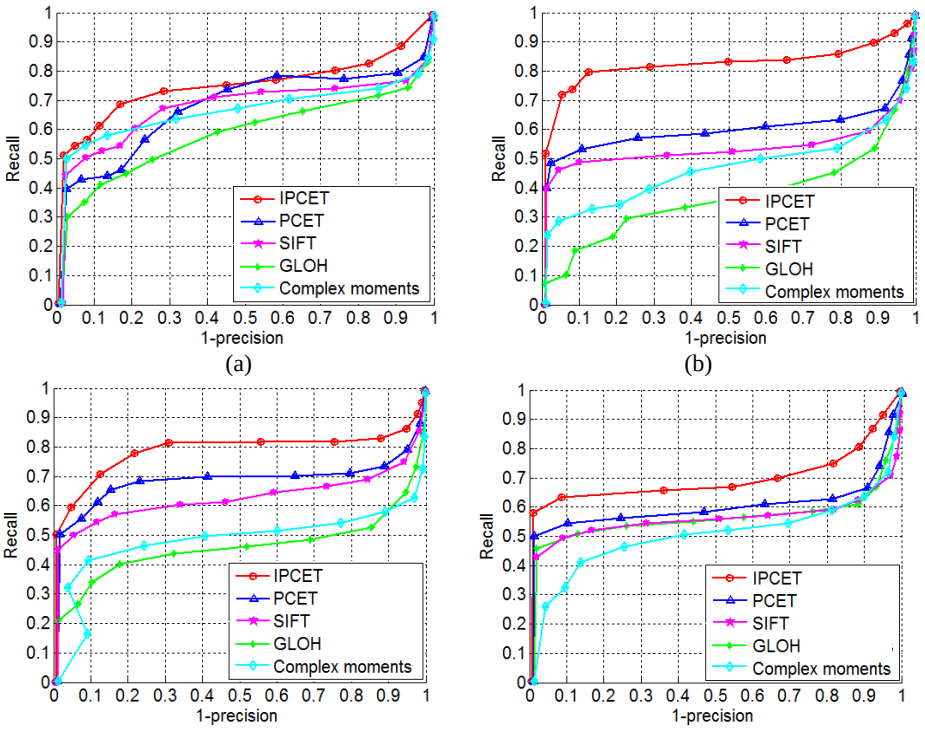


Fig. 6. Precision-recall curves performance evaluations under different photometric transformations, with overlap error threshold O_t = 0.3; (a) image blur; (b) illumination; (c) JPEG; (d) noise.

(2) *Robustness under Geometric Transformations:*

(a) *Viewpoints*: Fig. 7 (a) shows the performance of descriptors if the viewpoint of camera is ranged from 10° to 50°. To eliminate the effects of the affine transformation, we use the Harris-Affine detector which extracts affine-invariant regions.

The descriptors are computed on point neighborhoods normalized with the locally estimated affine transformations. The performance of all descriptors is lower than for other image transformations, *i.e.* scale changes and rotation. Fig. 7 (a) shows the result of IPCETs is significantly better than other descriptors, whereas SIFT obtains a score similar to PCETs.

(b) *Scale*: Fig. 7 (b) shows the performance measures for the descriptors using the church scene [Fig. 4 (g)]. Scale changes lie in the range 2-2.5. We can observe that the performance of all descriptors is better than in the case of viewpoint changes. The regions are more accurate since there is less parameter to estimate. IPCETs obtain the best matching score.

(c) *Rotation*: To evaluate the performance for image rotation, we used images with a rotation in the range from 30° to 45°. The recall is less than 1 because many matching regions are obtained accidentally. The results are better than for viewpoint and scale changes. IPCETs descriptor is more robust than the other ones. PCETs comes second.

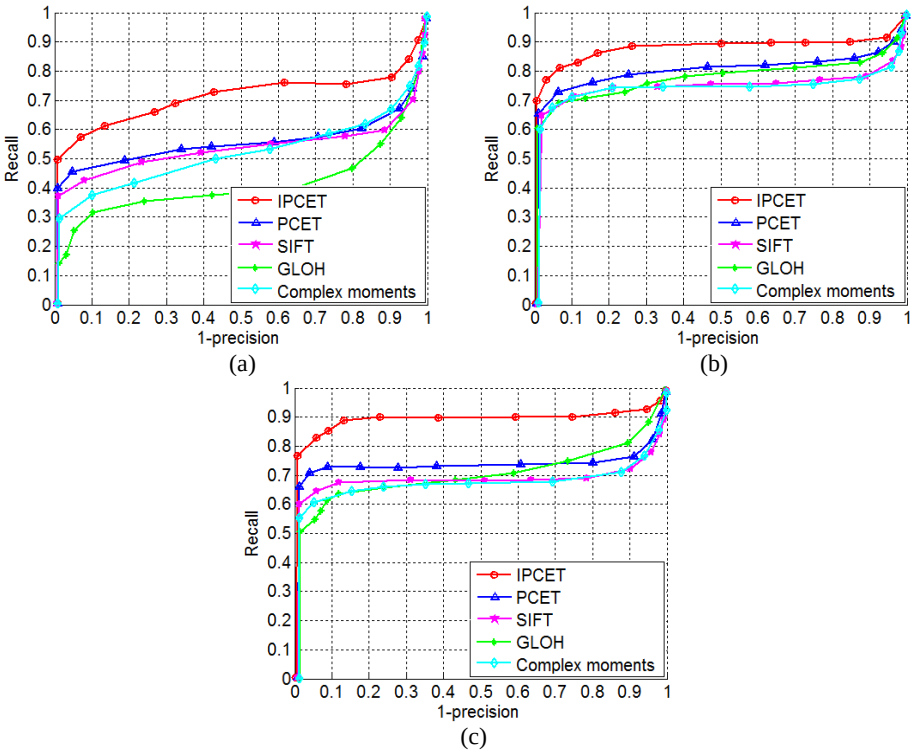


Fig. 7. precision-recall curves performance evaluations under different geometric transformations, with overlap error threshold O_t = 0.3; (a) Viewpoints; (b) Scale; (c) Rotation.

4.1.3 Matching example

Furthermore, we illustrate a matching example for images with a viewpoint change.

See Fig. 8. For the 43, 40, 17 nearest neighbor matches (displayed in blue line) and 4, 17, 31 false region matches (displayed in red line) obtained with the IPCETs, PCETs and complex moment descriptor, respectively. Fig. 8 (d) shows the absolute intensity differences of images using IPCET, Fig. 8 (e) shows the registration result using PCET, Fig. 8 (f) shows the registration result using complex moment. Root-mean-squared (RMS) difference between registered images when using complex moment is 12.856. When using PCET, the RMS difference between the images is 11.964, while RMS difference between images using IPCETs is 11.018. We can see that our algorithm produced more accurate registration results.

Table 4 presents the number of correct and false matches obtained with different descriptors. IPCETs obtains the highest correct number, a slightly lower score is obtained by PCETs. Complex moment achieves the lowest score. The number of correct matches varies from 43 to 17 and 59 to 21, respectively. There are approximately 2.5 times less correct matches for complex moment than for IPCETs. This clearly shows the advantage of IPCET-based descriptor.

Table 4. Matching example for the remote sensing image pairs.

Image	(Scale Angle)	Complex moment			PCET			IPCET		
		Correct	False	RMS	Correct	False	RMS	Correct	False	RMS
a	(1.3,15°)	17	31	11.53	40	17	7.94	43	4	5.45
b	(1, 45°)	21	19	7.15	58	16	5.84	59	3	5.12

4.1.4 Rotation angle estimation

The descriptor performance discrepancy can be attributed to the accuracy of the rotation angle estimation by the descriptors. The mean angle error alone is inadequate for assessing the accuracy of estimations because it does not contain any information about the variation of the estimation results. Therefore, we used the RMS error and overlap error as a criterion for comparison.

The RMS error is defined as:

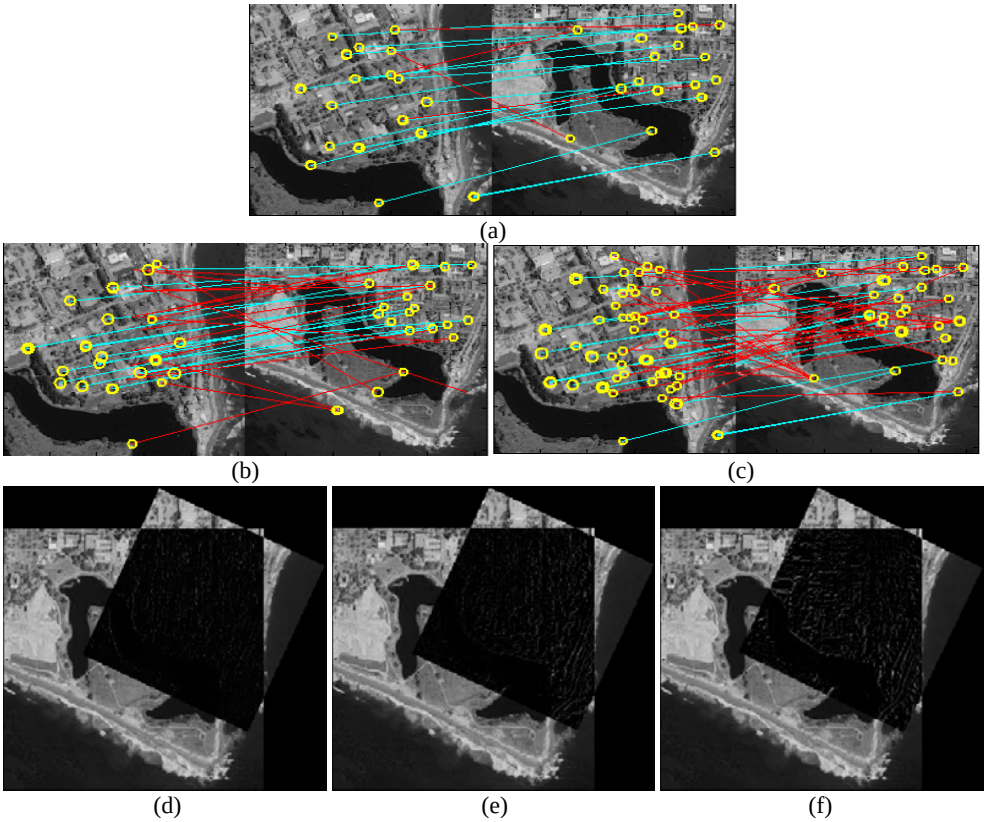
$$E_{RMS} = \sqrt{\frac{\sum_i (\theta_i - \theta_i^{'})^2}{Number\ of\ \ pairs}} \tag{22}$$

where θ_i and θ_i' are actual and the estimated rotation angle, respectively.

The overlap error is the represented as the overlap ratio between the numbers of estimated angles, with the errors that less than the value ε_e and the total number of image pairs

$$O = \frac{Number\ of\ \ pairs\ with \varepsilon < \varepsilon_e}{Number\ of\ total\ \ pairs}. \tag{23}$$

We have estimated the rotation angle with respect to the original pattern for the four



(a) (b) (c) (d) (e) (f)

Fig. 8. Correct matches (in green) and false matches (in red); (a) By IPCET; (b) by PCET; (c) by complex moment. Intensity differences of images after registered using; (d) IPCET; (e) PCET; (f) complex moment.

methods (SIFT, Shan and Moon-Chuen's method, Kim and Kim's method and proposed method) under the categories of rotation angle error $\varepsilon < 3°$, $\varepsilon < 6°$ and $\varepsilon < 9°$ under all transformations except viewpoint change. The rotation angle of the SIFT descriptor relies on dominant orientation of the gradient directions obtained from the interest region, while the moments based descriptor computes the image rotation angle via the phase difference. Here we have used the same number of moments: 25 moments were computed up to 8th order. Results are presented in Table 5.

We can observe from the Table 5 that all four methods show good performance for rotation and scaling images. More importantly, the overlap error for the proposed method is more than 89% while SIFT only has 40% to 89% overlap error when $\varepsilon < 9°$. The large rotation angle errors of SIFT are due to the error in determining the dominant orientation peaks.

From this graph, the overlap error for the proposed method is slightly worse than Kim's method for $\varepsilon < 9°$. However, Kim's precision rapidly decreases when the percentage of outliers increases. In contrast, the proposed IPCET method presents a lower overlap error but eliminates the outliers. This emphasizes the main advantage of IPCET

Table 5. RMS error and overlap error of rotation angle for all corresponding region pairs. 1:SIFT method, 2:Shan and Moon-Chuen's method, 3:Kim and Kim's method, 4: proposed method.

	Blur		Noise		JPEG		Rotation		Scaling	
	E_{RMS}	O	E_{RMS}	O	E_{RMS}	O	E_{RMS}	O	E_{RMS}	O
$\varepsilon < 3°$										
1	1.910	39.753%	1.884	42.867%	1.421	50.279%	1.336	54.738%	1.369	53.397%
2	1.445	79.823%	1.326	70.637%	0.923	93.616%	0.705	97.148%	0.704	97.452%
3	1.124	84.462%	1.084	92.901%	0.962	94.800%	0.682	99.674%	0.801	99.036%
4	0.998	89.735%	0.656	94.495%	0.626	96.611%	0.517	97.245%	0.497	97.657%
$\varepsilon < 6°$										
1	2.235	59.230%	2.191	61.140%	1.832	74.181%	1.643	81.973%	1.920	70.923%
2	1.740	89.873%	1.771	84.192%	1.135	96.173%	0.801	97.878%	0.776	97.910%
3	1.503	90.828%	1.326	95.671%	1.182	97.611%	0.681	99.875%	0.896	99.670%
4	1.137	91.917%	0.838	95.461%	0.721	96.447%	0.655	97.665%	0.497	97.657%
$\varepsilon < 9°$										
1	2.401	71.681%	2.272	80.466%	2.204	83.571%	2.012	89.566%	2.301	76.330%
2	1.846	91.233%	1.696	92.894%	1.187	96.533%	0.776	99.366%	0.857	97.910%
3	1.639	92.201%	1.489	96.512%	1.240	97.877%	0.681	100.00%	0.904	99.837%
4	1.137	91.917%	0.838	96.461%	0.721	97.447%	0.655	97.665%	0.497	97.857%

method: the algorithm removes more outliers. This property is important for many applications, such as estimating the transformation model.

4.2 Registration Performance

We illustrate the performance of the algorithm by using the IPCETs to determine corresponding points and then estimating the global transform using projective invariant as described in Section 3.2. We use a set of 4 aerial image pairs, and image pair contains noise, moving objects, 3-D structure building, or brightness changes. The results of the registration are show in Fig. 9. Each figure shows the matching results and the absolute intensity differences of images after registration. Notice that the difference image is not all black. This is because of the illumination changes between the two images. We can also find that high values show moving cars in the registration results of Fig. 9.

4.3 Time Complexity

The computational time of the proposed registration method is a function of image size, the number of region extraction, and the number of constructed descriptor feature vector. Given an $m \times n$ image, the computational time of the feature detection is on the order of nm. If M and N feature points are detected in two frames, the numbers of multiplications and additions required to compute IPCET moments up to order N are $O(P^2mn)$. The computational time of the cross-correlation matching method to find the correspondences is on the order of M^2N^2. If q matching points are found, the computational time of finding the best 4 matching points is on the order of q^4. We then transform the target

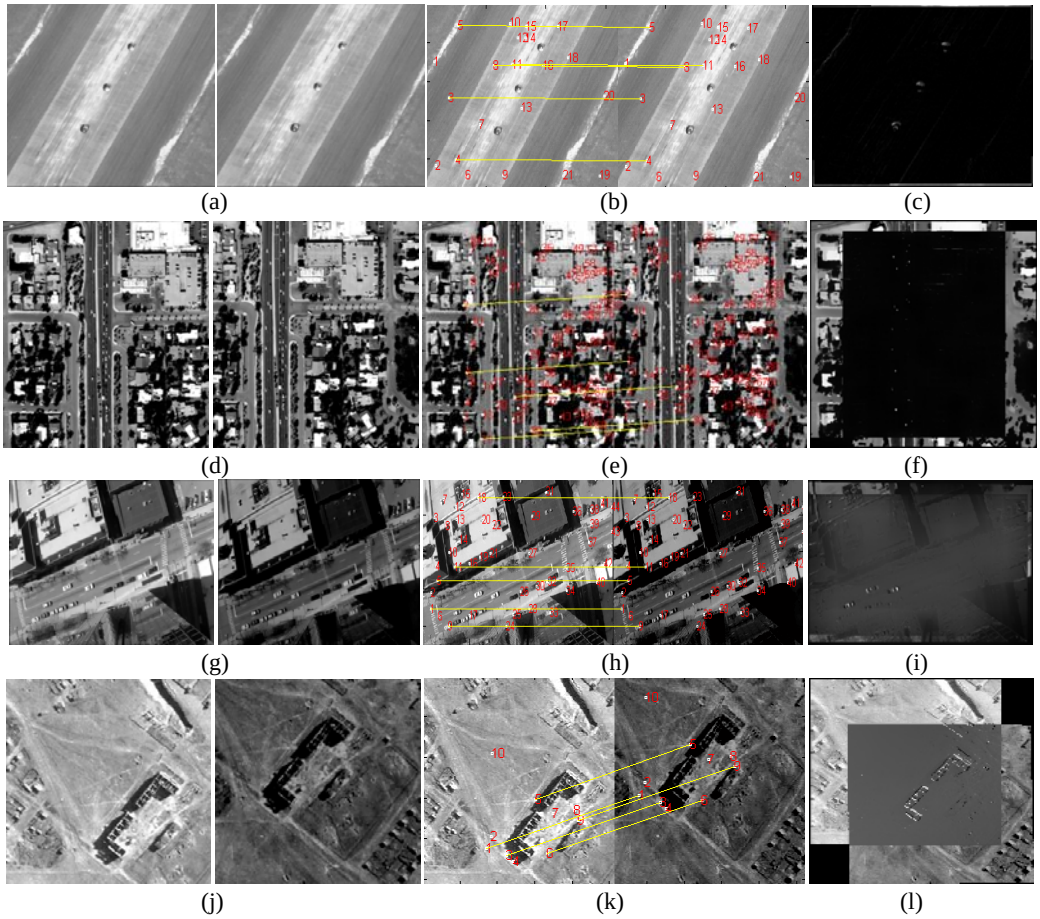


Fig. 9. Absolute differences registration results of planar background scene, complex urban scene and illumination change; (a) Two aerial images of planar background scene; (d) Two aerial images of complex urban scene; (g) and (j) Two aerial images of illumination change; (b), (e), (h) and (k) The matching points pair; (c), (f), (i) and (l) Registration.

image to the reference image using the best 4 matching points. Overall, the computational time of the proposed method is $O(nm) + O(P^2mn) + O(M^2N^2) + O(q^4)$. Theoretically speaking, the SIFT or SURF-based descriptor has a shorter feature matching time with respect to IPCETs descriptor. However, we can compute the distance between feature vectors using IPCET moments magnitude components firstly. If the magnitude-based distance satisfies the condition checking, the IPCETs descriptor needs further calculation of the cross-correlation difference to check if there is a rotation angle between two matching regions.

6. CONCLUSIONS

In this paper, we have developed a robust image registration algorithm that can be used for many of image stitching applications on mobile devices. Compared to com-

monly used moment-based, Filter-based and distribution-based descriptors, the IPCETs is free of numerical instability so that high order moments can be computed accurately. Therefore, IPCETs is more suitable for image registration. In this paper, the IPCETs is employed to design the registration codes and takes advantage of the phase information in the comparison process, which is robust to common photometric and geometric transformations. Moreover, it provides a method for estimation of the rotation angle between two matching regions that outperforms the robust estimator from Shan [20] and Kim [21]. The correspondence points were then used to estimate the parameters of a projective transformation to register images with impressive results.

REFERENCES

1. S. Gauglitz, T. Hollerer, P. Krahwinkler, and J. Rossmann, "A setup for evaluating detectors and descriptors for visual tracking," in *Proceedings of IEEE 8th International Symposium on Mixed and Augmented Reality*, 2009, pp. 185-186.
2. Y. S. Park, Y. I. Yun, and J. S. Choi, "A new shape descriptor using sliced image histogram for 3D model retrieval," *IEEE Transactions on Consumer Electronics*, Vol. 55, 2009, pp. 240-247.
3. M. Mayo and E. Zhang, "3D face recognition using multiview keypoint matching," in *Proceedings of International Conference on Advanced Video and Signal Based Surveillance*, 2009, pp. 290-295.
4. J. Liu and R. Hubbold, "Automatic camera calibration and scene reconstruction with scale-invariant features," in *Proceedings of IEEE International Symposium on Computer Vision*, Vol. 42, 2006, pp. 558-568.
5. B. Han and X. Lin, "A novel hybrid color registration algorithm for image stitching applications," *IEEE Transactions on Consumer Electronics*, Vol. 52, 2006, pp. 1129-1134.
6. J. Im, S. Lee, and J. Paik, "Improved elastic registration for removing ghost artifacts in high dynamic imaging," *IEEE Transactions on Consumer Electronics*, Vol. 57, 2011, pp. 932-935.
7. Z. Chen and S. K. Sun, "A zernike moment phase-based descriptor for local image representation and matching," *IEEE Transactions on image processing*, Vol. 19, 2010, pp. 205-219.
8. W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, 1991, pp. 891-906.
9. T. S. Lee, "Image representation using Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, 1996, pp. 959-976.
10. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, Vol. 15, 1997, pp. 415-434.
11. O. S. Kwon and Y. H. Ha, "Panoramic video using scale-invariant feature transform with embedded color-invariant values," *IEEE Transactions on Consumer Electronics*, Vol. 56, 2011, pp. 646-650.
12. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, pp. 1615-1630.

13. Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2004, pp. 506-513.
14. H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *Proceedings of the 9th European Conference on Computer Vision*, Vol. 3951, 2006, pp. 404-417.
15. S. Tabbone, L. Wendling, and J. P. Salmon, "A new shape descriptor defined on the radon transform," *Computer Vision and Image Understanding*, Vol. 102, 2006, pp. 42-51.
16. O. Pizarro and H. Singh, "Toward large-area mosaicing for underwater scientific applications," *IEEE Journal of Oceanic Engineering*, Vol. 28, 2003, pp. 651-672.
17. S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representant using local affine regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, 2005, pp. 1265-1278.
18. M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, 1962, pp. 179-187.
19. C. Y. Kim and O. J. Kwon, "A practical system for detecting obscene videos," *IEEE Transactions on Consumer Electronics*, Vol. 57, 2011, pp. 646-650.
20. S. Li, M. C. Lee, and C. M. Pun, "Complex zernike moments features for shape-based image retrieval," *Transactions on Systems, Man, and Cybernetics-part A: Systems and Humans*, Vol. 39, 2009, pp. 652-659.
21. W. Y. Kim and Y. S. Kim, "robust rotation angle estimator," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, 1999, pp. 768-773.
22. P. T. Yap, X. D. Jiang, and A. C. Chung, "Two-dimensional polar harmonic transforms for invariant image representation, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, 2010, pp. 345-351.
23. L. D. Li, S. S. Li, and G. H. Wang, "An evaluation on circularly orthogonal moments for image representation," in *Proceedings of International Conference on Information Science and technology*, 2011, pp. 394-397.
24. http://www.robots.ox.ac.uk/~vgg/research/affine.

Meng Yi (易盟) received the M.S. degree in Electrical Engineering from Northwestern Polytechnical University, Xi'an, China, in March 2008. Since 2009, he has been a Ph.D. of Electric Circuit and Systematic at Xidian University. Currently, he is currently a visiting doctoral candidate in Department of Electrical and Computer Engineering and Center for Automation Research, University of Maryland, College Park, Maryland, USA. His research interests include computer vision, pattern recognition, signal processing and biometrics.

Bao-Long Guo (郭宝龙) received the M.S. and Ph.D. degrees from Xidian University in 1988 and 1995, respectively, all in Communication and Electronic System. From 1998 to 1999, he was a visiting scientist at Doshisha University, Japan. He is currently a full Professor with the Institure of Intelligent Control and Image Engineering (ICIE) at Xidian University. His research interests include neural networks, pattern recognition, and image processing.

Chun-Man Yan (严春满) Ph.D. candidate in Circuits and Systems, with the Institute of Intelligent Control and Image Engineering, Xidian University. He received his M.S. degree from Lanzhou University in 2005. His research interests include MGA theory, image processing and pattern recognition.