http://ansinet.com/itj

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

ANSInet

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (7): 1345-1352, 2010
ISSN 1812-5638

# An Invariant Descriptor Design Method Based on MSER

Xu Sheng, Peng Qi-Cong and He Shan
UESTC-TI DSP Center, University of Electronic Science and Technology of China, 611731, China

**Abstract:** A novel approach of descriptor design for local appearance matching is presented. Because MSER region detector selects only the most stable regions which results in high repeatability, we choose it to detect image regions covariant to image transformation, which are then used as interest regions for computing descriptors. To get more distinctive feature vector to characterize the local image appearance, our descriptor consists of two main parts: (1) Affine Invariant Fourier Descriptor (AIFD) calculated based on MSER since AIFD possess invariant properties under translation, scaling, rotation and shearing and (2) color moments and GLCM based texture calculated from normalized patches. We deduce a fast patch normalization process which starts from a set of polygonal image regions output from MSER. Texture and color are the intrinsic properties of object and robust to affine transformation. We assessed our descriptor based on public image datasets which contains structured and textured scenes in different viewpoint as well as illumination change scenes. The experimental results showed that our descriptor obtained higher matching score comparing to several published descriptors.

**Key words:** Maximally stable extremal region, affine invariant fourier descriptors, color moments, gray level co-occurrence matrix

## INTRODUCTION

In a typical matching application, such as object recognition, image retrieval and object segmentation, the methods based on local appearance matching achieved many successes. Lots of approaches based on local features have been studied. Generally speaking, they have similar structures and steps (Mikolajczyk *et al*., 2005; Obdrzalek and Matas, 2006; Moreels and Perona, 2007; Tuytelaars and Mikolajczyk, 2008). First a set of interest image regions are detected serving as anchor locations. The number of obtained regions may be hundreds or thousands. This step is called as detector design. These regions are possibly overlapping, covariant to image transformation and can be repeatedly detected in images over large range variation. Once such elements of interest are found, the image appearance in their neighborhood can be encoded in an invariant way convenient for appearance similarity searching and recognition, which is know as descriptor design and is the focus of this study. Each descriptor is associated with a region and chosen to be invariant to viewpoint and illumination changes. Since, multiple regions are detected, the methods are robust to partial occlusions and cluttered background.

A number of researchers have designed and implemented methods for detecting affine covariant regions on images, including Harris-Affine detector (Mikolajczyk and Schmid, 2002), Hessian-Affine detector (Mikolajczyk and Schmid, 2004), the Maximally Stable Extremal Region detector (MSER) (Matas *et al*., 2004), Edge-Based Region (EBR) detector (Tuytelaars and van Gool, 2004), entropy-based region detector (salient regions) (Kadir *et al*., 2004). According to the comparison results for detectors' properties including repeatability, localization accuracy, robustness and efficiency, MSER has been demonstrated to be superior to other detectors in many aspects (Mikolajczyk *et al*., 2005; Tuytelaars and Mikolajczyk, 2008). This is also the reason we choose MSER detector as the base of the first step in our method.

Given MSER covariant regions, the following problem of how to design descriptors arises. Several techniques for describing local image regions have been researched. The state-of-the-art method SIFT (David, 2004) shows high detection and recognition accuracy in a general environment and is used for categorization and robot localization. However SIFT is designed for gray-scale images and as the number of objects to be recognized increases, the issue of scalability becomes more important (Kim and Kweon, 2008). Rahtu *et al*. (2005) used probability model to construct invariant features for images and proposed multi-scale auto-convolution transformation (MSA) method. The extracted features by MSA are robust to noise and affine deformation. But if

**Corresponding Author:** Xu Sheng, UESTC-TI DSP Center, University of Electronic Science and Technology of China, 611731, China

occlusion occurs in images that causes local information change, the values of MSA may change accordingly. Schaffalitzky and Zisserman (2002) proposed to use complex filters derived from the family $K(x,y,\theta)=f(x,y)e^{i\theta}$ where, $\theta$ is the orientation and f(x,y) can be a polynomial. These filters differ from the Gaussian derivatives by a linear coordinates in filter response domain. The shape context descriptor proposed by Belongie *et al.* (2002) is based on edges and comparable to SIFT. Edges are extracted with Canny filter, then their orientation and location are quantized into histograms in log-polar coordinates. Performance comparison and evaluation for several descriptors can be found by Mikolajczyk and Schmid (2005) and Burghouts and Geusebroek (2009).

In this study, we propose a new descriptor design approach for local appearance. Based on the output region of MSER detector, we get the most stable and nested regions which results in high repeatability. First we calculate the Affine Invariant Fourier Descriptors (AIFD) from the region polygon boundaries and then we extract the physical characteristics of objects including color and texture, that is to calculate the color moments and Gray Level Co-occurrence Matrix (GLCM) on the normalized image patches. Finally we combine the two kinds of feature vectors and get a more discriminative descriptor named CFCTD, acronym of Combination of Fourier, Color and Texture Descriptor. The MATLAB based evaluation result demonstrated the good performance for our descriptor.

## MSER BASED DETECTOR

The Maximally Stable Extremal Regions (MSER) are a watershed like segmentation introduced by Matas *et al.* (2004) which are defined by an extremal property of the intensity function in an image region and on its outer boundary.

A MSER region is a connected element of an appropriately thresholded image. The pixels inside the MSER have either higher (bright extremal) or lower (dark extremal) gray level than other pixels on its outer boundary. The set of all connected elements obtained through thresholding posses several desirable properties:

- Invariance to affine transformation of pixel values
- Since, only extremal regions whose support is virtually unchanged over a range of thresholds is selected, these regions are stable
- Since, no smoothing is involved, both very fine and very coarse elements can be detected, that is multi-scale detection
- Covariance with image continuous geometric transformations since pixels from a single connected element is transformed to another single connected element
- Given an image of n pixels, this set of extremal regions can be enumerated in O(nloglogn)

To implement MSER detection, the enumeration of extremal regions proceeds as follows: (1) Pixels are sorted by intensity. The computational complexity of this step is O(n) if the range of image values is small, typical in {0,...,255}, since, the sort can be implemented as BINSORT. (2) After sorting, pixels are marked in the image either in decreasing or increasing order and the list of the connected components and their areas is maintained in union-find manner with complexity O(nloglogn) that is almost linear and fast. (3) During the enumeration process, the area of each connected element is stored as an intensity function. Among the extremal regions, the maximally stable ones are the image parts where local binarization is stable over a large range of thresholds. The MSER stability definition is based on relative area change that is photometrical and geometrical invariant. Thus the MSER detection process is affine covariant.

The detection of MSER is related to threshold. Every extremal region is a connected element of a thresholded image. However, no global threshold is sought, all thresholds are tested and the stabilities of the connected elements are evaluated. The MSER detector output is not a binary image. Multiple stable thresholds exist for some parts of the image. In this case MSER detector may output systematically nested subsets. More details are given by Matas *et al.* (2004). Finally, we remark that the MSER described in this section and used in the experiments should be more precisely called intensity based MSER, since the different sets of extremal regions can be defined just by changing the ordering function.

An example of MSER is shown in Fig. 1a-c. Although, each view of the original object (the medicine Tylenol box) deformed greatly caused by view angles variation, a large number of similar regions were detected by MSER detector. Especially the edges of characters T, Y, L and O in the front of the box were shown clearly. And also a few characters on the top of the box could be distinguished. The MSER detector exhibits high repeatability and robustness which can be used as a useful foundation for object recognition and matching tasks.

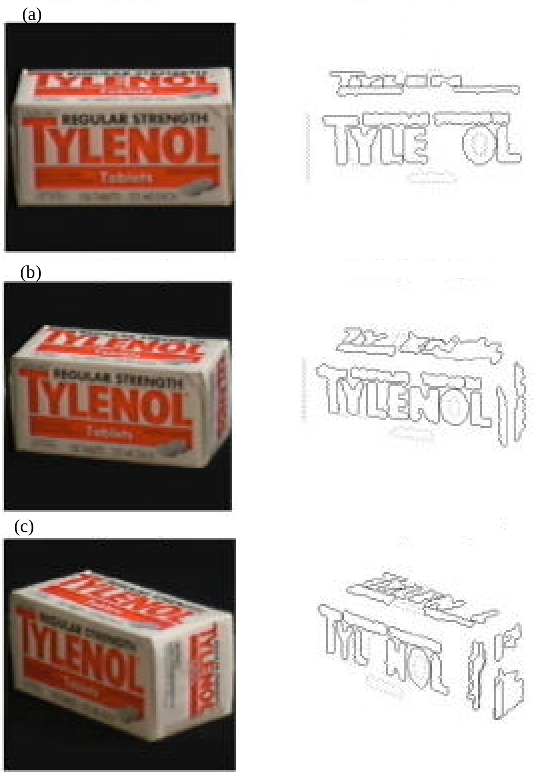

Fig. 1: MSER example: multiple views of original object are shown in left column; detected regions are shown in right column. (a) view angle = 0 degree, (b) view angle = 15 degree and (c) view angle = 45 degree

## THE IMPLEMENTATION OF PROPOSED CFCT DESCRIPTOR

Our CFCT Descriptor mainly integrates two kinds of information extracted from images, one is the MSER boundary information from which AIFD is calculated; the other is color and texture feature to represent the appearance of local normalized patches invariably.

**AIFD based on MSER:** In realistic application, due to the affection of distance, weather and camera visual angles, etc., we can observe 3D scenes showing different appearance in their 2D images, which generally can be approximated using the affine transformation. The affine transformation of an arbitrary curve $(x,y)^T$ in $R^2$ may be written as:

$$[x',y']^T = A[x,y]^T + B,\ \det(A) \neq 0 \tag{1}$$

where,

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},\ B = [b_1\ b_2]^T \text{ and } [x',y']^T$$

is the affine-transformed version. Affine Invariant Fourier descriptor (AIFD) is a kind of invariant under affine transformation and insensitive to shape distortion of objects. To guarantee the extracted features have invariant properties, naturally we consider using AIFD to avoid the inevitable distortion of object projection under different views.

To extract AIFD, first we should get the parameterization of object boundary and require (a) such parameterization is linear under affine transformation, (b) the parameterization function must yield the same parameterization independent of the affine parameters A, B and the initial point selection of the boundary. We can utilize the parameterization method given by Arbter *et al.* (1990) as:

$$t = \frac{1}{2}\oint_c \left|\det(p(\xi), p'(\xi))\right| d\xi = \frac{1}{2}\oint_c \left|x(\xi)y'(\xi) - y(\xi)x'(\xi)\right| d\xi$$

where, $p(\xi)$ is the boundary coordinates component $\{x(\xi), y(\xi)\}$ and $x(\xi)'$, $y(\xi)'$ is the first derivatives of $x(\xi)$, $y(\xi)$. If $B \neq 0$ in Eq. 1, we need to move the coordinate system to the area center of the boundary, defined by:

$$p_c = \frac{2}{3}\frac{\oint_c p(\xi)\det(p(\xi), p'(\xi))d\xi}{\oint_c \det(p(\xi), p'(\xi))d\xi}$$

Now the points on the boundary can be represented as $p = [u(t)\ v(t)]^T$ and then Fourier transform is applied to parameterized u(t) and v(t), respectively to get the Fourier series $F = [U_i, V_i]^T$, i = 0,1,2,...

Finally we can construct the affine invariable AIFD according the Fourier series F. Let,

$$I_k = \det\left[F_k \quad F_p^*\right],\ k \neq p$$

here, $F_k$, $F_p$ are two arbitrary Fourier series and fix p to a constant value, thus we get the affine invariable $Q_k$ which is independent to the affine parameters A, B and initial point selection, $Q_k = |I_k/I_p|$, k = 1,2,3.....

The outputs of MSER detector is a set of closed and might nested polygons which are suitable for calculating AIFD. For the extracted MSER polygon regions of the original Tylenol box and other two affine deformed one shown in Fig. 1, we calculate and compare the AIFD value $Q_k$ for character T and O separately. The results are shown in Fig. 2a and b. We can observe that the 3 AIFD curves of both characters in 3 view angles are similar with slight errors.

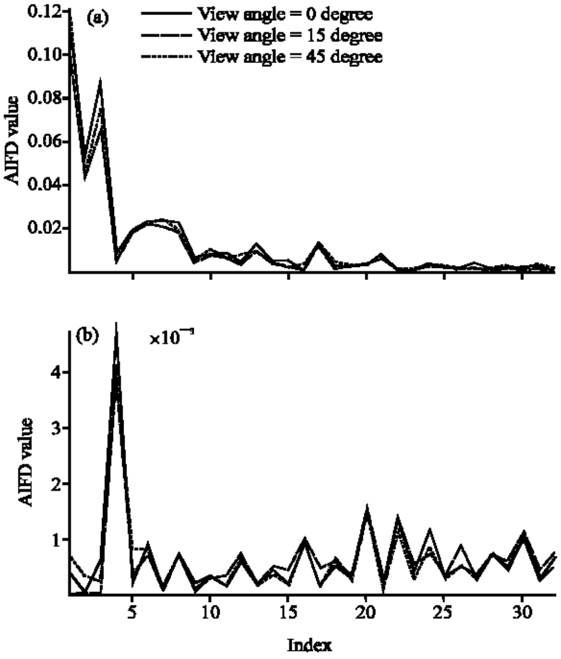

Fig. 2: MSER based AIFD value curves comparison in different views for character (a) T and (b) O in Fig. 1

**Color and texture computed on image patches:** The intrinsic physical information built in objects naturally is very useful for recognition. We propose to compute color moments and Gray Level Co-occurrence Matrix (GLCM) based texture features from images as another part of our CFCTD.

The MSER detector provides a set of arbitrary shaped regions of different size. To calculate the color moments and GLCM, we should derive a normalize area from the region first. To each polygonal region, we compute its several moments, including zero order algebraic moment (area), 1st order algebraic moments (center of gravity) and 2nd order central algebraic moments (covariance matrix). Based Green Formula (2), we deduce a fast method for the moments calculation.

$$\iint_{\Omega}\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)dxdy=\oint_{L}Pdx+Qdy \tag{2}$$

In Eq. 2, L is the polygon boundary of the region $\Omega$ in the clockwise direction. The area of $\Omega$ can be computed from Eq. 3.

$$|\Omega|=\mu_{00}=\iint_{\Omega}dxdy=\oint_{L}(ydx-xdy)/2 \tag{3}$$

For the closed polygon region with n vertexes $x_i$ and $y_i$, let, $\Delta x = x_{i+1}$-$x_i$ and $\Delta y = y_{i+1}$-$y_i$ , thus the integral of Eq. 3 can be replaced by summation:

$$|\Omega|=\mu_{00}=\sum_{i=0}^{n-1}\left(\int_{x_i}^{x_{i+1}}ydx-\int_{y_i}^{y_{i+1}}xdy\right)\Big/2 \tag{4}$$

Further let:

$$\begin{cases}x=x_i+t\Delta x\\ y=y_i+t\Delta y\end{cases},\ t\in(0,1)$$

we can rewrite Eq. 4 to Eq. 5:

$$\begin{aligned}|\Omega|=\mu_{00}&=\sum_{i=0}^{n-1}\left(\Delta x\int_0^1(y_i+t\Delta y)dt-\Delta y\int_0^1(x_i+t\Delta x)dt\right)\Big/2\\&=\sum_{i=0}^{n-1}\left(\Delta x(y_it+\frac{t^2}{2}\Delta y)\Big|_{t=0}^{1}-\Delta y(x_it+\frac{t^2}{2}\Delta x)\Big|_{t=0}^{1}\right)\Big/2\\&=\sum_{i=0}^{n-1}\left(\Delta x(y_i+\frac{1}{2}\Delta y)-\Delta y(x_i+\frac{1}{2}\Delta x)\right)\Big/2\\&=\sum_{i=0}^{n-1}(\Delta xy_i-\Delta yx_i)/2\end{aligned} \tag{5}$$

The centre of gravity of a closed polygon region is $\mu = (\mu_{10}, \mu_{01})^T$. Similar to region area computation, we arrive at Eq. 6.

$$\begin{aligned}\mu_{10}&=\iint_{\Omega}xdxdy=\oint_{L}(2xydx-x^2dy)/4\\&=\sum_{i=0}^{n-1}\left(\int_{x_i}^{x_{i+1}}2xydx-\int_{y_i}^{y_{i+1}}x^2dy\right)\Big/4\\&=\frac{1}{12}\sum_{i=0}^{n-1}\left(6\Delta xx_iy_i+3\Delta x^2y_i+\Delta x^2\Delta y-3\Delta yx_i^2\right)\end{aligned} \tag{6}$$

Also, $\mu_{01}$, the 2nd moments $\mu_{11}$, $\mu_{02}$ and $\mu_{20}$ can be computed in the similar way. Thus the 2nd algebraic moments can be combined and finally the covariance matrix can be constructed as Eq. 7.

$$\begin{cases}\mu'_{20}=\mu_{20}-\mu_{01}^2/\mu_{00}\\ \mu'_{02}=\mu_{02}-\mu_{10}^2/\mu_{00}\\ \mu'_{11}=\mu_{11}-\mu_{01}\mu_{10}/\mu_{00}\end{cases}\Rightarrow\ \Sigma=\begin{bmatrix}\mu'_{20}&\mu'_{11}\\ \mu'_{11}&\mu'_{02}\end{bmatrix} \tag{7}$$

A 2D affine transformation possesses 6 degrees of freedom that enough independent constrains should be applied. We adopt the method of Local Affine Frame (LAF) construction which can be found by Obdrzalek and Matas (2006). Once the covariance matrix is calculated, the region shape can be normalized using transformation:

$$T=\text{Cholesky}(\Sigma)/\sqrt{|\Sigma|}$$

Further the center of gravity of the region p, together with farthest point on the boundary q can determine (1) A de-skewed direction vector $u = T^{-1}(p\text{-}q)$, (2) the rotation angle $\phi = tg^{-1}(u_y/u_x)$ and (3) distance,

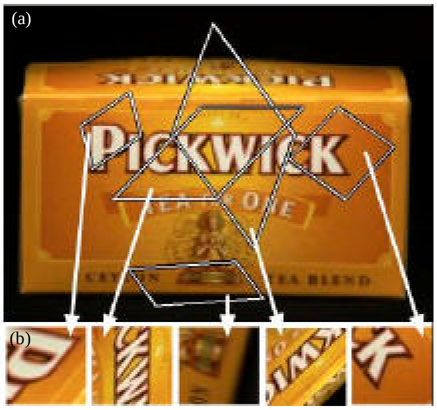

Fig. 3: (a, b) The normalized image patches using LAF construction

$$d = \sqrt{u_x^2 + u_y^2}$$

representing scaling. Finally, the transformation matrix A is constructed in the form:

$$A = \begin{bmatrix} d \cdot T\cos\phi & -d \cdot T\sin\phi & p_x \\ d \cdot T\sin\phi & d \cdot T\cos\phi & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

Thus, each pixel in the neighborhood determined by the LAF can be transformed into a square normalized region of 32×32 pixels by transformation matrix A. An example of image patches normalization process is shown in Fig. 3a and b. After normalization, these image patches possess good affine covariant properties. Based on these patches, now we can compute their color moments and GLCM texture features.

The color of objects is robust and insensitive to size and orientation of objects. To calculate color moments from the normalized patches, we use Eq. 8 to generate the 1st order moment η (the average color of an image), the 2nd central moment σ (image variance) and the 3rd central moment s (image skewness of each color channel) as proposed by Stricker and Orengo (1995) and used in 3D object recognition task (Sheng and Qi-Cong, 2009).

$$\eta = \frac{1}{|\Omega'|}\sum_i\sum_j p_{ij},\ \sigma = \left(\frac{1}{|\Omega'|}\sum_i\sum_j (p_{ij} - \mu)^2\right)^{1/2}, s = \left(\frac{1}{|\Omega'|}\sum_i\sum_j (p_{ij} - \mu)^3\right)^{1/3} \tag{8}$$

where, $|\Omega'|$ is the area of the normalized image patch, $p_{ij}$ is the patch pixel. The definition of Hue and Saturation components of HSI (Hue/Saturation/Intensity) color space are more close to the way of human being perceive color, thus Eq. is implemented in HSI space and outputs 9 moment values.

Another kind of import intrinsic information built in object is texture. Sometimes two objects share similar colors, however they also can be distinguished by their texture features in this case (Loum *et al.*, 2007). In this study, we propose to use the Gray Level Co-occurrence Matrix (GLCM) to estimate the 2nd order statistics of normalized patches. To reduce complexity, we compute only 4 statistical features including contrast, correlation, energy and homogeneity instead of 14 statistics extracted from GLCM as original suggested by Haralick *et al.* (1973). To use the spatial information of patches fully, GLCM are computed at 4 directions of horizontal, vertical and two diagonal (0°, 45°, 90° and 135°), that 16 texture features can be obtained in total.

## EXPERIMENTAL EVALUATION

The evaluation of performance of our CFCTD is conducted in this section and comparison with several published descriptors is provided also.

**Image data set:** Mikolajczyk studied the repeatability of various affine invariant detectors and created a publicly available database of images by Mikolajczyk *et al.* (2005), which depict planar objects, thus the homographies are known to determine the ground truth. The data set is available at www.robots.ox.ac.uk/~vgg/research/affine. We used a subset of this dataset shown in Fig. 4a-c. Each of the testing sequence contains 6 images with a gradual viewpoint or illumination transformation. The first row of scene named Graffiti in Fig. 4 contains homogeneous regions with distinctive edge boundaries. The second row wall contains repeated textures. These are referred to as structured versus textured scenes, respectively. The third row Light is used to test the descriptor performance under different lighting condition. All images are of medium resolution (approximately 800×640 pixels).

**Evaluation results:** The performance is compared for affine transformations and illumination changes. We calculate the matching score of our CFCTD and other descriptors as the measurement for comparison as suggested by Mikolajczyk *et al.* (2005).

The matching score is computed on the test data with a gradually increasing transformation between the reference image and others, shown in Fig. 4. The first column is used as reference image. Matching score evaluates how well does the descriptor represent a scene region, by comparing the number of corresponding

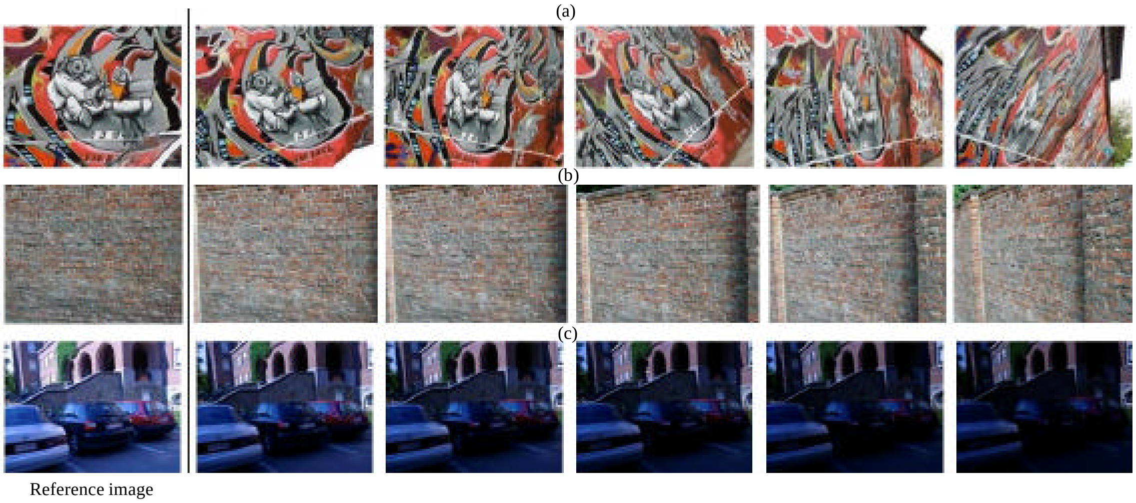

Fig. 4: Experimental image dataset with gradually viewpoint and illumination variation. (a) Graffiti, (b) wall and (c) light

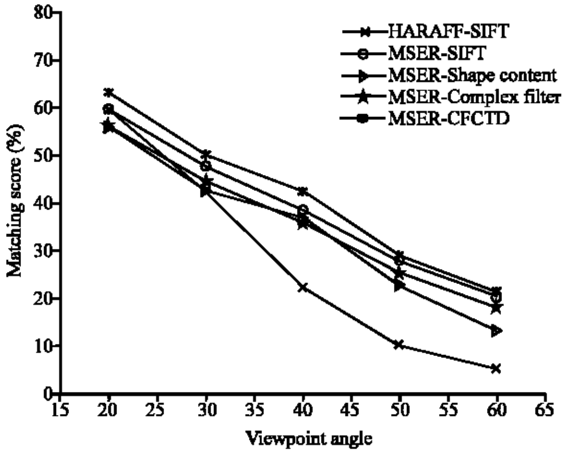

Fig. 5: Matching score comparison of descriptors, for the structured scene in viewpoint change condition

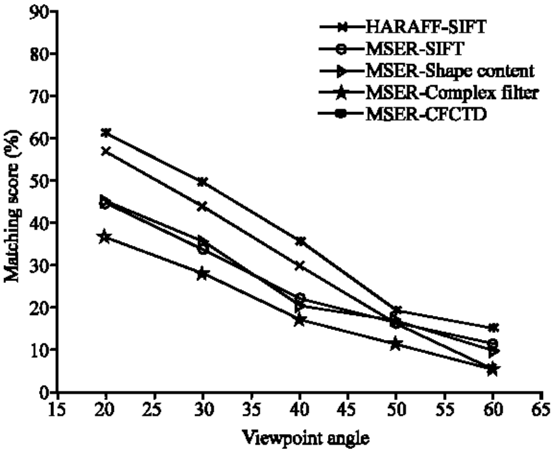

Fig. 6: Matching score comparison of descriptors, for the textured scene in viewpoint change condition

regions obtained with the ground truth and the number of correctly matched regions. The nearest neighbors in descriptors space are searched for match.

We compared our CFCTD to several published descriptors in the literatures, including SIFT (David, 2004), Shape content (Belongie *et al.*, 2002), Complex filter (Schaffalitzky and Zisserman, 2002). We used the Linux binaries provided publicly by related authors to compute matching scores. Because our CFCTD was computed from MSER and to make the experimental results more comparable, all other descriptors were computed based on MSER region, beside SIFT based on Harris-Affine detector (Mikolajczyk and Schmid, 2002). However, we should note that these descriptors can be computed on other regions detected by different detectors, thus the performance obtained by different descriptor/detector combinations may differ from the results in this study.

Figure 5 and 6 show the matching score of different descriptors measured across viewpoint angles. Note that there are also some scale and brightness changes in the test images. Each mark on a curve is the score representing between the reference image and another one with changed viewpoint. From Fig. 5 and 6, we can observe that the best results of matching score were obtained with our CFCTD descriptor for both scene types when view angles varied from 20° to 60° and the performance on Wall image set outperformed other descriptors with large margin under different view points due to the incorporated texture information. This result indicated the good affine invariant properties of our descriptor.

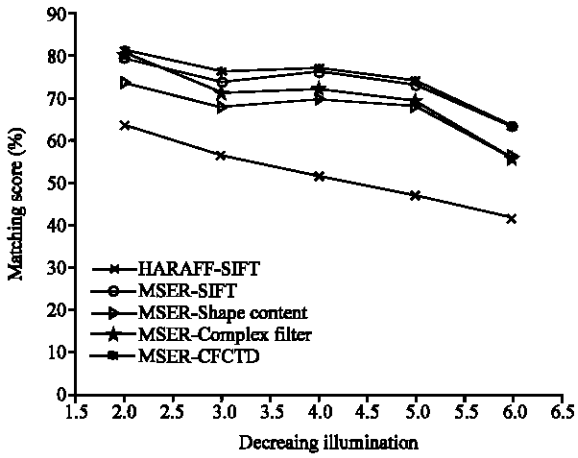

Fig. 7: Matching score comparison of descriptors, for the scene in lighting change condition

Again the measurement with the same scenario was carried out between the reference and other images in Light image set to evaluate descriptors performance under illumination change condition. The light changes are introduced by varying the camera aperture. Though the performance still decreased with the severity of the transformation, from Fig. 7 we can find that all curves were more horizontal than the curves in the Fig. 5 and 6, showing better robustness of illumination change. Although, our CFCTD obtained the highest matching score varying from 81.2 to 63.4% for this type of scene, our descriptor was only slightly superior to others.

## CONCLUSIONS

In this study a novel approach for descriptor design was introduced. The role of descriptor design is to well characterize the local image appearance around the location detected by the feature detector. We extract both the external boundary information of MSER regions to compute AIFD and use the color and texture of intrinsic properties of the normalized image patches. These synthetic information are rich and improve our matching task greatly. To normalize image patches from MSER regions, we deduce a fast procedure for moment computation to construct LAF.

The evaluation of performance of our CFCTD is performed in the context of object observed under different viewing conditions, including illumination changes and viewpoint changes from approximately 20 to 60 degrees. The comparison with other published descriptors used the same evaluation scenario and the same test data. The experimental results demonstrated the good performance of our descriptor.

As indicated in Fig. 7, the performance improvement for illumination change of our CFCTD is limit. We intend to add photometric normalization procedure and try other detector combination to enhance our CFCTD in future study.

## ACKNOWLEDGMENT

This study was supported by the 2009 Texas Instruments (TI) Innovation Funds.

## REFERENCES

Arbter, K. and W.E. Snyder, G. Hirzinger and H. Burhardt, 1990. Application of affine-invariant Fourier descriptors to recognition of 3-D objects. IEEE Trans. Pattern Anal., 12: 640-647.

Belongie, S., J. Malik and J. Puzicha, 2002. Shape matching and object recognition using shape contexts. IEEE Trans. Patt. Anal. Mach. Intell., 24: 509-522.

Burghouts, G.J. and J.M. Geusebroek, 2009. Performance evaluation of local colour invariants. Comput. Vis. Image Und., 113: 48-62.

David, G.L., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 60: 91-110.

Haralick, R.M., K. Shanmugam and I.H. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet., 3: 610-621.

Kadir, T., A. Zisserman and M. Brady, 2004. An Affine Invariant Salient Region Detector. In: Comptuter Vision (ECCV'04), Pajdla, T. and J. Matas (Eds.). Prague, Springer, Springer Berlin, Heidelberg, ISBN: 978-3-540-21984-2, pp: 228-241.

Kim, S. and I.S. Kweon, 2008. Scalable representation for 3D object recognition using feature sharing and view clustering. Pattern Recogn., 41: 754-773.

Loum, G., C. Theodore Haba, J. Lemoine and P. Provent, 2007. Texture characterisation and classification using full wavelet decomposition. J. Applied Sci., 7: 1566-1573.

Matas, J., O. Chum, M. Urban and T. Pajdla, 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image Vision Comput., 22: 761-767.

Mikolajczyk, K. and C. Schmid, 2002. An affine invariant interest point detector. Proceedings of the 7th European Conference on Computer Vision-Part I, LNCS 2350, May 28-31, Springer-Verlag, UK., pp: 128-142.

Mikolajczyk, K. and C. Schmid, 2004. Scale and affine invariant interest point detectors. Int. J. Comput. Vision, 60: 63-86.

Mikolajczyk, K. and C. Schmid, 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1615-1630.

Mikolajczyk, K., T. Tuytelaars, C. Schmid, A. Zisserman and J. Matas *et al*., 2005. A comparison of affine region detectors. Int. J. Comput. Vision, 65: 43-72.

Moreels, P. and P. Perona, 2007. Evaluation of features detectors and descriptors based on 3D objects. Int. J. Comput. Vision, 73: 263-284.

Obdrzalek, S. and J. Matas, 2006. Object Recognition Using Local Affine Frames on Maximally Stable Extremal Regions. In: Toward Category-Level Object Recogn, Ponce, J., M. Hebert, C. Schmid and A. Zisserman (Eds.). LNCS., 4170. Springer, Berlin, Heidelberg, ISBN: 978-3-540-68794-8, pp: 83-104.

Rahtu, E., M. Salo and J. Heikkila, 2005. Affine invariant pattern recognition using multiscale autoconvolution. IEEE Trans. Pattern Anal., 27: 908-918.

Schaffalitzky, F. and A. Zisserman, 2002. Multi-view Matching for Unordered Image Sets, or How Do I Organize My Holiday Snaps? In: Computer Vision (ECCV'02), Heyden, A. and G. Sparr (Eds.). Springer, Berlin, Heidelberg, ISBN: 978-3-540-43745-1, pp: 414-431.

Sheng, X. and P. Qi-Cong, 2009. A view-based approach to three dimensional object recognition. Inform. Technol. J., 8: 1189-1196.

Stricker, M.A. and M. Orengo, 1995. Similarity of color images. Proc. SPIE, 2420: 381-392.

Tuytelaars, T. and L. van Gool, 2004. Matching widely separated views based on affine invariant regions. Int. J. Comput. Vision, 59: 61-85.

Tuytelaars, T. and K. Mikolajczyk, 2008. Local invariant feature detectors: A survey. Foundat. Trends Comput. Graphics Vision, 3: 177-280.