Feature Based Registration for Panoramic Image Generation

Kawther Abbas Sallal[1], Abdul-Monem Saleh Rahma[2]

**[1] Computer Science Department, Babylon University,
Hila Babylon, 00964, Iraq**

**[2] Computer Science Department, University of Technology,
Baghdad, 00964, Iraq**

Abstract

The construction of large, high-resolution view is an active area of research in many fields like computer vision and computer graphics. This research is concerned with combining multiple images with overlapping region to produce a panoramic view by the detection of feature points for images with different camera motion in an efficient and fast way. Feature extraction is an important and critical step in panoramic view construction. This study presents techniques of corner detection, geometric moments and random sampling to accurately locate the important feature points in the images under noise, transformation, lighting, little viewpoint changes, blurring and compression circumstances. For each extracted feature in the image, a descriptor is calculated and the matching is done based on the homograph transformation. The results of experiments conducted on images taken by handheld camera and images taken from the internet. The results show that the proposed algorithm is fast and efficient compared with the panorama generation using the most famous SIFT method.

Keywords: *Feature extraction, motion estimation, registration, panoramic image.*

1. Introduction

Panoramic photography has received a growing interest. Among its numerous applications, such techniques may be used to efficiently represent video sequences, in terms of compression, enhancement, visualization. Recently have seen a growing interest in the mapping and visualization of the world's cities and sights. For example, systems such as Google Street View and Bing Maps enable users to browse street level imagery by presenting a panorama-based visualization of video captured at street level from a moving vehicle [1]. Constructing a full view panorama requires taking many regular images in order to cover the whole viewing space. Then, these images are aligned and composited into complete panoramic images using an image stitching algorithm. Image registration operation is very important to panorama generation.

Image Registration is the process of matching two or more images of the scene. This requires the estimation geometric transformations to align the images with respect to a common reference. Image registration is of great importance in all processing and analysis tasks based on the combination of data from sets of images. Image registration algorithms can be divided into two major categories: feature-based methods and area-based methods. Feature-based methods find relevant image features, known as control points, such as corners, point-like structures, line intersections, line ending points or high-curvature points that can be matched between two or more images. Once a sufficient number of points have been matched by correspondence on two images, a suitable geometric transformation can be computed and applied to align them. Area-based methods, also known as correlation based or template matching methods, work by finding correspondences between regions of the images without considering any features. Some of these algorithms are based on cross correlation in the spatial or frequency domain, or on mutual information. The correlation can be estimated locally, for example, for squared regions distributed over a regular lattice, or globally for the whole image. If two images are correlated, then the registration process continues by finding the parameters of a geometric transformation that maximizes cross correlation, and the images are aligned accordingly. The image registration process is one of the most complex and challenging problems of image analysis, where the extreme diversity of images and working scenarios make impossible for any registration algorithm to be suitable for all applications.

2. Related Works

Many advanced methods have been presented in recent years. [2] Proposed a method based on the scale invariance feature transform (i.e. SIFT) algorithm to stitch images captured by the video cameras together to form panoramic images. Based on the SIFT features and the KDT structure, the best pin first searching strategy is employed to match feature points. Then, in a post-processing pass, the author removes the mismatching feature points. Images captured by a surveillance camera are taken as the input to test the proposed method. A very popular approach is the panoramic image mosaic concept proposed by Szeliski [3] and its variants. With these approaches a rotation matrix is associated with each input image and to reduce registration errors a local as well as a global registration step is initiated. [4] Proposed the video serial images registration based on a feature based method algorithm. A matching method based on the improved algorithm can get better image fusion and image registration which is introduced for attaining more precise aggregate of matching points. In order to estimate the fundamental matrix which describes the whole geometry accurately and robustly, an improved SVD decomposition with weighted normalized fundamental matrix calculating method is proposed. In [5] the researchers using an algorithm based on Correlation. For Automatic Image Registration Applications, the features like edges are detected by using Sobel Edge Detection Algorithm. For matching the features, first Segmenting the image file in terms of different blocks and then applying the Hierarchical matching to create a pyramid of blocks. Finally, applying correlation based matching starting from the top level of the pyramid. Otherwise take a suitable pixel block size say about 32 x 32 pixel block from right image and search for the exact location of that 32 x 32 pixel block in the left image. [6] Proposes the feature based image mosaic approach. The mosaic image system includes features point detection, feature point descriptor extraction and matching. A RANSAC algorithm is applied to eliminate the number of mismatches and obtain a transformation matrix between the images. The input image is transformed with the correct mapping model for image stitching. In this paper, feature points are detected using steerable filters and Harris, and compared with traditional Harris, KLT, and FAST corner detectors. [7] Presents an approach for the panoramic view generation. First, salient features are robustly detected from the input images by a robust algorithm called Scale Invariant Feature Transform (SIFT). SIFT features are invariant to translation, rotation, image scaling and partially invariant to viewpoint, illumination changes and image noise. These features are matched between the successive images and hence image transformation is estimated. The image blending technique blends the images together to get a panoramic view without visible edge seam.

Our contribution is introducing a new direction for developing the used multimedia and devising new Media more arousing and attracting for viewers. A new method for invariant feature detection is introduced based on mixing local and global feature extraction methods. Some researchers worked on designing the panoramic image, but this work aims to develop an algorithm to produce the panorama in a simple and accurate way. The remainder of this paper is organized as follows. Section 2 describes the general system. Section 3 includes the methodology. Section 4 explains the steps for features detection and descriptors construction. Section 5 explains the feature matching operation and motion estimation. Section 6 discusses the experimental results and section 7 explains the conclusions.

3. Methodology

The panoramic image is constructed by alignment of multiple images with overlapping region. The alignment operation requires detection of similar regions in images. The similarities between images are determined by a new feature extraction method. The first step is capturing two or more images using handheld cameras. Then, the important features in each image are extracted using a local feature extraction method. Each extracted feature needs to calculate a descriptor that determines its invariability. The feature descriptor is calculated using a region feature extraction method. After that, the feature descriptors for both images are matched using a distance metric. The matching result is refined using second matching step based on transformation technology and the motion between the two images is estimated accordingly. The final step is using the estimated motion in the construction of the panoramic scene. The complete steps of the proposed system are shown in fig. 1 below.

A hand held camera with a 16 mega-pixel resolution is used. Since the images are taken by the user. To satisfy the overlapping requirement, the camera is fixed with some distance (d) and angle (Q) to obtain the overlapping between the captured images.

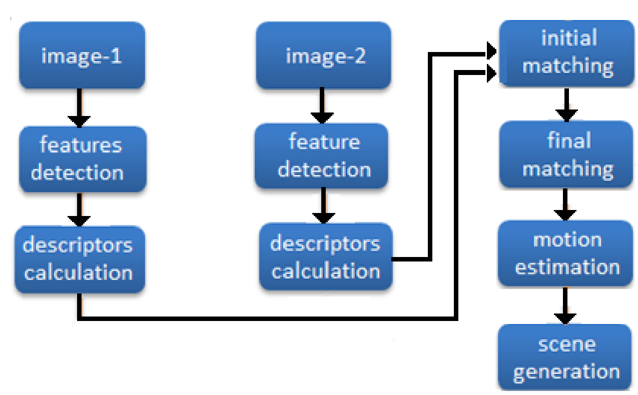


Fig.1 The general system

4. Features Detection and Descriptors Construction

4.1 Corner Points Detection Using Harris Detector

The 'corner' is a location in the image where the local autocorrelation function has a distinct peak. Corner point detection has found its application in various computer vision tasks. In this work, the Harris corner detector is proposed to extract corner information as the first step of the proposed algorithm. The Harris corner detector was proposed by [8]. Harris Corner Detector is one of the promising tools to analyze the corner points. This method not only solved the problem of the discrete shifts, but also it deals with the issue of directions with the advantage of the autocorrelation function and increased the accuracy of localization. Feature point extract by the Harris vertex arithmetic operator has a rotation and translation invariability and has a good robustness against noise and change of parameters during acquisition of data. Harris detector is based on the autocorrelation function or the image gradient [9]:

$$M = \exp^{-\frac{x^2+y^2}{2}} \otimes \begin{pmatrix} I_X^2 & I_X I_Y \\ I_X I_Y & I_Y^2 \end{pmatrix} \qquad (1)$$

Where $I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}$

Where, I_x and I_y denote the image gradients in the x and y directions. The feature point corresponds to the highest singular values, and it can be computed using the criterion:

$$h = DET(M) - k.\ Tr(M)^2 \qquad (2)$$

$k \in [0.04, 0.06]$

A 'corner' is said to occur when the two eigenvalues are large. Based on h, the pixels are classified as Corner pixel if $h > 0$, flat region pixel if $h \sim 0$ and edge pixel if $h < 0$. Then, set all h (x, y) below a threshold T to zero. The last step in corner detection is performing the non-maximal suppression to find local maxima. The flowchart of corner detection algorithm and results are shown in fig. 2.

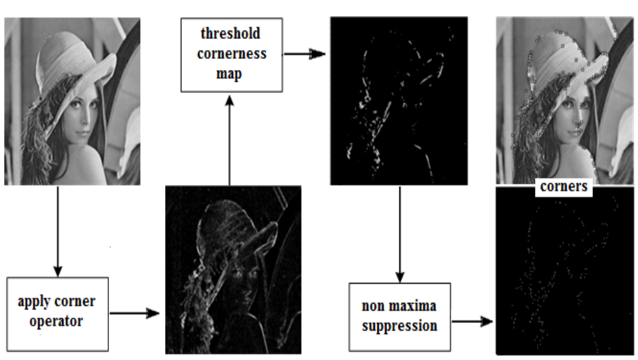


Fig.2 The steps of the corner point detector

4.2 Feature Descriptor Construction Using Moments

After detecting all image features using corner detection method, it is necessary to identify each feature point. Therefore, an efficient and simple description will be generated for each detected feature point. For each feature, a 25*25 circular window centered on this feature is determined. Then, the window is divided to 5*5 blocks. For each block the average of geometrical moments is calculated. The result is a descriptor of 5*5 matrix for each feature point.

Moment invariants are the most popular and widely used shape descriptors in computer vision derived by Hu. A 2-D continuous function $f(x, y)$ of the order $(p+q)$ is defined as [10]:

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy \qquad (3)$$

For p, q = 0, 1, 2…

The uniqueness theorem states that if $f(x, y)$ is piecewise continuous and has non zero values only in a finite part of x-y plane, moments of all orders exist and the moment sequence (m_{pq}) is uniquely determined by $f(x, y)$. Conversely, (m_{pq}) uniquely determines $f(x, y)$. The central moments can be expressed as [10][11][12]:

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)\,dx\,dy \qquad (4)$$

Where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

For a digital image, Equation (3.2) becomes

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \qquad (5)$$

$$\mu_{00} = m_{00}\ , \quad \mu_{10} = 0\ , \quad \mu_{01} = 0$$

The normalized central moments, denoted η_{pq}, are defined as

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad (6)$$

Where $\gamma = \frac{p+q}{2} + 1$

A set of seven invariant moments can be derived from the second and third moments:

$$\phi_1 = \eta_{20} + \eta_{02} \qquad (7)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \qquad (8)$$

IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (9)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (10)$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (11)$$

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

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (13)$$

The seven invariant moments, which are invariant to translation, scaling, mirroring and rotation, composed of the linear combination of the second-order and third-order central moments. Because of the seven moment invariants is relatively large, and to simplify comparison, making use of logarithmic methods. At the same time, taking into account the possible negative moment invariants situation, you have to take the absolute value before getting logarithm.

5. Feature Matching and Motion Estimation from Correspondences

Images can be in different situations or transformations that can be resulted during camera acquiring or by the applying image transformation techniques. According the complexity, these transformations are rigid, affine, piecewise affine and non-rigid. Rigid registration models are linear and only allow for translation, rotation and scale changes without any distortion. Affine transform is also linear and support overall distortions besides shearing and stretching. Piecewise affine and non-rigid models are nonlinear and allow for arbitrary local and global distortions. Like these transformations may effect on matching operation. Therefore, it is important to find a way to determine the true matches. The true matches can be determined if the points fit with a predefined model. The matching is done by computing the Euclidian distance between two descriptors depending on the second nearest neighbor technique as follows:

DES = D (i, j) / D (i, a), (14)
If DES < T then (i and j) is matched

Where D (i, j) is the distance between point (i) in the first image and point (j) in the second image and (a) is the second nearest neighbor point. The points (i) and (j) is matched if the value of DES is lower than the predefined threshold. The mismatched points can be regarded as outliers, which are the data that do not conform to the model. These outliers can rigorously disturb the estimated homograph, and consequently should be identified. Therefore, it's important to determine a set of inliers, which are the data whose distribution can be explained by some set of model parameters from the presented correspondences so that the homograph can be estimated in an optimal manner. Homograph is a concept in the mathematical science of geometry. A homograph is an invertible transformation of a projective plane to a projective plane that maps straight lines to straight lines. In the computer vision field, any two images of the same planar surfaces are related by a homograph. This is very important in computing the camera movement, like rotation and translation and other transformation between two images. In mathematical definition the homogeneous coordinates are used, because matrix multiplication cannot be used directly to perform the division required by the perspective projection.

$$x = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad x' = \begin{bmatrix} w'x \\ w'y \\ w' \end{bmatrix} \quad H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

Then: $x' = H.\ x$ (15)

The affine homograph is a special type of general homograph whose last row is defined as:

$h_{31} = h_{32} = 0$, $h_{33} = 1$.

The random sampling algorithm [13] is suggested in this work to be applied on the putative correspondences to determine the inliers. This algorithm was first introduced by Fischler and Bolles as a method to estimate the model's parameters in the presence of large amounts of outliers. It has been widely used in the computer vision and image processing for many different purposes. This algorithm includes two steps that are repeated in an iterative fashion. First, a set of points is randomly selected from the input dataset and the model parameters are computed using only the elements of this set as opposed to least squares, where the parameters are estimated using all the data available. In the second step the algorithm checks which elements of the full data set are consistent with the model instantiated with the parameters estimated in the first step. The set of such elements is called consensus set [14]. The algorithm ends when the probability of finding a better consensus set is below a certain threshold. In this work four points are randomly selected from the set of candidate matches to compute homograph. Then, select all the pairs which agree to the homograph. A pair (x; x'), is considered to agree to a homograph H, if Dist (H. x; x') <ε, for some threshold ε and Dist is the Euclidean distance between two points. The third step is repeating steps (1) and (2) until enough pairs are consistent with the computed homograph. The results from this step are the group of final matching features and the motion matrix that the second image is wrapped accordingly to generate the panoramic view.

6. Experimental Results and Discussions

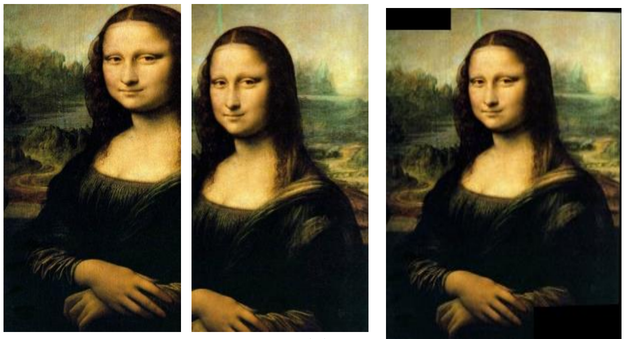
To evaluate the performance of the proposed algorithm, the experiments done on images taken by handheld camera and images taken from the internet in different situations and in comparison to the states of arts SIFT algorithm. For each image, the corners are extracted as shown in court image in fig. 3. The value of the used threshold must be sufficient to extract enough number of corners. After extracting the corner points in the two images, the descriptors for them are created by taking a 25*25 circular window around each point, divide the window to blocks of 5*5 each and finding the average value of moment invariants for them. Then, a very efficient algorithm is used for matching done by finding the second nearest neighbor. The very important factor here is the matching threshold because it must be chosen to get as more matching points as possible. The next step is eliminating the mismatch points or the outliers to get on only the true matches that represent the key feature points by using a random sampling algorithm. The number of iterations is an important factor here to obtain more matches between the images. The result of this step is shown in fig. 4. Finally, the panoramic scene is constructed according to the estimated motion from the previous step. The constructed panoramic view is explained in fig. 5. Other executions for the proposed method shown in fig. 6.

Fig. 3 Feature extraction by corner detector

Fig.4 The matching features between the two images

Fig. 5 The generated panoramic view

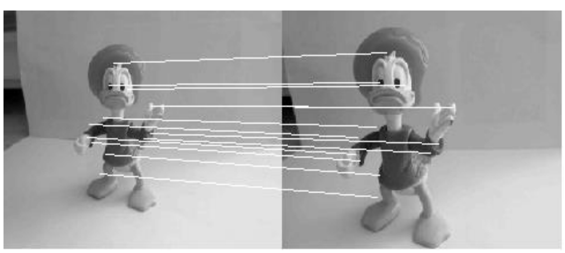
(a)



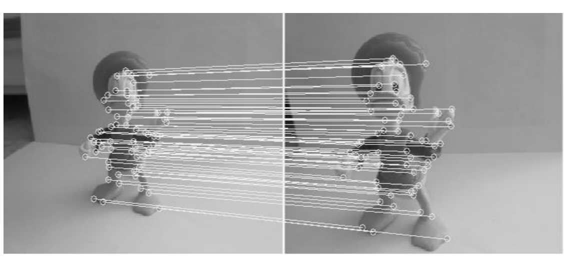
(b)

Fig. 6 Implementations of the proposed method

In comparison with the famous SIFT method; sift method is more complex than the proposed method. The proposed algorithm is accurate, sometimes more than SIFT method because in some cases SIFT method gives less number of matched feature points as explained in fig. 7. The number of extracted feature points in the proposed method is 63 while in SIFT method is 12. The accuracy of the proposed algorithm is coming from the using of the random sampling algorithm because it follows the object's motion to estimate the matched feature points under different circumstances like transformations ,lighten, noise, view point changes and compression.



(a)



(b)

Fig.7. The extracted feature correspondences. (a) The SIFT method (b) The proposed method

Fig. 8 and Fig. 9 explain the comparison in the no. of extracted features and the no. of matches after applying SIFT method and the proposed image on court image. Some of the transformation techniques are applied on one of the two images used for panoramic view generation to see the difference. As shown in the figures, and in experiments the proposed method gives us an efficient performance under different circumstances. The number of corners increases with the decrease in corner threshold. The corner threshold value used in experiments is 0.1.

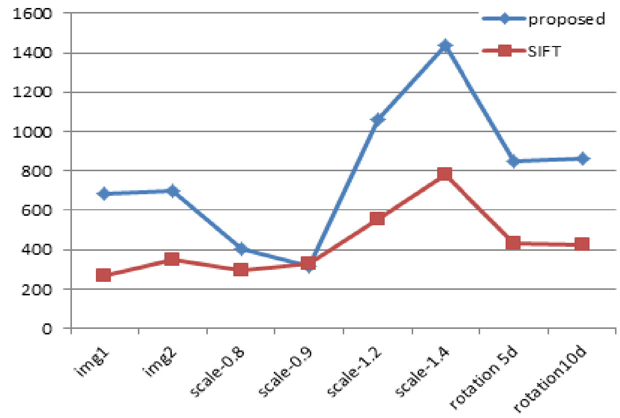


Fig.8 The No. of the extracted features using SIFT and the proposed method

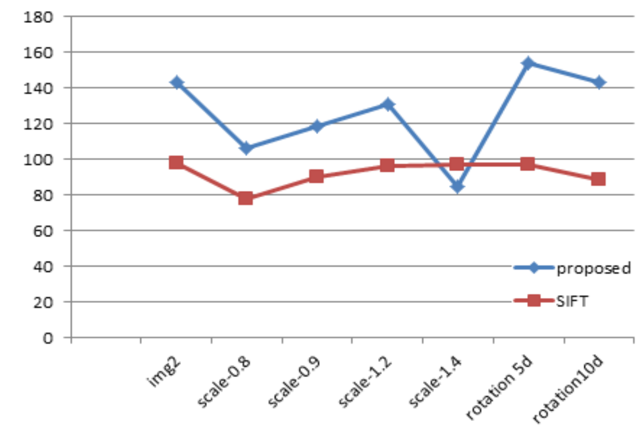


Fig.9 The no. of matched features using SIFT and the proposed method

7. Conclusions

Panoramic view construction is an important topic because it needs a fusion of image processing, graphics and computer vision techniques. Many researchers deal with generating panoramic image depending on the SIFT method but in this work, a panoramic image using a proposed feature extraction method is generated. The extracted features and constructed descriptors in the overlapping region of the images are based on corner points, geometrical moments and random sampling for feature matching and motion estimation from correspondences. The mixing of local feature extraction method represented by corner detector and region feature extraction method represented by geometrical moments and random sampling for feature filtering and motion estimation gives us a fast, efficient, accurate method and less complexity than the SIFT method.

References

[1] E. Zheng, R. Raguram, and P. F. George, "Efficient generation of multi-perspective panoramas", IEEE, International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 2011.

[2] L. Lin and G. Nan, "Algorithm for Sequence Image Automatic Mosaic based on SIFT Feature", WASE International Conference on Information Engineering, 2010.

[3] R. Szeliski, "Video mosaics for virtual environments", IEEE Computer Graphics and Applications, 1996, pp: 22–30.

[4] W. Bing and W. Xiaoli, "Video Serial Images Registration Based on FBM Algorithm", Research Journal of Applied Sciences, Engineering and Technology 5(17), 2013.

[5] R. Asha, K. Asha and M. Manjunath, "Image Mosaicing and Registration", International Journal of Computer Science Issues, Vol. 10, Issue 2, No 1, March 2013.

[6] M .V. Subramanyam and Mahesh, "Feature Based Image Mosaic Using Steerable Filters and Harris Corner Detector", I.J. Image, Graphics and Signal Processing, 6, 2013, pp:9-15.

[7] S. Prajkta, K. Krishnan, and P. Anita, "A Novel Approach for Generation of Panoramic View", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2) , 2011.

[8] C. Harris and M. Stephens, "A combined corner and edge Detector", In Alvey Vision Conference. 1988, Pp. 147–151.

[9] Y. Wang, Y. Chen, J. Li and B. Li, " The Harris Corner Detection Method Based on Three Scale Invariance Spaces", International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November 2012.

[10] M.K. Hu, "Visual pattern recognition by moments Invariants", IRE Trans. Information Theory, Vol 8. 1962, Pp.: 179- 87.

[11] R. Mohamed, Y. Haniza, , S. Puteh, Y.Ali, , S. Abdul R., R. Mohd, Y. Sazali, D. Hazri and M. Karthigayan, "Object Detection using Geometric Invariant Moment", American Journal of Applied Sciences. Vol 3 no.6. 2006, Pp.: 1876-1878.

[12] J. Bei and L.Chen, "Map Matching Algorithms Based on Invariant Moments", Journal of Computational Information Systems Vol. 7 (16). 2011, Pp.: 5668-5673.

[13] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", Communications of the ACM, vol. 24. No. 6. 1981,Pp.: 381–395.

[14] M. Zuliani, "Computational Methods for Automatic Image Registration", PHD thesis. University of California, USA,2006.

Kawther Abbas Sallal received the B.S. and M.Sc. in computer science from University of Technology, Baghdad, Iraq in 2003 and 2006 respectively. She is now preparing for PHD degree in computer science in Babylon University, Babylon, Iraq. Her interests include image processing, computer vision, video processing, computer graphics, biometrics and data security.

Abdul-Monem Saleh Rahma is currently a professor and head of the computer sciences Department, University of Technology, Baghdad, Iraq. He received his M.Sc. degree from Brunel University, UK in 1982 and his PHD degree from Loughborough University of Technology, UK in 1985. He published more than 100 papers in computer science and supervised on more than 24 PHD and 57 M.Sc. students. His research interests focus on image processing, video processing, computer graphics, biometrics and computer security.