Invariants Feature Points Detection based on Random Sample Estimation

Kawther Abbas Sallal
Ph.D Student
Computer Science Dept.
Babylon University

Abdul-Monem Saleh Rahma
Professor
Computer Science Dept.
University of Technology

ABSTRACT
Feature detection is the initial step in any image analysis procedure and is essential for the performance of computer vision applications like stereo vision, object recognition, object tracking systems. Research concerning the detection of feature points for different camera motion images in efficient and fast way. In this work, techniques of corner detection, geometric moments and random sampling are presented to simply and accurately locate the important feature points in image. For each extracted feature in image, a descriptor is calculated and based on the homograph transformation the matching is done. The results of experiments conducted on images taken by handheld camera and compared with the most famous SIFT method. The results show that the proposed algorithm is accurate, fast, efficient and robust under noise, transformation and compression circumstances.

Keywords
Corner detection, geometrical moments, random sampling, features extraction, matching

1. INTRODUCTION
In the field of recognition, object tracking, multi-view geometry, motion based segmentation, mapping, indexing and retrieval, image features is always a key factor. An image feature consists of interest points that are appeared by abrupt changes in the signals from acquired images. Interest point detectors extract image features which are distinctive in their neighborhood and reproduced in corresponding images. Local features have been shown to be well suited to matching and recognition as well as to many other applications as they are robust to occlusion, background clutter and other content changes. The difficulty is to obtain invariance to different conditions. In the case of significant transformations, feature detection has to be adapted to the transformation, as at least a subset of the features must be present in both images in order to allow for correspondences. Features which have proved to be particularly appropriate are interest points. Many solutions for features extraction have been presented over the past few years.

Many of the existing methods search for local extrema in the scale-space representation of an image. This idea was introduced by Crowley &Parker [1]. In this approach the pyramid representation is computed using difference-of-Gaussian filters. A feature point is detected if a local extrema is present and if its absolute value is higher than a threshold. In other work (Alvarez &Morales) [2] an affine invariant algorithm for corner detection was proposed. They apply affine morphological multi-scale analysis to extract corners. For each extracted point they build a chain of points detected at different scales, but associated with the same local image structure. The final location and orientation of the corner is computed using the bisector line given by the chain of points. The main drawback of this approach is that an interest point in images of natural scenes cannot be approximated by a model of a perfect corner, as it can take any form of a bi-directional signal change. This approach cannot be a general solution to the problem of affine invariance but gives good results for images where the corners and multi-junctions are formed by straight or nearly straight edges. [3] Proposes to use the Laplacian-of-Gaussian (LoG) and searches for maxima of scale. The scale-space representation is built by successive smoothing of the high resolution image with Gaussian kernels of different size. [4] Proposed an efficient algorithm for object recognition based on local extrema in the scale-space pyramid built with difference-of- Gaussian (DoG) filters. The input image is successively smoothed with a Gaussian kernel and sampled. The difference-of-Gaussian representation is obtained by subtracting two successive smoothed images. Thus, all the DoG levels are constructed by combined smoothing and sub-sampling. The local extrema in the pyramid representation determine the localization and the scale of the interest points. The common drawback of the DoG and the LoG representation is that local maxima can also be detected in the neighborhood of contours or straight edges, where the signal change is only in one direction. These maxima are less stable because their localization is more sensitive to noise or small changes in neighboring texture. A more sophisticated approach [5], solving this problem, is to select the scale for which the trace and the determinant of the Hessian matrix (H) simultaneously assume a local extremum. The affine shape estimation was used for matching and recognition by [6]. He extracts interest points at several scales using the Harris detector and then adapts the shape of the point neighborhood to the local image structure using iterative procedure. The points as well as the associated regions are therefore not invariant in the case of significant affine transformations Furthermore, there are many points repeated at the neighboring scale levels, which increases the probability of false matches and the complexity. [7] Implemented a feature-based stereovision solution using moment invariants as a metric to find corresponding regions in image pairs that will reduce the computational complexity and improve the accuracy of the disparity measures that will be significant for the use in small robotic vehicles. [8] found an extended Harris detector to scale-space and propose a novel method - Harris-like Scale Invariant Feature Detector (HLSIFD). HLSIFD uses Hessian Matrix which is proved to be more stable in scale-space than Harris matrix. HLSIFD suppresses edges smoothly and uniformly, so fewer fake points are detected. [9] Presented an enhancement to SIFT algorithm. They highlight an issue with how the magnitude information is used and this issue may result in similar descriptors being built to represent regions in images that are visually different. To address this issue, they proposed a new strategy for weighting the descriptors depending on orientation histogram modification. The results show that Symmetric-SIFT descriptors built using the proposed strategies can lead to better registration accuracy than descriptors built using the original Symmetric-SIFT technique. [10] Proposed an improved Harris corner detection method for effective image registration. This method effectively avoids corner clustering phenomenon occurs during the corner detection process, thus the corner points detected distribute more reasonably.

Our contribution is introducing a new direction for developing the feature points detection and description algorithm using fusion of techniques for features extraction. Many researchers worked on enhancing different approaches for extracting features, but this work aims to develop an efficient, fast, accurate and simple algorithm to extract strong feature points in image by generating descriptors using corner detection, moment invariants and random sampling approaches.

The reminder of this paper is organized as: Section (2) Harris corner detector. Section (3) includes the feature descriptor construction using geometrical moments. Section (4) explains the features matching and inliers estimation from correspondences. Section (5) explains the experimental results and the conclusions in section (6).

2. CORNER POINTS DETECTION

The 'corner' is defined as 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, Harris corner detector is proposed to extract corner information as first step of the proposed algorithm.

The Harris corner detector was proposed by Harris, & Stephens, 1988 [11]. 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 dealt with the issue of directions with the advantage of the autocorrelation function and increased the accuracy of localization. Feature point extract by Harris vertex arithmetic operator has rotation and translation invariability and has good robustness against noise and change of parameters during acquisition of data. Harris detector is based on the autocorrelation function or image gradient [12]:

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

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

Where, I_x and I_y denote the image gradients in the x and y directions.

A feature point corresponds to the highest singular values, and it can be computed using the criterion:

$$h = DET(M) - k.\ Tr(M)^2 \tag{2}$$

$k \in [0.04, 0.06]$

A 'corner' is said to occur when the two eigenvalues are large. On the basis of 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.1.

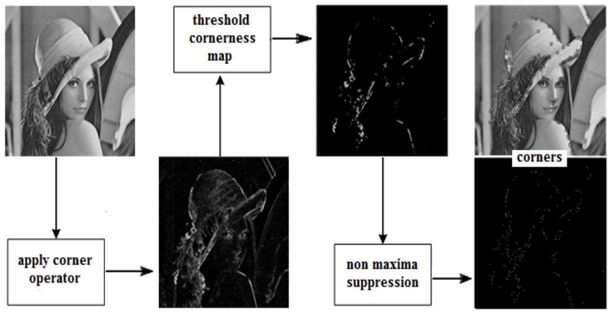


Fig.1. Flowchart and results for corner point detector

3. CONSTRUCTION OF FEATURE DESCRIPTOR USING MOMENTS

After detecting all image features using corner detection method, it is necessary to identify each feature point. Therefore, an efficient and simple descriptor will be generated for each detected feature point. For each feature, a 25*25 circular window centered at this feature is determined. Then, the window is divided to 5*5 blocks. For each block the average of geometrical moments are calculated. The result is a descriptor of 25 elements 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 order $(p+q)$ is defined as [13]:

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

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 xy plane, moments of all order 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 [13][14][15]:

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

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

For a digital image, Equation (3.2) becomes

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

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

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

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad (6) \qquad \text{Where } \gamma = \frac{p+q}{2} + 1$$

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

$$\phi_1 = \eta_{20} + \eta_{02} \tag{7}$$

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

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

(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 in order to facilitate comparison, making use of logarithmic methods. At the same time, taking into account the possible negative moment invariants situation, you have to get absolute value before getting logarithm.

4. FEATURE MATCHING AND INLIERS ESTIMATION

Images can be in different situations or transformations that can be resulted during camera acquiring or by the applying image transformation techniques. According the order of complexity, these transformations can be rigid, affine, piecewise affine and non-rigid or elastic. Rigid registration models are linear and only allow for translations, rotations and uniform scale changes without any distortion. Affine models are also linear and support overall distortions represented as shears and stretches. Piecewise affine and elastic 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 following:

$$DES = D(i,j) / D(i,a), \quad (14)$$

If DES < T then (i and j) is match

Where D (i ,j) is the distance between point (i) in first image and point (j) in second image. (a) is the second nearest neighbor point. The points (i) and (j) is matching if the value of DES is lower than the predefined threshold. The mismatched points can be considered as outliers "which are the data that do not fit the model“. These outliers can severely disturb the estimated homograph, and consequently should be identified. The goal then is 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 from a projective plane to a projective plane that maps straight lines to straight lines. In the field of computer vision, any two images of the same planar surface in space are related by a homograph. This is very important in computing the camera motion 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.

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

Then: $x' = \mathrm{H}x$ (15)

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

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

The random sampling algorithm [16] 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 is basically composed of two steps that are repeated in an iterative fashion. First a set of points are 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 dataset are consistent with the model instantiated with the parameters estimated in the first step. The set of such elements is called Consensus Set [17]. The algorithm terminates 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 with the homograph. A pair (x; x’), is considered to agree with a homograph H, if dist (Hx; 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 a sufficient number of pairs is consistent with the computed homograph.

5. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed algorithm, the experiments done on images taken by handheld camera and in different situation like translation, rotation, scaling and projection and in different lightning conditions and in comparison to the state of arts SIFT algorithm. First, for each image the corners are extracted and the corners on the borders region are eliminated as shown in fig.2 (a). The value of the used threshold must be sufficient to extract enough number of corners. Therefore, it taken between 0.1 to 1.5. after extracting the corner points in the two images, the descriptors for them is created by taking a 25*25 circular window around each point, dividing it 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. In these experiments the threshold taken between 0.5 to 1 and the result is as in fig.2 (b). The final step is eliminating the mismatch points or the outliers to get on only the true matches that represent the key feature points by using random sampling algorithm. The number of iterations is important factor here to obtain a biggest number of matches between the images. In our experiments the number of iterations used is between 2000 to 13000 iteration. The result of this step is shown in fig.2(c).

Fig.3. explains the impact of threshold values on the number of extracted feature points. As shown in this figure when the

corner threshold decreased the more feature points we can have. But in matching descriptors and random sample thresholds they must be increased to have more feature points.

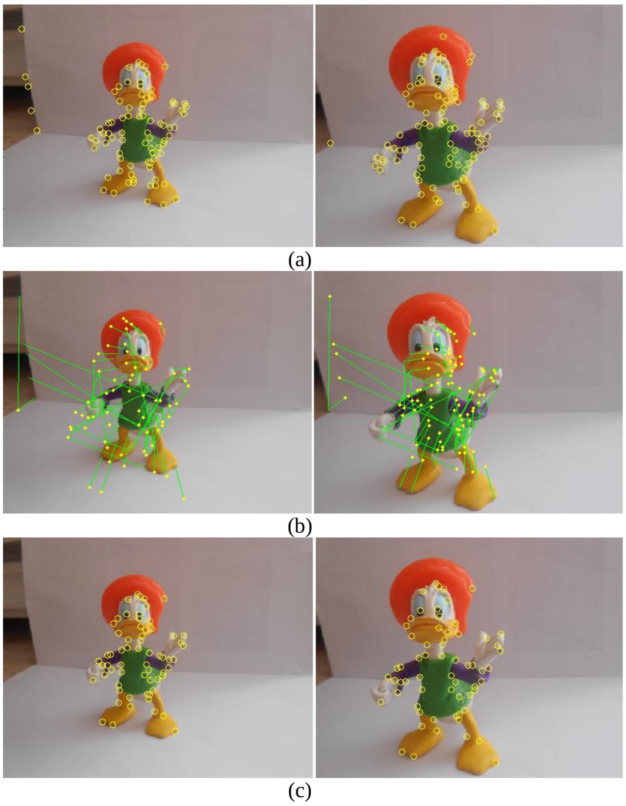


Fig.2. feature extraction result: (a) the extracted corner points. (b) matching result using second nearest neighbor (c) the final features extraction using random sampling

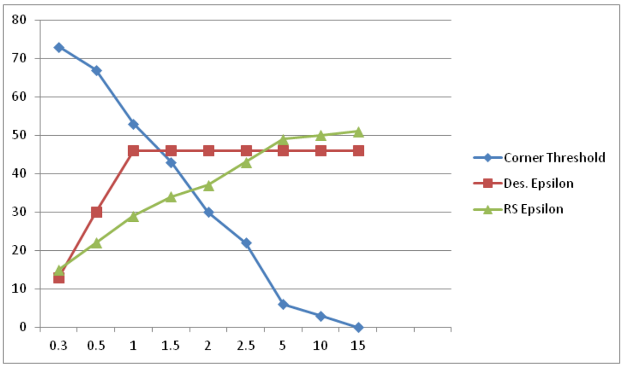


Fig.3. the impact of threshold values on number of extracted feature points (raw=threshold value and column=number of feature points)

Other implementations on the proposed algorithm for different samples taken with handheld camera are illustrated in fig.4. Some samples in fig.5 illustrate the results of implementing the proposed algorithm on compressed images. The number of extracted matching feature points is effected by the amount of noise and the amount of transformation. As shown in fig.6, salt and pepper and speckle noises are applied with different degrees on sample image. The number of extracted feature points is decreased with the increase in noise degree.

Fig.7 shows applying different degrees of rotation and scaling on sample image. We notice that in scaling operation, the number of extracted feature points is not effected and approximately the same. But in rotation operation, when the rotation angle is increased the number of extracted feature points is decreased.

Fig.4. Results of the proposed algorithm

Fig.5. The results of the proposed algorithm on the compressed images

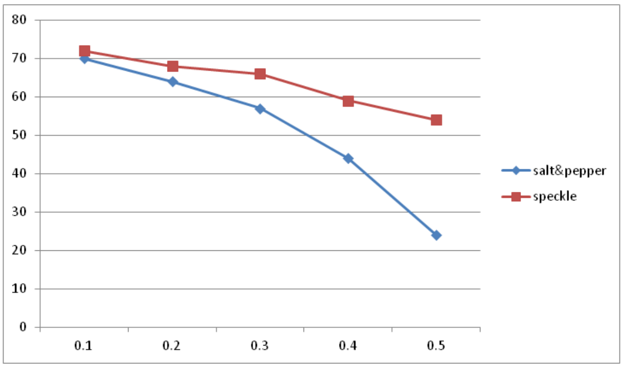


Fig.6. the number of matching points with different degrees of noises (raw= noise degree, column=number of points)

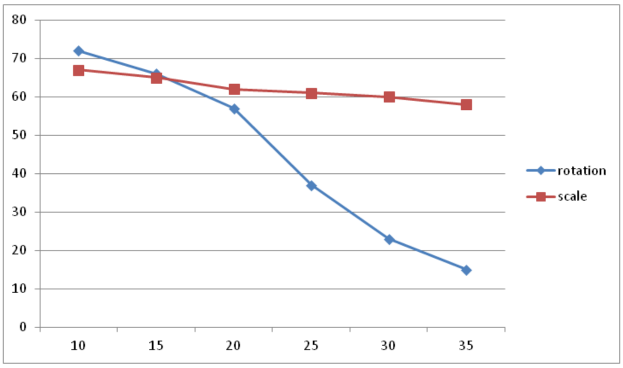
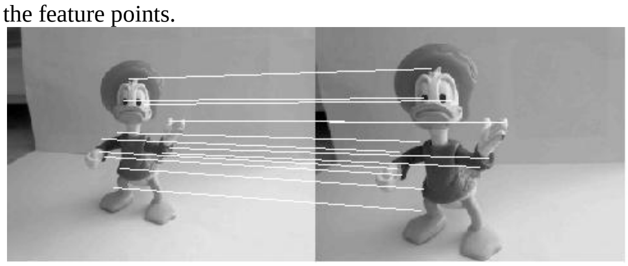
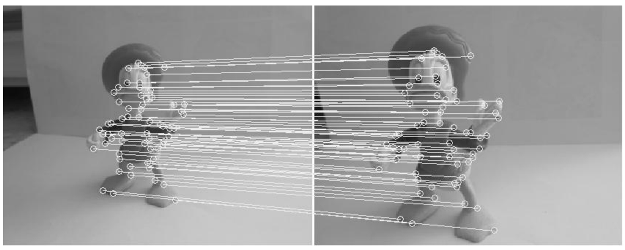


Fig.7. The number of matching points with different degrees of rotation and scale (raw= rotation and scale degree, column=number of points)

In comparison with the famous SIFT method; sift method is more complex than the proposed method. The proposed algorithm is simple and fast. The accuracy of the proposed algorithm is coming from the using of random sampling algorithm because it follows the objects motion to estimate the feature points.



(a)



(b)

Fig.8. The extracted feature correspondences. (a) The SIFT method (b) The proposed method

6. CONCLUSIONS

Intensity in the image does not change arbitrarily, but there may be a change in overall contrast due to changes in illumination or camera parameters. This may make the operation of extracting sufficient features for matching a difficult operation. Therefore, an efficient algorithm is proposed to solve this problem. Through the use of nonlinear combination of geometric moments we can obtain one group of scale invariance, translation invariance and rotation invariance of moment invariants. The traditional moments method has a lower computational cost and also limited to the affine or any simpler model.

In the other side, Harris corner detector is chosen due to its good results with nature of image such as the robust to noise, rotation change, light change, and the repeatability of detected points. Mixing these two techniques with random sampling technique in a smart manner gives us an efficient and robust method. The results of the proposed algorithm show that it has a high performance with noise, compression and different transformations. Therefore, it can be used then in many of computer vision applications.

7. REFERENCES

[1] Crowley, J. and Parker, A.1984. A representation for shape based on peaks and ridges in the difference of low pass transform. IEEE Transactions on Pattern Analysis and Machine Intelligence. 6(2):156–170.

[2] Alvarez, L. and Morales, F. 1997. Affine morphological multiscale analysis of corners and multiple junctions. International Journal of Computer Vision. 2(25):95–107.

[3] Lindeberg, T. 1998. Feature detection with automatic scale selection. International Journal of Computer Vision. 30(2):79–116.

[4] Lowe, D.G. 1999. Object recognition from local scale-invariant features. In Proceedings of the 7th International Conference on Computer Vision. Kerkyra. Greece. pp. 1150–1157.

[5] Mikolajczyk, K. and Schmid, C. 2002. An affine invariant interest point detector. In Proceedings of the 7th European Conference on Computer Vision. Copenhagen. Denmark. vol. I. pp. 128–142.

[6] Baumberg, A. (2000). Reliable feature matching across widely separated views. In Proceedings of the Conference on Computer Vision and Pattern Recognition. Hilton Head Island. South Carolina. USA. pp. 774–781.

[7] Prashan, P. & Farzad, S. 2008. Feature based stereo correspondence using moment invariant. 4th International Conference on Information and Automation for Sustainability. Sustainable Development Through Effective Man-Machine Co-Existence. IEEE Region 10 and ICIAFS. Colombo. Sri Lanka. pp. 104-108.

[8] Yinan, Y., Kaiqi, H. and Tieniu, T. 2009. A Harris-like Scale Invariant Feature Detector. 9th Asian Conference on Computer Vision, Xi'an, September 23-27. Springer Berlin Heidelberg. pp. 586-595.

[9] Tanvir, H., Wei, T., Guojun, L. and Martin L. 2010. An Enhancement to SIFT-Based Techniques for Image Registration. Dicta. International Conference on Digital Image Computing: Techniques and Applications. pp.166-171.

[10] Yijian, P., Hao, W., Jiang, Y. and Guanghui, C. 2010. Effective Image Registration based on Improved Harris Corner Detection. International Conference on Information, Networking and Automation (ICINA). 18-19 Oct. pp. 93-96. IEEE.

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

[12] Malik, J., Dahiya, R. and Sainarayanan, G. 2011. Harris Operator Corner Detection using Sliding Window Method. International Journal of Computer Applications (0975 – 8887)Volume 22– No.1. pp. 28-37.

[13] Hu M.K. (1962). Visual pattern recognition by moments Invariants. IRE Trans. Information Theory, vol 8. pp:179- 87.

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

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

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

[17] Zuliani M. (2006). Computational Methods for Automatic Image Registration. PHD thesis. University of California, USA.