

A NEW APPROACH OF LOCAL FEATURE DESCRIPTORS USING MOMENT INVARIANTS

¹Lee-Yeng Ong, ¹Siong-Hoe Lau and ²Voon-Chet Koo

¹Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia

²Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia

Received 2014-08-05; Revised 2014-09-29; Accepted 2014-10-14

ABSTRACT

Moment invariants have been widely introduced in recognizing planar objects for a few decades. This is due the robustness of moment function in distinguishing the original identity of object under various two Dimensional (2D) transformations. A set of moments computed from a planar images, represents the global description of an object's shape and geometrical features of an image. Since global descriptor utilizes the information of a whole object or shape to describe the features of an object, it does not tolerate occlusion. If there is a mixture of regions that do not belong to the object of the interest, an additional task of segmentation is required to isolate the object for recognition. Hence, moment invariants are proposed to be employed as local descriptors for object recognition since local descriptors do not suffer from the drawbacks caused by image clutter and occlusion. A new approach of local feature descriptors using moment invariants is presented. The preliminary framework is divided into three different stages. Interest points are firstly detected in the entire image. The local descriptors are then produced by applying moment invariants on the region around the interest points. Cross-correlation is finally carried out for feature matching.

Keywords: Geometric Moments, Feature Matching, Local Descriptor

1. INTRODUCTION

Since the introduction of geometric invariants in 1962, moment invariants have been applied in object recognition, shape analysis, image description and matching (Flusser *et al.*, 2009). The invariants are able to provide descriptive information of an object for distinguishing its identity from another object. Although the object undergo 2D transformations (translation, scale, rotation and skew), the identification task remain invariance. Due to the promising result, moment invariants are further extended to new areas, such as hand gesture recognition, image registration, fingerprint verification, image retrieval and action classification (Almoosa *et al.*, 2008; Chen *et al.*, 2013; Costantini *et al.*, 2011; Li *et al.*, 2012).

Moment function interprets an object (in a 2D image) as a 2D intensity distribution, which provides global features of an object: Total area, coordinates of centroid and orientation. The performance of global features is seriously affected when region of the object is partially occluded by another object. This situation is commonly happen in natural images, where multiple objects are mixed in a scene. Segmentation task has been widely used to overcome this limitation by separating the object of interest from the scene. The segmented region, however, might not represent the intensity distribution of the whole object when partial occlusion took place.

Meanwhile, local features do not suffer from the drawbacks caused by image clutter and occlusion. A local feature is an image pattern extracted from a particular region of an object. It represents the

Corresponding Author: Lee-Yeng Ong, Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia

descriptive information that is associated with the change of intensity distribution in the image pattern (Shvarts and Tamre, 2012). Local features are normally extracted from the region around the key points within an object. These features are distinctive and recognizable even though parts of the object are occluded. Therefore, geometric moment invariants are proposed to be employed as a new approach of local descriptors. This approach is able to maintain the ability of moment invariants in providing unique and distinguishable features in a natural image or sequence of frames.

This paper presents a preliminary framework on selecting the feature points, formulating the invariance feature descriptors and matching descriptors in a sequence of consecutive frames. The proposed framework is divided into three different stages. Interest points are firstly detected in the entire image. The local descriptors are produced by applying moment invariants on the region around the interest points. Cross-correlation is finally carried out for feature matching.

1.1. Related Works

In the history of object recognition, many early-developed approaches are based on global features. Moment invariants are one of the earliest and widely used methods. The main reason is due to the robust performance of invariance property in different transformations. Indeed, the recognition rate of global features is affected when foreground objects are mixed with background scene in natural images (Tuytelaars and Mikolajczyk, 2008). In the current world that equipped with video surveillance system, there is a tremendous input of natural images sent for daily recognition task. Global features are no longer sufficient for recognizing object that is partially occluded or part of object is out of the field of vision.

In order to overcome this limitation, a few regions (blobs) with reliable description are extracted from the image. The extracted regions contain descriptive information that is corresponding to different subparts of the image. A string of vectors is then formed with the description of blobs. The recognition is performed by matching similarities between subparts of a foreground object even in changing background and partial occlusion (Krolupper and Flusser, 2007). One of the famous approaches in extracting local features is SIFT. Lowe (2004) method transforms an image into a multi-scale sampling of image patches centered on the

interest points. Each of the feature vectors is invariant to image scale and rotation. Lowe suggested four stages of filtering method for SIFT, which includes scale-space extrema detection, keypoint localization, orientation assignment and keypoint descriptor. The resulting features are used by nearest-neighbors algorithm to identify the best-matched object in an image. Since SIFT is able to generate a large number of local features, object is still recognizable in substantial level of occlusion.

On the other hand, some researchers introduced the combination of both independent algorithms from local features detection and description. The most recent combinations between FAST detector with BRIEF descriptor or BRISK descriptor offer a much more suitable alternative for real-time applications (Miksik and Mikolajczyk, 2012). This is due to the outstanding result of FAST detector in several comparison studies (Rosten and Drummond, 2006; Miksik and Mikolajczyk, 2012; Sens *et al.*, 2012). As compared to other existing detectors, FAST feature detector achieves a nearly constant of 2 ms runtime per image with respond to an increasing number of features. Rosten and Drummond (2006) have proven that FAST-9 is the most reliable detector with shortest runtime and low processing power. A fast and reliable detector is definitely in need for producing an efficient combination with feature descriptor.

2. PROPOSED FRAMEWORK AND METHODOLOGY

From the discussion of Shvarts and Tamre (2012), local feature is the descriptive information selected from a specific region of an object to avoid drawbacks of image clutter and occlusion. The proposed framework of formulating a set of suitable local features with the moment invariants function is shown in **Fig. 1**. A set of feature points is initially selected from an input image. The invariance descriptors are then formulated from the neighborhood region of each feature point, which indirectly build up a unique identity for an object. Since every object is recognized with a unique descriptor, it can be used to locate an object in the consecutive video frames.

2.1. Methodology of Feature Points Detection

There are several types of local invariant features discovered by researchers in decades ago. Image properties, such as points, edge or small image

patches are extracted as a local feature and converted into descriptors (Tuytelaars and Mikolajczyk, 2008). Several algorithms have been continuously developed in the early literature of image processing, for finding corner points at the extrema of functions computed on the shape. Since 1980, Moravec introduced corner detection algorithm for robot navigation. The algorithm was further improved by Harris and Stephens (1988; Shi and Tomasi, 1994) for invariance detection. More simple and efficient algorithms, such as SUSAN and FAST are recently developed for improving the computational time of corner detection (Smith and Brady, 1997; Rosten and Drummond, 2006).

The algorithm of FAST detector is built on the basic concept of SUSAN detector. According to Rosten and Drummond (2006), a corner point can be concluded if there is sufficiently large set of pixels in the circular neighborhood, significantly brighter or darker than the central point. A circle of 16-pixels is initially formed by a fixed radius around the central point, as illustrated in **Fig. 2**. The selected pixel from north (1), south (9), east (5) and west (13) locations of the circle are compared with a threshold for classifying brighter, similar and dark categories. If there are at least three of the pixels brighter or darker than the threshold, the central point is concluded a corner point. Otherwise, the test criterion is continued to be applied on the remaining pixels in the circle.

2.2. Methodology of Invariance Descriptor Formulation

2.2.1. Invariance Features from Moment Function

A set of moments computed from a planar images, represents global description of the object shape and geometrical features of the image. When applying to images, simple properties of the image which include area of an image, centre of mass and orientation information can be found via moment functions. The properties of an image can be generated from a geometric moments with the general definition given as:

$$G_{ij} = \int \int_{\zeta} x^i y^j f(x, y) dx dy, \quad i, j = 0, 1, 2, 3, \dots \quad (1)$$

The moment function in Equation 1, G of order $(i+j)$, consists of monomial functions in the image region of ζ , for 2D density distribution, $f(x, y)$. Geometric moments were the first moment function that was used to derive a set of invariant descriptors. Hu (1962) presented a set of invariant descriptors from geometric moments. The presented set is able to recognise images, no matter in translation, scaling and

rotation transformations. Since then, Hu's publication has been extensively referenced in nearly all moment related literature for the past few decades.

In order to achieve translation and scale invariants, geometric moments are defined with respect to the image centroid (x_0, y_0) as the origin, i.e., Equation 2 and scale factors are eliminated. Translation and scale invariants are listed in Equation 3. By following Hu's method in deriving rotation invariants with theory of algebraic invariants, Equation 3 is used to formulate a set of functions that are invariant with respect to translation, scale and rotation changes in an image. The invariant functions are given in Equation 4 (Mukundan, 1998):

$$C_{ij} = \int \int_{\zeta} (x - x_0)^i (y - y_0)^j f(x, y) dx dy, \quad (2)$$

$i, j = 0, 1, 2, \dots$

$$\eta_{ij} = \frac{C_{ij}}{C_{00}^{\rho}}, \quad \rho = \frac{i + j + 2}{2} \quad (3)$$

$$\begin{aligned} M_1 &= \eta_{20} + \eta_{02}, \\ M_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \\ M_3 &= \eta_{20}\eta_{02} - \eta_{11}^2, \\ M_4 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \\ M_5 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \end{aligned} \quad (4)$$

An example of feature descriptor that is generated from the translation, scale and rotation invariants for a sample image is illustrated in **Table 1**. The sample image consists of an alphabet 'F' with size 100×100 pixels. Based on the result of feature descriptor computed from Equation 4, only minor variation that is less than 0.001 occurred among the same invariance function in different transformations. The feature descriptor of the sample image is further compared with other images, by using the same alphabet with the almost similar font types and recorded in **Table 2**. The result shows a larger difference between the images of similar font types, as compared to **Table 1**. Therefore, geometric moment invariants is capable of building a unique identity of a specify image although it undergoes several geometrical transformations.

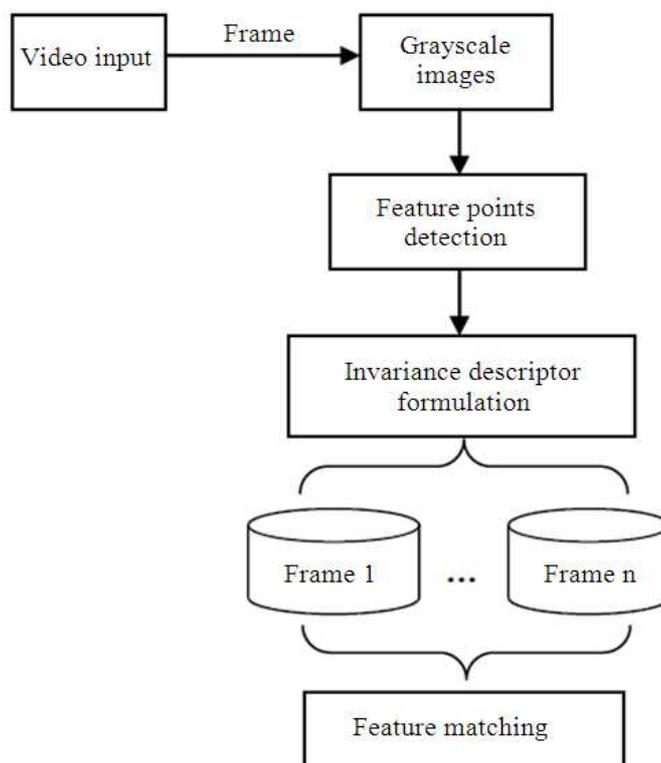


Fig. 1. The proposed framework of formulating a set of local descriptor

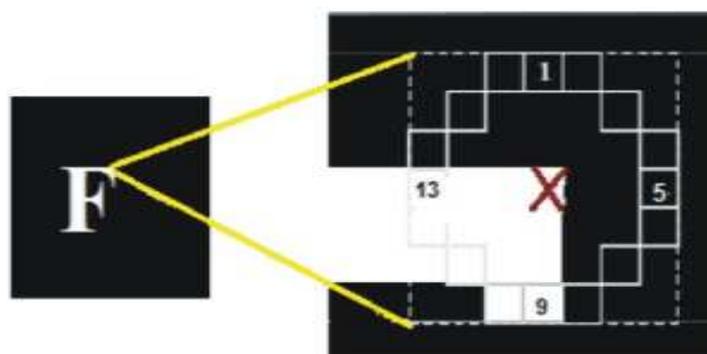


Fig. 2. Illustration of pixel selection in the circle of FAST feature point detector

Table 1. Feature descriptor of sample image in different transformation

Transformation	Feature descriptor				
	M ₁	M ₂	M ₃	M ₄	M ₅
Original image	2.85760	6.30000	6.44820	9.27250	10.53200
Reflected against x-axis	2.85520	6.30160	6.44120	9.27120	10.53650
Scaled by 200%	2.85400	6.30860	6.43550	9.27640	10.53860
Rotated by 90°	2.85450	6.30550	6.43790	9.27050	10.54090
Variation (δ/μ)	0.00055	0.00061	0.00085	0.00028	0.00035

Table 2. Feature descriptor of different images

Font type	Feature descriptor				
	M ₁	M ₂	M ₃	M ₄	M ₅
Times new roman	2.8576	6.3000	6.4482	9.2725	10.5320
Arial	2.8997	6.3683	6.5380	9.3618	10.4418
Century	2.7750	6.2108	6.2591	8.8579	10.5737
Garamond	2.7758	6.2542	6.2498	8.9943	10.4415
Average difference (+/-)	0.0688	0.0677	0.1591	0.2607	0.07413

2.2.2. Formation of Proposed Local Descriptor

Once a feature point has been detected, local descriptor is formulated from the neighborhood of the feature point. A region of 10×10 pixels around the feature point is extracted and computed with the invariance function from Equation 4. For each feature point, p , the proposed Invariance Descriptor (ID) is derived as Equation 5:

$$ID(p) = [M_1 \ M_2 \ M_3 \ M_4 \ M_5] \quad (5)$$

Since each invariance descriptor constitutes to the identity of a frame, Fr is formed with all the extracted invariance descriptor from the feature points.

$$Fr(n) = [ID(1) \dots ID(p)]^T, \ n, \ p = 1, 2, 3, \dots \quad (6)$$

2.3. Methodology of Feature Matching

The methodology of matching the proposed local descriptors among frames is shown in Fig. 3. After the invariance descriptors of all feature points are formulated from Equation 6, the descriptors are ready to be used for discovering the matching pairs from consecutive frames. In order to determine the matching pairs, the descriptors $ID(p)$ of a set feature points from the previous frame, $Fr(n-1)$ have to be related in a certain criterion with the target feature sets from the current frame, $Fr(n)$. The linear correlation coefficient is chosen to measure the association between the descriptor sets, $ID(p)$ from $Fr(n-1)$ and $Fr(n)$. Based on the Pearson's correlation coefficient, r from Equation 7, the pair of descriptors that experienced the positive coefficient closest to 1 indicates a strong association between the descriptors. Thus, the pair of descriptors with largest coefficient is shortlisted as matching pairs.

$$r = \frac{(Fr_{n-1} - \overline{Fr_{n-1}})(Fr_n - \overline{Fr_n})}{\sqrt{(Fr_{n-1} - \overline{Fr_{n-1}})^2 (Fr_n - \overline{Fr_n})^2}} \quad (7)$$

Instead of using all the shortlisted pairs, only those highly reliable pairs are remained to improve the matching performance. An efficient way for evaluating

matching pairs as the reliable pairs is by using RANdom Sample and Consensus (RANSAC) algorithm. RANSAC algorithm estimates the possible homographies that elaborate the relation between descriptor pairs in different frames (Hartley and Zisserman, 2004). During the estimation, the less reliable pairs or considered the outliers are rejected. This iterative method ends up with homographies that are estimated from the inliers.

3. RESULTS AND DISCUSSION

The proposed framework is tested with a sequence of four sample frames, where each frame consists of 512×512 pixels. This sequence of images is retrieved from Hartley and Zisserman (2004), as shown in Fig. 4. The sample grey-scale images are captured at the corridor of a building and experienced several transformations, which include translation, scale and rotation.

Based on the proposed framework in Fig. 1, feature points are initially detected from each frame with FAST-9 detector. The results of detected feature points are highlighted in the image sequence, as shown in Fig. 5. It is noted that mostly of the feature points from the current frame are still detectable in the next frame although it involved transformation changes. Invariance feature descriptors are subsequently formulated from the neighborhood of each feature points. The neighborhood area of 10×10 pixels within the feature point is selected for descriptor computation. In order to find out the matching keys, the set of invariance descriptors from the current frame is correlated with the descriptor from the next frame. The largest coefficient represents the stronger matching pairs but not necessary the reliable matching pairs. Fig. 6 shows the output of matching pairs selected from the largest coefficient. However, some less reliable matching pairs are not associated to the correct points in the latter frame. An iterative method, RANSAC is used to estimate a suitable model of homography between the descriptor pairs. At the same time, the less reliable pairs that have been considered the outliers of the model would be rejected. Fig. 7 shows the result of the finalized reliable matching pairs across 2 consecutive frames.

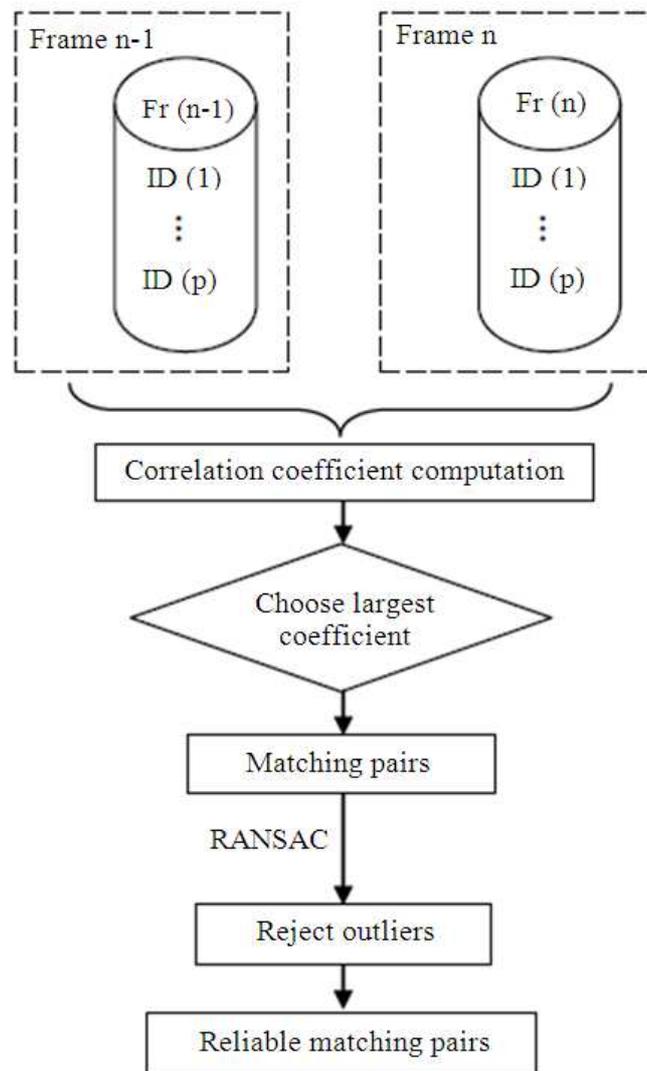


Fig. 3. The methodology of feature matching





Fig. 4. A sequence of sample images that is transformed into different sizes, orientations and positions

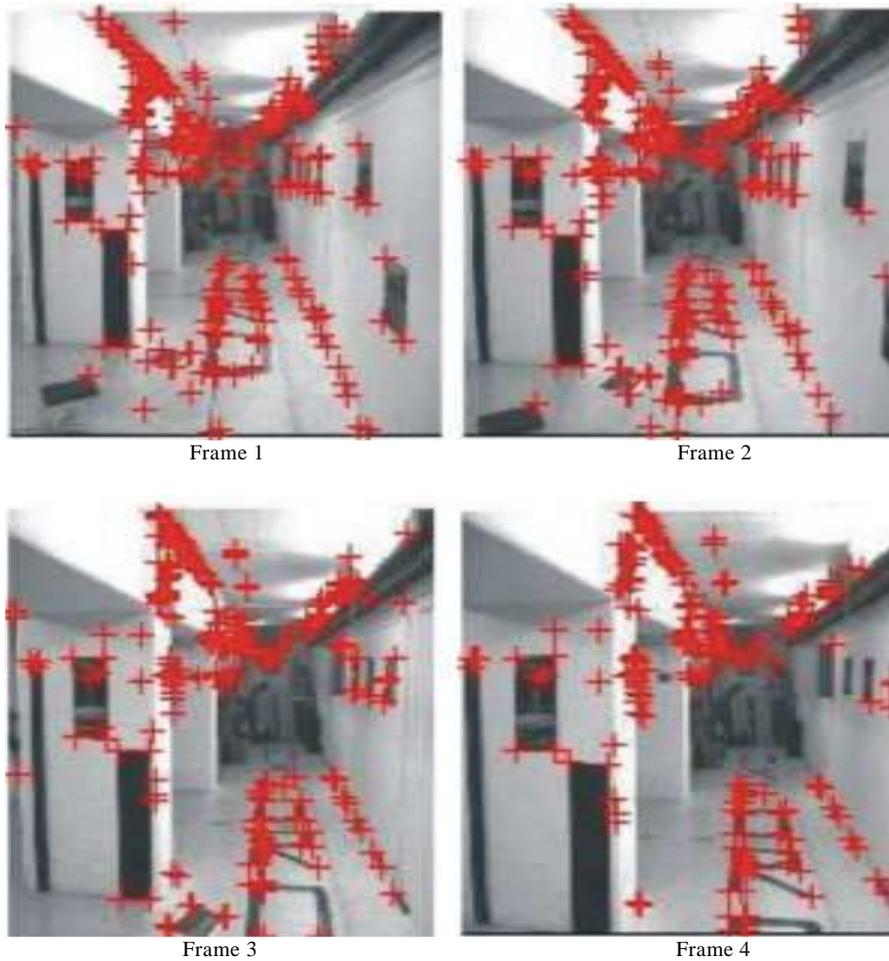


Fig. 5. Detected feature points with FAST detector on a sequence of images

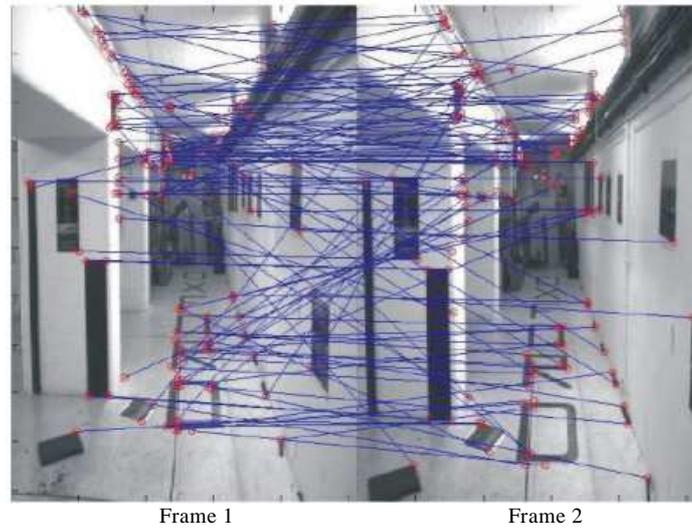


Fig. 6. Matching pairs computed from correlation coefficient between feature descriptors from frame 1 and 2

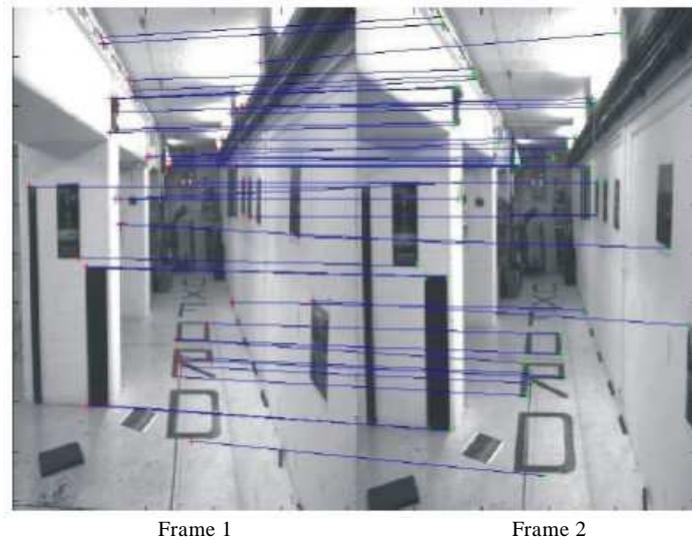


Fig. 7. Result of reliable matching pairs between feature descriptors from frame 1 and 2

The performance of the feature matching is evaluated by obtaining the percentage of correct feature matching between two frames. The details of total matched points, false matching and percentage of matching accuracy are listed in **Table 3**. The result of feature matching in two consecutive frames, such as frame 1 and 2, frame 2 and 3 and frame 3 and 4, have reflected a promising matching result of 88 to 90%. In addition, feature matching in alternate frames (frame 1 and 3 and frame 2 and 3) have also achieved the result of 83% and

above. These verified the usefulness of geometric moment invariants as local feature descriptors. An additional testing is done for feature matching in the situation of two and more missing frames. However, the matching accuracy reduced tremendously. In the situation between frame 1 and frame 4, there are several obvious transformations (rotate and zoom in) took place. It caused the huge changes in feature point detection and descriptor formulation, which leads to the increment of false matching.

Table 3. Result of feature matching between frames

Frame	Feature matching		
	Matched points	False matching	Accuracy (%)
1 and 2	46	5	89.13
2 and 3	34	4	88.24
3 and 4	31	3	90.32
1 and 3	26	3	88.46
1 and 4	10	6	40.00
2 and 4	24	2	83.33

4. CONCLUSION

A new approach of local feature descriptors using moment invariants is presented and tested in the proposed framework. The proposed descriptors are invariant to changes in scale, rotation and translation in consecutive frames and also alternate frames. This approach can be served as a new contribution for features tracking in image warping, locating moving objects in surveillance video and indoor robot navigation system.

5. ACKNOWLEDGEMENT

The research work was supported by the Fundamental Research Grant Scheme (FRGS) of Ministry of Education Malaysia.

5.2. Author's Contributions

All authors equally contributed in this work.

5.3. Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

6. REFERENCES

- Almoosa, N.I., S.H. Bae and B.H. Juang, 2008. Toward robust moment invariants for image registration. Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, Mar.-Apr. 31-04, IEEE Xplore Press, Las Vegas, NV , pp: 1009-1012. DOI: 10.1109/ICASSP.2008.4517783
- Chen, R., S. Wang, L. Gong and C. Liu, 2013. Hand gesture recognition for human-computer interaction using moment invariants and neural network classifier. Lecture Notes Comput. Sci., 8102: 661-667. DOI: 10.1007/978-3-642-40852-6_66
- Costantini, L., L. Seidenari, G. Serra, L. Capodiferro and A. Del Bimbo, 2011. Space-time Zernike moments and pyramid kernel descriptors for action classification. In: Image Analysis and Processing-ICIAP, Maino, G. and G.L. Foresti, (Eds.), Springer Berlin Heidelberg, ISBN-10: 978-3-642-24087-4, pp: 199-208.
- Flusser, J., B. Zitova and T. Suk. 2009. Moments and Moment Invariants in Pattern Recognition. 1st Edn., John Wiley and Sons, Chichester, ISBN-10: 0470684763, pp: 312
- Harris, C. and M. Stephens, 1988. A combined corner and edge detector. Proceedings of Alvey Vision Conference, University of Manchester, pp: 1-6. DOI: 10.5244/C.2.23
- Hartley, R., and A. Zisserman, 2004. Multiple View Geometry in Computer Vision. 2nd Edn., Cambridge University Press, Cambridge, ISBN-10: 1139449141.
- Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Trans. Inform. Theory, 8: 179-187. DOI: 10.1109/TIT.1962.1057692
- Krolupper, F and J. Flusser, 2007. Polygonal shape description for recognition of partially occluded objects. Pattern Recognit. Lett., 28: 1002-1011. DOI: 10.1016/j.patrec.2006.12.021
- Li, C., J. Li, B. Fu and X. Yang, 2012. Fingerprint verification based on DFB and Hu invariant moments. J. Computat. Inform. Syst., 4: 1407-1414.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision, 2: 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94
- Miksik, O. and K. Mikolajczyk. 2012. Evaluation of local detectors and descriptors for fast feature matching. Proceedings of the International Conference on Pattern Recognition, (CPR' 12), CMP, Prague, Czech Republic, pp: 2681-2684. DOI: 10.1.1.301.6783

- Moravec, H. P. 1980. Obstacle avoidance and navigation in the real world by a seeing robot rover. Dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Carnegie-Mellon University, Pennsylvania.
- Mukundan, R., 1998. Moment Functions in Image Analysis: Theory and Applications. World Scientific, Singapore, ISBN-10: 9810235240, pp: 150.
- Rosten, E. and T. Drummond. 2006. Machine Learning for High-Speed Corner Detection. In: Computer Vision-ECCV, Leonardis, A., H. Bischof and A. Pinz (Eds.), ISBN-10: 978-3-540-33832-1, Springer Berlin Heidelberg, pp: 430-443.
- Senst, T., B. Unger, I. Keller and T. Sikora, 2012. Performance evaluation of feature detection for optical flow tracking. Int. Conf. Pattern Recognit. Applic. Meth., 2: 303-309.
- Shi, J. and C. Tomasi, 1994. Good features to track. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 21-23, IEEE Xplore Press, Seattle, WA, pp: 593-600. DOI: 10.1109/CVPR.1994.323794
- Shvarts, D. and M. Tamre, 2012. Local and global descriptors for place recognition in robotics. Proceedings of the 8th International DAAAM Baltic Conference, Apr. 19-21, Tallinn, Estonia, pp: 1-6.
- Smith, S.M. and J.M. Brady, 1997. SUSAN-a new approach to low level image processing. Int. J. Comput. Vision, 23: 45-78. DOI: 10.1023/A:1007963824710
- Tuytelaars, T. and K. Mikolajczyk, 2008. Local invariant feature detectors: A survey. J. Foundat. Trends Comput. Graph. Vision, 3: 177-280. DOI: 10.1561/06000000017