Capability of Modified SIFT to Match Stereo Imagery System

Omar Abusaeeda[1*], Salah Naas[2] , Nasar Aldian Shashoa[3]

[1] abossada1@gmail.com, [2] naas.salah@gmail.com, [3] dr.naser.elec@gmail.com

[1] Department of Computer Networks, Faculty of Information Technology,
Azzaytuna University, Libya
[2, 3] Department of Electrical & Communication Engineering, Faculty of Engineering,
Azzaytuna University, Libya
*Corresponding author email:
Received: 00 April 2018 / Accepted: 00 May 2018

ABSTRACT

This paper presents an improved version of SIFT method for extracting invariant features from images that can be used to solve the correspondence problem between different views of an object or scene in an image. Scale invariant feature transform (SIFT) has recently gained substantial attention in the computer vision community to address the problem. Corresponding features in sequential pairs of images, at various different angular separations, were identified by applying a scale invariant feature transform (SIFT). Due to limitation in the standard SIFT; some of matches are considered false matches. Epipolar-line and disparity window criteria were introduced to enhance the performance of SIFT. Experiments revealed that considerable number of unfaithful matches were removed when new criteria are introduced. Future work will focus on improving the SIFT technique; to rectify the negative matches in order to obtain better matching result.

Keywords: SIFT, Image matching, positive matches.

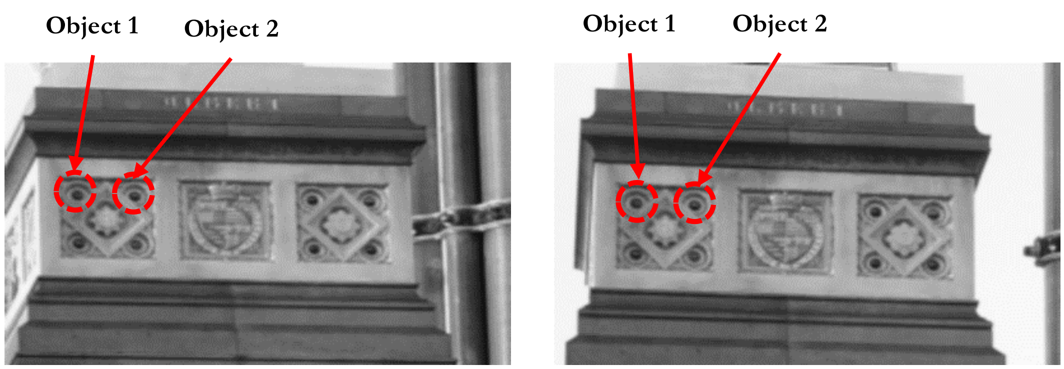
1 Introduction

A typical image matching method begins with detecting points of interest, then selects a region around each point, and finally associates a descriptor with each region. Correspondences between two images may then be established by matching the descriptors of both images. SIFT is proposed by David Lowe in 2004 [1] to extract features of interest from images that can be used for reliable matching between different views of an object. The features are invariant to image scaling and rotation and partially invariant to change in 3D viewpoint and additional noise. Over recent years, SIFT has played a significant role in various computing applications such as object recognition, 3D modelling and video tracking.
Feature matching can be defined as the process of matching corresponding points between two or more images of the same scene. Feature based methods match special features of two images, such as corners or edges to produce a sparse disparity map [2,3]. This method matches more features, rather than matching textured regions in the two images [4,5]. Feature based

methods provide more precise positioning for the matching results and are more reliable than correlation-based matching when good image features can be extracted from the scene [6]. Feature based methods are widely used in wide-base stereo image matching [7,8]. Correspondences between two images is established by matching the descriptors of both images. Numerous variations exist on the computation of interest points matching. It can be traced back to the work of stereo matching using a corner detector [9,10], which was later improved by Harris and Stephens [11,12]. Consequently, the Harris corner detector has since been extensively used for various other image-matching tasks. The approach was presently expanded to match Harris corners over a large image range by using a correlation window around each corner to select likely matches. Moreover, Harris corners were used to select interest points, but rather than matching with a correlation window, they used a rotationally invariant descriptor of the local image region. This allowed the matching of features under arbitrary orientation change between the two images. Additionally, it was demonstrated that multiple feature matches could accomplish general recognition under occlusion and clutter by identifying consistent clusters of matched features [1]. The local feature approach was extended to achieve scale invariance and more distinctive features whilst being less sensitive to local image distortions such as 3D viewpoint change [13]. In recent times, there has been an inspiring body of work on extending local features matching. Most recently, there has been an impressive effort on expanding the approach of local feature descriptor [14,15]. While this method is not completely affine invariant, a different approach is used in which the local descriptor allows relative feature positions to shift extensively with only small changes in the descriptor [16]. This approach produces descriptors, which are consistently matched across a substantial range of affine distortion. It also makes the features more robust against changes in 3D viewpoint. This approach not only has the advantages of extracting more efficient feature, but it also able to identify larger numbers of features. Furthermore, Principal Components Analysis SIFT (PCA-SIFT) was introduced. This technique accepts the same input as the standard SIFT descriptor. The advantage of this approach is the size of the descriptor. It produces a more compact descriptor in comparison to standard SIFT. On the other hand, it tends to blur the edges around the objects [17]. Another local feature descriptor named, Speeded-Up Robust Features (SURF) was proposed [18]. SURF is mainly designed for real time application where the speed is the main concern. SURF performance is similar to SIFT but it is not invariant to rotation and illumination changes [5]. The choice of methods is informed by the computer vision application under consideration. It has been demonstrated recently that features identified by SIFT are highly distinctive and invariant to image scales and rotations, and partially invariant to a change in illumination [19]. It is indicated that using multiple images might help to solve some problems associated with stereo matching. However, more information may also carry the risk of increased uncertainties. Repeating features is a common problem encountered by stereo matching algorithms that apply feature-based method for visible light images.

1.1 Repeating Features

The images in Figure 1 present a good example of repeated features, which are commonly found in stereo pairs, where Left view and Right view are the images obtained at different views. To study the effect of local similarity, consider the repeating features Object 1 and Object 2, illustrated in the two views in Figure 1.



Left view **Right view**

Figure 1: Repeating *features extracted from two different views*

By applying the stereo matching criterion on this pair of images, on one hand, and as these features are obvious and strong, the opportunity to match them is significant. On the other hand, unless special care is taken, Object 1 in the Left view would have an equal opportunity to match with Object 1, Object 2 or other similar features in the Right view. As a result, an error in matching could occur. This finding is exacerbated when considering overlapping structures commonly found in cluttered visible light images.

1.2 Keypoint Matching

The SIFT algorithm adopts the fast nearest-neighbour method to identify the best match for a particular feature from a database of features. Since the keypoint is described by its descriptor, the nearest neighbour is defined as the keypoint with minimum Euclidean distance for the invariant descriptor vector [20]. Nevertheless, numerous features from an image will not match correctly in the derived keypoints database for the reason that they were not detected in the training images. Lowe [1] mentioned to discard all matches in which the distance ratio between closest neighbour to that of the second closest neighbour is greater than 0.8. This ratio removes 90% of the false matches on the other hand it discards less than 5% of the correct matches. Even though 90% of false matched are discarded, the remaining 10% of false matches might be a problematic for a particular image application. To maximise the potential applicability of SIFT, additional boundary conditions of search for corresponding are proposed. These added criteria will tighten the support of the standard SIFT.

2 Materials and Methods

In our work, additional bounding criteria of a disparity window and an epipolar line constraint are employed. The former is defined as the intersection of the epipolar plane with the image plane, while the latter concerns the nominally zero vertical disparity. i.e. the epipolar line is along the image y-axis (vertical in the display), while the disparity window is along the image x-axis (horizontal in the display). Both criteria are shown in Figure 2.

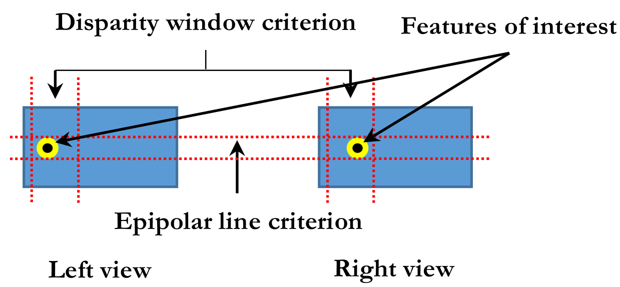


Figure 2: Two *perspective views and their corresponding features of interest*

Taking into account various practical fluctuations a tolerance of ±1 pixel deviation in the y-axis coordinate position (vertical in the display) is employed to accommodate a practical epipolar line criterion [19]. To further limit the search space a disparity window criterion is introduced. The window size in pixels is determined by the angle separation between views.

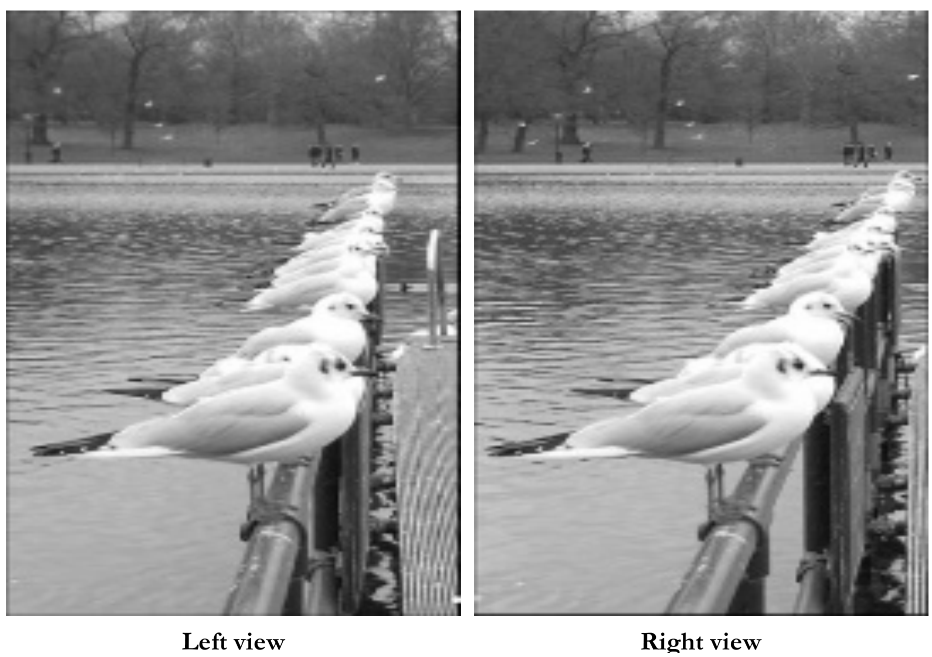


Figure 3: Corresponding *pair of images employed in this study*

It should be noted that for comparative purposes the new criteria are applied to corresponding pairs that have already satisfied the standard SIFT criteria. Care was taken when stereo images were selected to ensure overlapping. Repeating features were also taken into consideration to ensure that the standard SIFT is supported by the new added criteria. An example of stereo image are shown in Figure 3.

3 Results and Discussion

Matched features are categorised into two groups; negative and positive matches. The positive matches are the matches that satisfy the standard SIFT, epipolar and disparity window criteria while the negative matches satisfy the standard SIFT criterion but violate either the epipolar line or disparity window criteria. Figures 4 represents the positive matches indicated by horizontal green colour lines connecting the corresponding pairs, while the negative or erroneous matches are presented in Figure 5 and are shown as red colour lines.

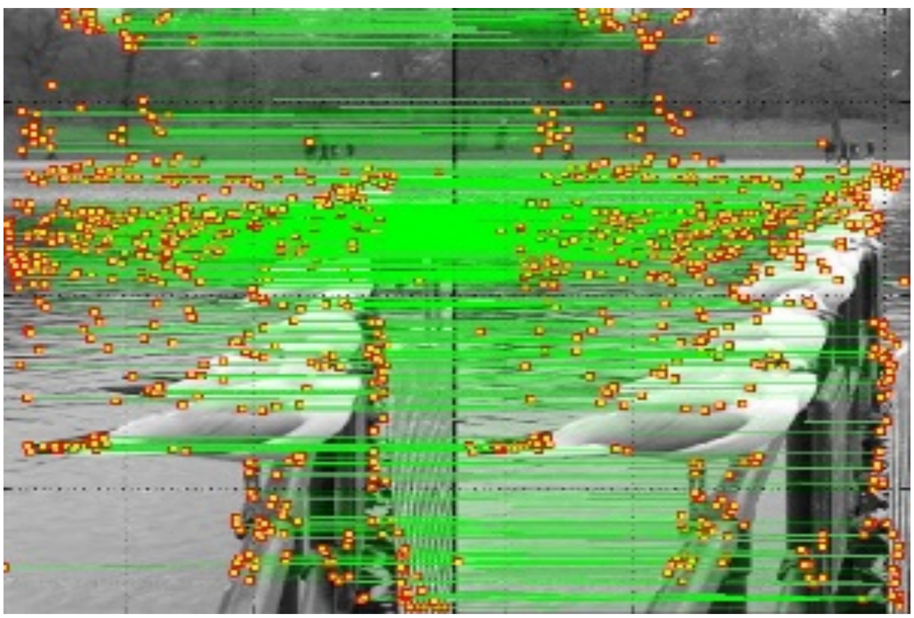


Figure 4: Proposed *matching criteria results positive matches only*

The application of the criteria tighten the support of standard SIFT. The increase in negative matches (and the corresponding decrement in positive matches) is the expected consequence of logically „ANDing‟ the criteria. It is important to note that only the matches in Figure 4 (shown as green bars) which meet standard SIFT, epipolar line and disparity window criteria are considered as positive matches. The matching procedure described above has been repeated for 180 stereo pairs and the numbers of positive and negative matches for each matching criterion are presented in Figure 6.

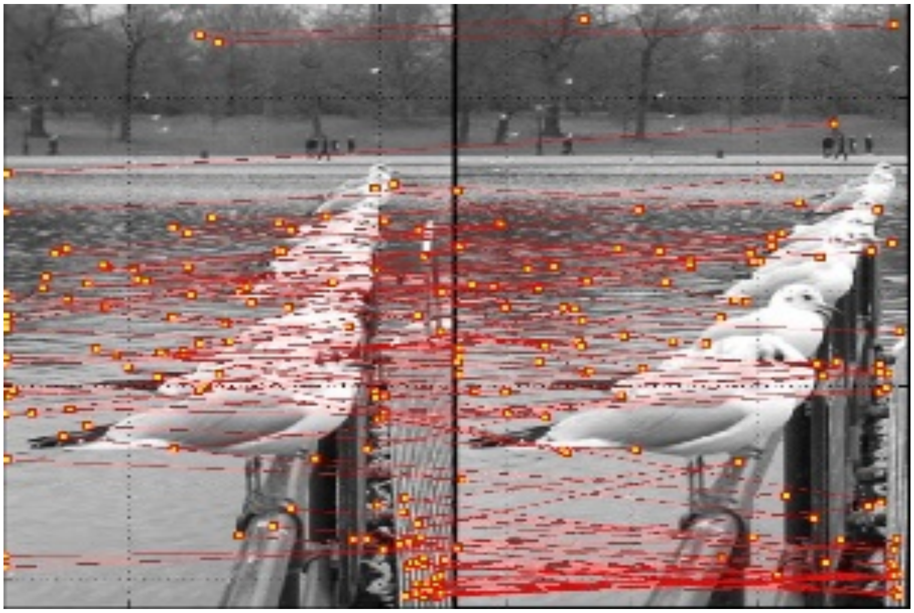


Figure 5: Proposed *matching criteria results negative matches only*

The bar chart in Figure 6 has been plotted to demonstrate the effectiveness of new criteria in rejecting incorrect matches. The first bar in black represents the average number of matches generated by the 180 stereo pairs, which corresponds to matches that have met the standard SIFT criteria.

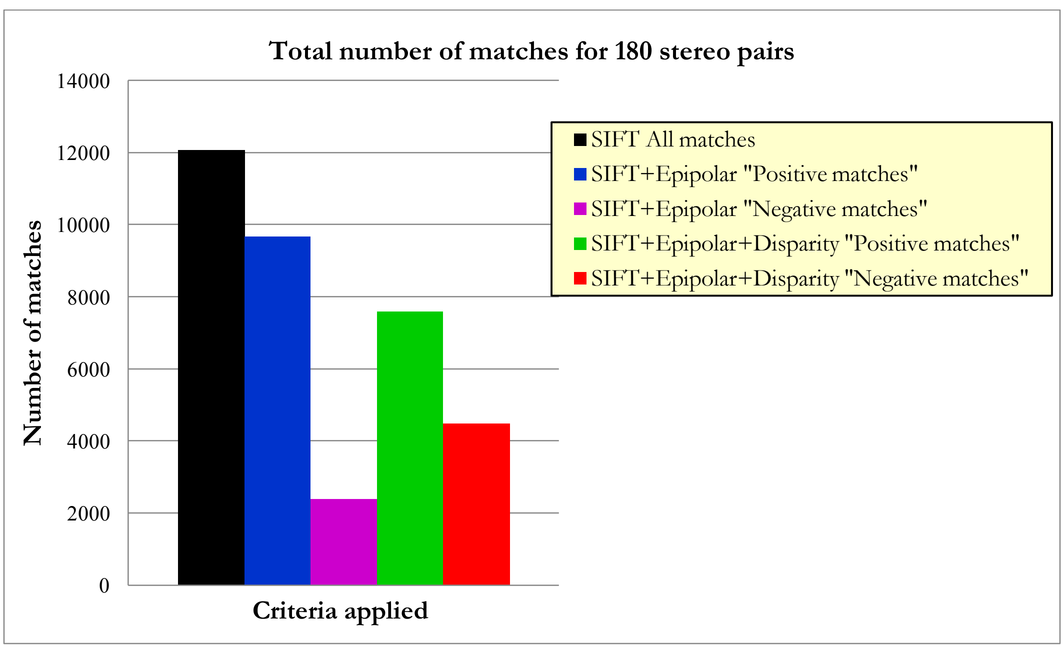


Figure 6: Total *number of matches for 180 stereo pairs*

Similarly, matches that have satisfied the epipolar line criterion, matches that have failed the epipolar line criterion, matches that have fulfilled the epipolar and disparity window criteria, and matches that have violated either or both new added criteria have been computed and plotted sequentially in the bar chart revealed in Figure 6. The proposed criteria have demonstrated that they can remove 37.2% of unfaithful matches i.e. 19.8% are attributed to the epipolar line criterion and a further 17.4% attributed to the disparity window criterion.

4 Conclusions and Future Work

The material presented in this paper assessed the performance of optimized SIFT when dealing with stereo pairs. The potential of SIFT to locate correspondences in stereo pairs is established and quantified for 180 stereo pairs. The performance of SIFT is significantly enhanced by applying two additional criteria namely; a disparity window and an epipolar line constraint. . Each pair of images is analysed twice to accommodate either perspective view as the reference view. Experiments revealed that around 37% of unfaithful matches were removed. The appropriateness of the additional criteria is supported by the matching results organised in Figure 4, 5 and 6.

Solving the correspondences problem is an ill posed problem in computer vision applications. It has been established that performance of optimized SIFT significantly reduces when presented with spatially simple images. Therefore, a rigorous analysis of the SIFT parameters to increase the robustness and density of keypoints would enhance the fidelity the matching result. Also, it might be worthwhile to combine the optimized SIFT algorithm with other feature matching techniques so more keypoints are generated.

References

[1] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision,* vol. 60, no. 2, pp. 91-110, 2004.

[2] Y. Chen and J. Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, pp. 1252-1267, 2002.

[3] M. Brown, R. Szeliski and S. Winder, "Multi-image matching using multi-scale oriented patches," *in Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 1, pp. 510-517, 2005.

[4] D. Bespalov, W. C. Regli and A. Shokoufandeh, "Local feature extraction and matching partial objects," *Comput. -Aided Des.,* vol. 38, pp. 1020-1037, 2006.

[5] A. Baumberg, "Reliable feature matching across widely separated views," *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000.

[6] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," *in European Conference on Computer Vision,* vol. 4, pp. 128–142, 2002.

[7] A. Klaus, M. Sormann and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *International Conference on Pattern Recognition,*vol. 18, pp. 15-18, 2006.

[8] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vision Comput,* vol. 22, pp. 761-767, 2004.

[9] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1615-1630, 2005.

[10] C. B. Perez and G. Olague, "Learning invariant region descriptor operators with genetic programming and the f-measure," *in International Conference on Pattern Recognition*, 2008, .

[11] C. Harris, "Geometry from visual motion," 1993.

[12] C. Harris and M. Stephens, "A combined corner and edge detector," *in Alvey Vision Conference*, pp. 50, 1988.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," *in Iccv*, pp. 1150, 1999.

[14] P. Scovanner, S. Ali and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," *in Proceedings of the 15th International Conference on Multimedia*, pp. 357-360, 2007.

[15] P. Moreno, A. Bernardino and J. Santos-Victor, "Improving the SIFT descriptor with smooth derivative filters," *Pattern Recog. Lett.,* vol. 30, pp. 18-26, 2009.

[16] Z. Qi, R. Ting, F. Husheng and Z. Jinlin, "Particle Filter Object Tracking Based on Harris-SIFT Feature Matching," *Procedia Engineering*, vol. 29, pp. 924-929, 2012.

[17] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, .

[18] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vision Image Understanding*, vol. 110, pp. 346-359, 2008.

[19] Z. Feng, B. Yang, Y. Chen, Y. Zheng and T. Xu, "Features extraction from hand images based on new detection operators," *Pattern Recognit*, 2010.

[20] A. Fusiello and L. Irsara, "Quasi-euclidean uncalibrated epipolar rectification," in *International Conference on Pattern Recognition (ICPR),* 2008.