Asian Journal of Computer and Information Systems (ISSN: 2321 – 5658)
Volume 04– Issue 01, February 2016

Image Registration Techniques Based on the Scale Invariant Feature Transform for Aerial Surveillance Images

Vaithiyanathan V[1], Divya Lakshmi K[2] and Radha Rammohan S[3,*]

[1] Professor, School of Computing, SASTRA University,
Thanjavur, India

[2] Research Scholar, School of Computing, SASTRA University,
Thanjavur, India

[3]Assistant Professor, Salalah College of Applied Sciences,
Sultanate of Oman

[*]*Corresponding author's email: radha-mohan.sal [AT] cas.edu.com*

ABSTRACT—*Aerial surveillance plays a vital role in continuous monitoring of areas where human intervention is of higher risk. Image registration is a process to bring several images of a scene to a single co-ordinate system. It is widely done using the local features extracted from the images. Local features of an aerial surveillance image is also used in several computer vision tasks such as scene matching, object identification and object tracking. The features identified must be repeatable across several transformations on the image. This paper aims to find the most repeatable local feature from the existing state of art affine invariant features : Harris Affine, Hessian Affine. They are described using the Scale Invariant Feature Transform (SIFT). The dataset comprises aerial images captured from low altitude unmanned aerial vehicles using thermal and visual sensors. The ground truth between the images of a scene is estimated manually using control points, in the form of a homograph matrix. The repeatability of the various features are estimated using the ground truth estimated. The Hessian Affine feature outperforms in performance compared to the other affine invariant feature detector making it a potential candidate for the various computer vision tasks on aerial surveillance images.*

Keywords— local features, aerial surveillance images, image matching, affine invariance

---

1. INTRODUCTION

Aerial surveillance involves capturing of images from airborne sensor over a specified region for the purpose of monitoring the region for specific purposes such as intrusion detection or development planning. The purposes are achieved through various computer vision processes such as object tracking, mosaicing, scene matching and change detection. The computer vision processes combine the information from the overlapping images captured by the airborne sensors from a scene over a period of time. The variations between the images of a scene poses a challenges for the algorithms that combine information from the images.

Variations can be classified as- Geometric, Photometric, Occlusion. Geometric variations are caused by the spatial transformation of the pixels of the image. There geometric variations are modeled mathematically by means of transformation functions. There are different types of geometric functions - Rigid, Similarity, Affine, Projective, Elastic. The transformations involve the operations of rotation, translation, uniform or non-uniform scaling. The transformations can be uniform throughout the image or non-uniform. Occlusion is a variation where the coverage of an object varies between the images. Shadows or clouds in the images that hides portions of objects in the image also causes occlusion. The above said variations are overcome by encoding the image using local features.[1]

2. METHODOLOGY

Local features are key points and their local neighborhood in the image. There are different techniques for choosing the key points, detecting their neighborhood region, and encoding the neighborhood region. [2]

2.1. Choosing Key Points

There are different measures for choosing the key points in the image. Corner based and blob based are two of the important measures. Corner based measures are based on the first derivative and blob based measure are based on second derivative. Harris, Hessian, Plessey, Kitchen and Rosenfield are few of the measures for key point selection. The above said measures are sensitive to scale changes. Tolerance to scale variations is incorporated by means of simulating the image over different scales by means of Gaussian blurring and identifying key points over the scale space. Harris-Laplace, Hessian-Laplace, Difference of Gaussian, Laplace of Gaussian are some of the measures that incorporate the concept of scale space generation and hence are tolerant to scale variations. The Harris and Hessian Affine are the cornerness measures used in the experiments and are given in (1) to (4).

$$Ha(x,y) = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{1}$$

$$HarrisMeasure = Det(Ha) - \alpha Trace^2(Ha) \tag{2}$$

$$He(x,y) = \begin{bmatrix} I_{xx} & I_{xy} \\ I_{xy} & I_{yy} \end{bmatrix} \tag{3}$$

$$HessianMeasure = Trace^2(He) / Det(He) \tag{4}$$

2.2. Detecting Neighborhood Region

Neighborhood region of a chosen key point should be small enough to handle occlusion and large enough for it to be unique. Affine invariant local features find elliptical region around the key point that are invariant to affine transformation that involves rotation, translation, uniform and non-uniform scaling. Elliptical region is iteratively computed till the ratio between the eigen values of the second moment matrix over the region equals to one.

2.3. Encoding Neighborhood Region

The affine invariant local features normalize the elliptical neighborhood region to circular region using the second moment matrix obtained while detecting the neighborhood region. The circular region obtained is encoded using the gradient orientation histogram which is the descriptor of the Scale Invariant Feature Transform (SIFT) . The circular region is thus transformed to vector of size 128.[3]

The key point along with the descriptor represent the local feature of the image. The local feature thus detected should be repeatable across different transformation, unique within the image. The size of the local feature should be favorable for real time matching. The computation time of the feature must meet the timing constraints of the application for which the local feature must be computed on the fly such as in real time scene matching.

3. EVALUATION

The performance of the local features- Harris Affine and Hessian Affine are measured over dataset of visual and thermal images. The executables of the feature detection algorithms are taken from the author's web site and is used for the evaluation procedure.

3.1. Performance measures

The local features derived are validated using the performance measure of repeatability and number of correspondences. Repeatability checks for the presence of the local feature in the corresponding location in the transformed image. The number of correspondences performance measure, compares the similarity of the descriptors in the local features that are present in the corresponding location.[4]

3.2. Evaluation data set

The data set comprises ten pairs of images captured from low altitude unmanned aerial vehicle. Each pair of image comprises a thermal and visual images of a scene.[5] The view point between the images also vary drastically. The ground truth of transformation between each image in a pair is manually estimated by means of control points in the form of homograph matrix. Figure 1 shows sample image pair from the data set and the images registered with the help of the homograph matrix estimated using the control points.

Figure 1 Image pair from dataset and images registered using Homograph estimated

3.3. *Results*

The image pairs were sorted according to the scale and angle variations between the image pairs. Curves of repeatability and number of correspondences were obtained for the local features of Harris affine and Hessian affine which is shown in Figures 2 to 5.

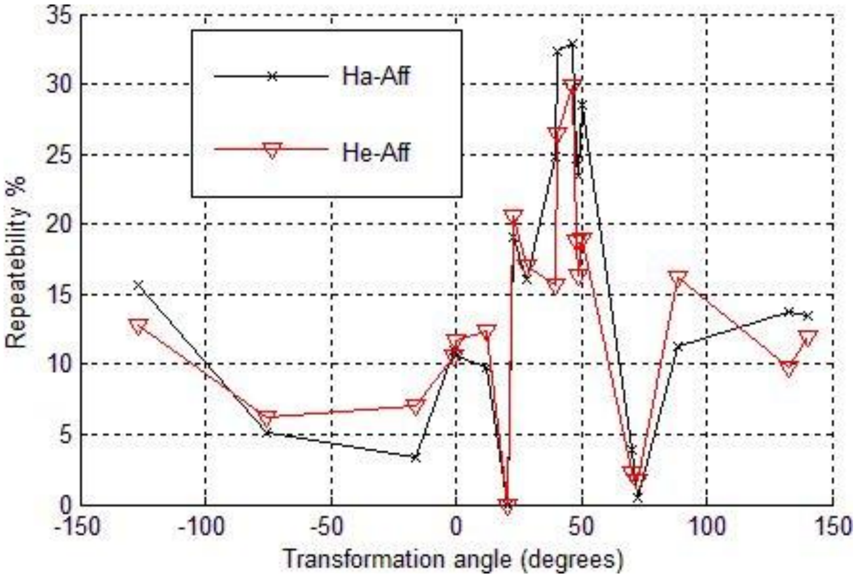

Figure 2 Transformation Angle Vs Repeatability

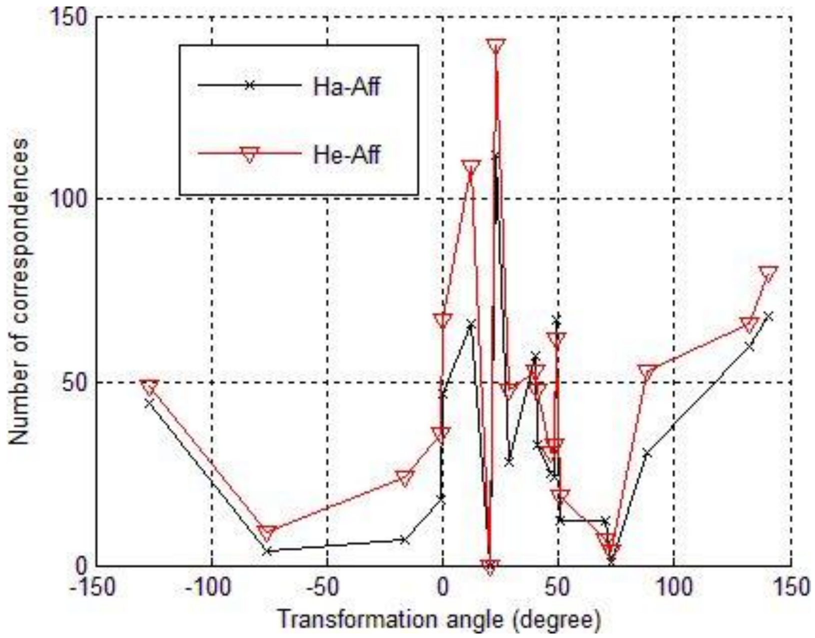

Figure 3 Transformation Angle Vs Number of Correspondences

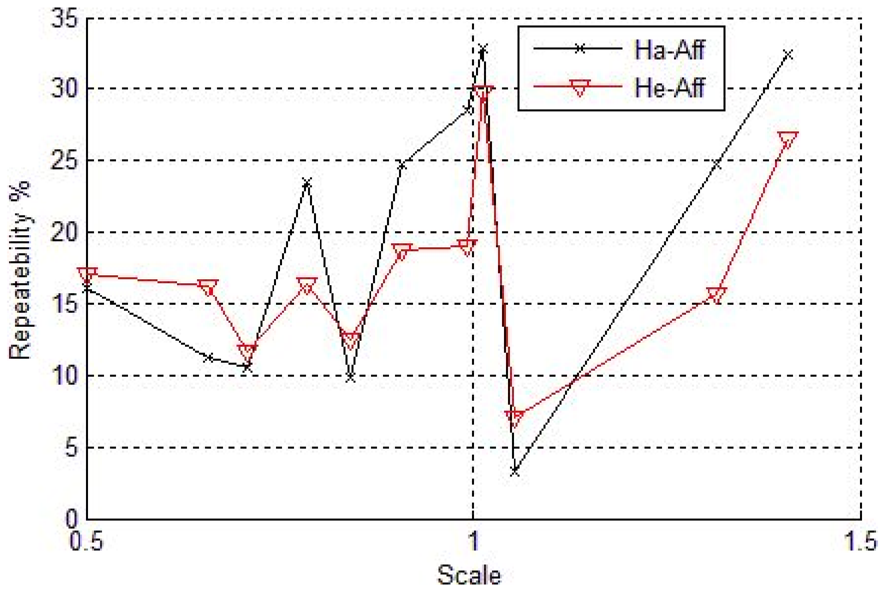

Figure 4 Scale Vs Repeatability

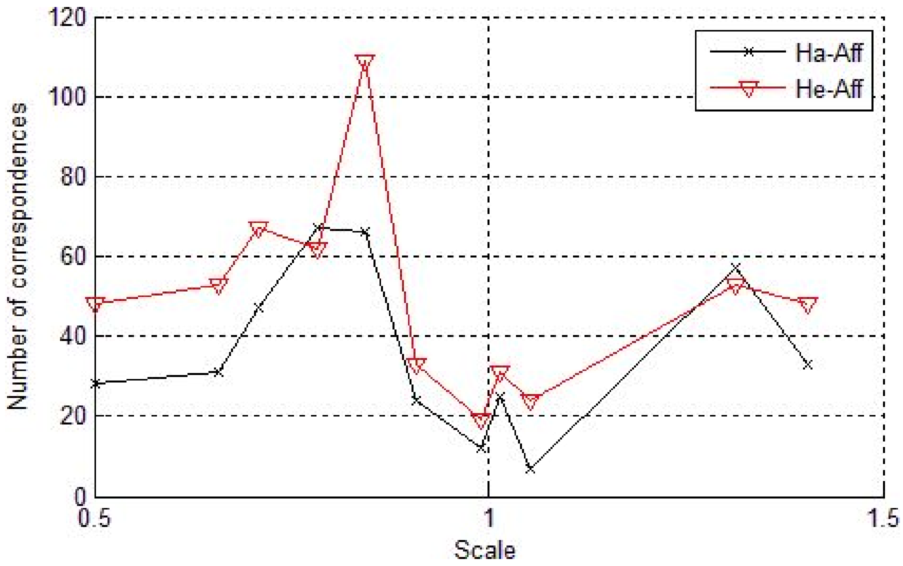

Figure 5 Scale Vs Number of Correspondences

4. CONCLUSION

It is observed from the performance curves that the Hessian Affine local feature outperforms the Harris Affine Local feature for most of the scenario in the aerial images. The blob based cornerness measure of Hessian Affine is thus found to be more effective for identifying the key points in the aerial images captured for the purpose of surveillance when compared to the corner based measure in the Harris Affine. Future work will be to improve the local features for higher repeatability with lesser time complexity for computing the local feature.

5. ACKNOWLEDGEMENT

This material is based on the research work financially supported by the Council of Scientific and Industrial Research (CSIR), the premier research and development organization in India, under the Senior Research Fellowship scheme.

6. REFERENCES

[1] Lisa Gottesfield Brown, " A study of image registeration techniques", ACM Computing Surveys, vol. 24, no. 4, pp. 325-376, December 1992.

[2] Krystian Mikolajczyk, and Cordelia Schmid, " Scale & Affine Invariant Interest Point Detectors", International Journal of Computer Vision, vol. 60, no. 1, pp. 63-86, October 2004.

[3] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November 2004.

[4] Krystian Mikolajczyk, and Cordelia Schmid, "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 27, pp. 1615-1630, October 2005.

[5] Saeed Yahyanejad, and Bernhard Rinner, "A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 104, pp. 189-202, June 2015.