Volume 5, Issue 5, May 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Invariance Analysis and Comparison of Object Detection Algorithms

Aaditya Rakesh*
Deptt. of Mathematics, IIT-Delhi
India

Dr. Savita Goel
Computer Services Centre, IIT-Delhi
India

Abstract— Object detection algorithms encounter various challenges in the form of transformations and ambient distortions(like blur, noise, illumination) which lead to significant amount of failures in successful object recognition. This paper discusses the factors involved in developing invariance with regards to these aspects in object detection algorithms. It presents an exhaustive analysis of two object detection and recognition algorithms: Scale Invariant Feature Transform (SIFT) and Affine-Scale Invariant Feature Transform (ASIFT). Focus is laid on studying their performance with regards to affine transformations.

Keywords— Detector, Descriptor, SIFT, ASIFT, Keypoint, Feature

I. INTRODUCTION

With ever growing amounts of data in the internet world, it has become imperative to search through it and find meaningful relationships so that computational capacities can be leveraged for human benefit. In this regard, processing of visual data using mathematical models and computational structures has become a pivotal field to extract information and interpret it meaningfully. Object detection and recognition systems play a fundamental role in this direction. This field has become an indispensable tool in research in various fields ranging from medical science, astronomy, satellite imaging to more anthropocentric processes like automobiles, visual data storage and related issues. Detector and Descriptor can be regarded as the fundamental building blocks of all well established object detection algorithms functioning today. The detector is used in the first stage and computes various properties of the image to isolate and extract certain specific candidate points in input images, which are distinctive with regards to their neighbourhood in response to various distortions and environmental transformations. The descriptor is then employed to attach an invariant description to each key point which is unique enough to characterise that point in different images. Matching can hence be conducted by comparing the descriptors in the images within certain range of deviation. Filtering is done at various levels to reject points which aren't distinctive enough and hence are liable to give erroneous results in the matching stage. The two key operations should be developed to be as invariant as computationally possible for successful results.

Lowe[1] presented Scale Invariant Feature Transform(SIFT) which is a full scale invariant object detection method. It extracts distinctive feature keypoints from the image data which is relatively invariant to scale, rotation and partially invariant to illumination.

Affine invariant Methods can be considered as a generalisation of the scale-invariant methodologies to non-uniform scaling, with a different scaling factor in two orthogonal directions and without preserving angles. This impacts the localisation and the scale and also shape of local structures. Hence, scale-invariant detectors don't give good results in the case of significant affine transformations. The SIFT detector, proposed by Lowe is hence deductively less invariant to affine transforms than the Hessian-Affine[5] and the Harris-Affine detectors [4]. However, when combined with the SIFT descriptor, its overall affine invariance turns out to be comparable.

Morel[3] proposed an affine invariant method to cover the parameters ignored by SIFT to correct find matches and resist affine transformational changes. Affine-SIFT build upon SIFT to simulates the rotation of camera's optical axis. Rather than normalising all the six affine parameters as performed in MSER, Harris-Affine and Hessian-Affine and some more detectors, ASIFT *simulates* three parameters and *normalises* the rest. The objective of this paper is to perform a critical analysis and study the performance of the two algorithms under changes in various parameters of operation. In the first section, an overview of SIFT and ASIFT object recognition algorithms has been explained. After discussing their different aspects, a critical performance analysis has been done for both of the algorithms. Analysis has also been performed by varying different like parameters like blur level, rotation and viewpoint angle to study their effect on the algorithms . The conclusion elucidates the various observations and discusses further research avenues.

II. OVERVIEW OF ALGORITHMS

Scale invariant Feature Transform

The biggest advantage of Scale invariant Feature Transform(SIFT) is scale invariance. The fundamental objective of detecting keypoint locations robust to scale change can be proficiently computed by searching for features across all possible scales, using a continuous iterative methodology for creating a scale space(to model zoom), the Gaussian function being the optimal operator for this process. Figure 1 gives a representation of the methodology for creation of a scale space.

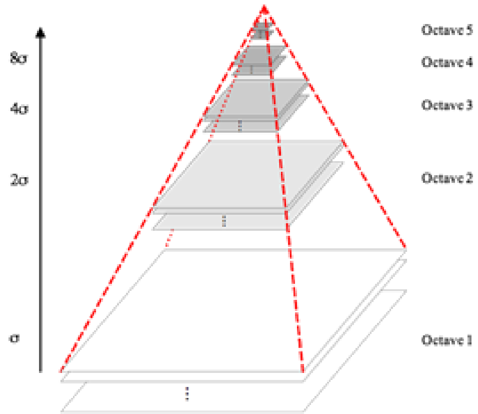

Fig. 1 Scale-space pyramid Construction

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)}$$

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)$$

The entire methodology comprises of the following procedures:(1) Scale Space Extrema Detection, (2) Candidate Keypoint Filtering (3) Creation of Orientation Histogram, (4) Formulation of Image Descriptor and (5) Keypoint matching. For its operations, SIFT makes use of the grayscale information of image data. It employs the Difference of Gaussian(DoG) region detector[2] on account of its rotation, translation and scale invariance. This is followed by rejection of unwanted keypoints to reduce computational load in the upcoming stages. This is done by rejecting low contrast points(using thresholding) and deleting points on edges using Hessian matrix(due to strong response of Gaussian function to edges in the image). Also, sub-pixel accuracy is achieved by employing taylor series interpolation of second order to find or true matches. A descriptor based on the gradient orientation distribution in the neighbourhood region of the detected feature is employed to aid in matching of these keypoints across varied images. Figure 2 explains the formation of an orientation histogram for a keypoint by computing the magnitude and orientations in its neighbourhood. SIFT uses a description based on local information skipping the use of global information and features. In this, it uses a 128-dimensional vector to describe each keypoint.An optimal match for each of the candidate keypoints is computed by using second nearest neighbour approach in a similar set of keypoints from the other image. This can be defined as a keypoint with minimum Euclidean distance for its attached description vector.

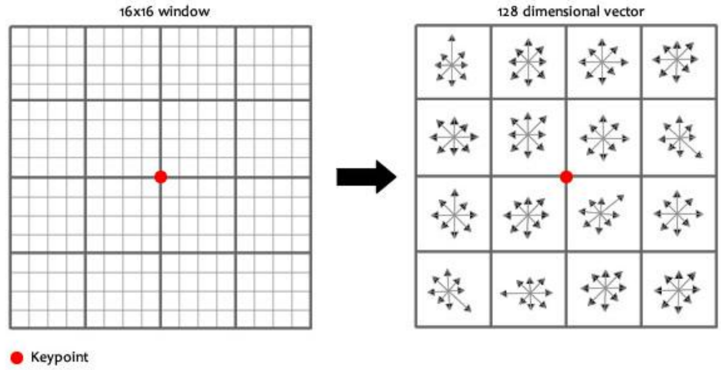

Fig. 2 Descriptor calculation based on Orientation Histogram

Affine Scale invariant Feature Transform

Affine Scale invariant Feature Transform (ASIFT) essentially simulates different image views that can be obtained by varying the parameters of the two camera axis orientations, namely the latitude and the longitude angles, which are ignored by the SIFT algorithm by making suitable assumptions. Then it covers the remaining four parameters by the SIFT method for the various image views. On a computational level, ASIFT simulates the two camera attributes, and applies SIFT which then simulates the scale and normalises the rotation and the translation. A significant attribute for judging the performance of affine recognition is the transition tilt. Transition tilt measures the degree of viewpoint change between different views. Figure 3 explains the various aspects when an image undergoes affine transformation.

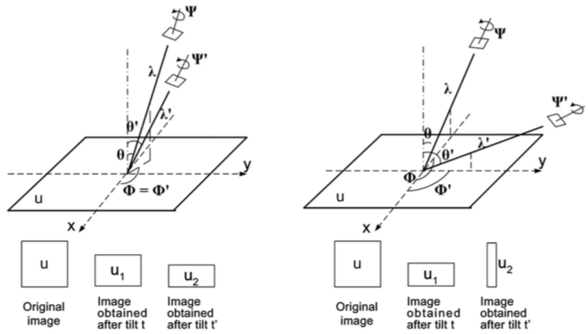

Fig. 3 Transition tilt in an image and its effect on affine parameters

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A = H_{\lambda}\, R_1(\psi)\, T_t\, R_2(\phi) = \lambda \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

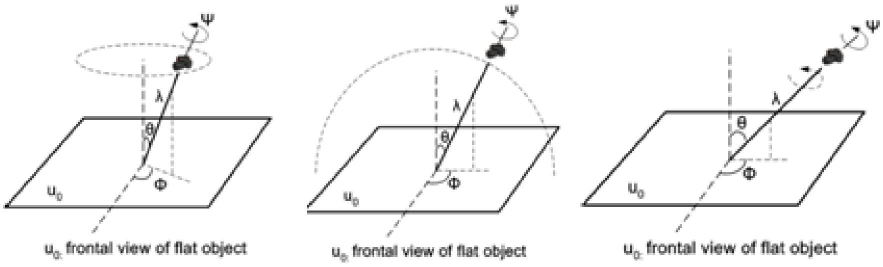

Fig. 4 Under affine transformation modelling and the effect of variation of varied parameters

Each image is transformed by simulating all possible affine variations that can be developed by changing the orientation of the camera optical axis. Figure 4 shows how the camera position change with variations in the affine parameters. The transformations are based on two characteristics: the longitude φ and the latitude θ. The images undergo φ-rotations followed by corresponding tilts with parameter $t = 1/\cos\theta$. A tilt by t in the direction of x is the operation $u(x, y) \rightarrow u(t.x, y)$. For computing tilt, directional t-subsampling in the x-direction is used. The rotations and tilt processes are used for a finite number of latitude and longitude angle values in a way that the results are close to any view that can be generated by using other values for the properties. The resulting image pairs are then matched using the scale invariant feature transform (SIFT) object recognition algorithm. Although ASIFT descriptors are calculated from the rotation and tilt transformed images, their nature in terms of information content and structure is the same as normal SIFT descriptors.

III. COMPARATIVE PERFORMANCE ANALYSIS

The analysis has been performed on 256x256 standard images. Both the algorithms are used to extract features and match keypoints. Various parameters are changed to see their impact on the results.

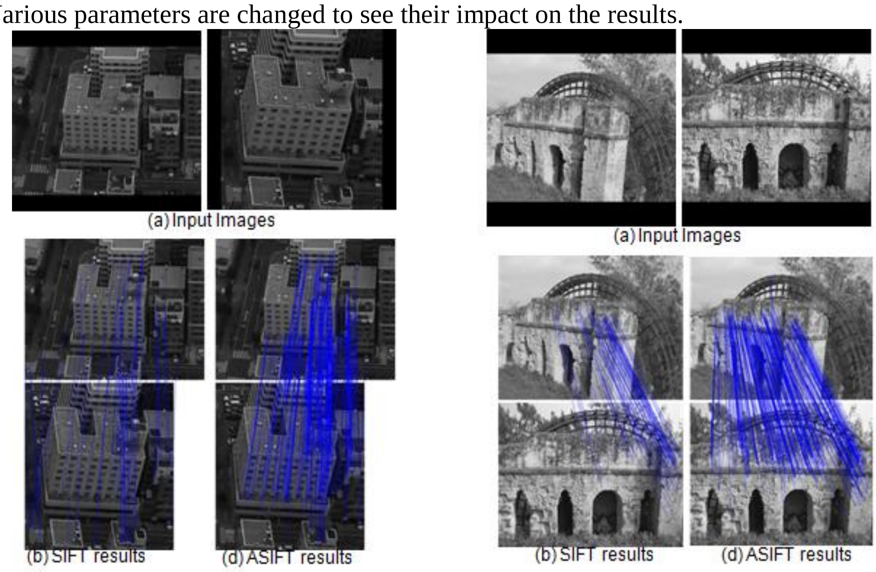

Fig. 5: SIFT and ASIFT performance on different structural images

Figure 5 demonstrates that in case of affine transformed images, ASIFT performs much better than SIFT in regards to number of matches. Also the correct matching rate of ASIFT is better as compared to SIFT.

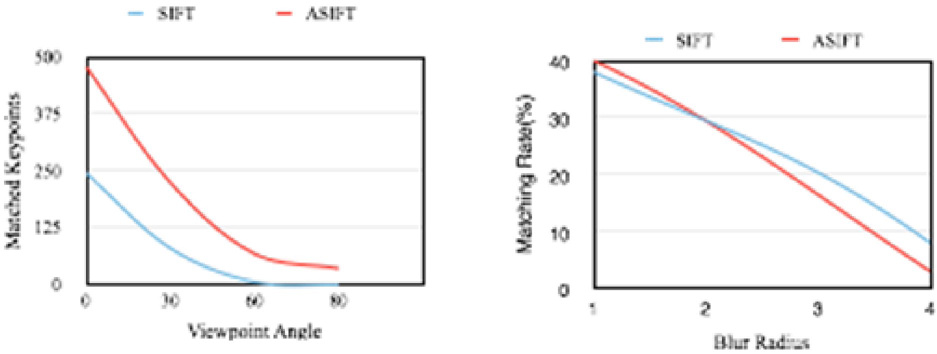

Fig. 6 SIFT and ASIFT performance with regards to change in viewpoint angle and the level of initial blurring in the two images

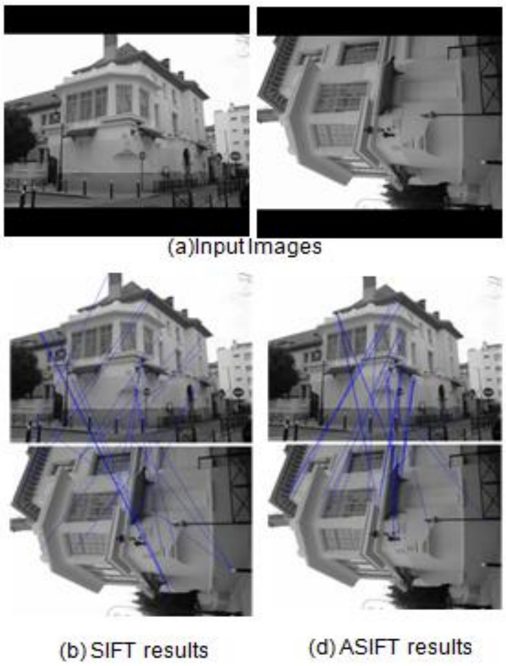

Fig. 7 Performance of the two algorithms under rotation change

Figure 6 depicts the performance levels of the two algorithms with change in viewpoint angle of the images and varying the blur level of the initial input image data. SIFT performs well until the change in viewpoint angle is under 30 degrees. The number of matches drastically comes down on further increasing the angle. ASIFT gives commendable performance even at an angle of 60 degrees after which its performance starts to decrease in a moderate manner. Both algorithms give a good and comparable performance for moderate levels of blur. Matching rate implies the ratio of number of matches to the total number of keypoints detected. This can be attributed to their robustness owing to their scale space representation and normalisation schemes. Figure 7 clearly depicts that the performance of both the algorithms is comparable under rotation change. This is on account of inherent rotational invariance in SIFT.

IV. Conclusions and Future Scope

From the results obtained, it is gathered that under similar conditions, SIFT has the optimal performance in regards to its predecessors. ASIFT gives better results than others in case of a large amount of affine transformation. Following conclusions can be drawn form the work carried out:

- Under drastic illumination change, these methods are unable to give effective results and fail to match the object in question.
- ASIFT and SIFT give similar results under normals blur levels. Their performances start to decrease moderately as the level of blur increases.
- ASIFT and SIFT give similar results under change of scale and rotation.
- ASIFT is quite robust to change in viewpoint. SIFT has, though lower than ASIFT, appreciable performance under low viewpoint angle change, but they fail when the change is substantial. The performance of SIFT starts deteriorating drastically when the viewpoint changes go beyond 30 degrees. The correct matching rate of ASIFT is, however, stable under different angles, even upto 65 degrees.

This work provides various avenues to build upon and take up this field further. A further evolved methodology for object recognition with full boundary detection can be developed by combining affine scale invariant feature transform (ASIFT) and a region merging algorithm. A robust region merging algorithm can be used to detect the object with full boundary in the other images based on ASIFT keypoints and a similarity measure for merging regions in the image. Also since most of the operations in our algorithms are computed independent of each other, there is immense scope of parallel processing to make these algorithms optimal for real time applications.

REFERENCES

[1] D. Lowe, "*Distinctive image features from scale-invariant keypoints*", International Journal of Computer Vision, Volume 2, No. 60, 2004 pp. 91–110.

[2] D. G. Lowe, "*Object recognition from local scale-invariant features*" in Proceedings of the International Conference on Computer Vision, 1999 pp. 1150–1157

[3] Jean-Michel Morel and Guoshen Yu, "*ASIFT: A New Framework for Fully Affine Invariant Image Comparison*", SIAM Journal on Imaging Sciences Volume 2, Issue 2, 2009 pp 438-469.

[4] Lindeberg, "*Feature detection with automatic scale selection*", International Journal of Computer Vision, Volume. 30, No. 2, 1998 pp. 79–116.

[5] T. Tuytelaars and L. Van Gool, "*Matching widely separated views based on affine invariant regions*", International Journal of Computer Vision, Volume. 1, No. 59, 2004, pp. 61–85.

[6] Tinne Tuytelaars and Krystian Mikolajczyk, "*Local Invariant Feature Detectors: A Survey*", Foundations and Trends in Computer Graphics and Vision Volume 3, No. 3, 2007 pp177–280.