IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

*INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING*
*Vol. 4, Issue 7, July 2016*

# Object Recognition using SIFT Keypoints

**Sahil Dalal[1], Preeti Meena[1]**

Department of ECE, Delhi Technological University, Delhi, India[1]

**Abstract:** This research paper presents a novel method for the identification of some object in a video using the distinctive invariant features from images. This method uses reliable matching between different views of an object or scene. The features shows a robust matching across a particular range of affine distortion, change in 3D viewpoint, addition of noise, invariant to image scale and rotation and change in illumination. In this, the recognition for the object proceeds by matching individual features to a database of features from known objects using a technique called as scale invariant feature transform. This approach to recognition can robustly identify objects among clutter and occlusion while achieving near real-time performance.

**Keywords:** Difference of Gaussian, Keypoint descriptor, Object, SIFT.

## I. INTRODUCTION

Object recognition is an important aspect of many problems in computer vision for 3D structure from multiple images, stereo correspondence, and motion tracking. This research paper describes image features that have many properties which make them suitable for detecting different images of an object or scene. The features are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint.

Large numbers of features can be extracted from typical images with efficient algorithms. In addition, the features are highly distinctive, which allows a single feature to be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition.

The development of image matching by using a set of local interest points can be traced back to the work of [1] on stereo matching using a corner detector. The Moravec detector was improved by [2] to make it more repeatable under small image variations and near edges. Harris also showed its value for efficient motion tracking and 3D structure from motion recovery [3], and the Harris corner detector has since been widely used for many other image matching tasks. While these feature detectors are usually called corner detectors. [4] showed that it was possible to match Harris corners over a large image range by using a correlation window around each corner to select likely matches. The ground-breaking work of [5] showed that invariant local feature matching could be extended to general image recognition problems in which a feature was matched against a large database of images. Earlier work by the author [6] extended the local feature approach to achieve scale invariance. Then, there has been an impressive body of work on extending local features to be invariant to full affine transformations [7]. Now, in recent years, wide range of techniques are utilized for object recognition. These are color descriptors [8], genetic [9], unsupervised scale invariant learning [10], appearance information [11]. Some of the other techniques were also used in [12-17].

For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors. This paper will discuss fast nearest neighbour algorithms that can perform this computation rapidly against large databases.

This research paper is organised in the following sections as: Section II tells about the overview of the complete work using a block diagram. Section III tells about the different types of ECG signals used and the proposed method. Then in section IV, results are discussed followed by conclusion in section V.

## II. BLOCK DIAGRAM

Following are the major stages of computation used to generate the set of image features:

**Scale-space extrema detection:** The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

**Keypoint localization:** At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

**Orientation assignment:** One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

**Keypoint descriptor:** The local image gradients are measured at the selected scale in the region around each keypoint.

DOI 10.17148/IJIREEICE.2016.4742

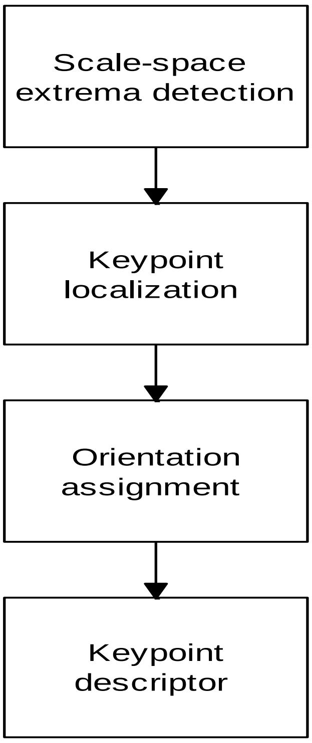

Fig.1. Block Diagram of the proposed work

These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

These steps are also shown in the form of a block diagram in the Fig.1.

This approach has been named the Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features.

## III.PROPOSED METHOD

A. Scale-space extrema detection

The scale space of an image is defined as a function, S(w, z, σ), that is produced from the convolution of a variable-scale Gaussian, g(w, z, σ), with an input image, i(w, z):

$$S(w, z, \sigma) = g(w, z, \sigma)*i(w, z) \qquad (1)$$

where ∗ is the convolution operation in g and i, and

$$g(w,z,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{w^2+z^2}{2\sigma^2}} \qquad (2)$$

To efficiently detect stable keypoint locations in scale space, we have proposed (Lowe, 1999) using scale-space extrema in the difference-of-Gaussian function convolved with the image, DoG(x, y, σ), which can be computed from the difference of two nearby scales separated by a constant multiplicative factor n:

$$\begin{aligned} DoG(w, z, \sigma) &= (g(w, z, n\sigma)-g(w, z, \sigma))*i(w, z) \\ &= S(w, z, n\sigma)-S(w, z, \sigma) \qquad (3) \end{aligned}$$

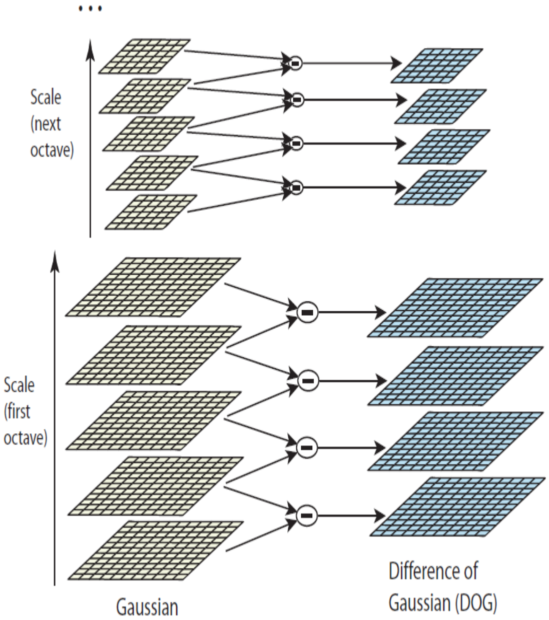

Fig.2. Difference of Gaussian in an image

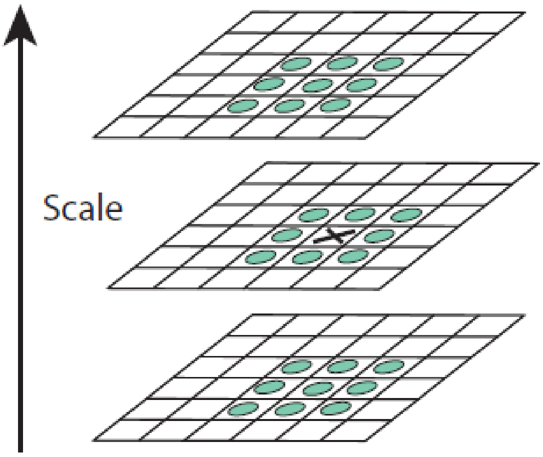

Fig.3. Detection of maxima and minima from DoG

B. Keypoint localization

Once a keypoint candidate has been found by comparing a pixel to its neighbours, the next step is to perform a detailed fit to the nearby data for location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

For fitting a 3D quadratic function to the local sample points to determine the interpolated location of the maximum, scale-space function, D(x, y, σ), is shifted as:

$$DoG(x)=DoG+\frac{\partial DoG^T}{\partial x}x+\frac{1}{2}x^T\frac{\partial^2 DoG}{\partial x^2}x \qquad (4)$$

where DoG and its derivatives are evaluated at the sample point and X = (w, z, σ)$^T$ is the offset from this point. The location of the extremum, $\hat{X}$ , is determined by taking the derivative of this function with respect to X and setting it to zero, giving

$$\hat{X} = -\frac{\partial^2 DoG^{-1}}{\partial x^2}\frac{\partial DoG}{\partial x} \qquad (5)$$

The function value at the extremum, DoG ( $\hat{X}$ ), is useful for rejecting unstable extrema with low contrast. This can be obtained by substituting equation (3) into (2), giving

$$DoG(\hat{X}) = DoG + \frac{1}{2}\frac{\partial DoG}{\partial x}\hat{X} \qquad (6)$$

For the experiments in this paper, all extrema with a value of |D( $\hat{X}$ )| less than 0.04 were discarded (as before, we assume image pixel values in the range [0,1].

C. Orientation assignment

By assigning a consistent orientation to each keypoint based on local image properties, the keypoint descriptor can be represented relative to this orientation and therefore achieve in-variance to image rotation. Fig.4 represents this orientation assignment.

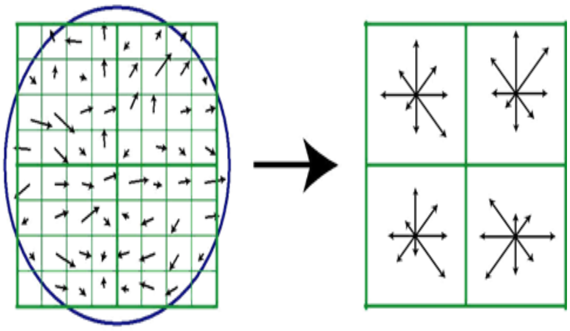

Fig.4. Orientation assignment in image gradient using keypoint descriptor

Following experimentation with a number of approaches to assigning a local orientation, the following approach was found to give the most stable results. The scale of the keypoint is used to select the Gaussian smoothed image, S, with the closest scale, so that all computations are performed in a scale-invariant manner. For each image sample, S(w, z), at this scale, the gradient magnitude, d(w, z), and orientation, φ(w, z), is pre-computed using pixel differences:

$$d(w,z) = \sqrt{(S(w+1,z)-S(w-1,z))^2 + (S(w,z+1)-S(w,z-1))^2} \quad (7)$$

$$\varphi(w,z) = \tan^{-1}\left(\frac{S(w,z+1)-S(w,z-1)}{S(w+1,z)-S(w-1,z)}\right) \quad (8)$$

D. Keypoint descriptor

The local image gradients are measured at the selected scale in the region around each keypoint.

These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. Fig.5 shows the object which is to be find out in the video.

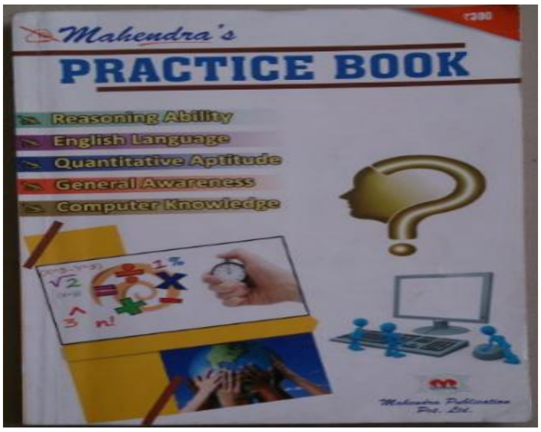

Fig.5. Object to be identified in the video

## IV. RESULTS & DISCUSSION

Now, in the results, frames of the video are shown in the Fig.6-12. In the figures, the object i.e., a book, whose keypoints are detected are shown in various positions and it can be observed that, using this method, the object is detected in all respects.

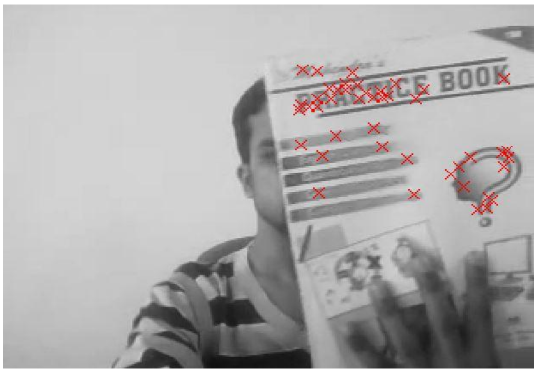

Fig.6. Frame 1

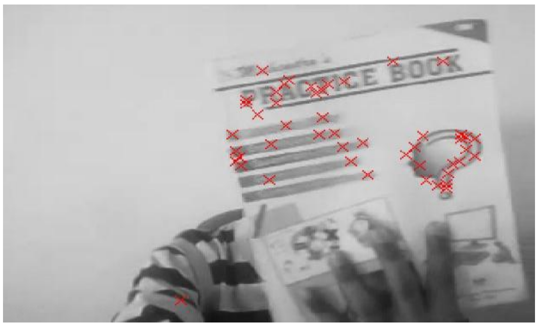

Fig.7. Frame 2

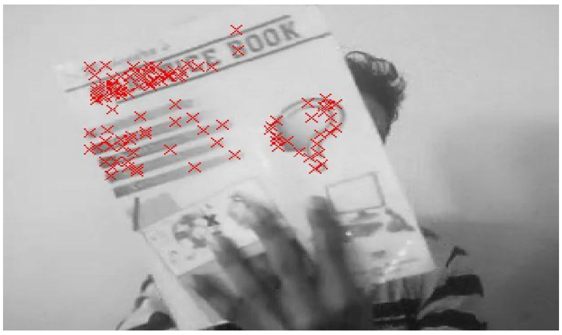

Fig.8. Frame 3

DOI 10.17148/IJIREEICE.2016.4742

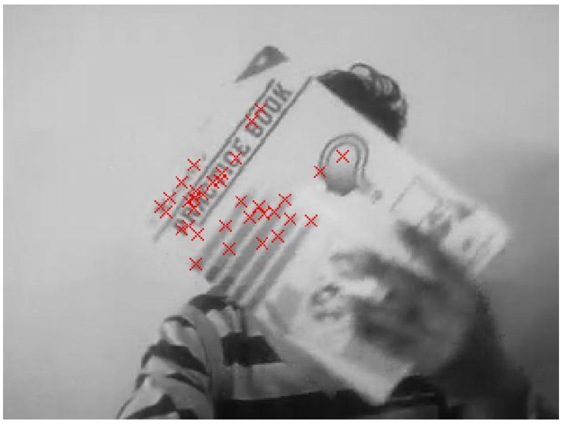

Fig.9. Frame 4

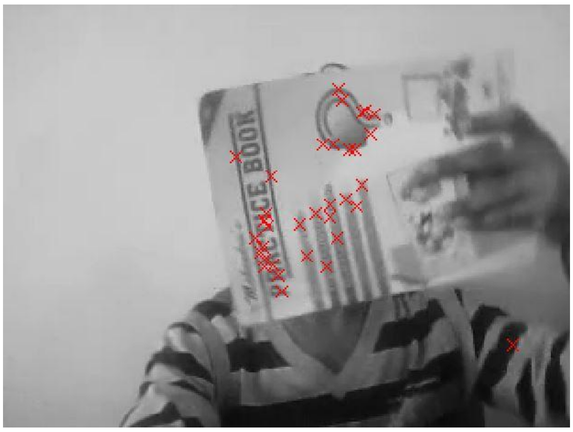

Fig.10. Frame 5

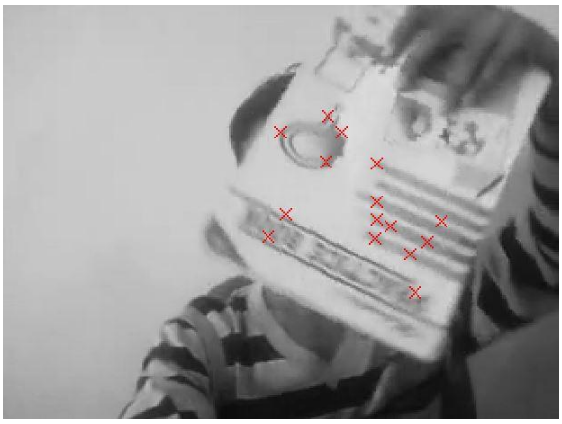

Fig.11. Frame 6

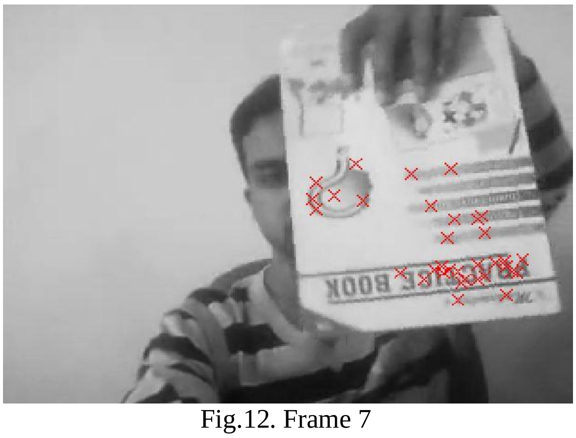

Fig.12. Frame 7

From these figures, recognition of the object is obtained with a successful recognition rate of more than 90%. Hence, the proposed technique of the novel method is helpful in detecting objects in videos.

## V. CONCLUSION

The SIFT keypoints described in this paper are particularly useful due to their distinctive-ness, which enables the correct match for a keypoint to be selected from a large database of other keypoints. These keypoints are then utilized to recognise an object in a video. And it can be concluded that the object is recognised efficiently. Hence, for the future work, recognition rate can be increased using some classifier and some more features of the object.

## REFERENCES

[1] Moravec, H. 1981. Rover visual obstacle avoidance. In International Joint Conference on Artificial Intelligence, Vancouver, Canada, pp. 785-790.

[2] Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151.

[3] Harris, C. 1992. Geometry from visual motion. In Active Vision, A. Blake and A. Yuille (Eds.), MIT Press, pp. 263-284.

[4] Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.T. 1995. A robust technique for matching two un-calibrated images through the recovery of the unknown epipolar geometry. Artificial Intelligence, 78:87-119.

[5] Schmid, C., and Mohr, R. 1997. Local grayvalue invariants for image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 19(5):530-534.

[6] Lowe, D.G. 1999. Object recognition from local scale-invariant features. In International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157.

[7] Baumberg, A. 2000. Reliable feature matching across widely separated views. In Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina, pp. 774-781.

[8] Koen E.A. van de Sande, Theo Gevers and Cees G.M. Snoek, “Evaluating Color Descriptors for Object and Scene Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 9, pp. 1582-1596, September 2010.

[9] George Bebis, Sushil Louis, Yaakov Varol and Angelo Yfantis, “Genetic Object Recognition Using Combinations of Views,” IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, pp. 132-146, April 2002.

[10] R. Fergus, P. Perona and A. Zisserman, “Object Class Recognition by Unsupervised Scale-Invariant Learning,” in Proceedings of CVPR, pp. 264-271, 2003.

[11] Zachary Pezzementi, Erion Plaku, Caitlin Reyda and Gregory D. Hager, “Tactile-Object Recognition from Appearance Information,” IEEE Transactions on Robotics, Vol. 27, No. 3, pp. 473-487, June 2011.

[12] Bjorn Ommer and Joachim M. Buhmann, “Learning the Compositional Nature of Visual Object Categories for Recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 3, pp. 501-516, March 2010.

[13] Jialue Fan, Xiaohui Shen and Ying Wu, “What Are We Tracking: A Unified Approach of Tracking and Recognition,” IEEE Transactions on Image Processing, Vol. 22, No. 2, pp. 549-560, February 2013.

[14] Joseph L. Mundy, “Object Recognition in the Geometric Era: A Retrospective,” J. Ponce et al. (Eds.): Toward Category-Level Object Recognition, LNCS 4170, pp. 3–28, 2006.

[15] Gyuri Dorko and Cordelia Schmid, “Object Class Recognition Using Discriminative Local Features,” Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence, October 2004.

[16] K. Matusiak, P. Skulimowski and P. Strumillo, “Object recognition in a mobile phone application for visually impaired users,” IEEE HSI 2013, Vol. 978-1-4673-5637-4, No. 13, pp. 479-484, June 2013.

DOI 10.17148/IJIREEICE.2016.4742