IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014.
www.ijiset.com
ISSN 2348 – 7968

Object Recognition using SIFT

Shivakanth #1**, Archana Mane** #2

[1] Asst. Prof. MVJ College of Engineering, Bangalore
Visvesvaraya Technological University India

[2] Asst. Prof. MVJ College of Engineering, Bangalore
Visvesvaraya Technological University India

Abstract— *Image matching is a fundamental aspect of computer vision, including object recognition or scene recognition. For the Human recognition system despite of different viewpoints and differences, it is easier to identify the images. This task is still a challenge for computer vision for extracting the scale – invariant and shift – invariant features from images to perform reliable object recognition. In this paper, the object recognition system which can resolve the difficulty of rotations of object, scale changes and illumination areresolved with the help of "SIFT algorithm". It is implemented with different phases such as scale space extreme detection, key point localization. The SIFT algorithm implemented with MATLAB, which is one of efficient tool to perform image processing.*

The recognition proceeds by matching individual features to a database of features from known objects using a fast nearest-neighbor algorithm, followed by a Hough transform to identify clusters belonging to single object, and finally performing verification through least-square solution for consistent pose parameters. This approach to recognition can robustly identify objects between clutter and occlusion while achieving near real-time performance.

Keywords*: SIFT - Scale Invariant Feature Transform, NSS – Nearest Neighbor Search, DOG-Difference of Gaussian.*

I. INTRODUCTION

Object recognition is one of the important research fields to realize cognitive ability of the computers, and is expected to be applied to Robot eyes or head mounted display. Recently, manual retrieval and classification of the image become difficult as volume of data becomes huge. Computerized object recognition system becomes prominence in such scenario. The problem in the object recognition is to deal with the rotations of the object, scale changes, and illumination changes. Moreover, there is the problem of occlusion that makes the object recognition difficult. SIFT was proposed by David Lowe as a robust feature for these problems, and the object recognition method.

The scale invariant feature transform (SIFT) algorithm that extracts featuresof an image in a manner that is stable over image translation, rotation, scaling, illumination and camera viewpoint. The SIFT has been selected as it is one of the most widely usedalgorithms for object recognition, that has been employed in many applications such as face/object recognition, robot localization and mapping, 3D-scene modelling, and action recognition. SIFT accepts an N × N image as input and produces aset of features. The input bandwidth of N2 pixels can be very high for large values of N.This algorithm is most widely used one for the imagefeature extraction. SIFT extracts image features that are stable over image translation, rotation and scaling and somewhat invariant to changes in the illumination and camera viewpoint.

II. FRAME WORK

The work methodology of this paper is as shown in fig.2.1. Two images are considered. One is input image and the other is stored image in the database. The key points of input image is obtained and for the image in database respectively. Both the images are concatenated and based on matching; it can be found whether the input image is present in database.

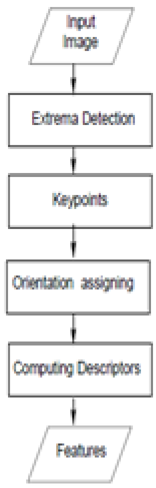

Fig. 2.1 work flow

2.1 Phases in SIFT algorithm

The proposed work has split into four major phases such as, Extrema detection, keypoint localization, Orient Assignment, key point descriptor generator. The SIFT algorithm is supposedly able to identify two objects as similar even the object is partly concealed in either one of the images, has changed orientation, or the object is viewed at different angles.

2.1 Extrema Detection

The first phase examines the image under various scales and octaves to isolate points of the picture that are different from their surroundings. These points are called extrema which is the potential candidates for image features.

2.2 Keypoint Localization

The Keypoint Detection, starts with the extrema and selects some of the points to be keypoints, that are a whittled down a set of feature candidates. This refinement rejects extrema, which are caused by edges of the picture and by low contrast points.

2.3 Orientation Assignment

Each keypoint and its neighborhood are converted into a set of vectors by computing a magnitude and a direction for them. It also identifies other keypoints that may have been missed in the first two phases; this is done on the basis of a point having a significant magnitude. The algorithm now has identified a final set of keypoints.

2.4 KeypointDescriptor Genration

Keypoint Descriptor Generation, takes a collection of vectors in the neigh-borhood of each keypoint and consolidates this information into a set of eight vectors calledthe descriptor. Each descriptor is converted into a feature by computing a normalized sumof these vectors.

III. THE SIFT ALGORITHM

The Nearest Neighbour search(NNS) algorithm detects the silmilarities between keypoints. The SIFT makes matching possible by generating the keypoint descriptors. It includes four computational phases.

3.1 Phase 1- Scale-space Extrema Detection

The first phase of the computation seeks to identify potential interest points. It searches over all scales and image locations. The computation is accomplished by using a Difference-of-Gaussian (DoG) function. The resulting interest points are invariant to scale and rotation, meaning that they are persistent across image scales and rotation.
Specifically, a DOG image D(x, y, σ) is given by

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma)$$

Where L(x, y, kσ) is the original image I(x,y) convolved with the Gaussian blur G(x, y, kσ) at scale kσ, i.e., L(x, y, kσ) = G(x, y, kσ) * I(x,y)

Hence, a DoG image between scales $k_i\sigma$ and $k_j\sigma$ is just the difference of the Gaussian-blurred images at scales $k_i\sigma$ and $k_j\sigma$. The figure 3.1 shows keypoints detected as extremas of DOG. For scale-space extrema detection in the SIFT algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave (an octave corresponds to doubling the value of σ), and the value of k_i is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave. Once DoG images are obtained keypoints are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate keypoint.

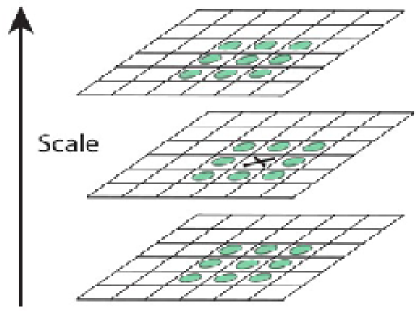

Figure 3.1 Keypoints are detected as extremas of DoG

3.2 Phase 2 - Keypoint localization

A large number of keypoint candidates are found from the extrema detection and so, this stage involves the rejection of keypoints with low contrast or poor localization along edges. First, for each keypoint candidate, interpolation of nearby data is used to accurately calculate position. The interpolation is accomplished by means of a quadratic Taylor expansion of the Difference-of-Gaussian scale-space function, D(x,y,σ) with the candidate key point as the origin.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T\frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

Where, D and its derivatives are evaluated at the candidate keypoint and X = (x,y,σ) is the offset from this point. To discard the keypoints with low contrast, the value of the second-order Taylor expansion D(X) is computed at the offset $\hat{\mathbf{x}}$ and it is determined by taking the derivative of this function with respect to $\mathbf{x}$and setting it to zero.

3.3 Phase 3 - Orientation assignment

For each of the keypoints identified in phase 2, SIFT computes the direction of gradients around. One or more orientations are assigned to each keypoint based on local image gradient directions. An orientation is assigned to each keypoint by building a histogram of gradient

ISSN 2348 – 7968

orientations (x; y) weighted by the gradient magnitudes m(x; y) from the keypoints neighborhood

$$m(x,y)=\sqrt{\left(L(x+1,y)-L(x-1,y)\right)^2+\left(L(x,y+1)-L(x,y-1)\right)^2}$$

$$\theta(x,y)=\tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right)$$

Where, L is a Gaussian smoothed image with a closest scale to that of a keypoint.

3.4 Phase 4 - Keypoint descriptor

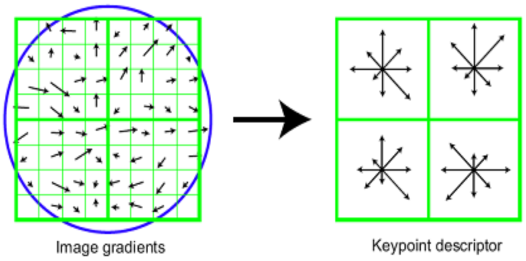

Figure 3.2: Image gradients and keypoint descriptors

The local image gradients are measured 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.

IV. MATCHING

When using the SIFT algorithm for object recognition, each keypoint descriptor extracted from the query image is matched independently to the database of descriptors extracted from all training images. The best match for each descriptor is found by identifying its nearest neighbor (closest descriptor) in the database of keypoint descriptors from the training images. Generally, many features from a test image do not have any correct match in the training database, because they were either not detected in the training image or they arose from background clutter. To discard keypoints whose descriptors do not have any good match in the training database, a subsequent threshold is used, which rejects matches that are too ambiguous. If the distance ratio between the closest neighbor and the second-closest neighbor, is below some threshold, than the match is kept, otherwise the match is rejected and the keypoint is removed. There are three possible image matches and they are:

- A match where the whole of one image matches the whole of another image.
- A part of one image matching the whole of another image.
- Part of one image matching part of another image.

V. IMPACT OF SUPPORTING SIFT

SIFT and SIFT-like GLOH features exhibit the highest matching accuracies (recall rates) for an affine transformation of 50 degrees. After this transformation limit, results start to become unreliable.

Distinctiveness of descriptors is measured by summing the eigenvalues of the descriptors, obtained by the Principal components analysis of the descriptors normalized by their variance. This corresponds to the amount of variance captured by different descriptors, therefore, to their distinctiveness. PCA-SIFT (Principal Components Analysis applied to SIFT descriptors), GLOH and SIFT features give the highest values.

SIFT-based descriptors outperform other local descriptors on both textured and structured scenes, with the difference in performance larger on the textured scene.

For scale changes in the range 2-2.5 and image rotations in the range 30 to 45 degrees, SIFT and SIFT-based descriptors again outperform other local descriptors with both textured and structured scene content.

Performance for all local descriptors degraded on images introduced with a significant amount of blur, with the descriptors that are based on edges, like shape context, performing increasingly poorly with increasing amount blur. This is because edges disappear in the case of a strong blur. But GLOH, PCA-SIFT and SIFT still performed better than the others. This is also true for evaluation in the case of illumination changes.

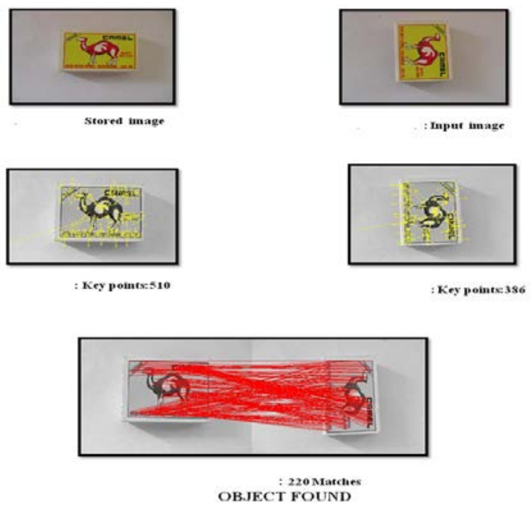

Figure: 6.1Results for Object recognition with different orientation

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The implementation of this algorithm is done in windows platform, with Matlab7.0 – Image processing toolbox. The object recognition flow is executed with the functions like MATCH, SIFT, TRANSFORMLINE,SHOWKEYS and

ISSN 2348 – 7968

APPENDIMAGE. Figure 6.1 shows output verified for different image oriention. Figure 6.2 shows output verified for different size of the input image. Figure 6.3 shows output verified for different illumination of the input image, stored image, key points for input image, key points for stored image and also shows the concatenated image so that object can be viewed whether it is present in stored images and also the matches between input and stored image through red line.

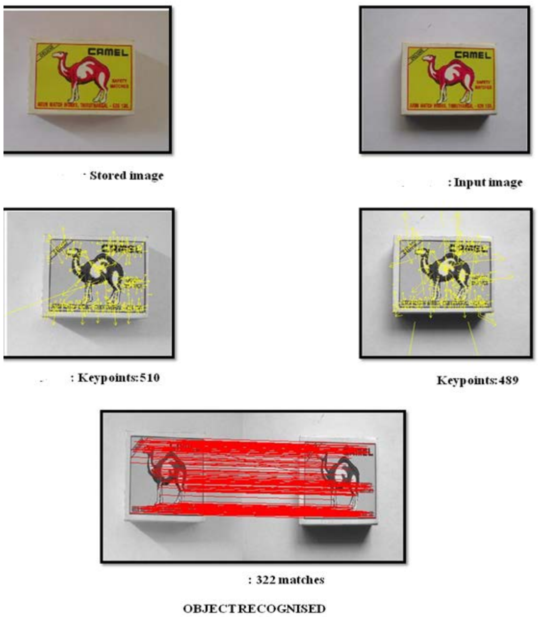

Figure: 6.2 Results for Object recognition with different size

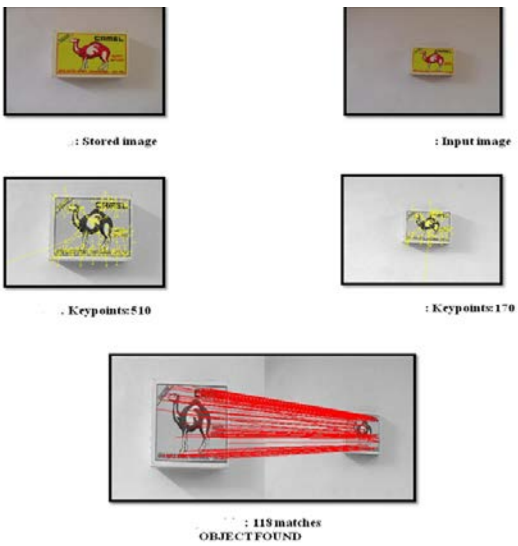

Figure: 6.3 Results for Object recognition with different illumination

VII. CONCLUSION

SIFT can be used to detect similar objects in two different images. The SIFT algorithm is supposedly able to identify two objects as similar even the object is partly concealed in either one of the images, has changed orientation, or the object is viewed at different angles. Implementation of such an algorithm could ease the computer vision. As SIFT shows many special features, which are unique in object recognition field, the algorithm could then fulfill the demand of product quality control and object separation in industries.

REFERENCES

1. Ballard, D.H., "Generalizing the Hough transform to detect arbitrary patterns," Pattern Recognition, **13,** 2 (1981), pp.111-122.

2. Basri, Ronen, and David.W. Jacobs, "Recognition using region correspondences," International Journal of Computer Vision, **25**, 2 (1996), pp. 141–162.

3. Beis, Jeff, and David G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces,"Conference onComputerVision and PatternRecognition, Puerto Rico (1997), pp. 1000–1006.

4. Crowley, James L., and Alice C. Parker, "A representation for shape based on peaks and ridges in the difference of lowpass transform," IEEE Trans. on Pattern Analysis and Machine Intelligence, **6,** 2 (1984), pp. 156–170.

5. Edelman, Shimon, Nathan Intrator, and Tomaso Poggio, "Complex cells and object recognition," Unpublished Manuscript, preprint at http://www.ai.mit.edu/ edelman/mirror/nips97.ps.Z

6. Grimson, Eric, and Thom´as Lozano-P´erez, "Localizing overlapping parts by searching the interpretation tree," IEEE Trans. on Pattern Analysis and Machine Intelligence,