INFORMATION PAPER
International Journal of Recent Trends in Engineering, Vol. 1, No. 2, May 2009

Performance Evaluation of Scale Invariant Feature Transform

Padmavati[1], Ajay Mittal[2], and Navdeep Kaur[3]
[1] Punjab Engineering College/Computer Science and Engineering Department, Chandigarh, India
Email: padmavati_khandnor@yahoo.co.in
[2] Punjab Engineering College/Computer Science and Engineering Department, Chandigarh, India
Email: ajay_mittal825@yahoo.com
[3] Punjab Engineering College/Information Technology Department, Chandigarh, India
Email: navdeepgill2007@gmail.com

***Abstract*— The SIFT algorithm produces keypoint descriptors. This paper analyzes that the SIFT algorithm generates the number of keypoints when we increase a parameter (number of sublevels per octave). SIFT has a good hit rate for this analysis. The algorithm was tested over a specific data set, and the experiments were conducted to increase the performance of SIFT in terms of accuracy and efficiency so as to provide its use in many real time applications for image recognition.**

We also shown that a keypoint is an image feature which is so distinct even though the image has gone through transformations i.e. given a keypoints in an image, if one scales the image to half the size, double the size, rotated to a particular degree, or object occlusion is done, the image can be recognized because some keypoints would still be identifiable. The algorithm is specifically tested for its feasibility for finding keypoint matches between two images.

***Index Terms*— DoG, SIFT, Keypoint**

I. Introduction

The drastic advances in both hardware and software technologies are making image processing applications technically and economically feasible.

A. *Why Local Features?*

A user is having an array of available recognition methods like geometry- or model-based object recognition, Appearance Based Methods and Recognition as a Correspondence of Local Features. But Ref. [8] Putting local features into correspondence is an approach that is robust to object occlusion and cluttered background in principle.

Ref. [1] a keypoint descriptor is a 128-dimensional vector that describes a keypoint. The reason for this high dimension is that each keypoint descriptor contains a lot of information about the point it describes. We shall have a closer look at what information the keypoint descriptors hold when we discuss the four phases of SIFT.

B. *Four computational steps*

The following discussion will be based Ref [2].

Step 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.

Step 2: Keypoint localization

For all interest points found in step 1, a detailed model is created to determine location and scale. Keypoints are selected based on their stability. A stable keypoint is thus a keypoint resistant to image distortion.

Step 3: Orientation assignment

For each of the keypoints identified in step 2, SIFT computes the direction of gradients around. One or more orientations are assigned to each keypoint based on local image gradient directions.

Step 4: Keypoint descriptor

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.

II. IMAGE RECOGNITION

To recognize a given object a user has an array of available algorithm choices; but Ref. [3] SIFT is an efficient algorithm in terms of performance. So we generated sets of keypoint descriptors for two images, by [4] after recognizing a given images i.e. detecting a key points in a given image one can begin matching two images.

A. *Complete match*

A match where the whole of one image matches the whole of another that means the two images is same.

B. *Partial match:*

Part of one image matching the whole of another image or part of one image matching part of another image.

C. *Characteristics of match*

The different matches will have different characteristics. That is, when we have complete match, we can expect a fairly large percentage of all keypoints matching. In partial match we would expect a large percentage match in the image with a whole match, and a small percentage match in the image partly matching. Or we would expect a fairly low percentage of keypoints matching in both images.

III. RESULTS AND DISCUSSION

We will now present some key point generation tests performed mainly to see what SIFT would do with images with varying number of sub level per octave.

A. Varying the parameter value (i.e. the number of sub level per octave)

Input image and compared images are same.

Figure 1. Input image (652X870)

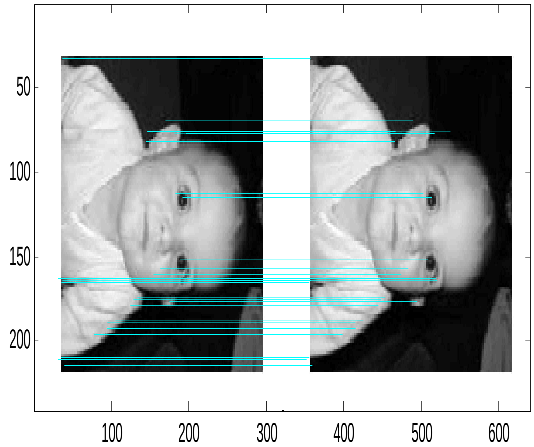

Figure 2. number of keypoints matched (49) with number of sub levels per octave(3)

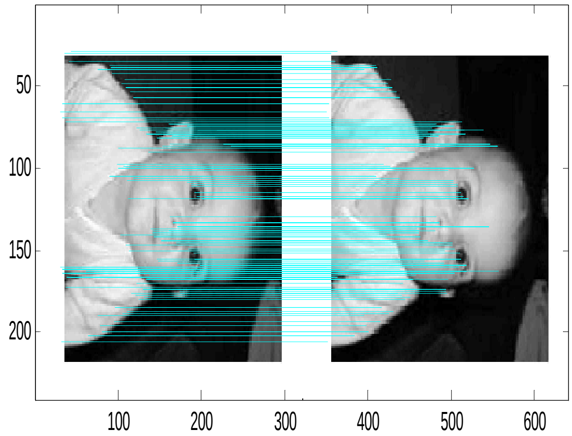

Figure 3. Number of keypoints matched (202) with number of sub level per octave (8)

Observations

Fig.1 is the input image with a resolution of 652X870. Since the original image had very few keypoints, we would expect the number of additional keypoints to be generated, when we increase the parameter i.e. the number of sub levels per octave. Fig. 2 as expected of SIFT, the algorithm identifies fairly few keypoints (only 49) for Fig. 1 with a parameter value 3 Ref. [1] . Suppose we increase the number of sub levels per octave then we will get more number of matched keypoints. In Fig. 3 number of keypoints matched (202) with parameter value 8 ,which has approximately 4 times more matching keypoints than Fig. 2. Table 1.1 shows the result as we increase the number of sublevel per octave parameter, the numbers of matching keypoints are increasing.

Keypoint Generation:

After proper testing it is observed that the number of keypoints found, when we set the parameter value 8 are approximately 4 times (i.e. 210) the number of keypoints found when we set the parameter value 3.

Quality of match:

After proper testing it is observed that the number of keypoints matched, when we set the parameter value 8 (i.e. 202) are approximately 4 times (i.e. 202) the number of keypoints that are matched when we set the parameter value 3(i.e. 49).

As we can see from the Table 1 even though we are increasing parameter value from 8 to 9(or any higher value), the number of keypoint found and keypoint match will decrease. So it is analyzed that when we set the parameter value 8 then it will generate highest number of keypoints. It is also observed that if we select parameter value 3 then we will get 100% match, which is calculated by the following formula,

Percentage= keypoint match / keypoint found

ACEEE

Table 1

Number of sub levels per octave	Keypoint found	Keypoint match	Percentage
1	2	2	50
2	27	23	85.19
3	49	49	100
4	59	61	96.72
5	88	86	97.73
6	90	88	97.78
7	188	178	94.68
8	**210**	**202**	96.19
9	122	114	93.44
10	136	124	91.18
…..…	……	…….	…..
15	131	120	91.6

B. Analyzing Scale Factor

Here compared images is scaling the input image by double or half the size

Figure 4. Input image (400X320) and total keypoints: 305

Figure 5: 200X160 (when the input image is scaled down to 2) and total keypoints: 265

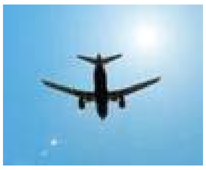

Figure 6. 100X80 (when the input image is scaled down to 4) and total keypoints: 238

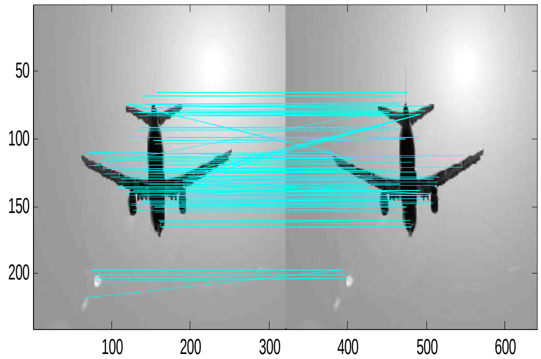

Figure 7. Total Keypoints matched 182 of Figure 1.4(400X320) and Figure1.4 (800X640)

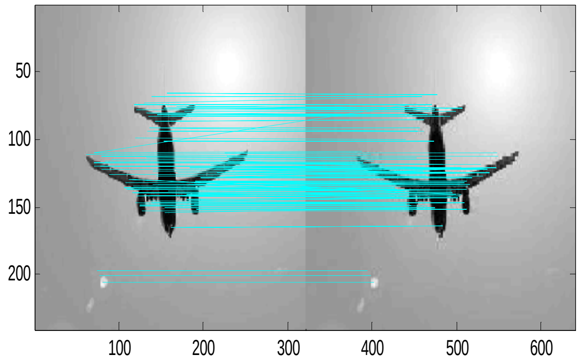

Figure 8. Total Keypoints matched 195 of Figure 1.4(400X320) and Figure1.5 (200X160)

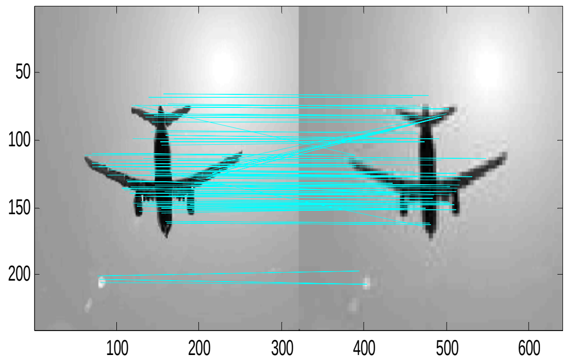

Figure 9: Total Keypoints matched 161 of Figure 1.4(400X320) and Figure1.6 (100X80)

Observations:
Observations are as under.

Keypoint Generation
After proper testing it is observed that the number of keypoints found in Fig. 4 i.e. the input image with a resolution of 400X320 is 305. when we set the compared image half the size of the input image i.e. Fig. 5 with the resolution of 200X160 and total keypoints which are generated are 265 which is approximately 86% of the input image. Fig. 6 when we set the compared image is scaled down to 4 of the input image ,the resolution is 100X80 and total keypoints which are generated are 238 which is approximately 78% of the input image. Table 2 shows the input image with various resolution, number of keypoint found and keypoint matched.

Quality of match:
In Fig.7 when we scale the input image by double the size, show the total number of keypoints matched are 182 i.e. approximately 61%, Fig. 8 when we scale the input image by half the size show the total number of keypoints matched are 195 i.e 73%.
So it is observed that changing scale of the image, keypoints would still be identifiable.

Table 2: Result Set-scaling

Resolution	Keypoint found	Keypoint match
400X320(Input image)	305	297
800X640	237	182
200X160	265	195
100X80	238	161

C. *Analyzing Occlusion Factor*

Figure 10. Input image (640X480) and total keypoints: 492

Figure 11. Input image (346X312) and Total keypoints: 389

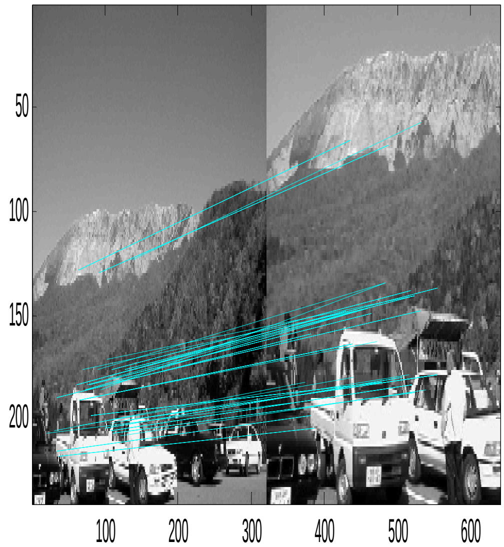

Figure 12. Total Keypoints matched of Figure 1.7 and Figure 1.8:99

Observations:
Observations are as under.

Keypoint Generation:
Fig. 10 is the input image with a resolution of 640X480 total keypoints generated are 492 and Fig. 11 is Occluded image with a resolution of 346X312 and total keypoints which are generated are 389, which is approximately 77% of the input image.
So it is observed that when a part of an object is occluded by other objects in the scene only features of that part are missed. As long as there are enough features detected in the un occluded part the object can be recognized because some keypoints would still be identifiable.

D. Analyzing Rotation Factor

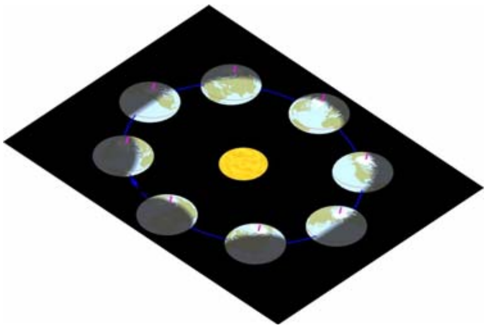

Figure 13.Image (990X990), degree and total keypoints: 275

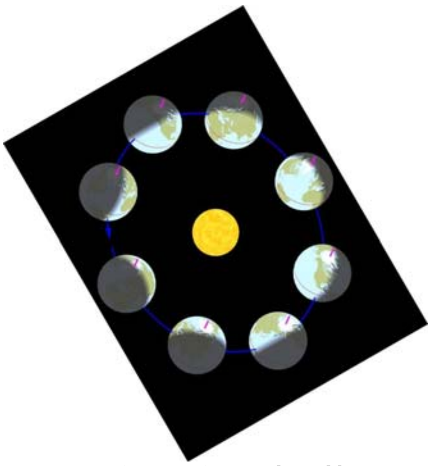

Figure 14. Image (920X993) ,rotated Total keypoints: 245

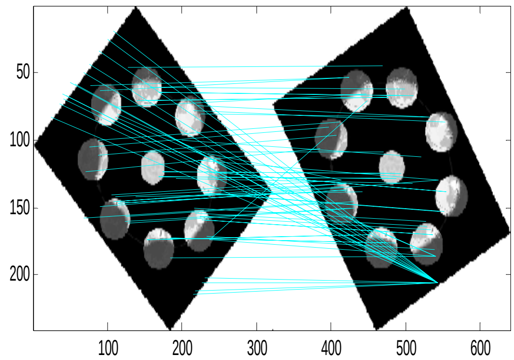

Figure 15. Keypoint match: 112

Observations:

Observations are as under.

Keypoint Generation:

Fig. 13 is the input image with a resolution of 990X990 total keypoints generated are 275 and Fig. 14 rotated the input image with a resolution of 920X993 and total keypoints which are generated are 245, which is approximately 89% of the input image.

Quality of match:

In Fig. 15 the total number keypoints that are matched are 112 which is approximately 50% of the Fig 14

IV. CONCLUSION

With the above discussion it is shown that SIFT is a robust algorithm for image recognition and matching. Through extensive testing, we have discovered when it will generate highest number of keypoints, under what circumstances SIFT, the image matching, object occlusion works well. We have also shown that SIFT is able to detect similarities between images, even though the images has gone through a transformation such as scaling and rotation.

V. FUTURE WORK

Ref. [7] SURF is faster as compared to SIFT, Ref. [6] using different set of detector and descriptors combination we can achieve object recognition with greater accuracy and as well we can improve the computation time.

VI. APPLICATION

Some potential application of the approach is to place recognition in the following applications,

A mobile device could identify its location by recognizing familiar locations.

A vehicle could identify its location by recognizing familiar locations.

Vision to blind: after recognizing keypoints of an object, matching these keypoints with the object i.e. stored in the database and if they match then converting this into voice (image to voice).

Pattern recognition, action recognition, real time image retrieval accuracy and matching speed etc.

REFERENCES

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

[2] David G. Lowe "Distinctive Image Features from Scale-Invariant Keypoints", 2004.

[3] Peter M. Roth and Martin" Survey Of Appearance-Based Methods For Object Recognition" Winter Technical Report, 2008.

[4] Lowe, D.G. Local feature view clustering for 3D object recognition. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, pp. 682-688, 2001.

[5] Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: ECCV. (2002) 128 – 142.

[6] Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: CVPR (2). (2004) 506 – 513.

[7]Herbert Bay, Tinne Tuytelaars, and Luc Van Gool "SURF: Speeded Up Robust Features".

[8] Jirı Matas and Stepan Obdrzalek "Object Recognition Methods Based On Transformation Covariant Features" Center for Machine Perception, Czech Technical University, Prague, 120 35, CZ