



Application of SIFT and SURF Detectors for Medical X-Ray Images

Authors

Prachi. G. Bhende¹, Dr A. N. Cheeran²

^{1,2}Department of Electrical Engineering, VJTI, Mumbai, India

¹Email: pgbhende@gmail.com, ²Email: ancheeran@vjti.org.in

Abstract

Detecting, identifying, and recognizing salient regions or feature points in images is a very important and fundamental problem in computer vision and robotics. For gray scale x-ray images, stable and repeatable salient features that are invariant to a variety of effects like rotation, scale changes, view point changes, noise, or change in illumination conditions can provide better classification results. This paper compares two different methods for scale and rotation invariant interest point/feature detector and descriptor: Scale Invariant Feature Transform (SIFT) and Speed Up Robust Features (SURF). It also presents a way to extract distinctive invariant features from gray scale X-ray images that can be used to perform reliable matching between different views of an object/scene.

Keywords: Feature detection, Feature matching, SIFT, SURF

INTRODUCTION

Recently, two promising approaches to detect salient regions in images are: In 2004, Scale Invariant Feature Transform (SIFT) by Lowe and Speeded Up Robust Features (SURF) by Bay et al. (2006) ^[1-2]. Both approaches do not only detect interest points or so called features, but also propose a method of creating an invariant descriptor. This descriptor can be used to uniquely identify the found interest points and match them even under a variety of disturbing conditions like scale changes, rotation, changes in illumination or viewpoints or image noise. Of course, there was and still is a variety of other methods available that claim to achieve the same goals. Comparative work done by Mikolajczyk and Schmid (2005) is a good reference for a list of different other interest point's detectors and descriptors and proves that SIFT outperforms other algorithms in terms of invariance ^[3]. Bay et al. (2006), in return claimed SURF to be superior to SIFT in terms of runtime efficiency while still yielding comparably good results with regards to feature point quality. This paper focuses on SIFT and SURF techniques to detect the robust regions (distinctive features) from gray scale X-ray images. First, "key points" are extracted from distinctive locations from the images such as edges, blobs, corner etc. Next, neighborhood regions are picked around every key point and distinctive feature descriptors are computed from each region. Feature descriptors are extracted from sample images and stored. This descriptor has to be distinctive and, at the same time, robust to noise, detection errors. Finally, the feature descriptors are matched between different images. Feature descriptor matching can be based Euclidean distance. X-ray images from Image Retrieval In Medical applications (IRMA) database is referred for experimentation purpose.

The overview of both the methods and their application to medical X-ray images is described in Section 2, in section 3 the experimental results are presented while in Section 4 the conclusion of the paper is discussed.

OVERVIEW OF METHODS

In 2004, Lowe presented SIFT for extracting distinctive invariant features from images that can be invariant to image scale and rotation [1]. It was widely used in image mosaic, recognition, retrieval etc. In 2006, Bay et al. introduced speeded up robust features technique (SURF), in which use of integral images for image convolutions and Fast-Hessian detector is applied [2]. Their experiments turned out that the latter was faster and that it worked well. Both approaches do not only detect interest points or so called features, but also propose a method for creating an invariant descriptor. This descriptor can be used to identify the found interest points and match them even under a variety of disturbing conditions, like scale changes, rotation, changes in illumination or viewpoints or an image noise. There are also many other feature detection methods, as edge detection, corner detection, etc. Different methods have their own advantages.

➤ **SIFT ALGORITHM OVERVIEW**

SIFT (Scale Invariant Feature Transform) algorithm was proposed by Lowe in 2004 to solve the image rotation, scaling, and affine deformation, viewpoint change, noise, illumination changes, also has strong robustness. The SIFT algorithm has four main steps: (1) Scale Space Extrema Detection, (2) Key point Localization (3) Orientation Assignment and (4) Description Generation.

1. Scale Space Extrema Detection -

The scale space of an image $I(x, y)$ is defined as a function $L(x, y, \sigma)$, that is produced from the convolution of $I(x, y)$ with a variable-scale Gaussian $G(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{1}$$

Where * is the convolution operation in x and y

To efficiently detect stable key point locations in scale space, it used a scale space extrema based on the difference-of- Gaussian function, $D(x, y, \sigma)$ which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k :

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k \sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k \sigma) - L(x, y, \sigma) \end{aligned} \tag{2}$$

The initial image is incrementally convolved with Gaussians to produce images separated by a constant factor k in scale space, shown stacked on the bottom. The scale space is divided into octaves (e.g., doubling σ) with an integer number of intervals, so $k = \sqrt{2}$. Adjacent image scales are subtracted to produce the difference-of-Gaussian images shown on the top. Next, resample the Gaussian image that has twice the initial value of $\sigma(1.6)$ by taking every second pixel in each row and column, and recursively iterate all the process [3-6].

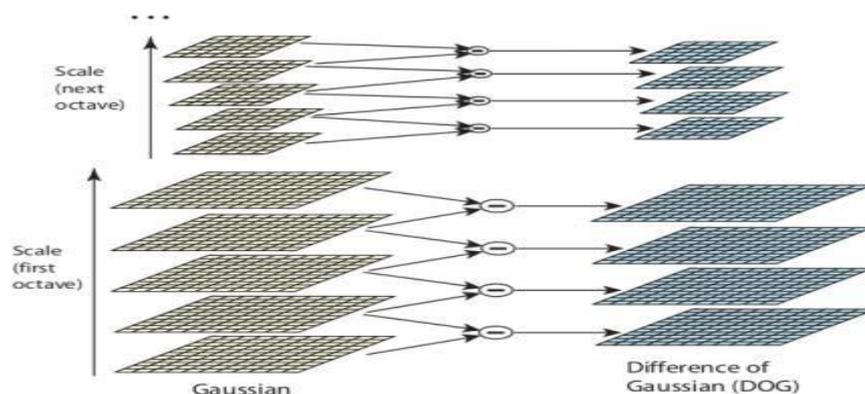


Figure 1: Scale –Space Detection using DOG [2]

Once this DOG's are found, images are searched for local extrema over scale and space. For eg, one pixel in an image is compared with its 8 neighbors as well as 9 pixels in next scale and 9 pixels in previous scales. If it is a local extrema, it is a potential key point. It basically means that key point is best represented in that scale. It is shown in below Figure 2.

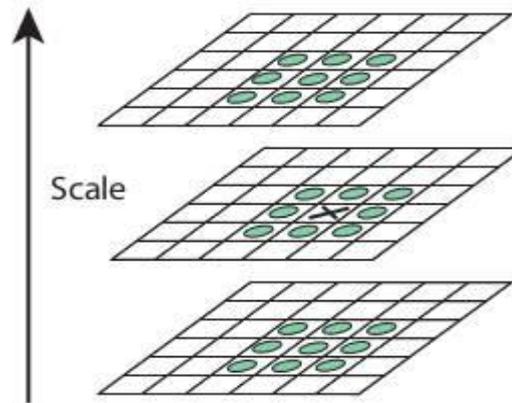


Figure 2: Scale Space Extrema Detection Process ^[2]

2. Key point Localization –

Once potential keypoints locations are found, they have to be refined to get more accurate results. Use of Taylor series expansion of scale space to get more accurate location of extrema is done, and if the intensity at this extrema is less than a threshold value (0.03), it is rejected. This threshold is called contrast threshold. DOG has higher response for edges, so edges also need to be removed. For this, a concept similar to Harris corner detector is used. A 2x2 Hessian matrix (H) is used to compute the principal curvature. As per the Harris corner detector, for edges, one eigen value is larger than the other. So here a simple function Trace/Determinant ($\lambda_1 \lambda_2 / \lambda_1 + \lambda_2$) ratio is applied. If this ratio is greater than a threshold, called edge threshold, that key point is discarded. So it eliminates any low-contrast key points and edge key points and what remain are strong interest points ^[7-9].

3. Orientation Assignment-

Now an orientation is assigned to each key point to achieve invariance to image rotation. A neighbourhood is taken around the key point location depending on the scale, and the gradient magnitude and direction is calculated in that region. An orientation histogram with 36 bins covering 360 degrees is created. (It is weighted by gradient magnitude and Gaussian-weighted circular window with σ equal to 1.5 times the scale of key point. The highest peak in the histogram is taken and any peak above 80% of it is also considered to calculate the orientation. It creates key points with same location and scale, but different directions. It contributes to stability of matching ^[10-13].

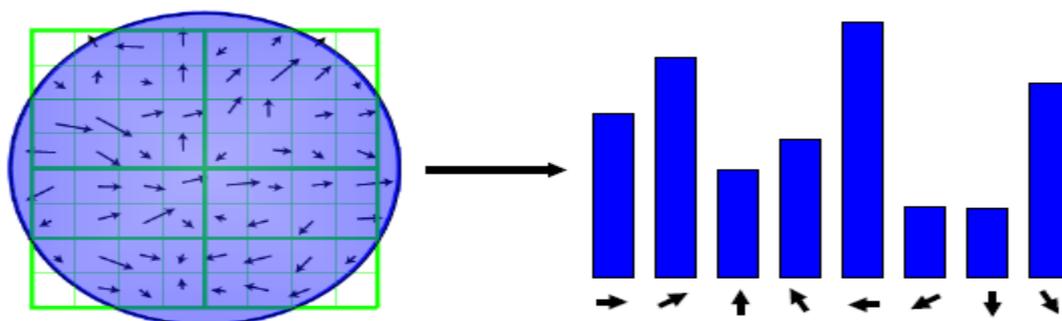


Figure 3: Gradient Histogram (36 bins)

4. Construction of SIFT Descriptor -

Now key point descriptor is created. A 16x16 neighbourhood around the key point is taken. It is divided into 16 sub-blocks of 4x4 size. For each sub-block, 8 bin orientation histogram is created. So a total of 128 bin values are available. It is represented as a vector to form key point descriptor. In addition to this, several measures are taken to achieve robustness against illumination changes, rotation etc [15-17].

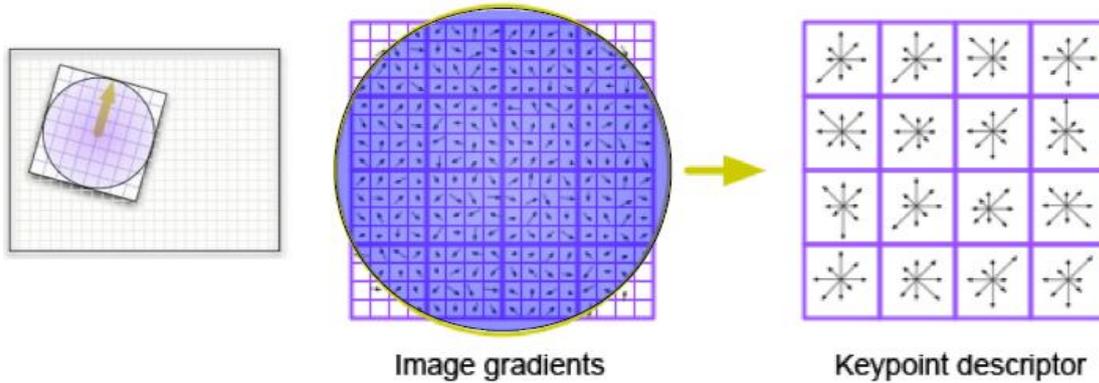


Figure 4: SIFT Descriptor [16]

5. Key point Matching-

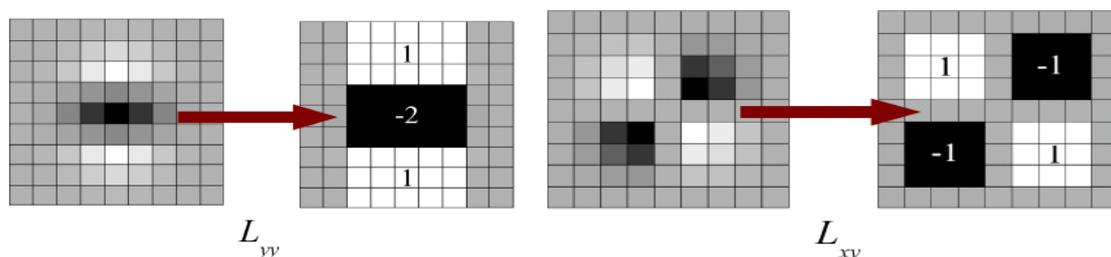
Keypoints between two images are matched by identifying their nearest neighbours. But in some cases, the second closest-match may be very near to the first. It may happen due to noise or some other reasons. In that case, ratio of closest-distance to second-closest distance is taken. If it is greater than 0.8, they are rejected. It eliminates around 90% of false matches while discards only 5% correct matches [18-21].

➤ **SURF Algorithm Overview**

In SIFT, Lowe approximated Laplacian of Gaussian (LOG) with Difference of Gaussian for finding scale-space. SURF goes a little further and approximates LOG with Box Filter. Figure5. shows a demonstration of such an approximation. One big advantage of this approximation is that, convolution with box filter can be easily calculated with the help of integral images. And it can be done in parallel for different scales. Also the SURF rely on determinant of Hessian matrix for both scale and location [22-23]. Herbert Bay et. al., introduced the local invariant interest points' detector-descriptor (SURF). SURF is invariant to common image transformations, rotation, scale change, illumination change and small change in viewpoint. SURF uses integral images (summed area tables), which are intermediate representations for the image and contain the sum of gray scale pixel values of image, to reduce computation time. The detector is based on Hessian matrix to make use of its good performance in computation time and accuracy Speed-up computations are possible by use of fast approximation of (i) Hessian matrix [H]and (ii) descriptor using “integral images”.

$$H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \tag{3}$$

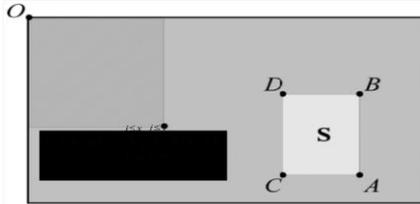
Where $L_{x \times x}(x, y, \sigma)$ is the Laplacian of Gaussian of the image. It is the convolution of the *Gaussian* second order derivative with the image. Approximate L_{xx} , L_{yy} , and L_{xy} using box filters. (box filters shown are 9 x 9 – good approximations for a Gaussian with $\sigma=1.2$)



The integral image $I_{\Sigma}(x, y)$ of an image $I(x, y)$ represents the sum of all pixels in $I(x, y)$ of a rectangular region formed by $(0,0)$ and (x, y) .

$$I_{\Sigma}(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \tag{4}$$

Using integral images, it takes only four array references to calculate the sum of pixels over a rectangular region of any size as shown below ^[24].

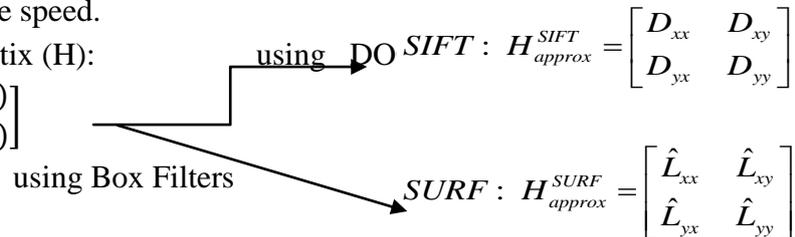


$$S = A + D - B - C$$

In SURF use of filters of larger size on the original image rather than smoothing the figure with different scale (Gaussian and subsampling like in SIFT) gives better result. Due to using integral images, filters of any size can be applied at exactly the same speed.

- Approximation of Hessian matrix (H):

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}$$



Instead of using a different measure for selecting the location and scale of interest points (e.g., Hessian and DOG in SIFT), SURF uses the determinant H_{approx}^{SURF} to find both.

Determinant elements must be weighted to obtain a good approximation:

$$\det(H_{approx}^{SURF}) = \hat{L}_{xx}\hat{L}_{yy} - (0.9\hat{L}_{xy})^2 \tag{5}$$

Once interest points have been localized both in space and scale, the next steps are:

- (1) Orientation assignment
- (2) Keypoint descriptor

Orientation Assignment : Following Figure 6 shows the circular neighborhood of radius 6σ around the interest point (σ = the scale at which the point was detected)

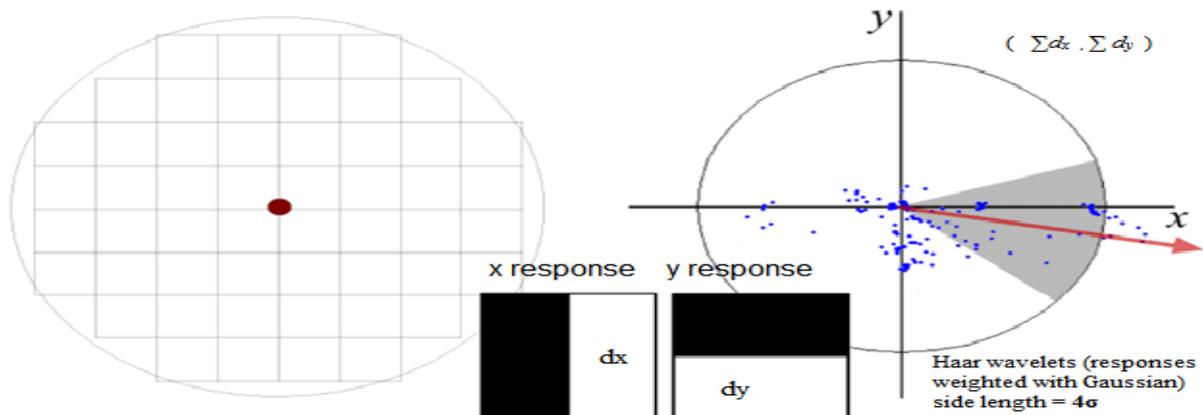


Figure 6: Orientation Assignment

- The image is convoluted with two first-order Haar wavelets.
- The filter responses at certain sampling points around the key point are represented as a vector in a two-dimensional space.

- A rotating window of 60° is used to sum up all vectors within its range, and the longest resulting vector determines the orientation.

Keypoint Descriptor-

For feature description, SURF uses Wavelet responses in horizontal and vertical direction (again, use of integral images makes things easier). A neighbourhood of size $20 \sigma \times 20 \sigma$ is taken around the key point where σ is the size. It is divided into 4×4 sub regions. For each sub region, horizontal and vertical wavelet responses are taken and a vector is formed like this, $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$. This when represented as a vector gives SURF feature descriptor with total 64 dimensions. Lower the dimension, higher the speed of computation and matching, but provide better distinctiveness of features [25].

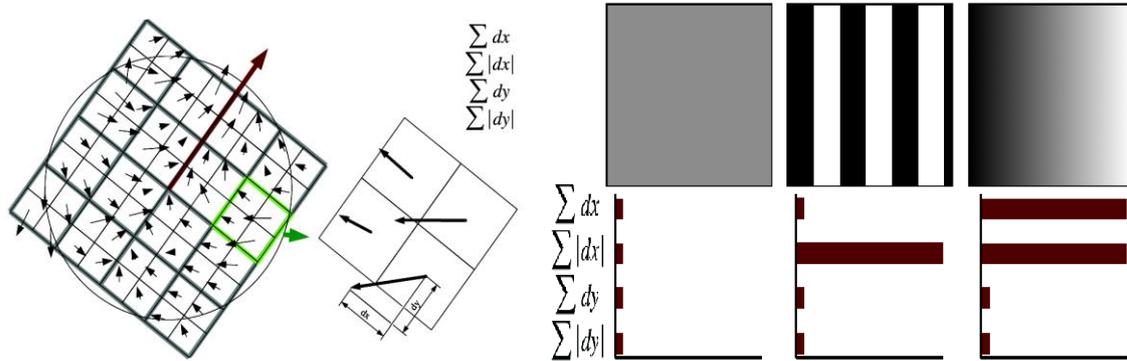


Figure 7: Key point Descriptor

The descriptor entries of a sub-region represent the nature of the underlying intensity pattern. Left: In case of a homogeneous region, all values are relatively low. Middle: In presence of frequencies in x direction, the value of $\sum |dx|$ is high, but all others remain low. If the intensity is gradually increasing in x direction, both values $\sum dx$ and $\sum |dx|$ are high [26].

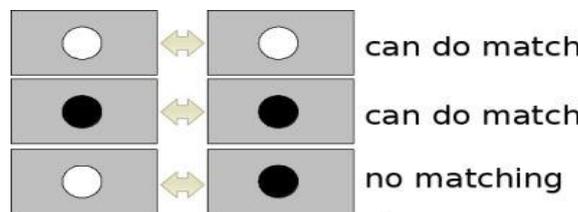


Figure 7: Key point Matching

In short, SURF adds a lot of features to improve the speed in every step. Analysis shows it is 3 times faster than SIFT while performance is comparable to SIFT. SURF is good at handling images with blurring and rotation, but not good at handling viewpoint change and illumination change.

EXPERIMENTAL RESULTS

To verify the effectiveness of the algorithm two images are taken as the experimental data as shown in figure 2(a) image1: Hand image (640x478, 153 KB) and 2(b) image2: Spine image (640x478, 127 KB). The experiments are performed on medical X-ray images from IRMA database. Features are detected in both images using SIFT and SURF algorithm. All Images are scaled by 1.2 times and rotated by 20 degrees for checking performance. Normalised Cross Correlation technique is used here for feature matching. We found the following matching feature time gap between SIFT and SURF algorithm application for image 1 and image 2.

Figure 2 (c) and 2(d) shows the detected features using SIFT in image1 and image2 respectively. Figure 2(e) and 2(f) shows results of SURF algorithm to image 1 and image2 respectively.



Figure 2(a) original hand x-ray image

2(b) original spine x-ray image ,

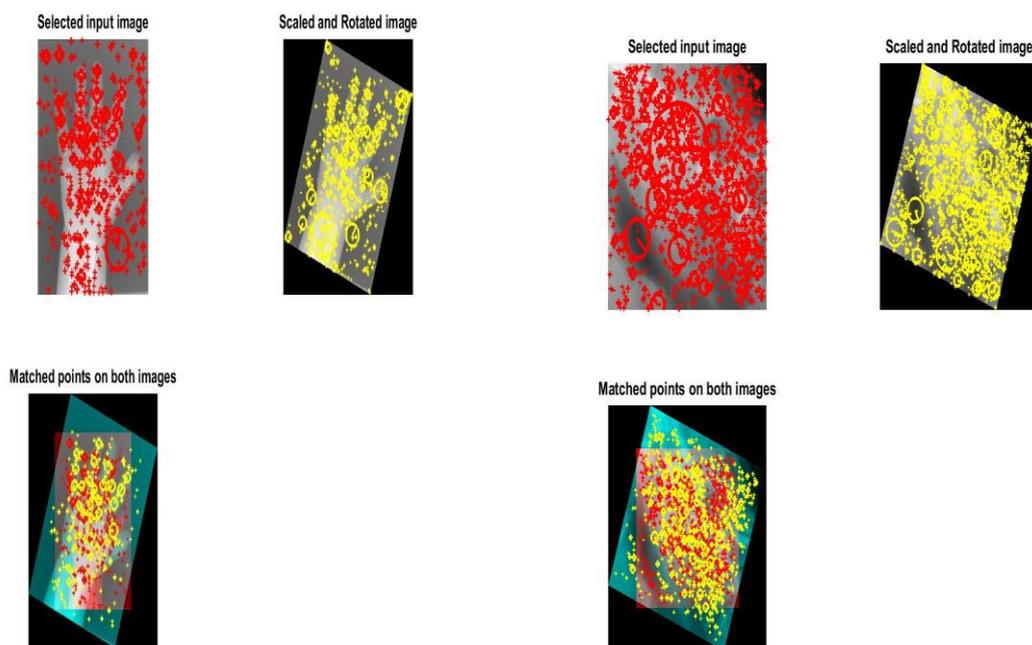


Figure2(c) SIFT features detected and matched original from original & rotated hand image

2 (d) SIFT features detected and matched from & rotated spine image

Below Figure 2(e) and 2(f) shows result of SURF algorithm applied to original and rotated hand and spine X-ray images.

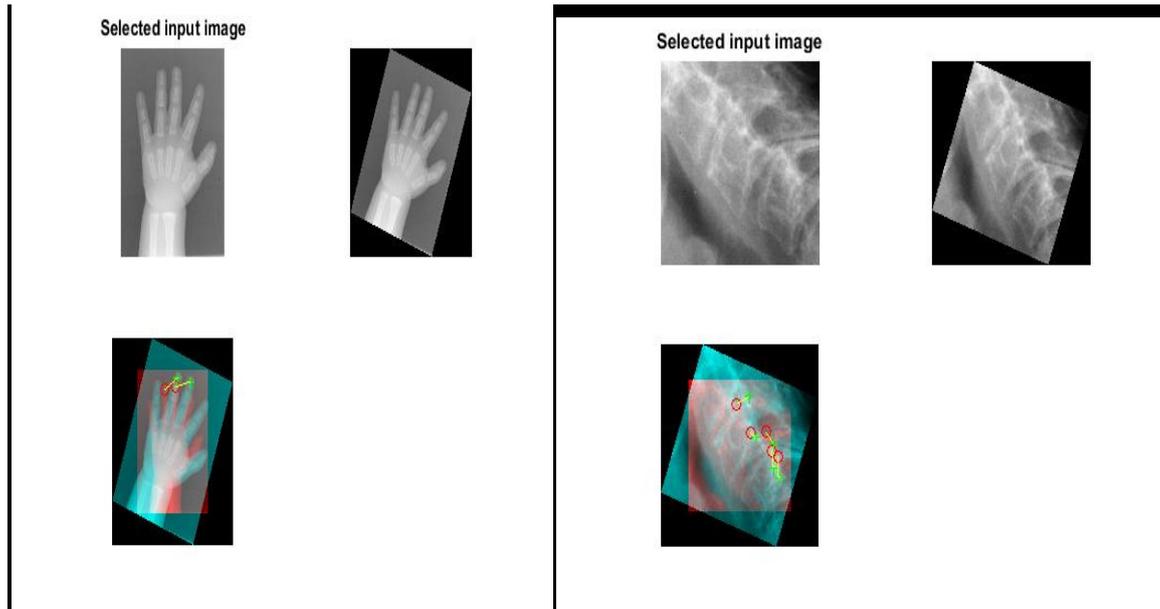


Figure 2(e) SURF features detected and matched from original from original & rotated hand image

2(d) SURF features detected and matched from & rotated spine image

The experimental results showing the detected and matched features for medical X-ray (hand and spine image) using the SIFT and SURF algorithms are summarized in Table 1. It seems that the description of the nature of the underlying image- intensity pattern is more distinctive than histogram based approaches. The simplicity and again the use of integral images make SURF descriptor competitive in terms of speed.

Table 1. Experimental results of SIFT and SURF algorithm

Algorithm	Image 1 Detected Features	Image 2 Detected Features	Rotated and Scaled Image 1	Rotated and Scaled Image 2	Matching Points of Image 1	Matching Points of Image 2	Total Execution time for Image 1	Total Execution time for Image 2
SIFT	391	883	618	1364	248	617	0.699848 seconds	1.315925 seconds
SURF	129	220	324	312	132	237	0.378608 seconds	0.424115 seconds

CONCLUSION

This paper has evaluated two feature detection methods for image registration. Based on the experimental results, it is found that the SIFT has detected more number of features compared to SURF but it is suffered with less speed in matching feature points. The SURF is fast and has good performance same as SIFT. It is found SURF to be a fast and scale and rotation-invariant interest point detector and descriptor. The important factor speed gain is due to the use of integral images, which drastically reduce the number of operations independent of the chosen scale. Future work includes merging these two descriptors with other feature extraction methods and analyzing the matching response time. Also other categories of medical X-ray images can be studied with these methods and can be useful dataset for retrieval applications in medical field.

REFERENCES

1. David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", In *International Journal of Computer Vision*, 60, 2, pp. 91– 110, 2004.
2. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features", In *proceedings of the ninth European Conference on Computer Vision*, May 2006.
3. Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors, *IEEE Transaction on Pattern Analysis & Machine Intelligence*, Vol.27(10),pp.1615–1630,2005. URL <http://lear.inrialpes.fr/pubs/2005/MS05>
4. T. Linderberg, "Feature detection with automatic scale selection", *IJCV*, 30(2),pp:79 - 116, 1998.
5. Seok-Wun Ha, Yong-Ho Moon, "Multiple Object Tracking Using SIFT Features and Location Matching", *International Journal of Smart Home*, Vol. 5, No. 4,pp. 17-26, October 2011.
6. G. Carneiro and A.D. Jepson, "Multi-scale phase-based local features", In *CVPR (1)*, pp. 736-743, 2003.
7. C. Harris and M. Stephens, "A combined corner and edge detector", In *Proceedings of the Alvey Vision Conference 1988*, pp. 147–151, 1988.
8. T. Linderberg and Bart M, Haar Romeny, "Linear scale-space" In *Geometry-Driven Diffusion*, Kluwer Academic Publishers ,pages 1–77, Dordrecht, Netherlands, 1994.
9. Luo Juan, and Oubong Gwun, "A Comparison of SIFT, PCA-SIFT and SURF", *International Journal of Image Processing (IJIP)*, Vol. 3, Issue 4, pp. 143-152.
10. [P. C. Cattin, H. Bay, L. Van Gool, and G. Szekely, "Retina mosaicing using local features", In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, October 2006.
11. F. Jurie and C. Schmid, "Scale-invariant shape features for recognition of object categories", In *CVPR*, volume II, pp. 90 - 96, 2004.
12. D. Lowe, "Distinctive Image Features from Scale-Invariant Key points", *International Journal of Computer Vision*, pp. 1-28, 2004.
13. Peer Neubert, Niko Sunderhauf, and Peter Protzel, "Fastslam using SURF features: An efficient implementation and practical experiences", In *Proceedings of the International Conference on Intelligent and Autonomous Vehicles, IAV07*, Toulouse, France, September 2007.
14. Stephen Se, David G. Lowe, and Jim Little, .Mobile Robot, "Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks", In *International Journal of Robotics Research*, 21, 8, pp. 735–758, 2002.
15. Nabeel Younus Khan, Brendan McCane, and Geoff Wyvill, "SIFT and SURF Performance Evaluation against Various Image Deformations on Benchmark Dataset", *International Conference on Digital Image Computing: Techniques and Applications*, pp.501-506, 2011.
16. Hongbo Li, Ming Qi And Yu Wu, "A Real-Time Registration Method Of Augmented Reality Based On Surf And Optical Flow", *Journal Of Theoretical And Applied Information Technology*, Vol. 42, No.2, pp. 281-286, August 2012.
17. C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints", *International Workshop on Statistical Learning in Computer Vision*, Prague, 2004.
18. M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality- sensitive hashing scheme based on p-stable distributions", In *Symposium on Computational Geometry*, pp. 253-262, 2004.
19. R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning", In *CVPR (2)*, pp. 264-271, 2003
20. L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "General intensity transformations and differential invariants", *JMIV*, Vol. 4(2), pp. 171-187, 1994.

21. W. T. Freeman and E. H. Adelson, "The design and use of steerable filters", *PAMI*, Vol. 13(9), pp. 891- 906, 1991.
22. J. Goldstein, J. C. Platt, and C. J. C. Burges, "Redundant bit vectors for quickly searching high-dimensional regions", *In Deterministic and Statistical Methods in Machine Learning*, pp. 137-158, 2004.
23. M. Grabner, H. Grabner, and H. Bischof, "Fast approximated sift", *In ACCV*, Vol.1, pp. 918-927, 2006.
24. Vini Vidyadharan, and Subu Surendran, "Automatic Image Registration using SIFT-NCC", *Special Issue of International Journal of Computer Applications (0975 – 8887)* , pp.29-32, June 2012.
25. T. Kadir and M. Brady, "Scale, saliency and image description", *IJCV*, Vol. 45(2), pp. 83 - 105, 2001.
26. Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", *In CVPR*, Vol. (2), pp. 506- 513, 2004.