Comparison of Feature Detection and Matching Approaches: SIFT and SURF

Darshana Mistry
PhD Student
Department of Computer Engineering
Technical Associate, eInfochips Training and Research
Academy, Ahmedabad

Asim Banerjee
Professor
Department of Electrical Engineering
Dhirubhai Ambani Institue of Information and
Communication Technology, Gandhinagar

Abstract

Feature detection and matching are used in image registration, object tracking, object retrieval etc. There are number of approaches used to detect and matching of features as SIFT (Scale Invariant Feature Transform), SURF (Speeded up Robust Feature), FAST, ORB etc. SIFT and SURF are most useful approaches to detect and matching of features because of it is invariant to scale, rotate, translation, illumination, and blur. In this paper, there is comparison between SIFT and SURF approaches are discussed. SURF is better than SIFT in rotation invariant, blur and warp transform. SIFT is better than SURF in different scale images. SURF is 3 times faster than SIFT because using of integral image and box filter. SIFT and SURF are good in illumination changes images.
Keywords- SIFT (Scale Invariant Feature Transform), SURF (Speeded up Robust Feature), invariant, integral image, box filter

I. INTRODUCTION

Image represent complex information in easy way. In today world, image and video are used in every way. Feature represent information of an image. Features can be point, line, edges, and blob of an image etc. There are areas as image registration, object tracking, and object retrieval etc. where require to detect and match correct features. Therefore, features are found such way which invariant to rotation, scale, translation, illumination, noisy and blur images. The search of interest points from one object image to corresponding images is very challenging work. It should be such that same physical interest points has found in different views.

There are many algorithms are used to detect and match features as SIFT (Scale Invariant Feature Transform), SURF (Speeded up Robust Feature), FAST, ORB etc. SIFT and SURF are most robust and used method for feature detection and matching.

Features are matched based on finding minimum threshold distance. Distance can be found using Euclidean distance, Manhattan distance etc. If distances of two points are less than minimum threshold distance, that key points are known as matching pairs. Feature points are applied to find homography transformation matrix which are found using RANSAC.

II. SIFT (SCALE INVARIANT FEATURE TRANSFORM)

Main steps to detect and matching feature points in SIFT are as [1] [6]:

A. Scale-space extrema detection

For finding uniqueness feature, the first stage searches over scale space using a Difference of Gaussian (DoG) function to identify potential interest points that are invariant to scale and orientation. The scale space of an image is defined as L (x , y , σ) (equation 1) which is produced the convolution of a variable-scale Gaussian G(x , y , σ)(equation 2) with an input image I(x, y):

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

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

To detect efficient and robust key points, scale space extrema are found from the multiple DoG. D(x, y, σ) which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k (equation):

$$D(x,y,\sigma) = \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{3}$$
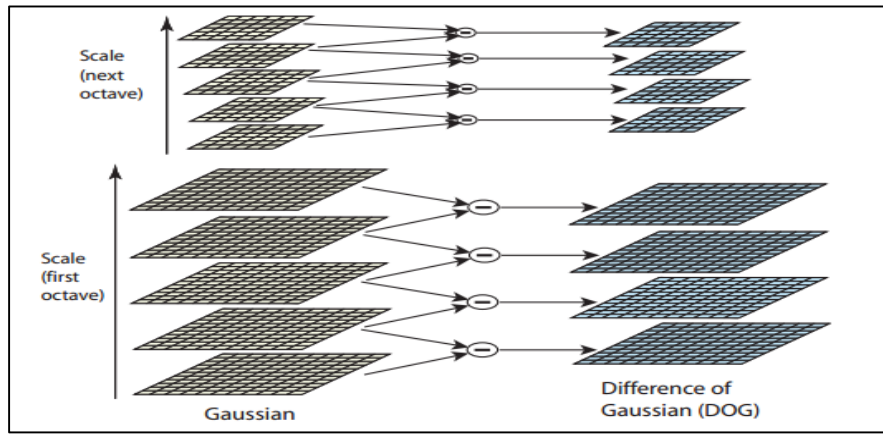
Fig. 1: For each octave of scale space, the initial image is repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated

B. Key Point Localization

At each feature point location, a detailed model is fit to determine location and scale. Key points are selected based on measures of their stability [6].

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \qquad (4)$$

$$\theta(x,y) = \tan^{-1}\left(\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x-1,y))}\right) \qquad (5)$$

C. Key Point Descriptor

The local image gradients are measured at the selected scale in the region around each key point. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

III. SURF (SPEEDED UP ROBUST FEATURE)

SURF's detector and descriptor are not only faster, but the former is also more repeatable and the latter more distinctive [2]. Hessian-based detectors are more stable and repeatable than their Harris based counterparts and observed that approximations like the DoG can bring speed at a low cost in terms of lost accuracy [2] [6].
There are main two steps in SURF:

A. Interest Point Detection

SURF concert original image to integral image. Integral Image which summed area tables is an intermediate representation of the image. It is the sum of intensity values of all pixels in input image. Image's rectangular region formed by origin O= (0, 0) and any point X=(x, y). It provides fast computation of box type convolution filters [2] [6].

$$I_\Sigma(X) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i,j) \qquad (6)$$

Based on integral image, there are only three operations (addition or subtraction) require for calculating sum of intensity values of pixels over any upright rectangular area.
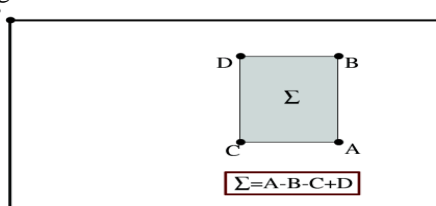

Fig. 2: Using integral images, it takes only three additions and four memory accesses to calculate the sum of intensities inside a rectangular region of any size

Integral image is convoluted with box filter. Box filter is approximate filter of Gaussian filter. The Hessian matrix $\mathcal{H}(X, \sigma)$, (as equation 7) where X=(x, y) of an image I, at scale σ is defined as follows:

$$\mathcal{H}(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \qquad (7)$$

Where $L_{xx}(X, \sigma)$(Laplacian of Gaussian) is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point X and similarly for $L_{xy}(X, \sigma)$ and $L_{yy}(X, \sigma)$.

B. Interest Point Description

In order to invariant to image rotation, we identify a reproducible orientation for the interest points. Because of that, first calculate the Haar wavelet responses in x and y direction within a circular neighbourhood of radius 6s around the interest point, with scale s (sampling step is depend on s) at which the interest point was detected. The size of wavelet which is scale depended and its side length is 4s. To compute the response in x or y direction at any scale only six operations are needed [2] [6].
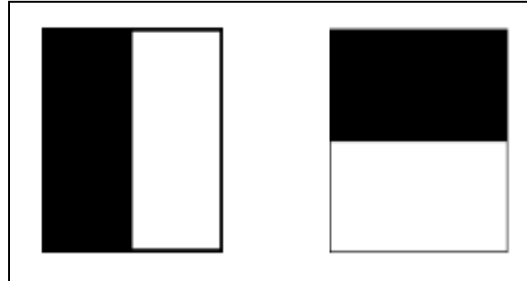

Fig. 3: Haar wavelet filters to compute the responses in x (left) and y direction (right). The dark parts have the weight -1 and the light parts +1

Once the wavelet responses are calculated and weighted with a Gaussian σ=2s centered at the interest point. The responses are represented as points in a space with the horizontal response strength along the abscissa and the vertical response strength along the ordinate. Find maximum the sum of all responses which is wavelet response in every sliding window (π/3 window orientation) (see figure 4). The horizontal and vertical responses within the window are summed. From these two horizontal and vertical, summed responses then yield a local orientation vector. The orientation of the interest point can be defined by finding the longest such vector over all windows.
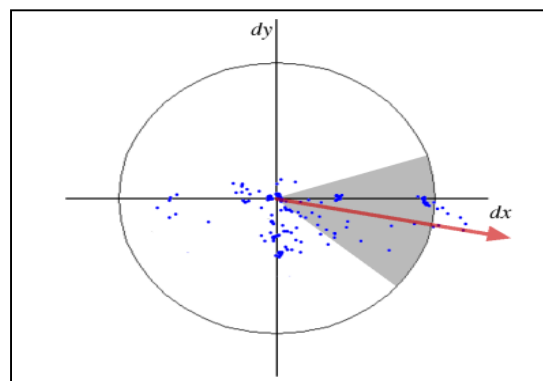

Fig. 4: Orientation assignment: A sliding orientation window of size π/3detects the dominant orientation of the Gaussian weighted Haar wavelet responses at every sample point within a circular neighborhood around the interest point.

To extract the descriptor, square region which size is 20s are constructed on interested points. Examples of such square regions are illustrated in figure 5.
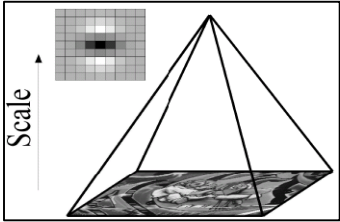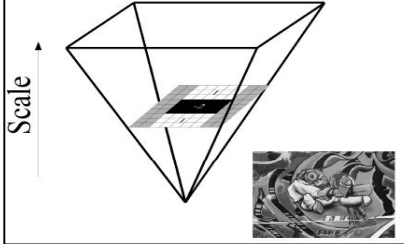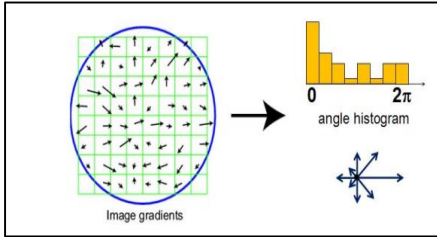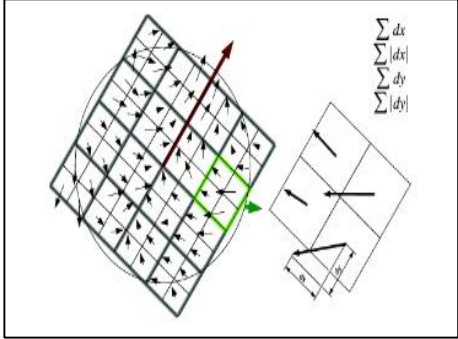

Fig. 5: Detail of the Graffiti scene showing the size of the oriented descriptor window at different scales

The wavelet responses dx and dy are summed up over each sub-region and form a first set of entries in the feature vector. In order to bring in information about the polarity of the intensity changes, extract the sum of the absolute values of the responses, |dx| and |dy|, each sub-region has a four-dimensional descriptor vector V for its underlying intensity structure $V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_x|)$. Concatenating this for all, 4 x 4 sub-regions, and this results in a descriptor vector of length is 64.

The wavelet responses are invariant to a bias in illumination (offset) and Invariance to contrast (a scale factor) is achieved by turning the descriptor into a unit vector.
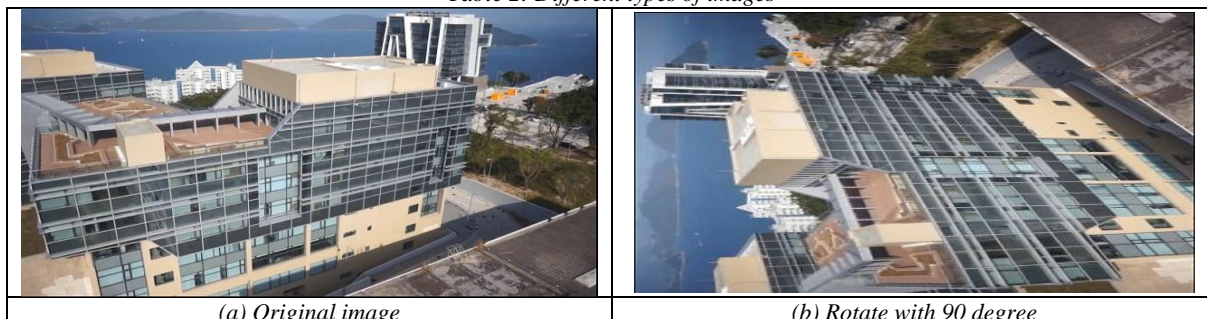
IV. DIFFERENCE BETWEEN SIFT AND SURF

Table 1: Difference between SIFT and SURF [1] [2] [3] [5]

	SIFT	SURF
Scale Space	*Difference of Gaussian (DoG) is convolved with different size of images with same size of filter.* Fig. 6. Fix filter is convoluted with down sampling images	*Different size of box filter(Laplacian of Gausian (LoG)) is convoluted with integral image.* Fig. 7. Fix image is convoluted with up sampling filters.
Key point detection	*Using of local extrema detection, apply Non maxima suppression and eliminate edge response with Hessian matrix*	*Determine the key points with Hessian matrix and Non Maxima suppression.*
Orientation	*Image gradiant magnitude and orientations are sampled around the key point location, using the scale of the key point to select the level of Gaussian blur for the image. Orientation of histogram is used for same.*	*A sliding Orientation window of size $\pi/3$ detects the dominant orintation of the Gaussian weighted Haar Wavelet responses at every sample point with in a circular neighbourhood around the interest points.*
Descriptor	*The key point descriptor allows for significant shift in gradiant positions by creating orientation histograms over 4 x 4 sample regions. The figure shows 8 directions for each orientation histogram, with the length of each arrow corresponding to the magnitude of the histogram entry.* Fig. 8. Orientation assignment	*An orientation quadratic grid with 4x4 square sub regions is laid over the interest point. For each square, the wavelet responses are computed from 5x5 samples. Descriptor of SURF is* $V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_x|)$ Fig. 9: Orientation assignments
Size of descriptor	*128 bits*	*64 bits*

V. DISCUSSION OF RESULT

In practical performance, OpenCV 2.4.9 is configured with Microsoft Visual Studio 2013 and Windows 8 OS. For data, HKHUST campus video is used [7]. The frame (1625) is used as original image. Different types of images are displayed in Table 2.
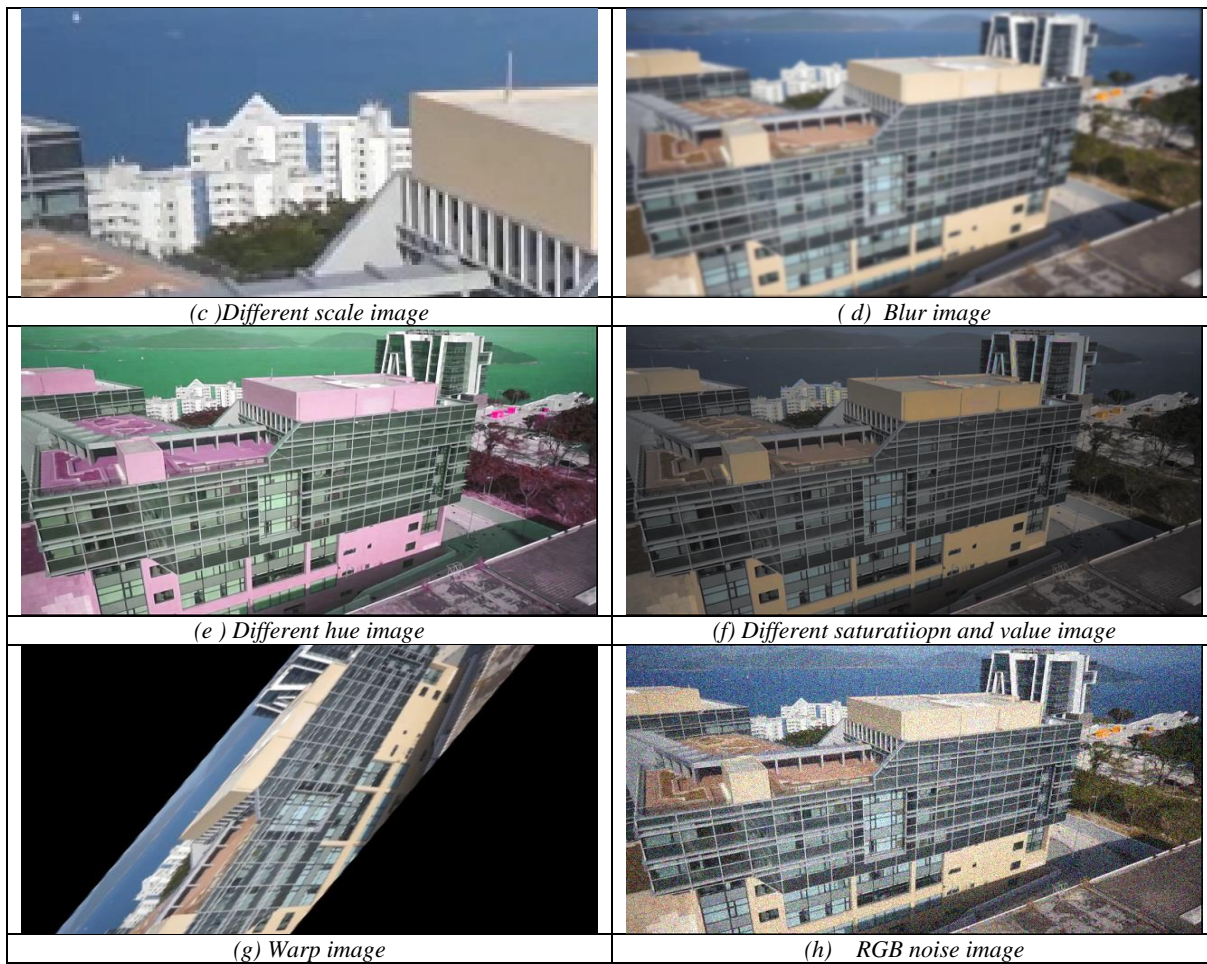
Table 2: Different types of images

(a) Original image	*(b) Rotate with 90 degree*

| *(c)Different scale image* | *(d) Blur image* |

| *(e) Different hue image* | *(f) Different saturatiiopn and value image* |

| *(g) Warp image* | *(h)   RGB noise image* |

Table 3: Feature detection and matching using SIFT and SURF

Image	SIFT				SURF			
	Feature extraction in image1	*Feature extraction in image2*	*Corresponding matching pairs after RANSAC*	*Time for extraction and matching pairs (ms)*	*Feature extraction in image1*	*Feature extraction in image2*	*Corresponding matching pairs after RANSAC*	*Time for extraction and matching pairs (ms)*
Image (a) and (b)	*3001*	*3043*	*215*	*5235.25*	*878*	*886*	*309*	*1354.27*
Image (a) and (c)	*3001*	*823*	*153*	*3511.85*	*878*	*325*	*32*	*989.488*
Image (a) and (d)	*3001*	*862*	*29*	*3626.46*	*878*	*356*	*217*	*1008.99*
Image (a) and (e)	*3001*	*823*	*153*	*4075.62*	*878*	*895*	*169*	*1215.27*
Image (a) and (f)	*3001*	*2164*	*479*	*4368.75*	*878*	*225*	*220*	*973.853*
Image (a) and (g)	*3001*	*1567*	*266*	*4263.84*	*878*	*1767*	*565*	*935.151*
Image (a) and (h)	*3001*	*3315*	*23*	*5223.4*	*878*	*992*	*57*	*1279.22*

SURF and SIFT are applying on original image with different types of images which mentioned in Table 2. In Table 3 explained how many features points are found in image1 and image2, corresponding matching pairs from two images and how much time are used for extracting and matching pairs of an image.

Table 3 information are represented as graph. Figure 10 displays about number of feature extraction in image using SIFT and SURF. Figure 11 displays about how many corresponding feature points are matched from detection of feature points using SIFT and SURF. Figure 12 displays about total time extraction for finding and matching feature points using SIFT and SURF. From Table 3 result displayed below points:

− SURF is provide good result with respect to matching feature pairs in rotation, blur image, different hue, different warp transformed and noise image than SIFT.
− SIFT is provide good result in different scale and different saturation and value image than SURF.
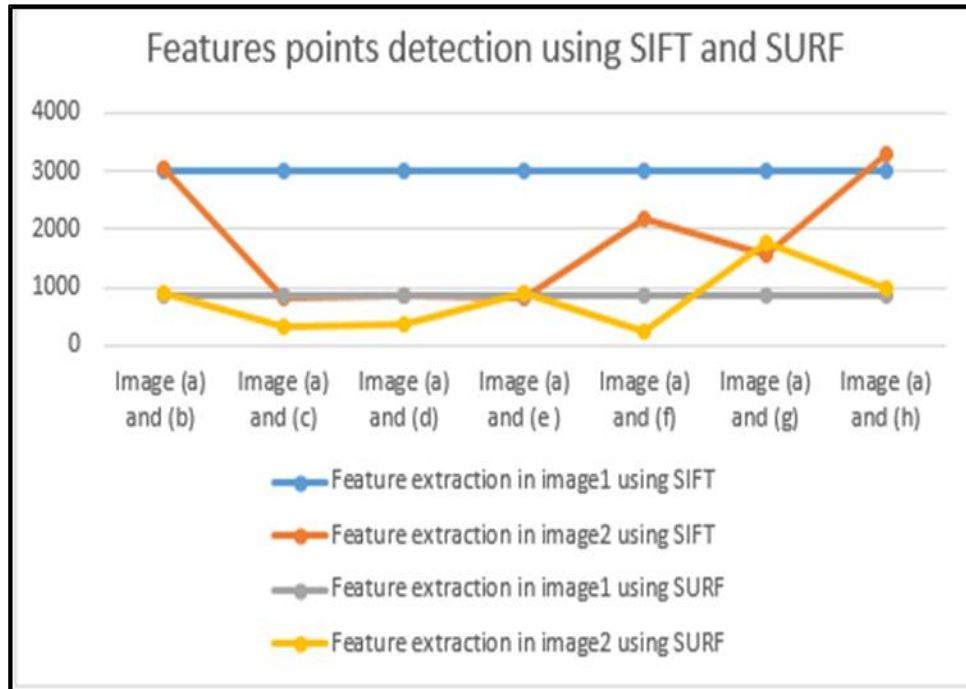− SURF is 3 times faster than SIFT.

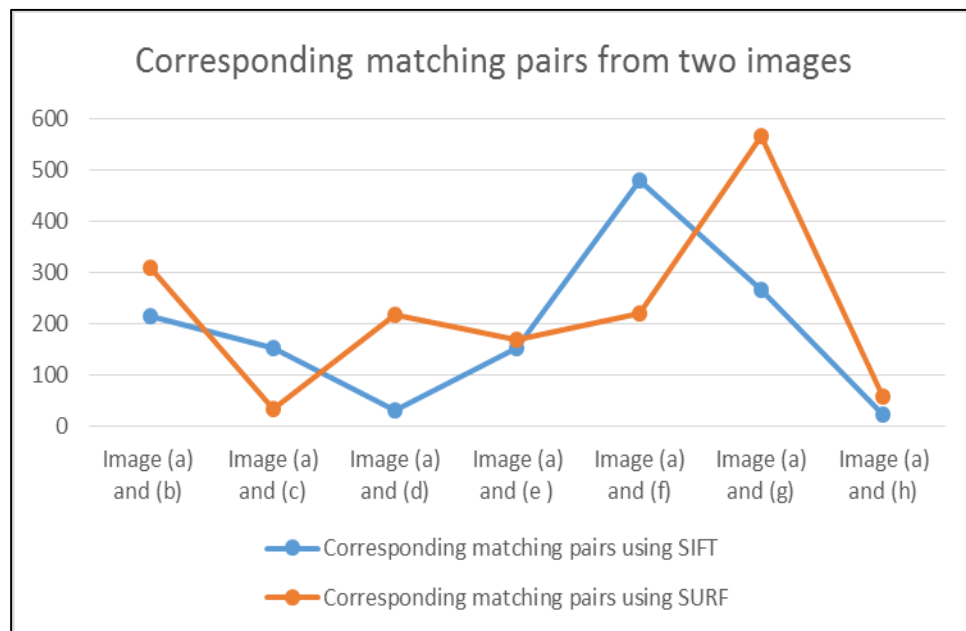Fig. 10: Feature detection in image using SIFT and SURF

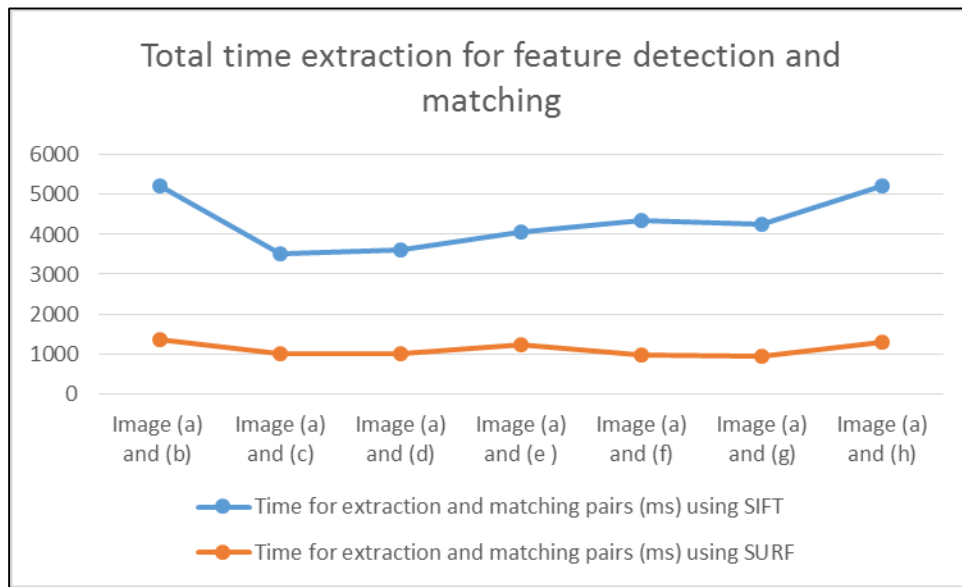Fig. 11: Corresponding matching pairs in two images using SIFT and SURF

Fig. 12: Total time extraction for feature detection and matching using SIFT and SURF

Table 4: Comparison of SIFT and SURF

Algorithm	Rotation	Scale	Blur	Illumination	Warp	RGB noise	Time cost
SIFT	Good	Better	Good	Good	Good	Good	Good
SURF	Better	Good	Better	Good	Better	Better	Better

VI. CONCLUSION

SIFT and SURF, both are robust method to find feature detection and matching. Both are invariant to rotation, scale, blur, Illumination, warping and noise area. SIFT has good result than SURF in scale area. SURF has good result in rotation, blur, warping, RGB noise and time consuming than SIFT. In illumination, both have same effect to find detect feature points.

REFERENCES

[1] David L.”Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision 60(2): 91-110, 2006.
[2] Herbert B., Andreas E., Tinne T. and Luc Van G.: Speeded up Robust Feature (SURF), Journal of Computer vision and image understanding 110 (3): 346-359, 2008.
[3] Herner. B., Tinee T., and Luc Van G.: “SURF: Speeded Up Robust Features”, Computer Vision- ECCV: 404-417, 2008.
[4] Luo J. and Oubong G.: A Comparison of SIFT, PCA-SIFT, and SURF. International Journal of Image Processing (IJIP) 3(4): 143-152, 2009.
[5] “Difference between SIFT and SURF”, https://www.quora.com/Image-Processing/Difference-between-SURF-and-SIFT-where-and-when-to-use-this-algo.23/11/2015.
[6] Utsav S., Darshana M. and Asim B.: Image Registration of Multi-View Satellite Images Using Best Feature Points Detection and Matching Methods from SURF, SIFT and PCA-SIFT 1(1): 8-18, 2014.
[7] HKHUST video data, “https://www.youtube.com/watch=OOUOPnLbjkI”.