SURF Based Image Matching from Different Angle of Viewpoints using Rectification and Simplified Orientation Correction

K.M.Goh, M.M.Mokji, S.A.R. Abu-Bakar

Abstract—Speeded-Up Robust Feature (SURF) is commonly used for feature matching in stereovision because of their robustness towards scale changes and rotational changes. However, SURF feature cannot cope with large viewpoint changes or skew distortion. This paper introduces a method which can help to improve the wide baseline's matching performance in term of accuracy by rectifying the image using two vanishing points. Simplified orientation correction was used to remove the false matching..

Keywords—affine; orientation; projective; SURF;

I. INTRODUCTION

IN a multiple camera system, process of matching corresponding objects or points from different viewpoints plays a substantial role as it will enable the extraction of information that correlate multiple view images such as overlapping pixels and object depth. Normally the matching process is applied to feature descriptors generated from all images captured from different viewpoint cameras. SIFT and SURF are the two most commonly used methods in generating feature descriptors as these methods were proven to be invariant to the image rotation, scale and noise[1]. In [2], SURF feature descriptors are matched by comparing features' contrast. In [3], a cross correlation matrix is built to identify the matching points by choosing the highest value in row and in column of current feature descriptors. In [4], orientation correction is introduced to improve numbers of matched points because features with similar appearance may be confused and thus wrongly matched. However, those methods are still vulnerable towards viewpoint changes [5-6] especially in the case of uncalibrated camera. In [5], the matching is done in the present of viewpoint calibration and depth image (Kinect style). In [7], PCA-SIFT are proposed to improve the correct match percentage but the matching rate still decreases when the viewpoint angle increased. In [8], using optical flow approach the matching is done in a series of frame in order to predict the location of the feature in the next frame. This method is applicable only when the object or the camera is moving.

K.M.Goh is with Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: kenneth86gkm@gmail.com).

M.M.Mokji is with the Department of Microelectronic and Computer Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: musa@fke.utm.my).

S.A.R. Abu-Bakar is with the Department of Microelectronic and Computer Engineering, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia (e-mail: syed@fke.utm.my).

Based on the literature, it has been shown that matching points at different angle and viewpoint is yet an issue to be improved. This paper proposed a method which is able to match points from different viewpoint angle in uncalibrated condition.

This paper is organized as follows. In part II, a brief theory of SIFT and SURF will be given. In part III, the theory on rectification using single vanishing point and two vanishing points will be presented. In part IV, the details of proposed technique will be discussed. Experimental results will be given in part V before the paper wraps up with the conclusion.

II. SCALE INVARIANT FEATURE TRANSFORM (SIFT) & SPEEDED-UP ROBUST FEATURE (SURF)

There are three main processes consist in these two methods, which are scale-space construction, orientation assignment, and descriptor generation. SIFT cannot be implemented in the real time application because of its heavy computation. SURF is the upgraded version of SIFT, in which the integral image was introduced to make the process faster.

A. Scale-Space Construction

In SIFT, the image is convolved with Gaussian filter with different scales. Then, the Hessian Matrix of the image is calculated. Feature point is identified by comparing the value of the current pixel in the Hessian matrix image with 8 surrounding pixels and 9 pixels at neighboring scale. In SIFT, the size of the image is expanded following scale while in SURF, the size of the image remains unchanged. To speed up the process, SURF uses approximation of the Gaussian filter where it is applicable with the integral image.

B. Orientation Assignment

This stage is important to make feature rotationally invariant. For SIFT, gradient magnitude and orientation of feature are computed in a certain region. Then, the orientation histogram is computed and the bin is weighted by gradient magnitude and a Gaussian window with a size of 1.5 of the scale. Three highest peaks are interpolated to get the dominant orientation. For SURF, Haar responses from the horizontal and vertical directions are used to convolve with feature's region where the size is 4 times larger than the scale. Integral images are applied here to make the process faster. Vertical and horizontal information are weighted with Gaussian filter and the orientation is calculated by shifting a segment of a circle covering 60 degrees with the increment of 10 degree every time. Summation of responses is computed in every segment and the longest summation will be the dominant orientation of the feature.

C. Feature Descriptor Generation

To generate descriptor, SIFT and SURF have different approaches respectively. In SIFT a region with 16 x 16 samples array is rotated according to the orientation calculated while in SURF, the region size is 20 x 20. The region for SIFT is divided into 16 subregions with even number of pixels and orientation histogram with 8 bins of each subregion is computed. 8 bins are represented as 8 vectors. There will be 128 elements descriptor vector for one feature in SIFT. The descriptor of SURF is generated by dividing region into 4 x 4 subregions. Horizontal information, *dx,* vertical information, *dy,* and their absolute values are computed for each subregion in order to get a vector of 4 elements. For the whole region, there will be 64 elements descriptor vector for each feature. The descriptor will be used for matching by comparing the Euclidean distance between descriptors. For more details and interested reader can refer to [1-2].

III. IMAGE RECTIFICATION USING SINGLE VANISHING POINT AND TWO VANISHING POINTS

Image rectification is a process of removing the affine factor and projective factor so that the object in the image will remain undistorted. In [15], the rectification is done by using few matched points in two images. In [9], the rectification is done using an ellipse and points. In [10], the rectification is done based on rotational angles of y and z axis of the input images with respect to the reference image. In [11], the rectification is done based on the estimated aspect ratio and the real height and width of the rectangular. In [12], the 8 corresponding points and epipolar geometry is needed for rectification. In [13-14], one or two vanishing points are used to rectify the image. The proposed rectification strategy in this paper follows with that in [13] because this method is simply based on lines available in images without doing any calibration.

A. Single Vanishing Point Rectification

In an image, a pair of parallel lines may meet in a point which is called as vanishing point as shown in Fig.1. This is called as projective factor.

Fig 1 All lines which are parallel in the real world meet at the vanishing point

A homogeneous point *(x,y,z)* is transformed by a matrix H_1 to become a perspective distorted point which is located at *(x₁,y₁,0,1)* where the matrix H_1 is represented in (1).

$$H_1 \cdot = \begin{bmatrix} 1 & 0 & 0 & p \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (1)$$

To get back the original location of each point, the perspective distorted points must be multiplied with the inverse H_1 matrix and the original location *(x,y)* is calculated by (2) and (3) where p_r is calculated using (4).

$$x = x_1 /(1 - p_r x_1) \qquad (2)$$

$$y = y_1 /(1 - p_r x_1) \qquad (3)$$

$$p_r = 1/ x_1 \qquad (4)$$

B. Two Vanishing Points Rectification

In certain conditions, an image may contain two vanishing points in different direction as shown in Fig.3.
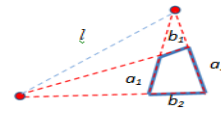

Fig. 2 A quadrilateral with two vanishing points in different direction

First, the projective factor needs to be removed. Two pairs of lines of a_1, a_2, and b_1, b_2 are extracted from the image where these pairs of lines are parallel in the real world, but not on the captured image. Then, two vanishing points are computed using cross product as given in (5) and (6) and a vanishing line *l* connecting the two points are obtained using (7). *l* contains the common three linear line elements which are (l_1, l_2, l_3). Then, a transformation matrix H_2 is established based on elements of the vanishing line as described in (8). The matrix is applied on image *J* to produce the affinely rectified image, J_a using (9) where the projective factor is removed resulting in two pairs of parallel lines a_1, a_2, and b_1, b_2 in both real-world and on the captured image.

$$v_1 = a_1 \times a_2 \qquad (5)$$

$$v_2 = b_1 \times b_2 \qquad (6)$$

$$l = v_1 \times v_2 \qquad (7)$$

$$H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \qquad (8)$$

$$J_a = H_2 J \qquad (9)$$

Next is to remove the affine distortions occurred in the image. In order to remove these distortions, dual degenerate conic theory is used. Basically the process needs to form an affine transformation matrix H_3 as shown in (10) where the elements of *A* can be obtained using (13)

To get H_3, two pairs of physically orthogonal lines d_1, d_2 and e_1, e_2 are extracted and fit into (11) where d_1 is orthogonal to e_1 and d_2 is orthogonal to e_2. Note that these two pairs of orthogonal lines cannot be parallel. To rectify the image, a pair of orthogonal lines must fulfill the property as shown in (11) where C is dual degenerate conic matrix which is represented as in (12). C is then used to compute S where A is the elements in H_3.

$$(d_n)^T C e_n = 0 \tag{11}$$

$$C = \begin{bmatrix} C_1 & C_2 & 0 \\ C_3 & C_4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{12}$$

$$S = AA^T = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix} \tag{13}$$

Matrix S is decomposed into UDU^T using singular vector decomposition (SVD) method since S is a symmetric matrix. A can be calculated using (14) and these elements are later replaced into H_3. With the matrix H_3, rectification process is completed using (15) with J_{rec} as rectified image.

$$A = U * \sqrt{D} * U^T \tag{14}$$

$$J_{rec} = H_3 J_a \tag{15}$$

Fig.3 shows an example of the rectification.
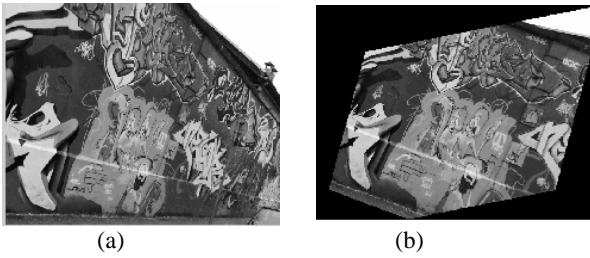

| (a) | (b) |

Fig 3. Image before rectification (a) and after rectification (b)

IV. PROPOSED METHOD

In this paper we proposed a method which can increase numbers of matched points from images taken from different viewpoint angles. Here image rectification is used to remove the perspective distortion of the image while simplified orientation correction technique is used to reduce false matched points. The overall matching process is illustrated in Fig.4.
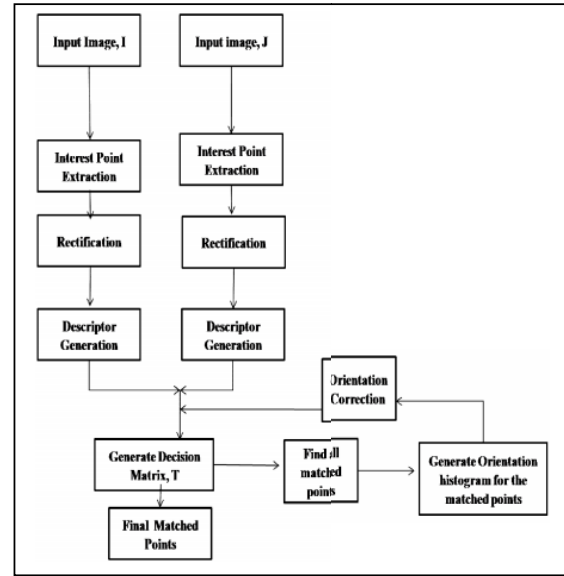

Fig 4 Process of the proposed method

The orientation correction will be looped for only once to keep the computation burden low. Two images at different viewpoints are chosen. Then, interest points are extracted from both images and the interest points are denoted as $P_z = \{ p_{z,1}, p_{z,2}, p_{z,3}, \dots, p_{z,M} \}$, where M is the total number of extracted interest points and $z = \{1,2\}$ for the case of matching 2 images. Then, rectification will be applied on both images for easy matching. For the rectification process, at least two parallel lines in real world are needed for single vanishing point rectification and four lines for two vanishing points rectification. In this work, these lines are extracted using Hough Transform. Single point rectification can only be applied onto the image with small viewpoint changes and without rotation by using (1) to (4). For large viewpoint changes, two points rectification are needed. For two points rectification, equation (5) to (15) are used to remove the projective and affine factors. After that, descriptors representing each of the interest point are generated based on SURF method. The descriptors are represented as $D_m = \{ d_{m,1}, d_{m,2}, \dots d_{m,64} \}$ where $m = \{1,2,\dots,M\}$. For the second image, m and M will be replaced with n and N respectively. Next, a decision matrix $T(m,n)$ with size $M \times N$ is formed where M and N are the number of features extracted from the two images respectively. The elements inside T are Euclidean of all features as denoted as (16).

$$T(m,n) = \sqrt{\sum_{k=1}^{64} (D_{m,k} - D_{n,k})^2} \tag{16}$$

To get these matched points, the method in [4] is adopted where point m and n are considered matched only if $T(m,n)$ is minimum at both column m and row n. In this paper, set of the matched points denoted as $\hat{p}$. Then, an orientation histogram is generated based on these matched points so that the rotational angle between the two images can be estimated by computing average orientation angle, $\hat{p}_{avg}$ at the highest bin of the

histogram. New descriptors are generated based on the Θ_{avg} to improve the number of correctly matched points.

We make use a simplified version of the technique implement in [4] to reduce the computation burden. Here, we do not generate new descriptor as it consumes heavy computation. Instead, we renew only the decision matrix T into T_{new} according to (16) where T_R is a set of false matched points and $T_R \subset p$. These false matched points are identified when their rotational angles are outside of ± 20 degree from the $_{avg.}$ Then, the matched points are again searched but now within the T_{new} based on the same minimum value on both row and column rules. Finally, the minimum value for the row and column is chosen to get the corrected matched points.

$$T_{new} = \left\{ x \mid x \in T \ \& \ x \notin T_R \right\} \qquad (17)$$

V. EXPERIMENTAL RESULT

To test the proposed method, we used two dataset for viewpoint change, which are Graffiti5 and Graffiti6. 7 images from Graffiti5 and 5 images from Graffit6 were chosen. Some of these images contain rotational changes. Fig.5 shows a few sample images from both datasets. Note that viewpoints angle of the images increases from top to bottom. Images in the first row are the reference image. These images are considered as complex images they have many regions that have similar appearance.
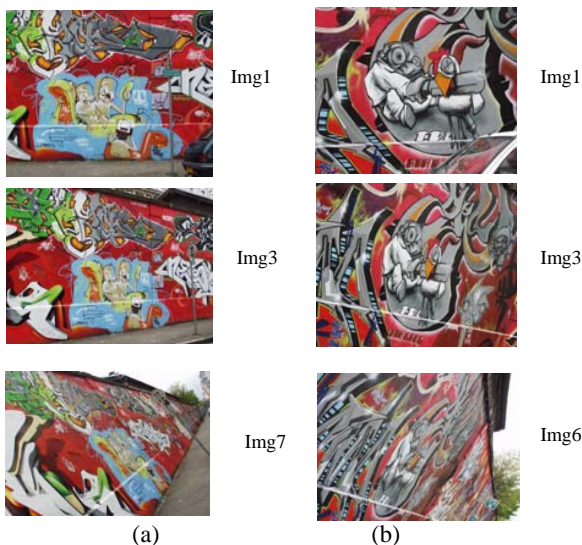

(a)   (b)
Fig 5 (a) images from dataset Graffiti5 and (b) images from dataset Graffti6

To compare the performance of the proposed method, we also performed tests using SIFT, SURF, SIFT with orientation correction, and SURF with orientation correction. The evaluation was measured based on number of correctly matched points. Besides, increments of percentages of correctly matched points over total matched points were also recorded. Results on number of correctly matched points are illustrated in Fig.6 for Graffiti5. From Fig.6, the number of correctly matched points decreases because number of feature points which supposed to be detected in the input images is getting less.

Increment in the number of correctly matched points in img2 and img3 is not significant because the change in the viewpoint angle between the reference image and input images is insignificant.
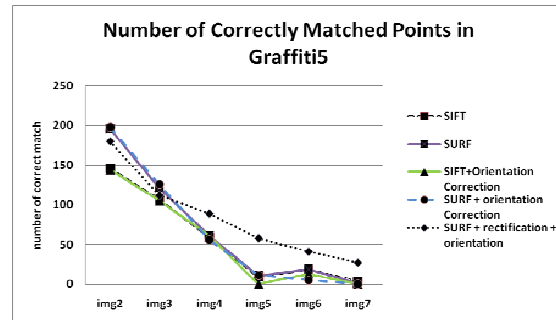

Fig. 6 Number of Correctly Matched Points in Graffiti5 in different angle of viewpoint

From img4 to img7, there has been a significant improvement in numbers of correctly matched points from 22 to 50 points. Percentage-wise this amounts to an increment from of around 18% to 44% over the total matched points. In img4, percentage of the increment was around 21% to 25%. In img5, percentage increase was from 35% to 44% while in img6, the increase was from 19% to 30%. In img7, in which the image was taken from the most extreme angle, the original SIFT and SURF were not able to match features because matching of the feature was based on the vertical and horizontal information, while both information were already distorted. However, the proposed method was able to improve the matching of about 25% to 29%. Fig.7 illustrates the number of correctly matched points in Graffiti6.
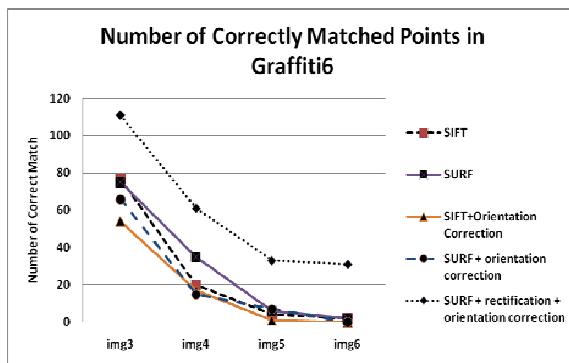

Fig. 7 Number of Correctly Matched Points in Graffiti6 in different angle of viewpoints

The proposed method also significantly detected more points in all images as shown in Fig.7. Numbers of correctly matched points of all images increases from 29 to 45 points, which amounts to percentage improvement from 20% to 42%. In img3, the increment range is 22% to 41%. In img4, the range is in between 24 % and 42%. The percentage increase for img5 is around 25% to 30% while for img6, the increments is in between 21% and 22%. Img6 is the most extreme angle image and only 0 to 2 points were correctly matched in all previous method.

With the proposed method, 31 points were able to be matched. Fig.8 and Fig.9 shows the matched points using SURF and matched points using proposed method in the img6 in Graffiti6.

Fig. 8 Matched Points using SURF

Fig. 9 Matched Points using proposed method. Correctly matched points significantly increase

VI. CONCLUSION

This paper introduced an approach for matching feature points from different viewpoint for stereovision. Since original SIFT and SURF are not able to handle the matching of images with perspective distortion, the rectification and orientation correction proposed in this paper able to reduce number of false matched points and subsequently improve the number of matching points. Totally 10 samples have been used to test the proposed method and the method generally increases the number of correctly matched points.

REFERENCES

[1] D. G. Lowe, "Distinctive image features from scale-invariant key points," International Journal of Computer Vision, pp.91-110, 2004, vol.60.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] H. Bay, T. Tuyteplaars, and L. V. Gool. " SURF: speeded up robust features." European Conference on Computer Version, 1:404-417, 2006.

[3] F.Zhao, Q.M. Huang, and W. Gao, "Image matching by normalized cross-correlation," Acoustics, Speech and Signal Processing, 2006.

[3] K.M.Goh, M.M.Mokji, S.A.R.Abu-Bakar, "Improved Rotational Matching of SIFT and SURF." Fourth International Conference on Digital Image Processing (ICDIP 2012), SPIE.

[4] Wang, Y., & Feng, J. (n.d.). Viewpoint Invariant Descriptor for RGB-D Images. Unpublished

[5] K. Mikolajczyk and C. Schmid. "Performance evaluation of local descriptors." IEEE transactions on pattern analysis and machine intelligence, 27(10):1615{30, Oct. 2005.

[6] K. Mikolajczyk and C. Schmid. "Performance evaluation of local descriptors." IEEE transactions on pattern analysis and machine intelligence, 27(10):1615{30, Oct. 2005.

[7] Juan, L., and Gwun, O. "A Comparison of SIFT , PCA-SIFT and SURF." International Journal of Image Processing IJIP 3, 143-152. 2009

[8] Clark, A.J.; Green, R.D.; Grant, R.N.; , "Perspective correction for improved visual registration using natural features.," Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference , vol., no., pp.1-6, 26-28 Nov. 2008.

[9] Feng Guo; , "Plane Rectification Using a Circle and Points from a Single View," 18th International Conference on Pattern Recognition, 2006. ICPR 2006., vol.2.

[10] Lingfeng Xu; Au, O.C.; Wenxiu Sun; Yujun Li; Sung Him Chui; Chun Wing Kwok; , "Image rectification for single camera stereo system," Image Processing (ICIP), 2011 18th IEEE International Conference on , vol., no., pp.977-980, 11-14 Sept. 2011.

[11] Zhengyou Zhang , Li-Wei He, "Whiteboard scanning and image enhancement"http://research.microsoft.com/enus/um/people/zhang/papers/tr03-39.pdf p.11

[12] Uncalibrated Stereo Rectification for Automatic 3D Surveillance". Ser-Nam Lim, Anurag Mittal and Larry S. Davis. IEEE ICIP 2004, Oct 24-27, Singapore

[13] Liebowitz, D. and Zisserman, A. "Metric Rectification for Perspective Images of Planes"

[14] G. Fangi, G. Gagliardini, E.S. Malinverni, Photointerpretation and small scale stereoplotting with digitally rectified photographs with geometrical constraints, in: Proceedings of the CIPA Congress 2001, September 18–21, 2001.

[15] Zilly, F. and Mueller, M. and Eisert, P. and Kauff, P."Joint Estimation of Epipolar Geometry and Rectification Parameters using Point Correspondences for Stereoscopic TV Sequences," 3DPVT, Paris, France, May 2010.