Single-Viewpoint Panorama Construction With Wide-Baseline Images using SIFT And SURF Features

Kripa S, Monisha Menon

[1](PG Student, Department of Electronics &Communication Engineering, Marian Engineering College
Trivandrum, India)
[2](Assistant Professor, Department of Electronics &Communication Engineering, Marian Engineering College
Trivandrum, India)

Abstract: *This paper presents an image stitching approach, which can produce visually plausible wide baseline panoramic images in which the inputs are taken from different viewpoints. This method got some variations from the previous methods, this approach uses Harris corner detector for the detection of corner point and the image stitching based on common stitching algorithms such as scale invariant feature transform (SIFT) combined with speeded up robust features (SURF) feature detection and random sample consensus (RANSAC) parameter estimation. It also uses homography similarity to identify if input images have enough correlation. This paper improves the composition and achieves good performance on misaligned areas.*

Keywords - *homogarphy, image acquisition, image conversion, RANSAC, SIFT, SURF, single view-point panorama.*

I. INTRODUCTION

Images are obviously an integral part of our daily life. Combining multiple photographic images with overlapping fields of view to produce one panoramic image or high-resolution image is called Image stitching or photo stitching. It can be performed commonly through the use of computer software; most approaches to the image stitching require nearly exact overlaps between the images and identical exposures to produce seamless results although some stitching algorithms actually benefit from differently exposed images by doing HDR (High Dynamic Range) imaging in regions of overlap. Some of the digital cameras can stitch their photos internally. Stitched images are used in applications such as interactive panoramic viewing of images, architectural walk-through, multi-node movies and other applications using images acquired from the real world.

Image stitching aims to combine multiple images to produce a high-resolution image. Image stitching has been researched in various fields, and there are many algorithms which handle image stitching efficiently. In this paper for the image stitching process the image acquisition and conversion is done firstly. Secondly detection of each corner of the images using Harris corner detection operator is done which efficiently predicts the features. Up to now common image stitching algorithm uses scale invariant feature transform (SIFT) [2] or speeded up robust features (SURF) [5] algorithm to detect feature points in images. Here we use SIFT combined with SURF to improve the performance. Thirdly from the detected feature points matching points are extracted. Finally estimate homography matrix using random sample consensus (RANSAC) [4] algorithm from matching point. Then overlap two images by using geometric transformation using homography matrix. Thus most of the images can produce combined image successfully. But in several environments, there are some problems that can confirm by eyes, such as misalignment. Misalignment is caused by difference of distance in image. The combined SIFT and SURF features can improve the performance of matching the baseline alignment, and make the alignment correction for linear and orientation alignment are preserved.

II. RELATED WORKS

1. General Image Stitching

Image stitching generally uses either SIFT or SURF algorithm to detect feature points each of two images. SIFT can detect more exquisite than SURF, but slower than SURF. Then approximated nearest neighbour (ANN) algorithm is used to extract matching points from this feature points. Estimate the homography matrix by using RANSAC parameter estimation algorithm where the matching is found. Then overlap an image which is transformed by homography matrix onto another image. Finally intensity difference between two images is reduced by multi band blending. Then image stitching is done. But, unnatural result is produced in some condition. Which arises in conditions that image contain various distance object. Because, feature points and matching points cannot focus on all distance points.

2. Scale Invariant Feature Transform (SIFT)

The SIFT algorithm (Scale Invariant Feature Transform) proposed by Lowe is an approach extract distinctive invariant features from images. In the work done by Tejasha Patil, Shweta, Mishra ,Poorva Chaudhari and Shalaka Khandale an improvement of the original SIFT algorithm provides more reliable feature matching for the purpose of object recognition is proposed. SIFT is used in computer vision problems based on feature matching including object recognition, pose estimation, image retrieval and for many other applications. However, in real world applications SIFT still needs improvement in the algorithm's robustness with respect to the correct matching of SIFT features. The main idea of SIFT is to divide the features extracted from both the images i.e. the test and the model object image; to various sub-collections before they are matched. In SIFT the features are divided into several sub collections considering the features arising from different octaves, that is from different frequency domains.

The major stages of computation used to generate set of image features are as follows:

➢ Scale-space extreme detection: The first stage of computation searches over all scales and image locations and is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation□

➢ Keypoint localization: Location is determined by fitting a detailed model and scale. Based on measures of their stability keypoints are selected.□

➢ Orientation assignment: For each keypoint location based on local image gradient directions one or more orientations are assigned. Further operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thus providing invariance to these transformations.□

➢ Key point descriptor: This approach transforms image data into scale invariant coordinates relative to the local features. In the region around each key point the local image gradients are measured at the selected scale. These are then transformed into a representation which allows for significant levels of local shape distortion and change in the illumination. □

➢ Many previous approaches use SIFT to find correspondences. For extremely challenging data, ASIFT can be adopted to obtain more feature matches.□

3. Speeded Up Robust Features(SURF)

SURF was first presented by Herbert Bay[5], et al., at the 2006 European Conference on Computer Vision. An application of the algorithm is patented in the United States. SURF is a patented local feature detector and descriptor in computer vision. It can be used for tasks such as object recognition, image registration and classification or three dimensional reconstructions. It is inspired partly by the scale-invariant feature transform (SIFT) descriptor. The standard version of SURF is several times faster than SIFT and is more robust against different image transformations while comparing with SIFT.

SURF uses an integer approximation of the determinant of Hessian blob detector to detect the interest point, which can be computed with 3 integer operations using an integral image which are pre computed. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest. These are computed with the aid of the integral image. SURF descriptors have been used to reconstruct 3D scenes, to locate and recognize objects, people or faces, to track objects and to extract points of interest.

An "upright" version of SURF (called U-SURF) is variant to image rotation and therefore faster to compute and better suited for application where the camera remains more or less horizontal.

4. Random Sample Consensus Algorithm(RANSAC)

The Random sample Consensus algorithm (RANSAC) proposed by Fischler and Bolles is a general parameter estimation approach designed to avoid outliers of large proportion in the input data. Many of the common robust estimation techniques like M-estimators and least median squares that have been adopted by the computer vision community from the statistics literature and works differently than RANSAC. RANSAC was developed from within the computer vision community. RANSAC is a resampling technique that generates candidate solutions by using the minimum number of observations (data points) which is required to estimate the underlying model parameters. As pointed out by Fischler and Bolles, unlike conventional sampling techniques that use as much of the data as possible in order to obtain an initial solution and then proceed to prune outliers, RANSAC uses the smallest set possible and proceeds to enlarge this set with consistent data points.

The basic algorithm is summarized as follows:

□ Select randomly the minimum number of points required to determine the model parameters.□

□ Then the parameters of the model are solved.□

□ Number of points from the set of all points fit with a predefined tolerance is determined.□

☐ If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold τ, using all the identified inliers model parameters are re-estimated and then terminate. ☐

☐ Else, repeat steps 1 through 4. ☐

III. IMAGE STITCHING

1. *Image Acquisition and Conversion*

In image processing, the action of retrieving an image from some source can be defined as image acquisition, usually a hardware-based source, so it can be passed through whatever processes need to occur afterward, this is called image conversion. Performing image acquisition in image processing is always the first step in the workflow sequence because, without an image, no processing is possible. The image that is acquired is completely unprocessed. Depending on the field of work, it can be processed.

2. *Corner Feature Point Detection*

Corner is the point in which testing function energy is very intense in any direction changes. Corners are one of the most important features of the data information. The result of image processing depends on the accuracy and quality of the corner detection, and can determine the outline features and important information of the image. Corner detection are used for optical flow velocity measuring, camera calibration, measurement and positioning motion estimation, etc., and has become an important tool for image processing. Corner detection can keep the useful image information. It can reduce data redundancy and improve the detection efficiency.

In order to take full advantage of the corner information in different scales, this paper uses Harris corner detector operation. Harris and Stephens improved upon Moravec's corner detector by considering the differential of the corner score with respect to direction directly, instead of using shifted patches that is the corner score is often referred to as autocorrelation function.

In this paper Harris edge detection is used to do the feature alignment or feature point detection [3] for that the RGB input image is first converted to gray scale image and then it is double précised. Normally each feature point is represented as a weighted sum of their four enclosing control vertices, and minimizing alignment errors of the warped points over all features which uses affine transformation for each grid. Instead of assuming affine transformation here Gaussian filter is used i.e. providing Gaussian transformation to the whole image. Assume two Gaussian filters g_1 and g_2 with standard deviation σ_d and σ_i respectively and is then convoluted so the center portion of the matrix is obtained. Consider taking an image patch over the area *(u,v)* and shifting it by *(x,y)*. Let the image is I and its partial derivative is I_x and I_y.

The Harris matrix is given as equation (1)

$$H = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \qquad (1)$$

The corner point is identified using the equation (2),

$$R = \frac{det(H)}{trace(H) + \epsilon} \qquad (2)$$

where ϵ is a small positive constant. The corner point close to the image border is not required since the image has to be stitched so that non maximal suppression with 3x3 window is done where the image is scanned and the pixels which are not part of the local maxima are set to zero and has an effect of suppressing all image information that is not a part of local maxima. In the previous works weighted value was used for the suppression.

3. *SIFT Feature Extraction*

The next step after feature point detection is the scale preservation where the variation in orientation is detected and is amended. Here feature point is done using Scale Invariant Feature Transform (SIFT). We already saw corner detectors like Harris corner detector. They are rotation-invariant, which means, even if the image is rotated, we can find the same corners. It is obvious because corners in rotated image also. A corner may not be a corner if the image is scaled. For example see the Fig 1. A corner in a small image within a small window is flat when it is zoomed in the same window. So Harris corner is not scale invariant thus we can improve the performance by using SIFT.
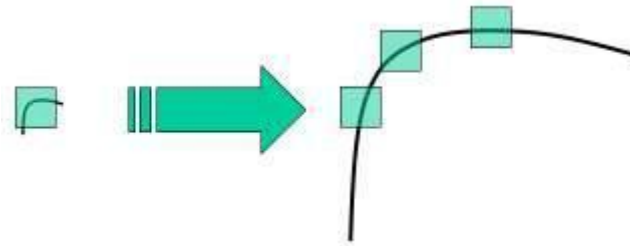
Fig.1. Corner points when image is scaled.

SIFT [2] is an algorithm to detect and describe local features in an image. SIFT key points of objects are first extracted from a set of reference images and stored in a database. In a new image an object is recognized by individually comparing each feature from the new images to this database and finding candidate matching features based on Euclidean distance of their feature vectors. Then the subsets of key points that agree on the object and its location, scale and orientation in the new image are identified to filter out good matches from the full set of matches.

Here first image is taken as the input and is considered as the first parameter then two more parameters like circle with a radius 10 since the SIFT descriptor used here is circle and an enlargement factor of 1.5 is defined. The output from the Harris corner detector i.e. the point detected from Harris corner detector is given to the SIFT descriptor. By using the SIFT, orientation can be found i.e. the angle variation in the images and the point where it varied can be detected. For the primary measurement of angle, number of angles and number of binaries is taken and from that number of samples is found. Let the number of angles be A, number of binaries be B and number of samples be S then S=B*B. Then the angle variation between $0°$ and $360°$ is found since all the point between this specified intervals have to be detected and for finding those points the angle is shifted by $45°$. To find the accurate point on where the angular variation occurs and to know how much it get varied SIFT array has to be created which needs the number of angles (A) and their locations.

Consider an edge image as 1 and then generate the de-Gaussian with this value. On the x-axis (I_x) and y-axis (I_y) of the corresponding image 2D digital filtering is done. From those values the gradient magnitude (I_{mag}) and the angle (I_θ) is found using the equation (3) and (4) respectively.
If needed assume I_θ as 0.

$$I_{mag} = \sqrt{I_x^2 + I_y^2} \qquad (3)$$

$$I_\theta = \tan^{-1}\frac{I_x}{I_y} \qquad (4)$$

Then a default grid i.e. the mesh grid centered at zero is applied to all the images in the interval, L of $\left[\frac{2}{B}; \frac{2}{B}; 2\right]$ on x-axis and $\left[L - \frac{1}{B} - 1\right]$ on y axis is created and the x-axis and y-axis of the grid is then reshaped based on the number of samples (S) and let it be g_x on x-axis and g_y on y axis.

An orientation image I_o is then defined with height, width and number of angles (A) of the image. Then a temporary array (tmp) is created. I_o can be further defined as, i.e. the column element of I_o is the dot product of the temporary array (tmp) created and I_{mag} and this is calculated for the total number of angles. This means that
the weight provided by the point which has angular variation is (5),

$$I_o = \text{tmp}* I_{mag} \qquad (5)$$

The Harris detector output is taken and its x value and y value is multiplied with the enlargement factor so that these values get enlarged by a factor of 1.5. Let the enlarged values of x-axis and y-axis be C_x and C_y respectively. For finding the co-ordinates of the sample point, increment the grid values g_x and g_y. Therefore the coordinate on x-axis changes to g_{xt} and g_{yt} where g_{xt} is the sum of the product of g_x with enlargement factor and C_x likewise g_{yt} is the sum of the product of g_y with enlargement factor and C_y. By using these values minimum and maximum value of x-axis and y-axis is found out i.e. negative infinity and positive infinity of mesh grid on both axis's is found. Thus mesh grid is defined from low value of x to high value of x in the x-axis and low value of y to high value of y in the y-axis. Then this new mesh grid is again resized for getting more accurate result. Later on the difference between the output from the Harris point detector and the output

from the shifted Harris point detector (i.e. shifted by enlargement factor) is taken and its absolute value on x-axis and y-axis is noted. This value is the weighted value. Therefore the distance measured gives the weight of the value then the weight corresponds to the x-axis and y-axis is taken i.e. the total weight depends upon the change in value on the x-axis and y-axis.
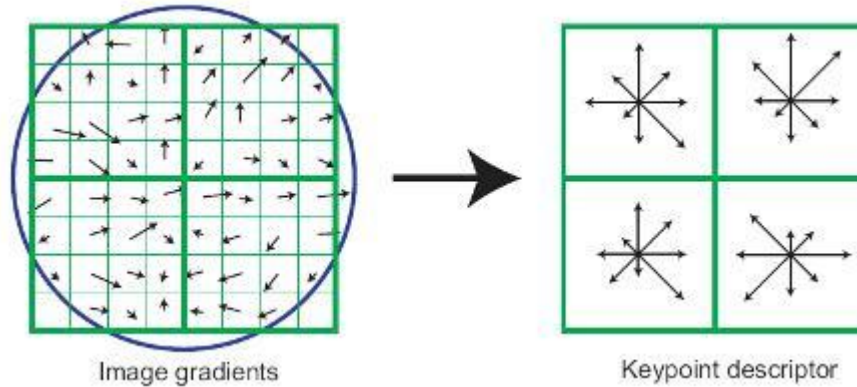

Fig. 2 (a) Image gradient (b) Key point descriptor

The orientation value I_o is then given to a temporary array (temp). The product of the temporary array and the total weight calculated on the above step is multiplied to get the current SIFT, this is done to provide more accuracy and has to find this value for all the images. After this step the current value should be reshaped i.e. corresponding to the value found new value should replace it. Then the descriptor should be normalized and suppress the large gradient. Finally renormalize to unit length.

4. SURF Feature Extraction

The next step is the SURF [5] detector which will enhances the output. SIFT mainly handle with the orientation whereas SURF can handle other constraints to. To detect interest points, SURF uses an integer approximation of the determinant of Hessian detector, which can be computed with 3 integer operations using a pre computed integral image. Its feature descriptor is based on the sum of the Haar wavelet response around the point of interest and are computed with the aid of the integral image. SURF descriptors have been used to reconstruct 3D scenes, to locate and recognize objects, people or faces, to track objects and to extract points of interest. The SURF algorithm is based on the same principles and steps as SIFT; but details in each step are different. The algorithm has three main parts like interest point detection, local neighbourhood description and matching. SURF uses square-shaped filters as an approximation of Gaussian smoothing. If the integral image is used, filtering the image with a square is much faster as in (6).

$$S(x, y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i, j) \qquad (6)$$

Using the integral image sum of the original image within a rectangle can be evaluated quickly, requiring evaluations at the rectangle's four corners. The output from the integral image is then taken and its first two points is assumed as the initial sample.

For extracting the interest point fast hessian is taken. Here the index map is filtered and then the response map is created for that we have to get the image attributes. First image and the second image are taken then it is divided by the initial sample. The new value is then propagated and computed from each point. Let the first image is taken as *W* and the second image is taken as *H* and the initial sample which is the cumulative sum is taken as *C*. The new values formed are *w* and *h* and it is defined in equation (7) and (8) respectively.

$$w = \frac{W}{C} \qquad (7)$$

$$h = \frac{H}{C} \qquad (8)$$

At each individual scale, the Hessian affine detector chooses interest points based on the Hessian matrix at that point is (9).

$$H(x) = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix} \qquad (9)$$

where L_{xx} is second partial derivative in the x direction L_{yy} *is second partial derivative in y direction* and L_{xy} is the mixed partial second derivative in the a and b directions. It's important to note that the derivatives are computed in the current iteration scale. As discussed in the Harris affine region detector article, the derivatives must be scaled appropriately by a factor related to the Gaussian kernel σ_I^2.

Interest points are those points that simultaneously are local extrema of both the determinant as in equation (10) and trace of the Hessian matrix as in equation (11) at each scale. The trace of Hessian matrix is identical to the Laplacian of Gaussians (LoG)

$$Det = \sigma_I^2(L_{xx}\, D_{yy} - L_{xy}^2 \qquad (10)$$

$$TR = \sigma_I\ (L_{xx} + L_{yy} \qquad (11)$$

these interest points based on the Hessian matrix are also spatially localized using Laplacian of Gaussians. Predictably, these interest points are called Hessian–Laplace interest points. Furthermore, using these initially detected points, an iterative shape adaptation algorithm is used by the Hessian affine detector to compute the local affine transformation for each interest point. The implementation of Hessian algorithm is almost identical to that of the Harris affine detector; and the above mentioned Hessian measure replaces all instances of the Harris corner measure. Then get extra response of the image is taken for that step size, border, lobe and size of the filter is taken. Then a normalization factor N is considered and is defined as in (12),

$$Normalization\ factor\ (N) = \frac{1}{filter\ size^2} \qquad (12)$$

where the filter size is doubled précised. Then the response component is computed and let it be D_{xx}, D_{yy}, D_{vv} and is normalized by multiplying each component with the normalized factor. Therefore those components changes as in equation (13), (14) and (15) respectively.

(13) $$D_{XX} = D_{xx} * N$$

(14) $$D_{YY} = D_{yy} * N$$

(15) $$D_{XY} = D_{xy} * N$$

Then the determinant of Hessian response and Laplacian is found by the equation (16) and (17) respectively

(16) $$Response = D_{XX}D_{YY} - 0.81D_{XY}^2$$

$$Laplacian = D_{XX} + D_{YY} \qquad (17)$$

Then the Laplacian and the response map is calculated.

If the interest point is empty then surf descriptor is used. Here we define the information about first 40 points then the size of the descriptor is defined. Orientation has to be found out for that the descriptor for different interest point is calculated. Get rounded interest point i.e. round the interest point to the nearest integer. In order to achieve rotational invariance, the orientation of the point of interest needs to be found. The Haar wavelet responses in both x- and y-directions within a circular neighbourhood around the point of interest are computed. The obtained responses are weighted by a Gaussian function centered at the point of interest, then plotted as points in a two-dimensional space, with the horizontal response in the abscissa as in equation (18) and the vertical response in the ordinate as in equation (19).

$$Response\ of\ x\text{-}axis = gaussian\ filter * haar\ x \qquad (18)$$

$$Response\ of\ y\text{-}axis = gaussian\ filter * haar\ y \qquad (19)$$

The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of size π/3. The horizontal and vertical responses within the window are summed. The two summed responses then yield a local orientation vector. The longest such vector overall defines the orientation of the point of interest. The size of the sliding window is a parameter that has to be chosen carefully to achieve a desired balance between robustness and angular resolution.

To extract SURF descriptor gets rounded interest point data. To describe the region around the point, a square region is extracted, centered on the interest point and oriented along the orientation as selected above. The interest region is split into smaller sub-regions, and for each one, the Haar wavelet responses are extracted for regularly spaced sample points. The responses are weighted with a Gaussian (to offer more robustness for deformations, noise and translation).

SURF descriptor significantly outperformed the base implementation of the Harris detector with square descriptor. SURF has closer performance as that of SIFT. SURF detector should outperform SIFT in many cases. SURF also runs very slowly. The detection step is quite fast but calculating the descriptors takes a very long time.

5. *Homography Prediction*

Algorithm 1 Outliers Rejection with Local Homographies using RANSAC

1:      **procedure** VERIFY(I_{src}, I_{dst}

2:      $B_{inlier} = \emptyset$

3:      **for all** $p \in I_{src}$ **do**

4:      Solve for H for $\{(p',q')|p' \in N(p)\}$

5:      **for all** $\{(p',q')|p' \in N(p)\}$ **do**

6:      **if** $|H(p')^2 < q'| < \gamma$ **then**

7:      $B_{inlier} \leftarrow (p',q')$

8:      **end if**

9:      **end for**

10:      **end for**

11:      **return** B_{inlier}

12:      **end procedure**

Then we have find the matching, for that the matching between image 1 and image 2 then matching between image 2 and image 3 is found using homography. In homography we define a 3x3 homography matrix H and the correspondence value is calculated where correspondence$_{12}$ is the image location and matched location of image 1 and image 2 likewise correspondence$_{23}$ is also calculated. Then the inliers have to be found out and the outliers have to be rejected for that RANSAC technique is used.

Random sample consensus (RANSAC) [4] is an iterative method used to estimate the parameters of a mathematical model from a set of observed data that contains outliers, when these outliers are to be accorded there will be no influence on the values of the estimates. Therefore, it can also be interpreted as an outlier detection method since this can detect the outliers and inliers seperately. It is a non-deterministic algorithm

because with a certain probability only it produces a reasonable result, with this probability increasing as more iteration is allowed.

The procedure is depicted in Algorithm 1 [1]. For image pair (I_i, I_i) firstly define the neighbouring sets

for each feature point in I_i and estimate the corresponding homographies. Each correspondence $(p^{'}, q^{'})$ is verified with several homographies as it is included in different neighbourhood sets. When it fits one homography, this correspondence are recognized as an inlier. After enumerating all feature points in I_i, inlier is obtained set B_1 for

(I_i, I_i). Then we swap I_i and I_i to get another inlier set B_2 by Algorithm 1.The final inlier set is $B_1 \cap B_2$.

6. Warping The Image

Image warping is a transformation which is applied to the domain of an image. Warping will modifies the geometrical properties of the image itself. The intensity of the warped image has to be same as the intensity of the original image at corresponding points. The images of features are compared and if it was matched with

another image it is stored and form as single combined frame of panorama. For constructing the panorama the output image is made as the same class of input image then assigns some input arguments. Later on linear interpolation is done. Thus the image is stitched in this step. The result produced with elapsed time of 60.053490 seconds.

IV. EXPERIMENTAL RESULT

Fig 3 shows a challenging scene example where three images are captured from different position and these images are read in the first stage. These images are selected from a file and return these filename as string. Then these images are made in same dimension.
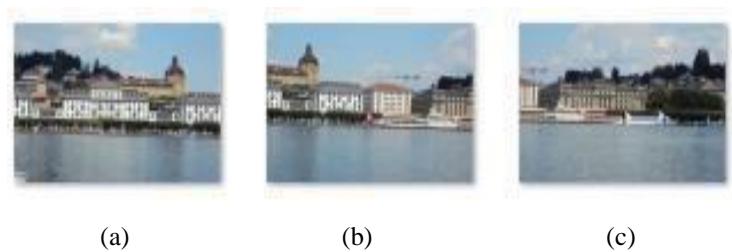

(a)   (b)   (c)

Fig. 3 Three input images captured from different positions

Fig 4 shows the Harris corner points. Below figure shows the corner point detection of the three input images and for that the input image is converted from RGB to gray.
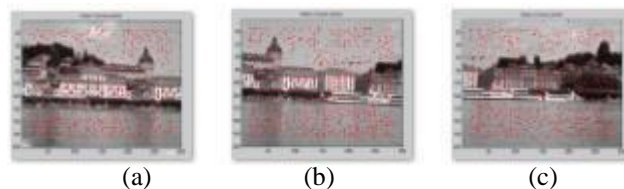

(a)   (b)   (c)

Fig. 4 (a) Harris corner point of image 1. (b) Harris corner point of image 2. (c) Harris corner point of image 3

Fig 5 shows the SURF feature combined with the matching technique. Here the image 1 and image 2 is firstly taken and is compaired for checking the matching then image 2 and image 3 is then taken and checked for its matching. Fig 6 is the inlier identification where the outliers are rejected. The homography is predicted using RANSAC by using this technique we can find out the outliers. RANSAC can estimate the parameters with high degree of accuracy.

Fig 6 shows the panoramic image here the image is stitched and this technique reduces the seams produced and improves the performance since SIFT and Surf is combined in this paper.
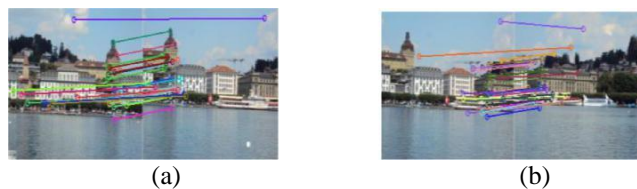
(a)   (b)

Fig. 5.(a) SURF features of image 1 and image 2 (b) SURF features of image 2 and image 3
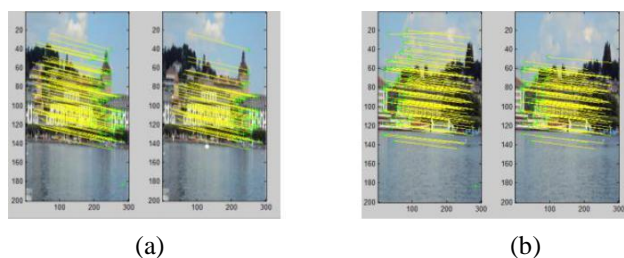

(a)   (b)
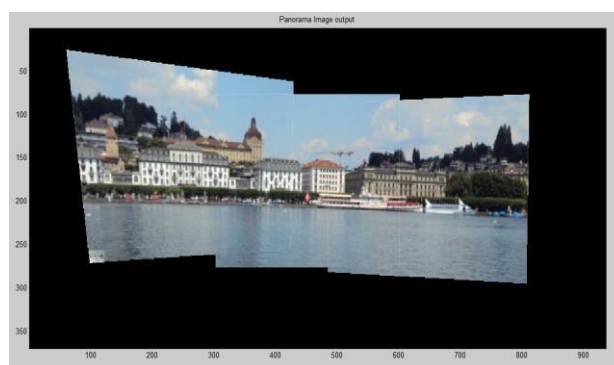
Fig. 6. (a) and (b) Inlier extraction

Fig. 7. Panoramic image

V. CONCLUSION

We have presented a new image stitching approach for wide-baseline images. Detection of each corner of the images using Harris corner detection operator efficiently predicts the features. Extracting the features of scale and alignment variation contents of certain searching range is done using SIFT descriptors. This feature extraction includes the feature alignment, scale preservation and line preservation. To improve the performance of matching the baseline alignment, we combine the SURF with SIFT and make the alignment correction for linear and orientation alignment are preserved. Multi-homography model has more degrees of freedom and can represent warping of a general smooth scene. The images of features are compared and if it was matched with another image it stored and form as single combined frame of panorama.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Guofeng Zhang, Yi He, Weifeng Chen, Jiaya Jia, and Hujun Bao, "Multi-Viewpoint Panorama Construction With Wide-Baseline Images," *vol. 25, no. 7, July 2016.*

[2]   D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2,91–110, 2004.

[3]   Ting Feng, Jie Yuan, "Feature point detection and matching of wide baseline image based on scale space theory and guided matching algorithm", *Multimedia Technology (ICMT) 2011 International Conference on*,538-542, 2011.

[4]   M A Fishler, R C. Bolles, "Random Sample Consensus A paradigm for model fitting with applications to image analysis and automated cartography", *[J]. Communications of ACM*, vol. 24, no. 6, pp. 381-395, 1981.

[5]   Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, June 2008.