■ 420

Image Mosaicing Technique for Wide Angle Panorama

G. Divya, Ch. Chandrasekhar
Department of Information Technology, Aditya Institute of Technology and Management (AITAM),
Tekkali, A.P., India
*Corresponding author, e-mail: divyacse.aitam@gmail.com, ch.chandrasekhar@hotmail.com

Abstract

Images are integral part in our daily lives. With a normal camera it is not possible to get a wide angle panorama with high resolution. Image Mosaicing is one of the novel techniques, for combining two or more images of the same scene taken in different views into one image. In the dark areas, the obtained image is a panoramic image with high resolution without mask. But in the case of lighting areas, the resultant image is generating mask. In order to gets wide angle panorama, in the existing system, extracting feature points, finding the best stitching line, Cluster Analysis (CA) and Dynamic Programming (DP) methods are used. Also used Weighted Average (WA) method for smooth stitching results and also eliminate intensity seam effectively. In the proposed system, to get feature extraction and feature matching SIFT (Scaled Invariant Feature Transform) algorithm used. In this process, outliers can be generated. RANSAC (Random Sample Consensus) is used for detecting the outliers from the resultant image. Masking is significantly reduced by using Algebraic Reconstruction Techniques (ART).

Keywords: image mosaicing, panorama, SIFT, RANSAC, homography, ART

1.   Introduction

Image Mosaicing is a technique which enables us to join together many small images into a one large image, from which more information can be gathered easily. This technique has been used since the development of photography to increase the field of view by pasting together many small images taken from the camera. For generating a comprehensible mosaic, these images have to be first aligned to a certain reference image and then pasted together. Based on the path in which the camera moves while taking the images, aligning these images together may produce distortions such as elongation, contraction, and viewing at an angle, in them. In this process, combines several images with overlapping field of view (FOV) to produce a panoramic image or a high-resolution image [1], as shown in Figure 1. Image mosaic technology is widely used in military and civilian fields, such as satellite remote sensing, UAV surveillance and searching, robot vision, medical probe, electronic image stabilization and virtual reality, etc.
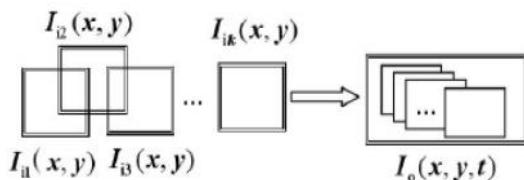

Figure 1. Sequential Images to be Mosaiced

There are 2 types of Image Mosaicing, such as Direct, Feature based [2]. Direct approach compares all pixel intensities of images with each other. It can provide very accurate registration and it has a limited range of convergence. These approaches include Fourier analysis techniques and also coarse to fine optimization of cost or objective functions.

Feature-based approaches have the advantage of being more robust against scene movement and are potentially faster, if implemented the right way. The biggest advantage,

however, is the ability to "recognize panoramas," i.e., to automatically discover the adjacency (overlap) relationships among an unordered set of images, which makes them ideally suited for fully automated stitching of panoramas taken by casual users. It is suitable for fully automatic Mosaicing. Feature detection and matching are used for image stitching from the input images which are having the features in both the images which are having corresponding points to find, and have to match those corresponding points to get a single panoramic image. The groups of features consider point features. Which includes methods using line intersections, high variance points, maximally distinct points with respect to a specified measure of similarity, and corners? With regard to feature detection, in most instances the core algorithms follow the definition of a point as a line intersection or as the centroid of a closed-boundary region. It has been found that corners form their own class of feature as the property of being a corner is hard to define mathematically.

There are two main approaches to finding feature points and their correspondences. The first is to find features in one image that can be accurately tracked using a local search technique such as correlation or least squares. The second is to independently detect features in all the images under consideration and then match features based on their local appearance. The former approach is to combine those images to get a single panoramic image. To get a high resolution panorama, the Image Mosaicing technique uses the following two algorithms. For the feature extraction and feature matching SIFT (Scaled Invariant Feature Transform) algorithm is used. For detecting the outliers from the resultant image RANSAC (Random Sample Consensus) algorithm is used.

2.　Literature Survey

Rafel C. Gonzalez et al [3], in their work image acquisition procedure and image restoration for the feature matching points to get a panoramic image. Li Jin et al [4], proposed in their work how feature matching is used to extract feature points effectively, by using SIFT algorithm. This method also proposes a reliable parameter estimation method, and the result is reliable to stitching a large image. Patil et al [5], proposed in their work the Scale Invariant Feature Transform (SIFT) algorithm can be applied to perform the detection and matching control points, for the image mosaic. SIFT algorithm providing more reliable feature matching for the purpose of object recognition within a single view. RANSAC is a re-sampling technique that generates candidate solutions by using the minimum number observations (data points) required to estimate the underlying model parameters. unlike conventional sampling techniques that use as much of the data as possible 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. Mallick [6] proposed feature based technique in their work for the Image Mosaicing. Firstly select the corresponding corners in the two images. RANSAC is used to estimate the homography relating the two images. The estimated homography is refined using Newton's non-linear method. A dynamic programming based blending algorithm was used to seamlessly blend the two images. To cut the overlapping regions of the images along the curve of minimum brightness in the difference image was a good idea, but its implementation is not as straight forward as in texture synthesis application. MalavikaBhaskaranand and Sandeep Bhat [7] proposed in their work, registration and mosaicking of images. Features in images are detected using the Scale- Invariant Feature Transform (SIFT). A nearest neighbor algorithm with Euclidean distance measure is used for establishing correspondences between images. The normalized Direct Linear Transformation (DLT) together with the Random Sample Consensus (RANSAC) algorithm is used to estimate the homography between the images. The images are then warped to a common coordinate system using the estimated homography. Alpha blending based on distance of the pixel from the image border is used to stitch the images into a smooth mosaic. Guido Bartoli [8] proposed the process of image registration as an automatic or manual procedure which tries corresponding points between two images and spatially align them to minimize a desired error, i.e. a consistent distance measure between two images. Features should be distinctive objects which are possibly uniformly spread over the images and easily detectable by using SIFT algorithm. The detected features in the reference and sensed images can be matched by means of the image intensity values in their close neighborhoods, the feature spatial distribution, or the feature symbolic description. For re-sampling the mapping functions constructed during the previous step are used to transform the sensed image and thus

to register the images. The transformation can be realized in a forward or backward manner. David G. Lowe [9] proposed in their paperfor image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors.

3.   Methodology
For Image Mosaicing the following methods are used, such as SIFT (Scaled Invariant Feature Transform) algorithm and RANSAC (Random Sample Consensus) algorithm.

3.1. SIFT Algorithm
a) Scale-space Extreme Detection
SIFT algorithm uses Difference of Gaussian (DoG) which is an approximation of Laplacian of Gaussian (LoG). Difference of Gaussian is obtained as the difference of Gaussian blurring of an image with two different, let it be υ and $k\sigma$. The Laplacian of Gaussian is:

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

This process is done for different octaves of the image in Gaussian Pyramid as shown in Figure 2.
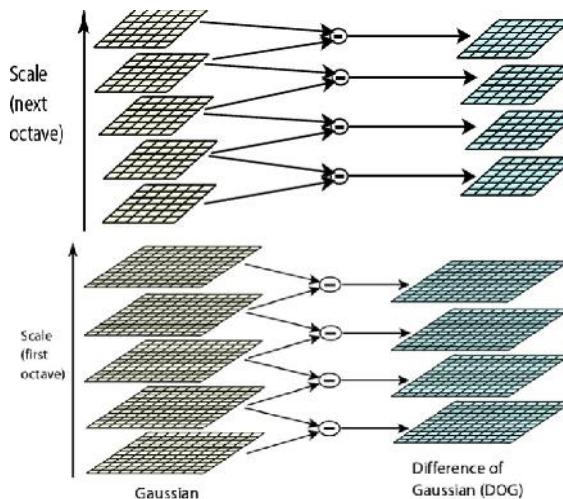

Figure 2. Gaussian Pyramid

The Gaussian blurring can be calculated by using Equation (2). Difference of Gaussian can be calculated by using Equation (3).

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

$$D(x,y,) = \left(G(x,y,k) - G(x,y,) \right) * I(x,y) \tag{3}$$

Once this DoG is found, images are searched for local extrema over scale and space. For eg, one pixel in an image is compared with its 8 neighbours 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 as shown in Figure 3.
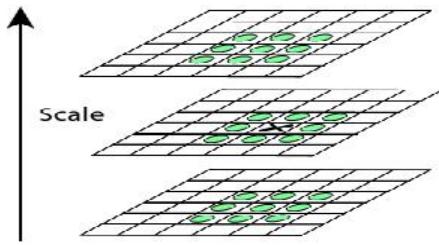
Figure 3. Scaling of Image pixels

Regarding different parameters of empirical data which can be summarized as, number of octaves = 4, number of scale levels = 5, initial, $\sigma = 1.6\ and\ k = \sqrt{2}$ etc as optimal values.

b) Key-point Localization

Once potential key points locations are found, they have to be refined to get more accurate results. Taylor series expansion as sown in the Equation (4), (5) and (6) of scale space to get more accurate location of extrema, and if the intensity at this extrema is less than a threshold value 0.03, it is rejected. So it eliminates any low-contrast key points and edge key points and remains strong interest points.

$$D(X) = D + \frac{\partial D^T}{\partial X}X + \frac{1}{2}X^T\frac{\partial^2 D}{\partial X^2}X \tag{4}$$

$$X = -\frac{\partial^2 D^{-1}}{\partial X^2}\frac{\partial D}{\partial X} \tag{5}$$

$$D(X^\wedge) = D + \frac{\partial D^T}{\partial X}X \tag{6}$$

c) Orientation Assignment

Now an orientation is assigned to each key point to achieve invariance to image rotation. A neighborhood is taken around the key point location depending on the scale, and the gradient magnitude can be calculated using Equation (7) and direction is calculated in that region by using Equation (8). An orientation histogram with 36 bins covering 360 degrees is created.

$$m(x,y) = \sqrt{(L(x+1,y)-L(x,y-1))^2 + (L(x,y+1)-L(x-1,y))^2} \tag{7}$$

$$\theta(x,y) = \tan^{-1}(L(x,y+1)-L(x,y-1)/(L(x+1,y)-L(x-1,y)) \tag{8}$$

d) Key-point Descriptor

Now key point descriptor is created. A 16x16 neighborhood around the key point is taken. It is divised into 16 sub-blocks of 4x4 sizes. 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.

e) Key-point Matching

Key points between two images are matched by identifying their nearest neighbors. 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.

3.2. RANSAC Algorithm [10]

Steps:

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

b) Solve for the parameters of the model.

c) Determine how many points from the set of all points fit with a predefined tolerance

d) If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold, re-estimate the model parameters using all the identified inliers and terminate.

e) Otherwise, repeat steps a) through d) maximum of N times.

The number of iterations, N, is chosen high enough to ensure that the probability p (usually set to 0.99) that at least one of the sets of random samples does not include an outlier. Let u represent the probability that any selected data point is an inlier and v = 1 – u the probability of observing an outlier. N iterations of the minimum number of points denoted m are required, where:

$$1 - p = (1 - u^m)^N \tag{9}$$

And thus with some manipulation,

$$N = \frac{\log(1-p)}{\log(1-(1-p)^m)} \tag{10}$$

3.3. Abstract Reconstruction Techniques (ART) [11]

In ART the image solution $c(x,y)$ is approximated by a weighted sum of basic functions, each of which is a shifted basic basis function $b(x,y)$ is:

$$c(x,y) = \sum_{i,j} c_{ij} \, b(x - x_i, \, y - y_j) \tag{11}$$

The most common basic function is a pixel that has a square support and is valued one inside the support and zero otherwise. The weights c_{ij} of the summation are the coefficients of the resulting approximated image vector.

In real applications, measure indirectly a finite number of line integrals. Let w_l be a measured integrals along a line l, then in the absence of noise for an image in the form of Equation (11) that:

$$w_l = \sum_{i,j} c_{ij} r_{ij}(l) \tag{12}$$

Where $r_{ij}(l)$ is the integral of $b(x-x_i \, , y-y_j)$ along l, which can be calculated from the geometry of data collection. When noise is present, as it is the case in real applications, the equality in Equation (12) becomes an approximation. In ART, a system of equalities, each of which in the form of Equation (12) is formed by considering all the lines along which data have been collected. An ART-type algorithm is essentially based on the following relaxation method for solving a consistent system of linear equalities. Let the system be:

$$R_c = w \tag{13}$$

With S unknowns and K equations.

For $1 \leq k \leq K$, let R_k be the transpose of the kth row of R and w_k be the kth component of w; then:

$c^{(0)}$ is the S-dimensional column vector of zeros (14)

$$c^{\sim(m+1)} = c^m + \lambda \, v^m R_{k_m} \tag{15}$$

With,

$$v^m = \frac{(w_{k_m} - R_{k_m}^t c_m)}{R_{k_m}^t R_{k_m}} \tag{16}$$

Where $k_m = m(mod\ k) + 1$ and $\lambda\ [0,2]$ is a relaxation parameter that controls how well the current equation is satisfied. The sequence $c^{(m)}$ converges to the unique minimum Euclidean norm solution of the system in Equation (13). In the inconsistent case, there is no convergence proof. However, ART-type algorithms are widespread, due to its relatively superior performance for various reconstruction tasks [12]. In practice   is kept low (about 0.1-0.01) to reduce noise fitting and/or to improve convergence speed. Another important parameter is the order in which the data is accessed. It has been observed that faster convergence can be achieved if a linear order (i.e., increasing or decreasing) is given up in favor of a non-sequential order (e.g., directions that are as orthogonal as possible to the previous ones).

Another reason for its popularity is the possibility of incorporating prior knowledge in the reconstruction process. For example, if the image is known to be non-negative, then after each iteration one can set the negative values to zero prior to the next iterative step. Such adjustment has been shown to improve the speed of convergence to a desirable reconstruction. The application of a transformation of the image in between two iterative steps. To be mathematically precise, in an ART-type algorithm with trick, the Equation (15) is replaced by:

$$\bar{c}^{(m+1)} = c^m + \lambda\ v^m R_{k_m} \tag{17}$$

$$c^{(m+1)} = T\ (\bar{c}^{(m+1)}) \tag{18}$$

Where T is the transformation that defines the trick.

4. Results
Case Study 1:
For the dark areas the same image taken from two different angles are given as input, as shown in Figure 4.
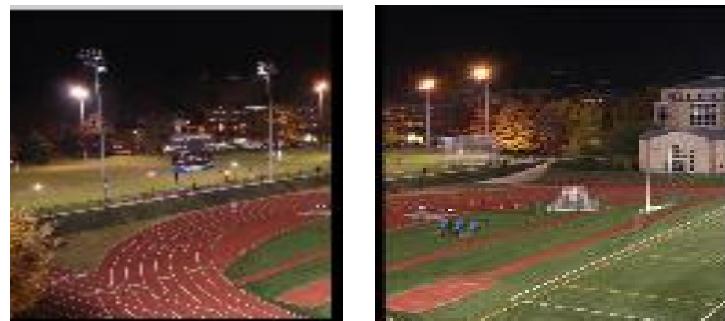
Figure 4. Two input images

The input images after homography warping, as shown in Figure 5.

Figure 5. Homography warping

For each descriptor in the first image, select its match to second image with lines joining the accepted matches, i.e., the feature matching as shown in Figure 6. For input image 1, total 1738 key-points found and for input image 2, 1428 key-points found. Out of these key points, the number of matches in both images is: 133.
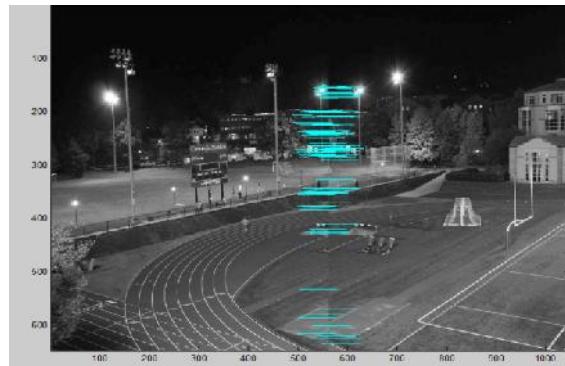

Figure 6. Feature matching

These matches are submitted to RANSAC which calculates a transformation that aligns the points in image1 and image2 and also return the inliers and best matched points, as shown in Figure 7.

Figure 7. Best matched points

Finally using the best matched points the stitch module stitches the both images gives an output which is panoramic view of both images as, shown in Figure 8.
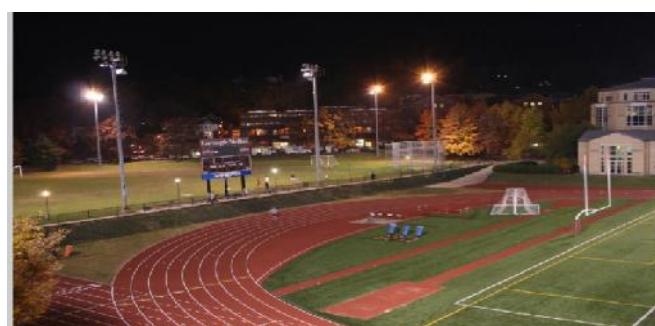
Figure 8. Final stitched image

From Figure 8, find that the output image is having more resolution and more clarity than the given input images.

Case Study 2:

For the lighting areas the same image taken from two different angles are given as input, as shown in Figure 9.

Figure 9. Two input images

The input images after homography warping, as shown in Figure 10.

Figure 10. Homography warping

For each descriptor in the first image, select its match to second image with lines joining the accepted matches. The feature matching, as shown in Figure 11. For input image 1, total 16820 key-points found and for input image 2, 4210 key-points found. Out of these key-points, the number of matches in both images is: 723 and unique matches are: 629.
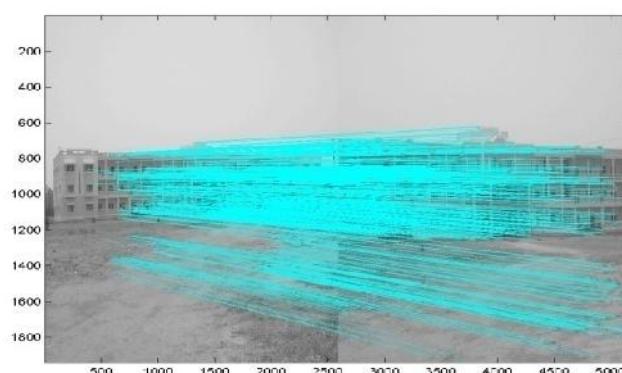

Figure 11. Feature matching.

These matches are submitted to RANSAC which calculates a transformation that aligns the points in image1 and image2 and also return the inliers and best matched points, as shown in Figure 12.

Figure 12. Best matched points

Finally using the best matched points the stitch module stitches the both images gives an output which is panoramic view of both images, as shown in Figure 13.

Figure 13. Final stitched image

From Figure 13, find that the output image is having more resolution and more clarity than the given input images.

5. Conclusion

The results obtained were reasonably good for the dark areas. But for the lighting areas, the mask is generating for the panoramic image. This may be due to non-ideal transformations and threshold values set in the homography algorithm.

In future work, the mask will be reduced by using Abstract Reconstruction Techniques for the lighting areas and also like to perform a more detailed, quantitative analysis of feature matching performance on large databases of panoramic images.

References

[1] Yan Gong, Hong Xie, Lei Yu. Research and Analysis of Key Technologies in Image Mosaic. *International Journal of Signal Processing, Image Processing and Pattern Recognition.* 2013; 6(5): 237-243.

[2]　Harshal JPRET. Image Mosaicing Approach And Evaluation Methodology. *International Journal of Advanced Technology & Engineering Research (IJATER)*. 2013; 1: 576-585.
[3]　Rafael C Gonzalez, Richard E Woods. Digital Image Processing.
[4]　Li Jin, Wang Yanwei, Liang Hng. *Image Mosaic Based on Simplified SIFT*. International Conference on Mechanical Engineering and Automation. 2012; 6: 90-95.
[5]　Tejasha Patil, Shweta Mishra, Poorva Chaudhari, Shalaka Khandale. Image stitching using mat lab. *International Journal of Engineering Trends and Technology*. 2013; 4(3): 302-306.
[6]　Satya Prakash Mallick. Feature Based Image Mosaicing. San Diego: University of California.
[7]　Malavika Bhaskaranand, Sandeep Bhat. Image Registration and Mosaicking. University of California.
[8]　Guido Bartoli. Image Registration Techniques. Universit adegli Studi di Siena. 2007.
[9]　David G Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision.* 2004.
[10] Konstantinos G Derpanis. Overview of the RANSAC algorithm. 2010; 1(2).
[11] Hstau Y Liao. A Gradually Unmasking Method for Limited Data Tomography. USA: Institute for Mathematics and its Applications.University of Minnesota Minneapolis.
[12] Sjors HW Scheres, Roberto Arabini, Salvatore Lanzavecchia, Rancesca Cantele, Twan Rutten, Tephen D Fuller, Jose´ M Carazo, Roger M Burnett, Carmen San Martı´n. Classification of single-projection reconstructions for cryo-electron icroscopy data of icosahedral viruses. *Journal of Structural Biology.* 2009; 151: 79-91.