International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

Image Stitching of Dissimilar Images

P. Sannidhi[1], Sathwik R. Gutti[2], M. R. Shamanth[3], R. Sai Charan[4], S. K. Parikshith Nayak[5]
[1,2,3,4]*Student, Department of Computer Science Engineering, Alva's Inst. of Engg. and Tech., Moodbidri, India*
[5]*Assistant Professor, Dept. of Computer Science Engineering, Alva's Inst. of Engg. and Tech., Moodbidri, India*

***Abstract*: Image stitching is done by combining multiple input images such that the ideal result is a single image that contains contexts from all the inputs as well as seamless transitions between contexts. It is most commonly used in the field of panoramic photography. Large dissimilarities in the input image set are very problematic in image stitching applications. These cause several objections in the resulting output, such as ghosting and distortion. This paper presents a fully automated image stitching process that is aimed at reducing objections and increasing stitching quality.**

The proposed implementation adapted existing methods with an objective of having am implementation that is more robust against the dissimilarities caused by perspective, illumination, and occlusion.

***Keywords*: image stitching, panorama.**

1. Introduction

Photographs are never exactly the same. Even in the exact same environment, in the right spot, with the right angle, and in the same lighting conditions, the image would contain, at least, minute differences throughout the scene. In reality, images of the same object would have varying perspectives, lighting conditions, and amount of occlusion. A panorama is an image that is created from several smaller images. This is done detecting and then making use of the similarities between images to join, or stitch, them in such a way that these images combine seamlessly. The study developed and tested an algorithm capable of combining two or more dissimilar images through existing stitching methods. For the purpose of this study, images are classified as dissimilar when they have slightly or extremely altered perspective, have different illumination or lighting conditions, and varying amount of occlusion.

After images have been acquired, preprocessing of images is mandatory before they can be stitched. For example, the images can be projected onto a geometrical surface which can be on either a spherical, cylindrical, or planar surface. Camera-made distortions must also be corrected first before stitching can be done further [2]. The course of the image stitching process can be divided into two steps: image registration and image merging. During image registration, fraction of neighboring images are tested and compared to see if there are similar details between the images that can help in the alignment. After determining which parts of the images are most likely to be aligned, the images are lined up to form a panoramic image. A scenic image is constructed after images are successfully fused together. Hence, the three main processes are as follows: image acquisition, registration and fusion [2].

The study did not cover stitching of images which are not photographs or those images that are the products of computer graphics. It also did not cover the stitching of images which are contextually different. These are images that contain totally different subjects.

Fig. 1. An example of image stitching. (top-left) and (top-right) are input images. (bottom) is the resulting image

2. Theoretical consideration

There are many existing image stitching algorithms that can create panoramas by stitching similar images or images with higher number of similar keypoints. Most image stitching methods use the following succession of steps in the stitching process: detection of keypoints, matching keypoints, aligning images, and blending images.

A. Keypoint detection

David Lowe's Scale Invariant Feature Transform (SIFT) is one approach in detecting the keypoints in an image. SIFT transforms image data into scale invariant and rotation invariant coordinates, and partially invariant to change in illumination and 3D camera viewpoint. In addition, SIFT provides robust detection even in the presence of affine distortion, resulting into distinctive key features [3]. In detecting the keypoints in an image, the stages that SIFT operator uses are as follows: local extrema detection, keypoint localization and orientation assignment [4]. One technique that can not only detect the keypoints but also match the detected keypoints is called Binary Robust Invariant Scalable Keypoints (BRISK). The method

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

ensures high quality keypoint descriptor and low computational requirements. In the detection process, using a saliency criterion, points of interest are detected in both the image and scale dimensions. The features are identified in octave layers of the image pyramid and in layers in-between in order to thoroughly increase the computing efficiency of the algorithm. At the neighborhood of each keypoint, appropriately scaled concentric circles are applied. The sampling pattern consisting of points lying on the concentric circles is used to obtain pairwise brightness comparison result [5].

B. Keypoint description

FREAK and BRISK are novel binary descriptors [6] that exceed the principles of modern industry. There are many cases where newly developed algorithm of feature description rarely surpasses its predecessors in case of effectiveness. The SIFT algorithm has been in the field for 10 decades and was also unmatched before the SURF algorithm was developed. Whenever a new descriptor is developed, many tests are immediately carried out for it to be instantly used in android camera operations.

The SURF descriptor uses an algorithm that is very similar to the SIFT descriptor [6]. The descriptor creates a grid around every feature point detected. Inside each grids, there are even smaller grids called sub grids. Inside each sub grids, the gradient is calculated into a histogram and the counts of the histogram is increased by the degree of the gradient; each is weighted by Gaussian.

BRISK descriptor creates several pattern points around the detected keypoint. BRISK is a 512-bit binary feature descriptor. It calculates the weighted Gaussian average of the selected pattern points around the keypoint. It associates the values of 1 or 0 depending on which pair is greater on each pair of Gaussian windows [5].

FREAK is also a binary descriptor just like BRISK. The improvement of FREAK is the sampling arrangement and method for pairing selection that the algorithm, BRISK, uses. On every keypoint detected, FREAK creates 43 pattern points around the keypoint. It then evaluates the weighted Gaussian of each of the 43 pattern points created nearby the keypoint [6].

C. Keypoint matching

After detecting the keypoints in each image, it is necessary to determine which keypoints match in order to create a basis for image alignment later on. There are several methods developed for matching of keypoints. Hausdorff distance and wavelet transform-based matching are two of these methods. An algorithm proposed by Iqbal, et al. [7] uses the Hausdorff measure in creating a fast search strategy to diminish the number of positions needed to be calculated. In another paper, Mong-Shu Lee [8] proposed the use of structure-based image similarity measurement called DTWT-SSIM in order to combine the shift-invariance advantage of dual-tree wavelet transform with the structure-preserving property of structural similarity metrics. The technique is proved to be effective in comparing edge maps in the presence of small noise in image [8].

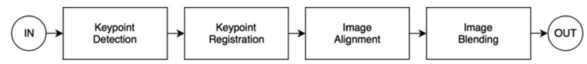

Fig. 2. An example of image stitching. (top-left) and (top-right) are input images. (bottom) is the resulting image

For faster matching of keypoints, an algorithm proposed by Schweitzer, et al. is helpful due to its application in robot vision. The method uses a dense calculation of haar wavelet response in building image descriptors for keypoint matching. A pre matching step based on bitmask operation is utilized in order to avoid heavyweight image descriptor comparisons [9].

D. Image alignment

In the process of matching keypoints of images, only distinctive keypoints are left that would help in aligning two or more images.

RANdom SAmple Consensus (RANSAC), a pre-alignment method, as proposed by Fischler and Bolles, is designed to deal with larger number of outliers in the input data. Unlike the conventional methods that use as much of data as possible in aligning images, RANSAC initially uses minimum number of data points in estimating the underlying model parameters and then proceeds to enlarge the data set using the consistent data points [10].

According to Liu and Zhang, the older methods of image stitching used the homography between two images to align them. Homography will not handle parallax so those methods involve the input images to have similar perspective or roughly planar. If those requirements are not met, it will not lead to alignment and resulting into a ghosting or broken image. For their own method, they have to align the images so that there is a local region in the overlapping area where it will be stitched [11]. In their paper, Liu and Zhang introduced a method to successfully align images regardless of significant variance in parallax [11]. The alignment technique used identifies SIFT feature points first and then matches them between the two images. After detection and matching, the paper employs their algorithm to search for a good alignment between the two images. The algorithm will randomly appoint a feature seed group; the seed group will expand to neighboring feature points to approximate a good alignment between the images. The alignment will go through an evaluation regarding its stitching quality. If the stitching quality made by the alignment is satisfactory, the algorithm will stop; otherwise it will repeat the steps distortions [11].

E. Image Warping

In image analysing, the warping of images is a significant stage. It is a function that deforms images by mapping between images.

The camera or viewing outlook gives optical distortion that needs to eliminate to have an image with reference grid such as map, or for the alignment of two or more images. Warping is a

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

couple of two dimensional functions which place a position in one image wherein the column number and row number are denoted to place into position in another image. In fin ing the right warp, there are a lot of ways to do it; however, the most used way is the compromise between insisting the distortion is smooth and achieving a good match. Smoothness can be certified by assuming a parametric form for the warp. There are a lot of existing methods for warping approach the use of thin-plate splines to produce smooth deformations [28].

Warping is necessary and sufficient to compensate for differences in image alignment that results from imperfect centration of the different lens systems and from minimal differences in magnification between objective lenses with same nominal magnification. In image warping, parameter estimation is hard since there is a large amount of computational load and the presence of local optima [28].

The homography of the images cannot account for parallax and results into having ghosting artifacts. In image compositions like seam cutting and blending are used to lessen the ghosting artifacts. Their method involves aligning of images first, then employ a seam- cutting algorithm to look for a seam to place aligned images together, and finally use a multi-band blending algorithm to create the final image stitching result. In order to create a seamless panorama of two dissimilar images, it is important to use image blending techniques.

The proposed algorithm by Whitaker for image blending relies on minimizing the difference metric which compares the level sets of the images. It results from a pair of differential equations that model multi-dimensional level set propagations.

The aforementioned method produces a more naturalistic appearance than the method of interpolation since it controls the shapes of image contours instead of intensity values [12].

Another algorithm proposed by Ho, et. al [13] focuses on colorization of combining images. The algorithm utilizes blending-weight diffusion rather than direct propagation of chrominance values in order to determine the priority order of color propagation in image colorization [13].

3. Proposed algorithm

The proposed process uses the conventional image stitching pipeline, as illustrated in Fig. 2. The techniques used are mostly modified existing processes.

A. Keypoint detection

Feature points are detected from the image. Scale Invariant Feature Transform (SIFT) detector is used for keypoint detection. The first stage is scale-space extrema detection in which after searching over all scales and image location, difference-of-Gaussian function is used in order to determine potential feature points that are invariant to scale and orientation. Next, features are localized through rejecting those kepoints lower than a certain threshold and those that slide along the edges. Orientation is then assigned to each feature point [1].

B. Keypoint description

After detecting the keypoints, a descriptor is used in order to describe the feature points. Feature descriptors contain interesting information about the keypoints. The information is placed into a feature vector which is used in differentiating one feature from another. In later processes, the descriptor information is used for accurate matching of keypoints.

In the process, a 16x16 window is created around the keypoint. The window is broken down into sixteen windows, each having 4x4 bins. The magnitude and orientation of each bin in the 4x4 window is calculated through Eq1 and Eq2, respectively.

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (1)$$

$$\theta(x, y) = \tan^{-1}\frac{L(x+1, y) - L(x-1, y)}{L(x, y+1) - L(x, y-1)} \quad (2)$$

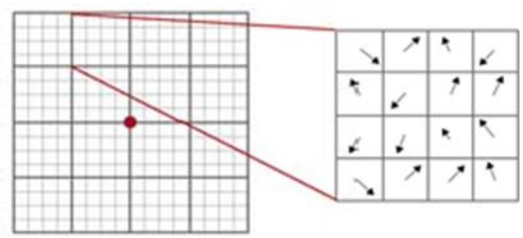

Fig. 2. Magnit3de and orientation of one of the 4x4 bins around the keypoint

A histogram is created based on the magnitude and orientation of all the bins. Instead of using an 8 bin histogram like the SIFT descriptor, we used only 4 bin histogram. Fig. 4. depicts the 4 bin histogram.

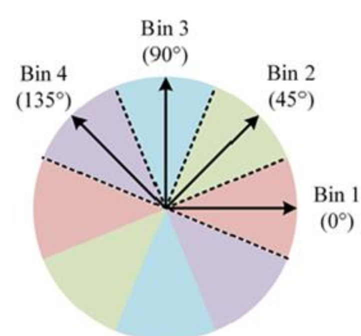

Fig. 4. The histogram divided into 4 bins

Fig. 5. Night image of the scene

The reason behind using only 4 bins is that at times, colors

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-5, May-2019
www.ijresm.com | ISSN (Online): 2581-5792

of objects in an image are affected by the color of the surroundings, as depicted in Fig. 5 and 6. Because of different times of day, the colors vary in the images.

Fig. 6. Evening image of the scene

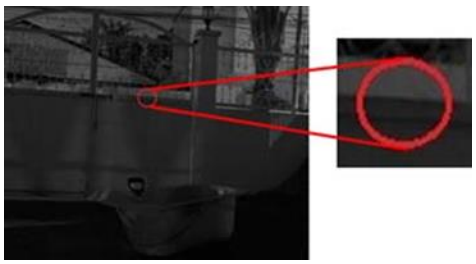

Fig. 7. Counterpart grayscale image of Fig. 5.

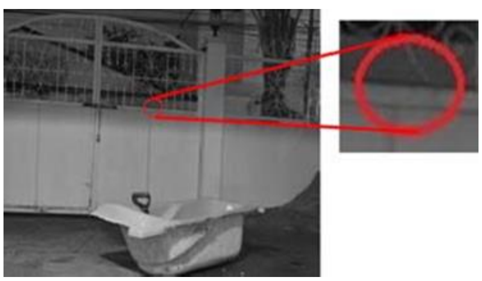

Fig. 8. Counterpart grayscale image of Fig. 6.

Due to different lighting conditions, the direction of gradient in Fig. 5 is opposite of the transition of gradient in Fig. 6, as depicted in the encircled portions. Aside from the shown area, there are several other areas in the images which contain opposite direction of gradients.

Aside from the taking into consideration the calculated magnitude and orientation of the pixels, the distance between the keypoint and the pixel in the window is also needed. Gaussian weighting function is used for this purpose.

The image is then normalized. The magnitude of the pixels is truncated to 0.2, then the image is normalized again. A 64-bit feature vector is then created in order to describe a keypoint.

Fig. 9. Input image

(a) (b)

Fig. 10. (a) SIFT descriptor stitched result (b) Proposed descriptor stitched result

Through the proposed descriptor, less number of matches are detected. In Fig. 9, the number of matches when the SIFT descriptor is used are 3022, while the number of matches when the proposed descriptor is used are 1527. Hence, the weakness of the proposed descriptor is that there are less number of matches which can be used. Yet, the resulting stitched image of the SIFT descriptor is the same with the resulting stitched image of the proposed descriptor, as shown in Fig. 10 (a) and Fig. 10 (b), respectively. This is because the later techniques used for image alignment overcome the limitation of the descriptor.

For lighting-varying images, the proposed descriptor is, at times, better than the SIFT descriptor.

Fig. 11. Input Image

(a) (b)

Fig. 12. (a) SIFT descriptor stitched result (b) Proposed descriptor stitched result

Fig. 11. is the input images. Fig. 12 (a) is the resulting stitched image when SIFT descriptor is used, while Fig. 12 (b) is the resulting stitched image when proposed descriptor is used. Based on the results, it is evident that the proposed descriptor performed better.

C. *Alignment*

Image alignment uses a modified Local homography method. Local homography was introduced in [Parallax tolerant...]. The method takes a number of candidate homographies, calculated from a local matching keypoint group, and scores these based on how well they support seam blending. This method, complemented with seam blending, allows for increased robustness for cases of large perspective difference.

D. Warping

Warping increases alignment between matching keypoints by applying a finer-grained transform than the homography. The method used here calculates warping vectors by smoothing and interpolating the sparsely distributed matching keypoint offset vectors. The result is increased alignment in the overlapping region.

E. Blending

Seam blending is a blending technique that uses graph cuts to segment the overlapping region and providing a smoother cut from one to the next. Other algorithms cut seams by labeling the segments to be either one image or the other [cite some]. The proposed techniques modify this such there exists a blending region that increases the ability to seamlessly blend non-like colored areas, while still avoiding strong ghosting effects.

Color adjust blending was a technique developed that re-colors areas that the seam blend was not able to blend well enough, in terms of color. It does this by first detecting areas that, after the application of seam blending, show a large color gradient difference in an area that, in the input images, had none. It then segments these areas and applies a weighted mask to re-color the image such that those areas are of perceptually equal color.

4. Experiment and discussion

A survey of 30 participants was conducted to score the output quality of the proposed process compared with APAP, Harris-RANSAC, and our own implementation of SIFT. The proposed algorithm scores slightly higher than the rest as shown in Fig. 13.

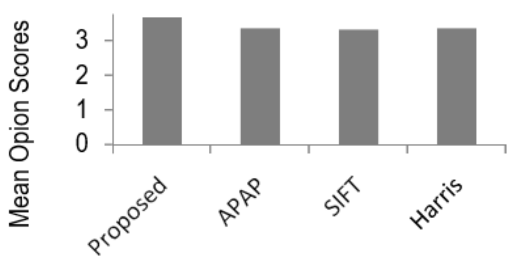

Fig. 13. Comparison of mean opinion scores of APAP, SIFT, Harris, and the proposed implementation

5. Conclusion

The paper presented a method for stitching dissimilar images. We observed that images with significant dissimilarity often cannot be stitched well without producing unnecessary artifacts. We then developed a method that can stitch dissimilar images with minimal errors. We used an existing keypoint detector, namely SIFT. We then used a descriptor in order to find information about the keypoints. Instead of using 8-bin histogram, we just used 4 bins in order to improve the performance of the algorithm for lighting varying images. After creating description of keypoints, other processes, such as matching, local homography, warping, seam blending and color adjusting, followed.

Our experiments on images with dissimilarities, such as perspective, occlusion and lighting, depicted the effectiveness of our algorithm. Through the results of the survey, the success of our algorithm is further proved.

Future studies may look at how to lessen the time of the stitching process given larger dimensions of images. Hence, aside from focusing on the quality of the stitched images, the future studies may consider also the time of the processing of an image in order for the algorithm to be convenient in whatever application to be used.

Image stitching is an extensive area of research which has many possible applications. For future researchers, we recommend developing stitching algorithms for specific applications, such as for medical technology. Various fields are in need of such algorithms and technologies which can help the people and the society.

References

[1] D. Lowe, "Distinctive Image Features from Scale- Invariant Keypoints," Department of Computer Science, University of British Columbia, vol. 1, p. 28, 2004.
[2] Y. Chia and Chen, "Image Stitching - Comparison and New Techniques," Computer Science Department of the University of Auckland CITR at Tamaki Campus, 1998
[3] D. Lowe, "Distinctive Image Features from Scale- Invariant Keypoints," Department of Computer Science, University of British Columbia, vol. 1, p. 28, 2004.
[4] B. Hau and C. Hsin, "Affine Invariant Local Features based on Novel," International Conference on Communications and Electronics, p. 6, 2012.
[5] S. Leutenegger, M. Chli and R. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," Computer Vision, IEEE, p. 8, 2011.
[6] Schaeffer, "A Comparison of Keypoint Descriptors in the Context of Pedestrian Detection: FREAK vs. SURF vs BRISK," Stanford University, California.
[7] Z. Zhijia, H. Shabai and S. Zelin, "A Fast Strategy for Image Matching using Hausdorff Distance," in Intemational Conference on Robotics,Intelligent Systems and Signal Processing, Shanghai, 2003.
[8] M. Lee, M. Chen and D. Jwo, "Images Matching based on Edge Maps and Wavelet," International Congress on Image and Signal Processing, vol. 3, p. 5, 2010.
[9] M. Schweitzer and H. Wuensche, "Efficient Keypoint Matching for Robot Vision using GPUs," IEEE 12th International Conference on Computer Vision Workshops, p. 8, 2009.
[10] K. G. Derpanis, "Overview of the RANSAC Algorithm," Department of Computer Science and Engineering, York University, no. 1.2, p. 2, 2010.
[11] Liu, F. and Zhang, F., "Parallax-tolerant Image Stitching," Computer Vision and Pattern Recognition, IEEE, p.3662-3669, 2014.
[12] R. Whitaker, "A Level-Set Approach to Image-Blending," IEEE Transactions on Image Processing, vol. 9, no. 11, p. 13, 2000.
[13] K. Ho and Y. Chung, "Blending Weight Diffusion for Image Colourisation," IET Image Processing, vol. 1, no. 1, p. 9, 2012.
[14] A. Vedaldi and B. Fulkerson. "VLFeat: An open and portable library of computer vision algorithms." Proceedings of the international conference on Multimedia. ACM, 2010. pp. 1469-1472.