Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 19 (2013) 420 – 427
Procedia Computer Science

The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013)

IMAGE STITCHING WITH COMBINED MOMENT INVARIANTS AND SIFT FEATURES

A.Annis Fathima[a]*, R.Karthik[b], V.Vaidehi[c]*

[a]Research Scholar, [b]Student, Department of IT, Anna University,Chennai, India.
[c]Professor, Department of IT, Anna University,Chennai, India.

Abstract

Image stitching is used to combine multiple photographic images from camera network with overlapping field of view to produce panoramic view. With image stitching, the view is enlarged and the amount of information increases with the no. of images that are stitched. In the existing methods, the whole images from the adjacent views are considered thus leads to increase in both time and computational complexity. In this paper, an approach for image stitching using invariant moments combined with SIFT features is presented to reduce the time and computational complexity. It is observed that only a small portion of the adjacent view images are overlapped. Hence, the proposed method aims in detecting overlapping portion for extracting matching points. The overlapping regions are determined using gradient based dominant edge extraction and invariant moments. In the deduced region, the SIFT (Shift Invariant Feature Transform) features are extracted to determine the matching features. The registration is carried on with RANSAC (Random Sample Consensus) algorithm and final output mosaic is obtained by warping the images. The proposed approach results in reduced time and computational when compared to existing methods.

Keywords— Panorama; Image stitching; Invariant Moments; SIFT Features.

1. Introduction

The panorama is a wide angle view of physical space and gives a seamless representation of the complete scene. In computer vision, a panoramic image is created by the process of Image stitching. Image stitching combines multiple images from the camera network with overlapping fields and the resultant image will have much higher information than independent images. It is performed to overcome the loss of complete information about the scenario and to avoid any redundant information

* A. Annis Fathima,V.Vaidehi
E-mail address: annisfathima@yahoo.com, vaidehi@annauniv.edu

1877-0509 © 2013 The Authors. Published by Elsevier B.V.
Selection and peer-review under responsibility of Elhadi M. Shakshuki
doi:10.1016/j.procs.2013.06.057

from the neighbouring cameras in the network. This technique is widely applied for the purpose of surveillance, video summarization, remote sensing etc. When a region is photographed using camera network, the adjacent cameras capture its field of view with the overlapped region between the cameras. The common or overlapped data may exist in both the images and also the geometric alignment of an image with respect to the other may not be the same.

Image stitching is achieved using image features obtained between corresponding images taken from multiple viewpoints. Block matching and feature-point matching are the two basic ways to identify the matching region from the input images. Block matching algorithms calculate the correlation between regular-sized blocks generated in sequential images. Such methods include either NCC (Normalized Cross-Correlation) or phase correlation using an FFT (Fast Fourier Transform) [1]. These methods involve a series of complex calculation and it is very sensitive to the slight distinction between images [2]. Feature based methods extract distinctive features from each image and matches these features to establish a global correspondence. Features are generally distinct, easy to extract, and not affected by the camera's perspective. [3]. A feature involves various discontinuities in an image such as edges, interest or corner points, ridges etc. Geometrical features including lines, edges and corners are frequently used for detecting targets [4]. The performance of the edge detectors deteriorates rapidly when edges are blurred and noisy. Unlike edges, corners provide considerable information in the description of the shape of an object. A simple feature detector named SUSAN computes the fraction of pixels within a neighborhood which have similar intensity of the center pixel [5]. It avoids computation of derivatives but its efficiency decreases when the image is more susceptible to noise. The FAST detector, introduced by Rosten and Drummond in [6-7] compares pixels only on a circle of fixed radius around a point. A point is classified as a corner only if one can find a large set of pixels on a circle of fixed radius around the point are all significantly brighter or darker than the central point. But the major limitation of FAST detector is that multiple features are detected adjacent to one another.

The Harris corner detector is robust against the changes in rotation, illumination and noise. However this operator fails when disposing images with larger scale changes [8]. Speeded-Up Robust Features (SURF) is an efficient scale and rotation-invariant feature detector [9]. It allows a good compromise between feature complexity and robustness to common deformations. But it is not viewpoint and illumination invariant.

Mikolajczyk made a study on various feature descriptors and it was found that Scale Invariant Feature Transform (SIFT) generally performed the best [10]. SIFT extract distinctive invariant features from images that can be used to perform reliable matching between different view of an object or scene. Using SIFT, the image information is transformed into local feature coordinates that are invariant to shift, rotation, scale, noise, occlusion, partially invariant to illumination changes.

After finding correct corresponding points in the reference image and the target image to be matched, the next process is to calculate the relationship between the two input images. This relationship can be defined as the homography matrix which relates the pixel coordinates between two images. Two widely used approaches for estimating the homography relationship are RANdom SAmple Consensus (RANSAC) and Least Median of Squares (LMedS). LMedS is robust as it can select feature point matches when the original sample has up to 50% of poorly matched points. However, the calculation speed is rather very slow [11]. The RANSAC algorithm is another method to estimate a model in the presence of outliers. It works very well even in situations where more than 50% of the data points are outlier [12].

An approach for image stitching using a random corner method is proposed in [14]. The features are extracted from the images using the Harris corner detector and registration was carried out using RANSAC algorithm. Automatic image registration is proposed by Gonçalves et.al. [15]. In the initial step the foreground objects were extracted through segmentation, characterized and then matching is done with the statistical based rotation and translation parameter estimation. But the drawback of this method is the huge computational complexity associated with the segmentation stage. An image mosaicking method for standing tree was put forward in [16] with SIFT features. This method suffers due to computational complexity, as the features are extracted from the entire image instead of only the matched region. One of the major challenges associated with the existing methods is the computational complexity involved in the feature extraction stage. In this paper, an approach to overcome the computational complexity in image stitching is devised. In the first step, the overlapped regions from the images are identified and the feature extraction is carried on, thus reducing the amount of computation.

The paper is organized as follows: proposed system describing the matching region selection, extraction of SIFT features and Homography computation are given in section 2. The implementation details and results are discussed in section 3 and the paper concludes in section 4.

2. Proposed Work

A feature based image stitching approach is presented in this paper. Initially the overlapping region from the input images is determined using combined gradient and invariant moment method, to reduce the processing area for feature extraction. The SIFT features are extracted for feature point matching. These features are matched to determine the correspondence between the images. Then registration is performed using RANSAC algorithm. Finally the images are warped into a single frame to produce the image mosaic.

2.1. Selection of matching regions

Selecting the matching region is a challenging task as the input images vary by viewpoint, geometric variations such as translation, rotation etc. Block based methods like NCC works fine if the images are not subjected to geometric variations. Simple phase correlation algorithm is employed for determining the matching if the input images vary only by translation. It fails when the images are subjected to rotation. Segmentation and clustering based methods were tried to determine the matching region. But these methods fail to give acceptable results. It also imposes a huge time complexity for processing. One way to reduce the time complexity is to use only the dominant information from images like edges. This could significantly reduce the matching cost. Traditional edge detection methods like sobel, canny, prewitt and roberts operators are applied in determining the edges. These approaches failed to yield better results under varying environmental and illumination conditions. Gradient based edge detection produced better results for natural color images. After determining the dominant edges from the images, images are partitioned into equal sized blocks and compared against each other for finding the similarity. Simple intensity based comparison of images is time consuming and also it gives poor results if the images vary by rotation.

Invariant moment has been widely used as feature for recognition in many areas of image analysis. There are a variety of applications which involve the invariant moments as its property does not change with rotation, scale and translation [13]. Hence invariant moments are computed for each block and it is compared with the target image blocks for finding the similarity.

The moment invariants are moment-based descriptors on shapes, which are invariant under general translational, rotational, scaling transformation. The moments of a binary image f(i,j) are given by Eq.1.

$$m_{pq} = \sum_i \sum_j i^p \, j^q \tag{1}$$

where p and q define the order of the moment. Zero order moment are used to represent the binary object area. Second-order moments represents the distribution of matter around the center. The invariant moments of lower order are enough for most registration tasks on remotely sensed image matching. If $f(x, y)$ is a digital image, then the central moments are defined by the Eq. 2.

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p \left(y-\bar{y}\right)^q f(x,y) \tag{2}$$

Moments η_{ij} where $i+j \geq 2$ can be constructed to be invariant to both translation and changes in scale by dividing the corresponding central moment by the properly scaled $(00)^{th}$ moment, using the following formula.

$$\eta_{ij} = \frac{\mu_{ij}}{\mu 00^{\left(1+\frac{1+j}{2}\right)}} \tag{3}$$

From moments η_{ij}, seven Invariant moments are calculated with Equations 4 to 10.

$$\mathrm{M1} = \eta_{20} + \eta_{02} \tag{4}$$

$$\mathrm{M2} = (\eta_{20} - \eta_{02})^2 + 4\eta^2{}_{11} \tag{5}$$

$$\mathrm{M3} = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{6}$$

$$\mathrm{M4} = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \tag{7}$$

$$\mathrm{M5} = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{8}$$

$$\mathrm{M6} = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \tag{9}$$

$$\mathrm{M7} = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \tag{10}$$

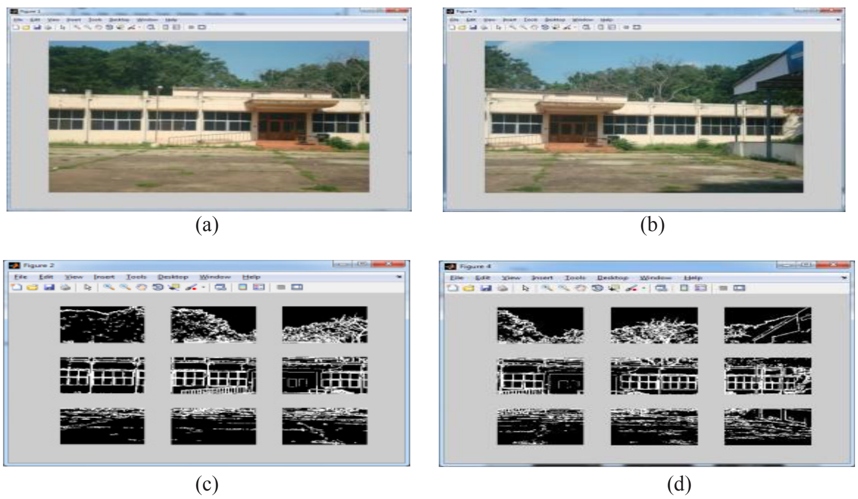
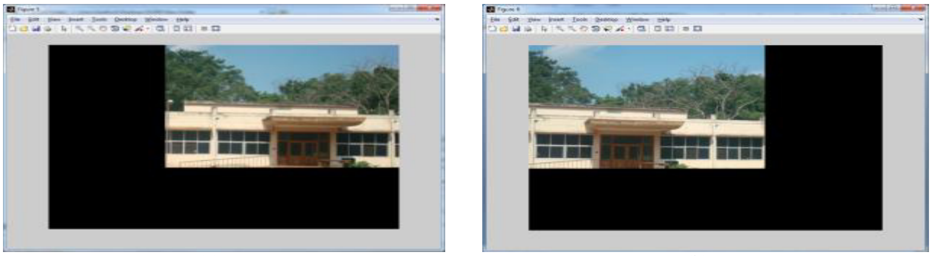


Fig. 1. (a) and (b) Input images; (c) and (d) Dominant edges extracted and divided into blocks



(a) (b)

Fig. 2. Selection of matching region from (a) input image 1 (b) input image 2

The seven invariant moments M1 to M7 mentioned in from Eq. 4 to 10 are typical unchanged properties of images under rotation, scale and translation. After determining the dominant edges of the images, the images are partitioned into equal sized blocks as shown in Fig.1. The absolute difference between the moment values obtained from input images is presented in Table 1. The blocks with the minimum absolute difference is considered and taken up as the matching regions for the following steps of extracting SIFT features. The results for matching region for both the images are shown in Fig. 2 (a) and (b).

2.2. SIFT feature extraction

SIFT algorithm is used for extracting the features from the matching regions. It consists of four steps namely scale-space extrema detection, Keypoint localization, orientation assignment and a keypoint descriptors. The details about these steps are clearly discussed in [9]. SIFT features are extracted from the selected region of the input image and the resulting matching key points are displayed in Fig. 3. The matched feature point of the first image is connected to the corresponding feature point of the second image by means of a line to indicate the correspondence between the images.

Table 1. Absolute Difference of moment values between image blocks

Block	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1	1.676499	2.539358	1.521297	2.289308	2.469128	2.502751	1.768629	2.048087	2.073778
2	0.689153	0.233216	1.609551	0.134288	0.162986	0.196610	0.552005	0.290354	0.295141
3	0.852266	0.070101	1.772665	0.276879	0.142490	0.073752	0.715142	0.422396	0.395864
4	0.738266	0.191656	1.658667	0.113784	0.123730	0.148548	0.601057	0.308396	0.281866
5	0.862629	0.145105	1.783031	0.190572	0.021009	0.058198	0.725175	0.432758	0.406230
6	0.752530	0.264131	1.671916	0.082187	0.100852	0.177553	0.615816	0.321414	0.295115
7	0.522245	0.883194	0.975623	0.671674	0.817001	0.846648	0.569306	0.461691	0.460809
8	0.334504	0.634766	1.218861	0.420323	0.565450	0.598220	0.465097	0.217985	0.209458
9	0.531012	0.427025	1.450576	0.151817	0.328716	0.362390	0.386169	0.183194	0.195640

Fig. 3. Matching feature points from the input images

2.3. Homography Computation

After finding corresponding points in the reference image and the target image to be matched. The next process is to calculate the relationship between them by transformation model and it is defined as a matrix called homography matrix. This relates the pixel coordinates between two images. Random Sample Consensus (RANSAC) method is used to obtain the inline points and transformation matrix. The homography matrix estimated using RANSAC algorithm is applied to the target image in order to get aligned with the reference image. The resultant transformed image is displayed in Fig. 4. The target image that should be aligned with the reference image is subjected to warping process and the resultant stitched image is presented in Fig. 5.

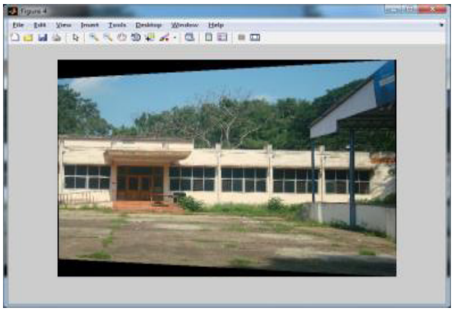


Fig. 4. Transformed target image after homography estimation.

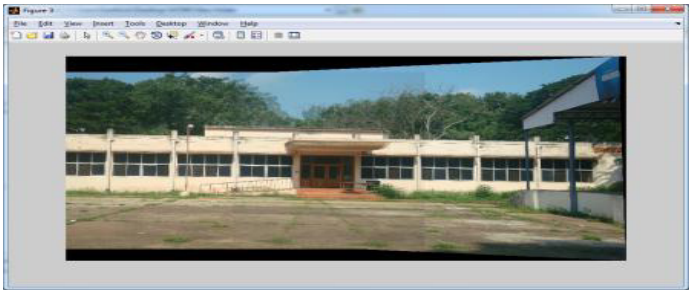


Fig. 5. Final stitched image mosaic

3. Experimental Results

The proposed work frame is implemented using MATLAB7.1 in Intel(R) i3, 2.3GHZ with 4GB RAM. The sample database with 15 set of images are created for the experimental purpose. In the stitching process, the first step is the identification of the matching region by using the gradient based method for determining the dominant edges and invariant moments are derived. The absolute differences between the blocks of the adjacent view images are tabulated and the resultant overlapped portions of the images are given in the Section 2.1. The results for the matching point deduction and the homography estimation are provided in Section 2.2 and 2.3 respectively.

The framework is applied for the dataset and sample image results are presented in Fig. 5 and 6. The performance of the proposed work is analyzed in terms of no. of features detected and the total time taken for stitching the images. The proposed approach is compared against the existing approach [16] utilizing SIFT features alone for feature extraction. The corresponding observations for the best, worst and average cases are presented in Table 2 and 3. Table 2 gives the comparison on no. of features detected as matching points in both the existing and proposed approach. The percentage reduction in the no. of features for matching point computation is also tabulated. These observations clearly indicate that the proposed work frame is capable of stitching the images with less no. of features. Table 3 gives the comparison of time taken for complete process of stitching with both the existing and proposed approaches is tabulated.

(a)

(b)

(c)

Fig. 6 (a) and (b) input images (c) stitched image

Table 2. Comparison on no. of features detected as matching points.

Image set		***No. of features detected***		***Percentage reduction***
		Existing	Proposed	
Set 1	Image 1	12405	1653	86.6
	Image 2	11853	1554	86.9
Set 2	Image 1	10795	2212	79.5
	Image 2	8758	1895	78.4
Set 3	Image 1	4228	665	84.3
	Image 2	9630	1508	84.3

Table 3. Comparison on time taken for image stitching.

Image set	***Total time taken in sec***		***Percentage reduction***
	Existing	Proposed	
Set 1	88.6	47.45	46.4
Set 2	72.09	44.17	38.7
Set 3	57.14	43.65	23.6

4. Conclusion

This paper proposed an approach for image stitching using invariant moments and SIFT features has been proposed in this work. It was observed that nearly 70% of the features do not lie in the overlapping region of the image. These features incur unnecessary computations for feature extraction and matching. To address this problem, overlapping regions are selected from the input images by comparing the similarity of invariant moments at region level. This step has significantly reduced computations for feature extraction and matching. Experiments show that the proposed work frame is highly effective in terms of computational and time complexity reducing 83% of features and 35% in time respectively.

References

[1] Oh-Seol Kwon and Yeong-Ho Ha. Panoramic video using Scale Invariant Feature Transform with embedded color-Invariant values. *IEEE transactions on Consumer Electronics* 2010;**56**:792 - 798.

[2] Jan Yang, Linlin Wei, Zhiwei Zhang and Hongmei Tang. Image Mosaic Based on Phase Correlation and Harris Operator. *Journal of Computational Information Systems* 2012;**8**:2647-2655.

[3] Richard W. Pazzi, Azzedine Boukerche , Jing Feng and Ying Huang. A Novel Image Mosaicking Technique for Enlarging the Field of View of Images Transmitted over Wireless Image Sensor Networks, *Mobile Netw Appl* 2010 **15**:589–606.

[4] X. Dai and S. Khorram. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. *IEEE Transactions on Geoscience and Remote Sensing* 1999, **37:** 2351–2362.

[5] S. M. Smith and J. M. Brady. SUSAN — A new approach to low level image processing. *International Journal of Computer Vision* 1997; **23:**45–78.

[6] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. *Proceedings of the International Conference on Computer Vision* 2005:1508–1511.

[7] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. *Proceedings of the European Conference on Computer Vision* 2006:430–443.

[8] Chenguang Guo, Xianglong Li, Linfeng Zhong and Xiang Luo. A Fast and Accurate Corner detection based on Harris Algorithm. *Intelligent Information Technology Application* 2009:**2**, 49-52.

[9] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 2004;**60**:91-110.

[10] Mikolajczyk, K. A comparison of affine region detectors. *International Journal of Computer Vision* 2005; ***65:***43–72.

[11] Xiangru Liand and Zhanyi Hu. Rejecting Mismatches by Correspondence Function. *International Journal of Computer Vision* 2010;**89**:1-17.

[12] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 1981;**24** :381-395.

[13] J. Flusser and T. Suk. Rotation Moment Invariants for Recognition of Symmetric Objects. *IEEE Transactions on Image Processing* 2006;**15**:3784-3790.

[14] Deepak Kumar Jain, Gaurav Saxena and Vineet Kumar Singh, Image mosaicking using corner techniques, *International Conference on Communication Systems and Network* Technologies, Rajkot, May 2012:79-84.

[15] Hernâni Gonçalves, José Alberto Gonçalves and Luís Corte-Real, HAIRIS: A Method for Automatic Image Registration Through Histogram-Based Image Segmentation, *IEEE Transactions on Image Processing*, 2011, **20**:776-789.

[16] Dianyuan Han, A standing tree image mosaic algorithm based on SIFT, *International conference on Computer science and electronics* engineering, Hangzhou , China, 2012, 1:353-356.