BACKGROUND MOSAICING FOR SCENES WITH MOVING OBJECTS

Udhav Bhosle, Subhasis Chaudhuri and Sumantra Dutta Roy

Department of Electrical Engineering
Indian Institute of Technology Bombay
Powai,Mumbai -400076.
{udhav,sc,sumantra}@ee.iitb.ac.in

ABSTRACT

The general problem of mosaicing is to create a single seamless image by aligning a series of spatially overlapped images. The result is an image with a field of view greater than that of a single image. This paper proposes a framework for creating a panoramic view representing the background of a given image sequence, by discarding the foreground objects. Our system performs motion estimation and segmentation on the input video stream, and extracts the background. We extract features from video frames, and use a novel Geometric Hashing-based method for fast and automatic image registration and mosaicing.

1. INTRODUCTION

Image mosaicing overcomes the limitations of the limited field of view of a camera, by aligning and pasting frames in video sequences, which enables a more complete view [1] Three major issues are important in image mosaicing: Image alignment or Image registration, Image cut and paste and Image blending [2, 3]. In addition to automating the process, an additional problem is that of moving objects in video frames. In this paper, we present a novel method to deal with these issue.

The estimation of motion parameters in the image sequence is a very important problem for mosaicing. The primary focus in the literature has been on motion of a single object in the scene. However, in most practical situations the motion field is not homogeneous as there may be several objects undergoing different motions. Estimation of image transform parameters can be biased by moving objects because moving regions of the image indicate a transformation different than the transformation due to camera. For example, direct minimization of pixel intensity difference has been widely used to register images [4, 5]. However, moving regions of high constract contribute significant residual to minimization, producing biased results. In feature based registration, features arise on the boundary between foreground and background objects. These features move unpredictable with respect to the rest of the image, producing unreliable results [6].

Computing the motion of several moving objects in image sequence involves simultaneous motion analysis and segmentation. This task can become complicated when image motion changes significantly between frames, as with camera vibrations. To handle the registration of background motion and the segmentation of foreground objects, several approaches can be used. In [7] local motion is computed for each image directly from the video sequence with optical flow. Then the image is partitioned into regions that have a coherent local motion with an iterative refining framework. Optical flow is also used in [8], to directly partition image into two regions, using a clustering technique. The bigger region is assimilated to the background , and its motion can be computed precisely using the region mask, and a parametric model. In [9] the dominant motion of the image pair is first evaluated to find a rough estimate of the background motion. Then a background mask is computed, by segmenting the aligned images on local motion intensity. The first motion estimation may be modified by some moving objects. For this reason a refinement is performed by computing extra alignments and segmenta-

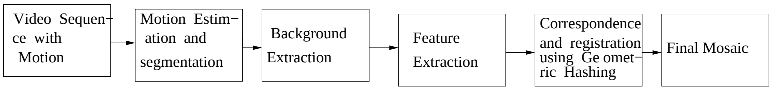

Figure 1: Block diagram of the proposed system

tion, where each registration is computed only for the previously segmented region. In this article, we deals with panoramic mosaic of background scene without any restrictive assumptions on the specific camera movements. The first task is motion estimation and segmentation. Objects with a motion different from the background motion are not taken into account for compositing the mosaic. The problem is to extract the background . Second task is correspondence and registration. We use a feature based method for image registration. Matching features across images has exponential time complexity. We reduce this to the polynomial-time. This speed up the matching process in addition to automating it. The rest of the paper is organized as follow. Section 2 describes Motion estimation and segmentation, Section 3 discuss Geometric Hashing. Section 4 describe panoramic Image mosaicing. We conclude the paper in Section 5.

2. SYSTEM DESCRIPTION

2.1. Motion Estimation and Segmentation

We chose to use a dominant motion approach with an associated segmentation. The background is expected to occupy the main part of the image, so the dominant motion effectively represents its motion. As the background is supposed to be static in the real world its apparent motion is just the effect of the camera movement. Further it is the furthest object in the image, so the image need only a foreground/background segmentation. Figure 1 shows block diagram of the proposed system.

2.2. Assumed Background Properties

The background is expected to have the following properties.

(1) It is situated behind the rest of the scene.

(2) Appearance of the scene remains constant over the time, the only changes in the grey levels are due to global motion.

(3) Background pixels occupy the main part of the image.

The dominant motion approach is not biased by foreground objects for translation, by using the hierarchical method exposed in [10]. But it is less robust as soon as higher order models(affine, planar)are used. In the segmentation, pixels are classified as moving or stationary using simple analysis based on local normalized difference. Regions having uniform intensity may be interpreted locally both as moving and as stationary and intensity difference caused by motion is also affected by the magnitude of the gradient in the direction of the movement. Therefore, rather than using a simple grey level difference as a motion measure for classifying the pixels, the grey level difference normalized by the gradient magnitude is used as a local motion measure. For classifying the pixels, multiresolution scheme is used. For every pixel, at each resolution level, both Motion measure and reliability of motion measure are calculated [10].

First all pixels at the lowest resolution level are initialized as ”unknown, to be moving or stationary”. If the computed motion measure is high(i.e. pixel is moving)or if low with high reliability (i.e. pixel is stationary), then the motion measure of the pixel at that resolution level is set to be new computed motion measure. Otherwise, if the local information available at the current resolution level doesn’t suffice for classification, then the motion measure form the previous lower resolution level is maintained. This algorithm yields a continuous function, which is an indication to the magnitude of the displacement of each pixel between the two images. Taking a threshold on this function yields partitioning of image to moving and stationary regions. By extracting the background, feature are found using Harris corner detector. The corresponding between images is found using Geometric Hashing.

3. GEOMETRIC HASHING

Image alignment requires matching M points in one image with N points in another. As such, this process has an exponential time complexity, $O(M^N)$. Lamdan *et al.* [11] propose geometric hashing as a fast method for 2-D object recognition using an affine assumption where M object points are to be matched to N image points, We generalize this idea for image alignment (the first step in image mosaicing), according to the specific transformation between two images – Euclidean, affine, or the most general projective case.

A 2-D transformation requires K basis points (K = 3 for Euclidean and affine, 4 for projective). We can select ordered pairs of K basis points from the first image in $\binom{M}{K} \times K!$ ways (this is $O(M^K)$). For each such basis, we compute the coordinates of the remaining $M-K$ ($O(M)$) points. A *hash table* stores these coordinates, indexed by the basis points. We repeat the process for the second image. Matching rows of coordinates between hash tables of the two images has *quadratic* time complexity. We can reduce this to *linear* is we sort each row in the hash tables. Hence, the problem of matching image features reduces to $O(M^{K+1}N^{K+1}) \times$ the row matching time. This has polynomial time complexity, an improvement over the exponential time complexity required for a naive feature match. We show the application of Geometric Hashing to panoramic background mosaic.

4. PANORAMIC IMAGE MOSAICING

A commonly used camera model is [12]:

$$\lambda \mathbf{p} = \mathbf{A} \left[\begin{array}{c|c} R & T \end{array} \right] \mathbf{P} \tag{1}$$

relating the coordinates of a 3-D point in the world coordinate system $\mathbf{P} = [X \; Y \; Z \; 1]^T$to its image point $[x \, y \, 1]^T$. λ is a projective constant. Here $\mathbf{A}$ denote matrix of internal camera parameters, R denote a rotation matrix and T, a translation vector. We can relate the image coordinates to the (non-homogeneous) coordinates of the 3-D points in the camera coordinate systems using $\lambda \mathbf{p} = \mathbf{AP}$ and$\lambda' \mathbf{p}' = \mathbf{A}'\mathbf{P}'$. For two cameras looking at the same point 3-D point $\mathbf{P}$

$$\mathbf{P}' = R\,\mathbf{P} + T \tag{2}$$

For panoramic image mosaicing, $T = 0$. So $\lambda' \mathbf{A}'^{-1} \mathbf{p}' = \lambda R \, \mathbf{A}^{-1} \mathbf{p}$. Hence, we have

$$\mu \mathbf{p}' = \mathbf{Hp} \tag{3}$$

$$\mu \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

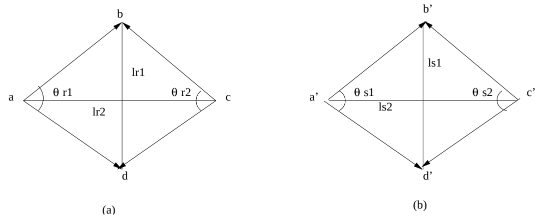

Figure 2: (a,b,c,d) basis quadruplet in reference image (left) and (a',b',c',d') basis quadruplet in second image(right).

H is a 3×3 invertible, non-singular homography matrix. The above homography matrix represents a 2-D to 2-D projective transformation. Homographies and points are defined up to a nonzero scalar. So eight parameters are to be found out. The above equation can be written as

$$x' = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1} \quad y' = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1} \tag{5}$$

Every point correspondence gives two equations, thus to compute H, we need four point correspondence. For a pair of corresponding points, it can be written as

$$h_1 x + h_2 y + h_3 - h_7 x x' - h_8 y y' = x'$$
$$h_4 x + h_5 y + h_6 - h_7 x y' - h_8 y y' = y'$$

Therefore, we use a projective basis for our geometric hashing-based scheme. We consider projective bases defined by pairs of four non-collinear projective points, using the canonical frame construction of [13]. This method considers mappings from the four non-collinear points to the corners of a unit square. Thus, we have $\binom{m}{4} \times m!$ possible choices for the basis vectors. *It is important to note that the relative change of successive camera positions is often kept small to maximize the number of corresponding points between images.* We use this to advantage in a novel Geometric Hashing-based method to further reduce the time complexity of alignment(details are available in our previous work [2, 3]).

Algorithm :

(1) Represent the reference image by the sets of corners.

(2) For every quadruplet (of which three must be non-collinear), find the angles (θ_1, θ_2) formed by two linearly independent vectors and lengths(l_1, l_2) between two end points as shown in Figure 2.

(3) For the second frame of the scene, for every quadruplet find the corresponding (θ, l) values.

(4) for every quadruplet in the second image, find the difference between angle $\theta s1_{(j)}$ and angle $\theta r1_{(i)}$ and

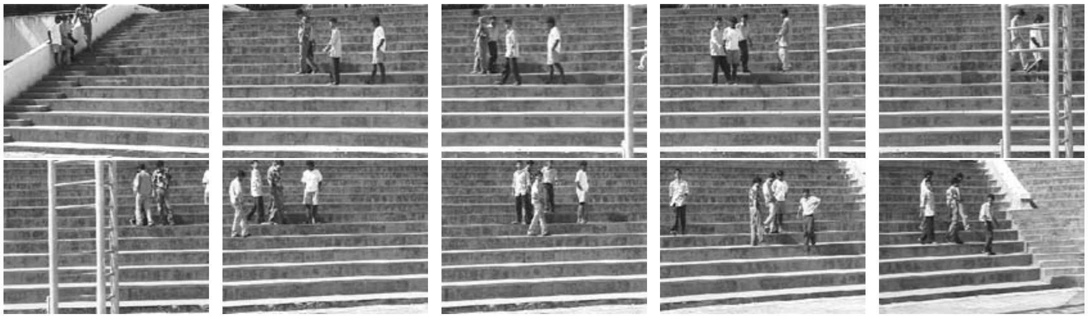

Figure 3: Some of the images of video sequence with multiple moving components

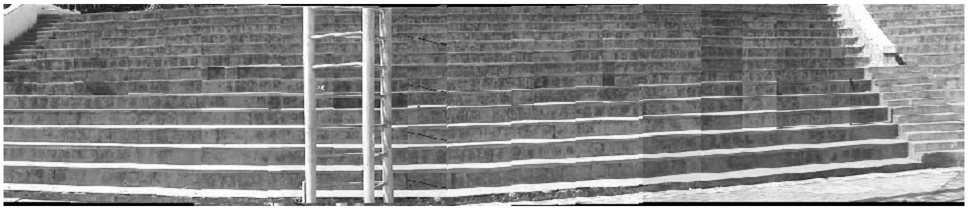

Figure 4: A panoramic background mosaic created from 25 frames of the SAC, IITB, Powai

difference between $\theta s2_{(j)}$ and angle $\theta r2_{(i)}$ of all quadruplet in the reference image:

$$\delta\theta1_{(i,j)} = \mid \theta s1_{(j)} - \theta r1_{(i)} \mid, \; \delta\theta2_{(i,j)} = \mid \theta s2_{(j)} - \theta r2_{(i)} \mid$$

Similarly, calculate the difference in lengths as

$$\delta l1_{(i,j)} = \mid ls1_{(j)} - lr1_{(i)} \mid, \; \delta l2_{(i,j)} = \mid ls2_{(j)} - lr2_{(i)} \mid$$

where $i = 1,2,3...\binom{M}{4}$; $j = 1,2,3...\binom{N}{4}$. out of $\binom{M}{4} \times \binom{N}{4}$ combinations, few most likely correct pairs can be identified through two passes. We can discard the quadruplets which give angle difference more than the threshold. The pairs of quadruplets with small difference in $\theta1$ and $\theta2$ will be considered for comparison based on lengths. By sorting based on $\delta l1$ and $\delta l2$, choose pairs with minimum value of $\delta l1$ and $\delta l2$. So, the pair with least values of $\delta\theta1, \delta\theta2, \delta l1, \delta 12$, considered as a right candidate. This means a quadruplet in the reference image matches with quadruplet in the second image. Even in the absence of any invariance in parameters θ and l, the above constraints can be safely used as the relative change in these parameters is very small due to dense time sampling of images. The required transformation can be obtained from a pair of matched quadruplet or estimated from more matched vertices by using least square error(LSE) estimation method. By finding transformation between two frames, the second frame is transformed with respect to first one and they are combined to form mosaic. Here, reference image is selected and all other images are registered with respect to the reference image, and mosaic is created. In this case, the region in the overlapping area is taken form one of the images, so there is no effect of blurring in mosaic image. Figure 3 shows some of the images of video sequence with multiple moving components. Figure 4 shows panoramic background mosaic. We show another example in Figure 5 and Figure 6.

5. CONCLUSION

This paper presents a new method to produce a background mosaic from a video sequence. Our method involves two steps. One is segmentation of forground and background region, second is alignment and pasting of images onto the mosaic. First step relies on the assumption that the background is the dominant object in the source image, so that a dominant motion detection approach is used. For the alignments of images we use the method based on Geometric Hashing. This gets over the problem of exponential time complexity in matching features across images. We show results in the support of proposed strategies.

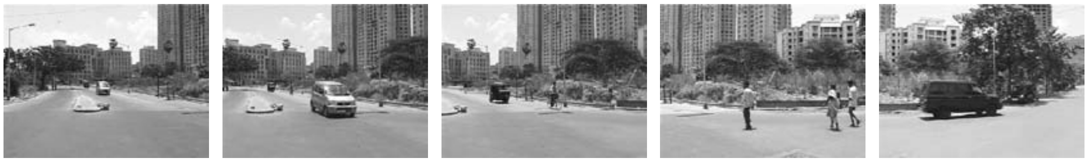

Figure 5: Some of the images of video sequence with multiple moving components of powai lake complex

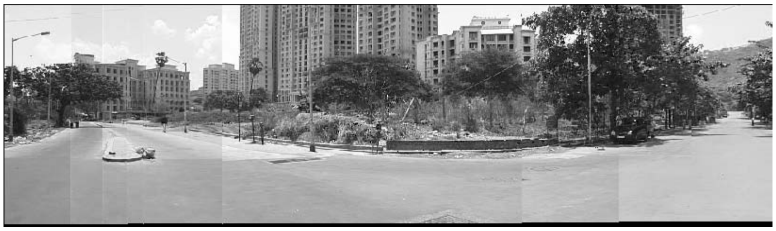

Figure 6: A Panoramic background mosaic created from a set of 13 frames of powai lake complex, Mumbai

REFERENCES

[1] S. Peleg and J. Herman, "Panormaic Mosaic by Manifold Projection," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, April 1997, pp. 338 – 343.

[2] U.Bhosle, S. Chaudhuri, and S. Dutta Roy, "The Use of Geometric Hashing for Automatic Image Mosaicing," in *Proc. National Conference on Communication (NCC02)*, January 2002, pp. 533 – 537.

[3] U. Bhosle, S. Chaudhuri, and S. Dutta Roy, "A Fast Method For Image Mosaicing Using Geometric Hashing," *IETE Journal of Research: Special Issue On Visual Media Processing*, pp. 317 – 324, May - August 2002.

[4] R.Szeliski and H.Yeung, "Creating Full View Panoramic Image Mosaic and Environment in *Computer Graphic Procedding, Annual Conference Series*, 1991, pp. 251 – 258.

[5] S.E. Chen, "Quicktime VR, an Image Based Approch to Virtual Environment Navigation," in *SIGGRAPH*, August 1995, pp. 29 – 38.

[6] J.Davis, "Mosaics of Scenes With Moving Objetcs," in *Proc. IEEE Int. conf. on PAMI*, March 1998, vol. 20, pp. 354 – 376.

[7] J. Wang and E.Adelson, "Spatio-temporal Segmentation of Video data," in *SPIE:Image and video processing*, February 1994, vol. 2182.

[8] F. Moscheni, S. Bhattacharjee, and M. Kunt, "Spatio-temperoal Segmentation based on Region Merging," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 3, pp. 122 – 129, September 1998.

[9] M. Irani, B. Rousso, and S. Peleg, "Computing Occluding and Transparent Motions," *International Journal of Computer Vision*, February 1994.

[10] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical Model Based Motion Estimation," in *ECCV*, March 1992, pp. 237 – 252.

[11] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson, "Object Recogination by Affine Invariant Matching," *Pattern Recognition*, pp. 335 – 344, June 1998.

[12] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[13] C. A. Rothwell, *Recognition using Projective Invariance*, Ph.D. thesis, University of Oxford, 1993.