Fast Generation of Dynamic and Multi-Resolution 360° Panorama from Video Sequences

Zhigang Zhu[!+*], Guangyou Xu[!], Edward M. Riseman[+], Allen R. Hanson[+]
[!] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[+] Computer Vision Lab, Computer Science Department, University of Massachusetts at Amherst, MA 01003

Abstract

This paper presents a systematic approach to automatic construction of a dynamic and multi-resolution 360° panoramic (DMP) representation from image sequences taken during camera rotation and zoom. Although a simple 2D rigid motion model is used to estimate inter-frame motion parameters, our mosaicing methodology enables precise cylindrical panoramic mosaic. Moving objects are detected and separated from images based on motion information, and their accurate contours are extracted using a modified active contour algorithm. A multi-resolution representation is built for the more interesting areas by means of camera zooming. The DMP construction method is fast, robust and automatic, achieving 1 Hz on a 266MHz PC. No camera calibration, feature extraction, image segmentation or complicated nonlinear optimization processing is required in our algorithms. The construction of the panoramic representation can be used in virtual reality and very low bit-rate video coding.

Keywords: *Image-based VR, panoramic representation, dynamic mosaic, multi-resolution, object extraction*

1. Introduction

A panoramic representation of image sequences has a wide application scope, including virtual reality (VR), interactive 2D/3D video, tele-conferencing, content-based video compression and manipulation, and full-view video surveillance. For virtual reality, it has advantages of simplicity for rendering, photographic quality realism, and 3D illusion experienced by users. For video analysis and coding, it is superior to existing coding approaches in that it is a content-based representation with a very low bit-rate. In the sense of advanced human-computer interaction (HCI), these two categories will merge into a more general approach of interactive video (e.g. virtual conferencing), which adds the flexibility of synthesizing images with interactivity, selectivity, and enhanced field of view and resolution.

A wide field of view (FOV) lens, e.g. a fish-eye [1] or panoramic lens [2-6], can be a solution for generating panoramic presentations. Besides the expense of these specially designed image sensors, the image obtained will have substantial distortions, and mapping an entire scene into the limited resolution of a video camera compromises image quality. Constructing a panoramic representation by mosaicing image sequences captured by ordinary cameras, on the other hand, meets the requirements of many applications where high image resolution, low bit-rate, interactivity and photographic realism are needed.

Apple's QuickTime VR [7] captures a 360-degree panoramic image of a scene with a camera panning horizontally from a fixed position. The overlap in images are registered first by the user and then "stitched" together by the software in a best match. Similarly in [8] mosaics were constructed by registering and reducing the set of images into a single, larger resolution frame. However the final image mosaic is not a full 360-degree view. Plenoptic modeling [9] adds ranges (using a disparity map) to each panoramic image, thereby allowing reprojection from other viewpoints. The concept of plenoptic function is further explored by light field methods [10,11], which attempt to fully sample the plenoptic function within a subset of space. Clearly the generation of a full-view panorama is the foundation of these methods.

Shum & Szeliski [12] proposed a mosaic representation that associates a transformation matrix with each input image, rather than explicitly projecting all of the images onto a common surface (e.g., a cylinder). However the decomposition of the projective transformation matrix into rotation angles and the focal length is known to be very sensitive to image noise. Kang & Weiss [13] analyzed the error in constructing panoramic images and proposed a technique that has the advantage of not having to know the camera focal length *a priori*. In order to create a panorama, they first had to ensure that the camera is rotating about an axis passing through the nodal point, and the focal length of the camera cannot be changed throughout the rotation. Xiong and Turkowski [1] proposed a method to create image based VR using a self-calibrating fisheye lens. They take four pictures by rotating the camera 90 degrees around the nodal point and formulate the registration and self-calibration constraints as a single nonlinear minimization problem in which 34 parameters need to be determined. Manifold projection

[*] Contact Email: zhu@cs.umass.edu

[14] enable the fast creation of low distortion panoramic mosaics under a more general motion than the exact panning. The basic principle is the alignment of the strips that contribute to the mosaic, rather than the alignment of the entire overlap between frame. However the issues of full view panorama, independent object motion and camera zoom are not considered in this approach.

Static scenes are a common assumption in image mosaicing and image-based rendering [1, 7, 9-14], with the exception of a dynamic mosaic approach proposed by Irani et al [15] to describe dynamic events. However the accuracy of the contour of a moving object was not addressed, which is important for synthesis of fine detail of the dynamic events based on the mosaic representation. In our work we utilized a modified active contour method to extract the contour of the moving object. In the existing algorithms [17-19], only the intensity information was used. In order to detect and rapidly separate the dynamic and deformable objects from the scene, both motion and shape information will be utilized in our method.

We aim at the generation of realistic 2D/3D panoramas from video sequences with more general motion of a hand-held video camera. The construction of a layered 3D panorama from a vibrating translating camera has been reported in [19]. In this paper a new approach is proposed to automatically build a dynamic and multi-resolution 360° panorama (DMP), with good image quality, from a video sequence taken by a hand-held camera undergoing 3D rotation, zooming, and small translations. It should be noted that this is often the case for the general operations of a video camera by a cameraman. For construction of a realistic virtual environment, this requirement can be easily satisfied. Though the description of the DMP construction algorithms in this paper is mostly directed towards a scenario of virtual environment modeling, the same algorithms with slight modifications can be directly used in video analysis and coding, and in video surveillance.

This paper is organized as follows. In Section 2 the inter-frame motion model is derived and a pyramid-based motion detection algorithm is described. In Section 3 the image mosaicing and warping algorithm is presented in detail. The algorithm for moving object detection and segmentation from the background is presented in Section 4. Section 5 describes how to build the multi-resolution representation for the user selected "interesting" regions. Experimental results are given in the corresponding sections and a brief conclusion and discussion is provided in the last section.

2. Inter-Frame Motion Model

Let us make a basic assumption that the scene is static and all motions in the image are caused by the movement of the camera. The independent motion of other objects in the scene will be considered in Section 3 and Section 4. A coordinate system *XYZ* is attached to the moving camera; the origin O is the optical center of the camera (Fig. 1). *UV* is the image coordinate system whose origin is the intersection of the optical axis with the image plane. The camera motion has 6 degrees of freedom: three translation components and three rotation components. Since we use the camera as the reference coordinate system an alternative view is that the scene being viewed moves with 6 degree of freedom. Considering only an inter-frame case, we represent three rotational angles (roll, tilt and pan) by (α, β, γ) and three translation components by (T_x, T_y, T_z).

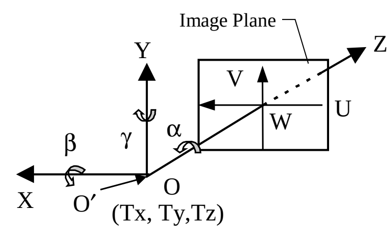


Fig. 1. Coordinate systems of the camera and the image

With the current frame at time t and the reference frame at the previous time t', a 3D point *(x, y, z)* with image coordinates *(u, v)* at time t will move from point *(x', y', z')* in the reference time t', with the image point *(u', v')*. Suppose the camera focal length *f* is *f'* before the motion. Under a pinhole camera model

$$(u,v) = (f\,x / z, f\,y / z)$$

the relation between the 3D coordinates is

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} a & b & c \\ i & j & k \\ l & m & n \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{1}$$

and

$$\begin{cases} u' = f' \dfrac{au + bv + cf + fT_x / z}{lu + mv + nf + fT_z / z} \\ v' = f' \dfrac{iu + jv + kf + fT_y / z}{lu + mv + nf + fT_z / z} \end{cases} \tag{2}$$

where $\boldsymbol{R}$ is the rotation matrix

$$\boldsymbol{R} = \begin{bmatrix} a & b & c \\ i & j & k \\ l & m & n \end{bmatrix} \approx \begin{bmatrix} 1 & \alpha & -\gamma \\ -\alpha & 1 & \beta \\ \gamma & -\beta & 1 \end{bmatrix}$$

and it can be approximated by the right-hand matrix for the rotation of successive frames. In order to construct the 360° panorama, panning is the dominant motion of the camera. Under pure 3D rotation (i.e. $T_x = T_y = T_z = 0$) , we have the following homogenous rotation transformation

$$\begin{cases} u' = f' \dfrac{au + bv + cf}{lu + mu + nf} = \dfrac{Au + Bv + C}{Lu + Mv + 1} \\ v' = f' \dfrac{iu + jv + kf}{lu + mu + nf} = \dfrac{Iu + Jv + K}{Lu + Mv + 1} \end{cases} \tag{3}$$

which is a special case of planar projective transformation. With four pairs of matched points in two successive images, eight parameters (A,B,C,I,J,K,L,M) can be solved from equation (3), and two images can be registered exactly to generate a planar mosaic with a larger field of view. However, the field of view is constrained to be less than 180° since the points in the direction of ±90° from the optical axis of the reference frame are mapped to infinity in the planar mosaic. A full-view cylindrical panorama can be constructed by decomposing f, f', α, β and γ from the eight projective parameters [12], but unfortunately the decomposition of the intrinsic and extrinsic parameters is very sensitive to noise. Thus we look for an alternative way to achieve the goal efficiently.

2.1. 2D rigid motion model is plausible

If the rotation angle is small, e.g., less than 5 degrees, between the successive frames, equation (2) can be approximated as

$$\begin{cases} u' \approx f' \dfrac{u+\alpha v-\gamma f+fT_x/z}{\gamma u-\beta v+f+fT_z/z} \\ v' \approx f' \dfrac{-\alpha u+v+\beta f+fT_y/z}{\gamma u-\beta v+f+fT_z/z} \end{cases} \tag{4}$$

Let

$$s = (\gamma u-\beta v+f+fT_z/z)/f' \tag{5}$$

we will have

$$\begin{cases} s\cdot u' = u+\alpha v-\gamma f+fT_x/z \\ s\cdot v' = -\alpha u+v+\beta f+fT_y/z \end{cases} \tag{6}$$

Under 3D rotation that is dominated by panning motion, possibly with zooming and small translation, we have very small roll α, tilt β and $(T_x/z, T_y/z, T_z/z)$. Therefore a 2D rigid inter-frame motion model can be used

$$\begin{cases} s\cdot u' = u+\alpha v+T_u \\ s\cdot v' = v-\alpha u+T_v \end{cases} \tag{7}$$

where $s \approx fZ'/f'Z$ is a scale factor associated with zoom, and Z-translation; $(T_u, T_v) \approx (-\gamma f+fT_x/Z, \beta f+fT_y/Z)$ is the translation vector representing (pan/X-translation, tilt/Y-translation); and α is the roll angle. This motion model is also plausible if the scene is far away. Given more than 2 pairs of corresponding points between two frames, we can obtain the least square solution of motion parameters, s, T_u, T_v and α, in equation (7). The errors of approximation are especially small for the narrow vertical strip in the center of each image that will be used in our image mosaic algorithm (Fig. 2). This observation can be easily deduced by comparing equation (7) with equation (4) when β≈0, u≈0, and $(T_x/z, T_y/z, T_z/z) \approx 0$. If the image size is 384×288 and the equivalent focal length of the camera is 384 pixels, numerical analysis [21] shows that when all the three angles are less then 2 degrees, errors are of only 0~2 pixels in the central strip with the width w < 16 pixels (Fig. 2). The approximation is valid for image sequences taken by a hand-held video camera.

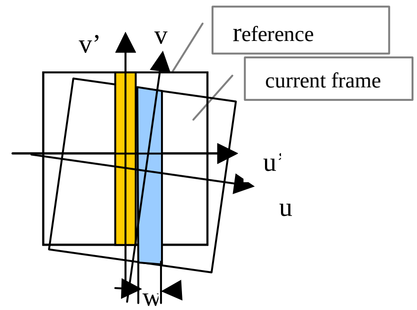


Fig. 2. Mosaicing strips

2.2. Inter-frame motion detection algorithm

The inter-frame image displacements are estimated by using a pyramid-based matching algorithm. The algorithm consists of the following steps [21]:

(1) *Generating the pyramids* for the current and the reference images. For computational efficiency, the final image displacements are only given for non-overlap image blocks of a given size, say 16×16, in the finest layer (original image) of the current frame. The matching process is carried out from coarse layers to fine layers, starting from a layer with certain image size, e.g. two times large as the matching block size.

(2) *Determining the image displacements.* For each block in a layer of the current frame, the correlation operation is carried out in an adaptive search window over the reference frame pixel by pixel. The maximum correlation is chosen in the search window, and if there are multiple best matches due to similar patterns in the search window, then the one with the smallest displacement is selected. The initial size of the search window is about half the image size in the first layer but it will be reduced in the finer layers, and after image warping (see Section 3). With this search strategy, the likelihood of mismatch will be reduced if the real motion displacements or the search windows are smaller. This is the basic principle for the re-matching-after-warping procedure described in the next section.

(3) *Calculating the belief value* of each match by combining the texture measure with the correlation measure. This step is important because the belief values will be used as weights in the parameter estimation. It should be noted here that the matching process will be affected by several kinds of noise, such as moving objects, illumination changes and camera zoom. We will present a method of two-embedded iterating cycles to deal with this problem.

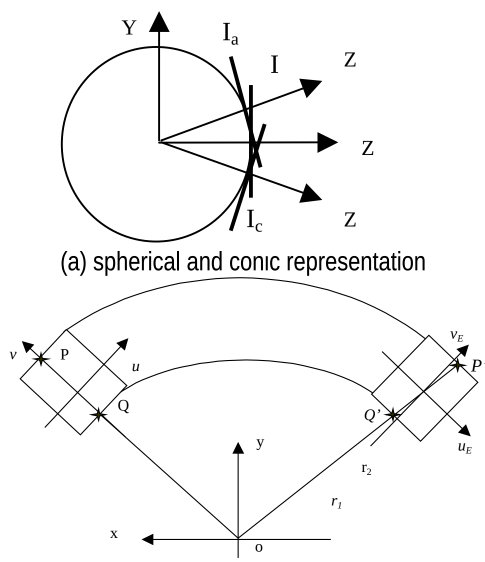
3. Image Mosaicing and Rectification

The relation between two frames from pure rigid 3D rotation is a strict planar projective transformation.

However, if we use planar reprojection, the field of view is limited to be less than 180 degrees. In an initial study, we first utilized a direct linear method similar to that in [12] to estimate camera parameters from projective transformation between two frames. The parameters include relative focal length, nodal point, aspect ratio, and the three inter-frame rotational angles of the camera. Theoretically it would be elegant if a cylindrical panorama can be constructed after the focal length and the three rotation angles have been decomposed. However, experimental analysis has shown that this decomposition is very sensitive to image noise and accuracy of the recovered motion parameters. Since the motion we consider in our domain is not a pure rotation, which make this difficult problem even harder, we adopt an alternative approach when camera panning is the dominant motion and the pan covers more than 360° around the viewpoint. The algorithm consists of the following three steps:

(1). *Estimating the 2D rigid transformation between two successive frames.* This step consists of two embedded iteration cycles [21]. The first (inner) iteration cycle is robust motion estimation based on the current motion vectors from image match. The 2D rigid transformation between two successive frames is estimated using a weighted least mean square method. Then an iterating process is carried out after modifying the weight of each match by the error between each measured image displacement and the one calculated from rigid transformation until the average error is below a certain tolerance level. Notice that this iterative process is only carried out on the current motion vectors without re-calculating them from the original images so the computation speed is very fast. The re-weighting process accounts for moving objects and other mismatches that are not consistent with the estimated rigid motion model. The second (outer) iteration cycle is for motion detection and estimation after warping the current frame using the calculated motion parameters. Then the difference between the warped image and the reference image provides residual errors for the motion model. If the residual is large then the residual motion displacements are estimated between the warped frame and the reference frame, and the inner iteration cycle runs again. Since the residual motion displacements are reduced, the probabilities of mismatches will be reduced hence the matching results will be improved. Experiments show that about two match cycles after rectification can achieve rather fine registration results.

(2) *Mosaicing the image frame by frame.* Any frame (e.g. the first frame, the center frame) is selected as the reference frame for the mosaic process, and the accumulating transformation parameters between each frame and this reference frame are calculated. Then images are warped and pasted frame by frame onto the final mosaic. If only one narrow vertical strip in the center of each frame is utilized, a 2D rigid transformation is sufficient to merge the successive frames. 2D rigid mosaicing approximately maps the image to a flattened conic surface, or sometimes a cylindrical surface, depending on the orientation of the optical axis (Fig. 3). The principle behind the conic mosaicing can be explained as follows. Suppose the central strip is represented in spherical coordinates. Then the 2D rigid transformation in equation (7) exactly describes the 3D rotation and zoom of the camera, even though error is introduced by the approximation of the circular arc by a planar strip in the real image. However if the roll and tilt angles are significantly smaller than the pan angle, then this error is negligible since the distortion is mostly in the vertical direction [21]. It also implies that the actual mosaic is a flattened conic surface rather than part of a sphere since the strip is planar. The cone is upward if the optical axis of the reference frame is slightly downward looking and vice versa (I_c and I_a in Fig.3 (a)). A true cylindrical panorama can be obtained only if the optical axis is strictly horizontal (I_b in Fig.3 (a)).



(a) spherical and conic representation

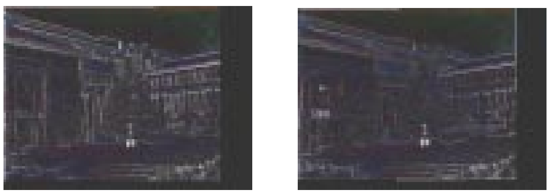
(b) flattened conic mosaicing and rectification

Fig. 3. The strip-mosaicing geometry

(3) *Rectifying the flattened conic mosaic to a flattened 360° cylindrical panorama.* This can be achieved by finding the correspondence of a (virtual) vertical edge in the head and tail of the conic mosaic. The correspondence is established automatically by matching the possible "connecting" frames in the image sequence with the first frame through the same pyramid-based matching strategy and selecting the frame with minimum difference with the first frame. To account for any illumination changes between the connecting head and tail frames, histogram specification from the frame in consideration to the first frame is performed. After the angle range, and the radii of inner and outer arcs of the flattened cone are computed

from the head-tail match, the re-projection of the conic mosaic to the cylindrical panorama can be determined (see Fig. 3(b)); details can be found in [21].

(a) the current frame (Frame no. 245)　　(b) the reference frame (Frame no. 0)



(c) difference image of inital matching, and (d) after re-matching

Fig. 4. A matching example from the 246-frame original image sequence (panning from right to left): the first and the last frame

It should be emphasized that no camera calibration or intrinsic camera parameters are needed, and the algorithm is completely automatic. Fig. 4 shows the matching process of the head and the tail frame from a 246-frame image sequence of the Library scene. The motion parameters from the initial match are $T_u = 49.07$, $T_v = 13.72$, $s = 1.00$ and $\alpha = 0.00$, while the motion parameters resulting from the second match are $T_u = 48.05$, $T_v = 13.74$, $s = 1.00$ and $\alpha = 0.00$. The second set of parameters result in a better registration result, which can be observed from the edges in the difference images between the two frames, especially the center strip of the image (e.g. the white lamp in front of the pine tree and the entrance near that tree), which will be used for mosaic. A mismatch correction example is given in [21].

Fig. 5 and Fig. 6 show the panoramas before and after cylindrical rectification and head-tail stitching. The conic panorama is constructed by simply pasting onto the mosaic one strip that is transformed (rotated, scaled and translated) from each frame with a width corresponding to the displacement of each frame in the mosaic. The original image sequence has 246 frames of 384 x 288 color images, so the average panning angle between two frames is about 1.5 degrees, which satisfies the condition in equation (7). The size of the rectified cylindrical panorama is 3494x323. If the compression ratio of the panorama in JPEG format is 20:1, the total compression ratio between the JPEG panorama and the original image sequence is about 500. Moreover new images of arbitrary viewing angles can be synthesized interactively, which is essential for applications of virtual reality and content-based video manipulation. In the case that there are moving objects in the scene, median values of the corresponding points in multiple frames are used to generate the conic panoramic background. The resulting image is slightly blurred since a wider strip is used in each frame (see Fig. 9c). The moving object extraction is presented in the next section.

Fig 5. Flattened conic mosaic (13% display scale). The original color image is 3806 x 773x24 bits. Notice the curved and uneven boundary created by the up-tilted angle and unstabilized motion of the hand-held camera, and error accumulating.

Fig.6 Flattened 360-degree cylindrical panorama (14% display scale). The original true-color image is 3494x323 x24 bits.

4. Moving Object Extraction

As the mosaic is being constructed, difference image between successive frames is analyzed. In practice, a difference image is calculated from three successive images for robustness. Then region analysis is carried out to determine those regions that may contain moving objects. In order to realize best figure-ground separation, the contour of the moving object in each region needs to be extracted. We apply an active contour model to extract contours from a noisy image [16, 17,18]. The basic idea of the active contour algorithm is to constrain the contour of an object onto a controllable continuous spline. The task is to minimize an energy function that takes into account both input image information and constraints on the continuity of the contour. Our modified active contour algorithm uses both motion and gradient cues of the images, and the control parameters are adaptively adjusted according to objects in the current image. The algorithm consists of the following four steps:

(1) A difference image is calculated from the current image and its predecessor and successor frames in the sequence. Regions with large residuals are detected through a region-grouping algorithm. Then, in each region, the difference value is threshold to a binary image, gaps and holes are filled using morphology-based method, and a larger scale grouping is used (if necessary) to generate a single mask for each moving object; this mask is then used as an initial contour in the following step.

(2) Control points are obtained from the initial contour, and curvatures at the control points are estimated. The points are evenly spaced although the space is adaptively changed according to the size of the initial contour. Then the parameters used in the energy function are automatically assigned according to the point spaces and the curvatures.

(3) The energy function is minimized using a dynamic programming approach to obtain the resulting contour.

(4) Each dynamic object is separated along its contour from the original frame and is labeled on the corresponding location of the panorama, and the dynamic sub-images of objects are represented individually.

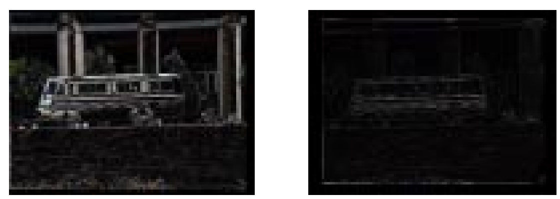
Fig. 7 shows the dynamic mosaic with the walking person pasted onto the mosaic every ten frames.

Fig. 7. Moving object detection and separation

5. Multi-Resolution Representation

In VR applications we want the ability to zoom and pan (under controlled motions) to enhance the visual realism; in image coding we need to handle the video sequence with zoom as well as pan. Therefore, we introduce a multi-resolution representation for each user-specified "interesting" portion of the panorama. Each of those regions on the panorama is labeled as a "zooming hot spot". This is similar to the sparse pyramid in [15], but our representation is more purposive and compact. The representation is constructed by physically zooming the camera when the more interesting regions of the scene are viewed. The zoomed frames are separated automatically from the original panning and zooming image sequence by observing the accumulating scaling (zoom) factor. An automatic registration between two zoomed frames is achieved in a manner similar to that for the panned frames, but the following step is to select representative frames as the components of a multi-resolution representation (instead of mosaicing the frames). It should be noted here that it is more difficult to accurately assess similarity in the zooming case than in the panning case, especially when the scale change is grater than 10%, since the scales of the match blocks are not the same in the two images. In this case re-match processing after warping (i.e. re-zooming) is vital for the accurate estimation of the scale parameter.

(a) the current frame (b) the reference (preceding) frame



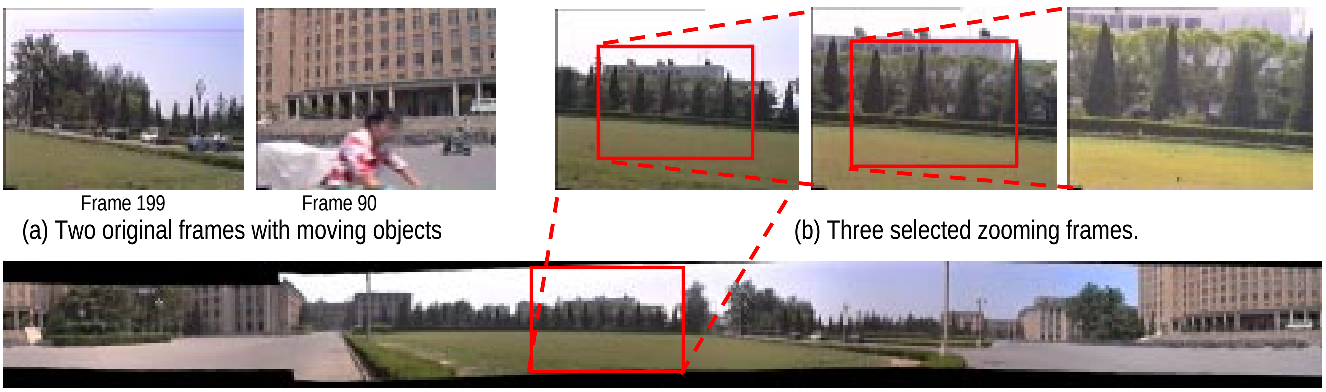
(c) difference image of initial matching and (d) after re-matching

Fig. 8. Iterative matching after image warping (zooming)..

Fig. 8 shows a matching example from the zoomed segment of the Main Building sequence shown in Fig. 9. The motion parameters from the initial matching process are $T_u = 7.29$, $T_v = -0.52$, $s = 1.03$ and $\alpha = 0.00$, while the motion parameters from the second (final) matching process are $T_u = 0.34$, $T_v = -0.99$, $s = 1.12$ and $\alpha = -0.00$. The second set of parameters results in a much better registration of the frames, as can be seen by comparing Fig.8(c) and Fig. 8(d). The zoom factor between Fig. 8a and 8b is 1.12. The reason for successful match is that every iteration brings the scale factor closer to the real one. Fig. 9(b) shows the selected zooming frames with 1.5 zooming factors between two selected frames for the Main Building image sequence. The rectangle in the mosaic and each selected zoom frame indicates the sub-region that corresponds to the next selected frame.

6. Conclusion and Discussion

The construction of the DMP (Dynamic and Multi-resolution Panorama) is fast, robust and automatic. The processing rate is about 1 frame per second for 384×288 color images using a Pentium II/ 266 MHz PC. A factor of 2 speed-up can be expected by algorithm optimization and MMX utilization. WE have tested our algorithms on many long image sequences (typically 100~600 frames) captured by a hand-held camera for both indoor and outdoor scenes, and a demo system of VR-based wandering through Tsinghua Library has been built using Netscape Plug-In [21]. Besides the most obvious applications such as virtual reality scene modeling and very low bit rate video coding, the DMP and the algorithm is also useful in other applications such as surveillance, change detection, video enhancement, indexing and manipulation.



(a) Two original frames with moving objects

(b) Three selected zooming frames.

(c) Cylindrical panorama after eliminating the moving objects (image size:3498x303). Notice that the zoom factor is changed.

Fig. 9. Multi-resolution panorama. The original image sequence has 561 frames, which consists of 3 zooming segments among the panning sequence. There are many moving objects (persons, bicycles) in the scene. Notice that most of the moving objects and noises (e.g. horizontal lines in frame 199) have been successfully filtered out (cars parked at the roadside remain in the mosaic).

Acknowledgments

This work was supported by the China High Technology Program under contract No. 863-306-ZD-10-22 and partially by DARPA under contract No. F30602-97-2-0032. Thanks are given to Dr. Yudong Yang, Mr. Heng Luo, and Mr. Qiang Wang at Tsinghua University for developing the early version of motion detection and image processing codes.

References

[1]. Y. Xiong, K. Turkowski, Creating image-based VR using a self-calibrating fisheye lens, *Proc.CVPR' 97*, 237-243.

[2]. S. K. Nayar, Omnidirectional video camera. *Prof. DARPA Image Understanding Workshop*, May 1997:235-241

[3]. V. Peri and S. K. Nayar, Generation of perspective and panoramic video from omnidirectional video, *Prof. DARPA Image Understanding Workshop*, May 1997:243-245

[4]. P. Greguass, Panoramic imaging block for three dimensional space, *U.S. Patent 4,566,763* (28 Jan 1986).

[5]. I. Powell, Panoramic lens, *Applied Optics*, vol. 33, no 31, Nov 1994: 7356-7361

[6]. Z. Zhu, E. M. Riseman, A. R. Hanson, Geometrical modeling and real-time vision applications of panoramic annular lens (PAL) camera, *Technical Report TR #99-11*, Computer Science Department, University of Massachusetts Amherst, February, 1999.

[7]. S. E. Chen, QuickTime VR - an image based approach to virtual environment navigation, *Proc. SIGGRAPH 95*, 29-38.

[8]. S. Mann, R. W. Picard, Video orbit of the projective group: a new perspective on image mosaicing, *Technical Report No.338*, MIT Media Lab Perceptual Computing Section,1995

[9]. L. McMillan and G. Bishop, Plenoptic modeling: an image-based rendering system, *Proc. SIGGRAPH 95*, 39-46.

[10]. S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, The lumigraph. Proc. *SIGGRAPH 96*, 43-54.

[11]. M. Levoy and P. Hanrahan. Light Field Rendering. *Proc. SIGGRAPH 96*, 31-42.

[12]. H.-Y. Shum and R. Szeliski, Panoramic Image Mosaics, *Microsoft Research, Technical Report, MSR-TR-97-23, 1997*

[13]. S. B. Kang, R. Weiss, Characteristics of errors in compositing panoramic images, *Proc.CVPR'97*, 103-109.

[14]. S. Peleg, J. Herman, Panoramic Mosaics by Manifold Projection. *IEEE Proc.CVPR'97*, 338-343.

[15]. M. Irani, P. Anandan, S. Hsu, Mosaic based representation of video sequence and their applications, *IEEE Proc ICCV'95*, 605-611.

[16]. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: Active contour models, IEEE *Proc. ICCV'*87, 259-269.

[17]. A. Amini, T. Weymouth, and R. Jain, Using dynamic programming for solving variational problems in vision, *IEEE Trans. PAMI*, 12(9), 1990,855- 867.

[18]. K. F. Lai, R. T. Chin, Deformable Contours: Modeling and Extraction, *IEEE Trans. PAMI*, 17(11), Nov 1995, 1084-1089

[19]. Z. Zhu, G. Xu, X. Lin, Constructing 3D natural scene from video sequences with vibrated motions, Proc. *IEEE VRAIS98*, 105–112.

[20]. J. Wang, E. H. Adelson, Representation moving images with layers, *IEEE Trans. on Image Processing*, 3(5),1994, 625-638.

[21]. Zhigang Zhu, Guangyou Xu, Allen R. Hanson and Edward M. Riseman, Fast construction of dynamic and multi-resolution 360-degree panorama from video sequences. *Technical Report TR #99-10*, Computer Science Department, University of Massachusetts Amherst, February, 1999.