Fast and Robust 2D Parametric Image Registration

Eun-Young Elaine Kang, Isaac Cohen and Gérard Medioni
Institute for Robotics and Intelligent Systems
School of Engineering
University of Southern California
{elkang, icohen, medioni}@usc.edu

Abstract

Camera stabilization is an important task for video analysis. In this report, we describe a method for estimating and stabilizing camera motion between two images via 2D image registration. Our approach is based on a pair-wise registration of images parameterized by 2D affine or projective transformation. Pair-wise image registration estimates transformation parameters for each pair of consecutive frames, and the transformation between arbitrary frames is computed by the concatenation of the pair-wise transformations. Pair-wise registration is based on 1) hierarchical parameter estimation and refinement 2) feature-matching 3) FFT(Fast Fourier Transformation)-based global matching 4) RANSAC-based parameter estimation.

1. Introduction

Stabilizing camera motion is a necessary task for video processing, such as video coding and video surveillance. This process enables us to focus on other scene analysis (eg. detecting and tracking moving objects). Here, we describe a fast and robust method to stabilize camera motion between two images. In general, motion stabilization for an arbitrary camera swath involves two technical issues: 1) robust and fast computation and 2) global registration. The first issue occurs due to the fact that many video processing applications demand real-time throughput. The second issues is because many video analysis tasks need to process multiple frames, which may not be consecutive in time, and global registration is required to process multiple frames seamlessly. We proposed a novel approach to solve these issues in [Kang00]. Per many readers' requests, in this report, we describe the technical details of our approach attacking the first issue, which is missing in the previous papers.

We achieve camera motion stabilization by a 2D image registration technique. The registration process is performed pair-wise. Pair-wise registration recovers the transformation for each pair of consecutive video frames. The method is based on feature-based hierarchical parameter estimation. It also uses the initial estimation based on a global matching in frequency domain and RANSAC outlier removal.

2D image registration methods have been intensively studied for the past several years in the academia and industry. Most of the methods use parametric approaches, which recover affine or projective or higher order, such as quadratic, parametric models by direct or feature-based error measurement. In the parametric approach, the initial parameters are estimated in various ways. And then, the parameters are iteratively refined by reducing and minimizing the errors between images. These errors are measured by considering all image intensities or features only. The former is called direct method and the latter feature-based method.

McMillan et al. generate a panoramic view for a rotating fixed camera and represent the images with a plenoptic function with respect to the camera center [McMillan95]. Moreover, they generate a view for a virtual viewpoint from a set of constructed panoramic view and simulate the depth by re-sampling the plenoptic function. Szeliski et al. also focus on the mono-centric mosaic, which has a fixed camera rotation axis, and show some extension for 8-parameter recovery [Szeliski94] [Szeliski97]. The method recovers the rotation parameters with focal length or 8-parameters by using the Levenberg Marquardt method. The limitation of McMillan's work is that it can only create mono-centric panoramic views and requires well-controlled camera system. Szeliski's work suffers from the local minima problem in case of large interframe motion.

Irani et al. recover the 2D quadratic transformation based on the direct image matching [Irani93] [Irani95] [Kumar95] [Irani96]. They also exploit the parallax information to process residual information after the image alignment resulting in 3D corrected mosaic images. Morimoto et al. address how to extract the dominant motion between two frames based on a 2D parametric model [Morimoto97] [Morimoto98]. In [Morimoto97], they employ 3D model based stabilization using Kalman filter to estimate the rotation between frames, and they de-rotate the camera sequence to compose mosaic. In addition, they show the accuracy of an image stabilization technique by using power signal-to-noise ratio [Morimoto98]. Sawhney et al. propose a mosaic method based on a similar minimization technique to Irani [Sawhney95] [Sawhney97] [Sawhney99]. But they take into account lens distortion and perform the minimization for all the parameters rather than pair-wise parameters. The first two groups take into account the 3D correction for mosaic construction, but they do not consider the case of large inter-frame motion nor and the accumulated error caused by pair-wise registration. The last group partly solves the accumulated error by estimating all transformation parameters at once. However, considering the number of parameters, that include 2D quadratic parameters and correction parameters for lens distortion for each pair of frames, the minimization technique does not converge unless a good initial estimation is given, and the process takes longer time to converge.

Peleg et al. introduce the manifold mosaic to minimize the image warping caused by the camera motion [Peleg97] [Rousso97] [Rousso98]. The manifold is adaptively defined by the optical flow. They use the 'slit' as a mosaic unit and stitch the slits. Unlike other methods, they handle the zoom-in/out case well by using pipe projection. This slit-based method is dedicated to create better mosaic images; therefore, it does not give a fairly straight forward method to recover parameters, which is useful for other operations, for example, inversely regenerating input video.

Our approach to stabilization is parameterized by 2D affine or projective transformation, and it is performed pair-wise. Pair-wise image registration estimates transformation parameters for each pair of consecutive frames, and the transformation between each frame and the reference frame is computed by the concatenation of the pair-wise transformation. Pair-wise registration is based on 1) hierarchical parameter estimation and refinement 2) feature-matching 3) FFT(Fast Fourier Transformation)-based global matching 4) RANSAC-based parameter estimation.

2. 2D Image Registration

We present a 2D parametric image registration method for robust and fast motion stabilization. The method is based on feature matching, initial parameter estimation based on global matching in frequency domain and hierarchical parameter refinement.

2.1 Motion Model

Before illustrating our registration method, we explain in which situation the 2D parametric motion can approximate the camera motion in 3D. The motion captured by the video stream is related to the camera motion in 3D. The mapping between two camera coordinate in 3D, (*X, Y, Z*) and (*X', Y', Z'*), can be represented by a rotation (*R*) and a translation (*T*) matrix shown in the following:

$$(X',Y',Z') = R(X,Y,Z) + T$$
$$= \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix}(X,Y,Z) + \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}$$

Figure 1 A Rigid Transformation in 3D

A point in (*X, Y, Z*) camera coordinate maps to (*X', Y', Z'*) by a rigid transformation. If all (X, Y, Z) are on a world plane, there exists a planar projective transformation between a world plane and a camera image plane.

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{31} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Figure 2 A Planar Projective Transformation between Two Planes in 3D

For a rigid camera motion of a 3D planar surface, there exist two planar projective transformations between a world plane and each camera image plane. There also exists a planar projective transformation between two images. In real world, a 3D planar surface transformation exists in scenes taken by a camera rotating about its axes (Pan-tilt) or tele-photo lens, where the depth of objects are much smaller than the distance between the object and the camera. Therefore the scenes taken by a pan-tilt camera or a tele-photo lens can be approximated by a 2D parametric transformation such as affine or projective transformations [Faugeras93][Hartley00].

Therefore, to align two images, we recover the 2D affine or 2D projective transformation (see Figure 3). For closely related views or far-distant scenes, we use affine parameters to approximate the transformation. But for images corresponding to sparse views separated by a significant camera rotation or closely-taken scenes, we use projective parameters. The following represents the affine and projective transformations between two images coordinates, one denoted as (*x, y*) and the other denoted as (*x', y'*).

$$\textit{Affine}$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

$$x' = ax + by + t_x, y' = cx + dy + t_y$$

$$\Pr ojective$$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} \sim \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

$$x' = \frac{(p_{11}x + p_{12}y + p_{13})}{(p_{31}x + p_{32}y + p_{33})}, y' = \frac{(p_{21}x + p_{22}y + p_{23})}{(p_{31}x + p_{32}y + p_{33})}$$

Figure 3**.** 2D Affine and Projective Transformation

2.2 Pair-wise Registration

Pair-wise registration starts from creating multi-resolution images for each frame. The estimation of projective parameters is highly dependent on the initial estimates. Unfortunately, an arbitrary initialization tends to lead to a local minimum [Szeliski94]. Such limitations can be overcome by estimating the initial parameters by using global matching in frequency domain [Chen94] [Reddy96][Davis98]. However the search for the parameters in the frequency domain is computationally expensive. The computation cost can be drastically reduced using a hierarchical approach. This coarse to fine registration technique is based on a several level pyramid representation of the frames that allows us to recover consecutively the translation of affine and projective parameters. This scheme allows us to refine the parameter estimation by matching feature points at different resolutions with an increasingly complex model. In the last stage of the parameter estimation, we improve the stability by using RANSAC.

3. Dominant Motion Matching

Unless good initial estimates for the parameters are given, the 8 projective parameters recovery method tends to suffer from local minimum problem while minimizing the errors. To select good initial parameters, we recover initial parameters in the frequency domain. The Fast Fourier Transform (FFT) method differs from other registration strategies because it searches for the optimal match according to information in the frequency domain.

Let I_1 and I_2 are the two input frames that differ only by a displacement (x_0,y_0).

$$I_2(x, y) = I_1(x - x_0, y - y_0)$$

The corresponding Fourier transform F_1 and F_2 will be related by

$$F_2(\xi,\eta) = e^{-j2\pi(\xi x0+\eta y0)} * F_1(\xi,\eta)$$

The cross-power spectrum of two frames *I* and *I'* with Fourier transforms *F* and *F'* is defined as

$$\frac{F(\xi,\eta)F'^*(\xi,\eta)}{|F(\xi,\eta)F'(\xi,\eta)|} = e^{j2\pi(\xi x0+\eta y0)}$$

where *F** is the complex conjugate of *F*.

The translation property of Fourier transform (Fourier shift theorem) guarantees that the phase of the cross-power spectrum is equivalent to the phase difference between the images. By taking the inverse Fourier transform of the representation in the frequency domain, we will have an impulse function; that is, it is approximately zero everywhere except at the displacement that is needed to optimally register the two frames.

In polar coordinates, the rotation between two images appears as translation. In the same way, the scale between two images appears as translation between two images in log-polar coordinate. Therefore, we get the scale and rotation parameters by converting the inputs to different coordinates and computing cross-power spectrums. More precisely, the input images are converted into log-polar coordinates. If there is a significant scale change, the system rectifies the images with respect to the scale. And then, the rectified images are converted into polar coordinates to recover the rotation parameters. This approach gives a stable initial approximation of each parameter for frames overlapping by at least 50% with each other.

4. Feature Matching

Recovering the parameters of the transformation is performed by minimizing the gray level error using Least Squares measure such as:

$$E = \sum \left(I_j(x, y) - I_j(M_{ij}(x', y'))\right)^2$$

where M_{ij} is affine or projective transformation. *(x, y)* and *(x',y')* are corresponding points.

To measure the error, we use a feature-based method [Cohen99]. Feature-based methods are relatively vulnerable to noise or moving objects. Also, if the features are not well distributed over the frame, the error measurement misleads the parameter estimation. The drawbacks of a feature-based approach are reduced by updating (adding and rejecting) features in each iteration of the parameter estimation and enforcing the distribution of selected features to be relatively uniform over the image. This feature based method enables the parameters to be recovered very fast compared to the direct method. For this reason, we use only features

Features can be defined as corners [Zoghlami97], high curvature points or lines and so on. In this proposal, the features are extracted by the Harris corner detector. The Harris corner detector computes the locally averaged auto-correlation matrix derived from the image gradients, and then computes the eigenvalues of the auto-correlation moment matrix to compute a corner ``strength'', minimum values of which indicate the corner positions.

After feature point extraction from two images, we select initial correspondences. The measure of correspondences is defined by either cross correlation (CC) or sum of squared differences (SSD). This correspondence scheme can be used in different levels of parameter estimation. In general, computing CC takes longer time than computing SSD. In our approach, the CC method is performed in the coarse level of resolutions with a large window size, and SSD is computed in the finest level of resolution with a small size of inputs.

$$CC(f,g) = \frac{\sum((f-\bar{f})(g-\bar{g}))}{\|f-\bar{f}\|\|g-\bar{g}\|}$$

$$\mathrm{SSD} = \sum_i (f - f_i)^2$$

5. Coarse-To-Fine Refinement

Although the registration in frequency domain can give good starting points for parameters, it is computationally expensive to convert each pair of input images to frequency domain and register them. To support efficient processing, we use hierarchical approach that performs registration and refines the parameters from the coarse-to-fine image resolution.

As the first step, we create Gaussian pyramids for each input frame [Burt83]. Then, the initial parameters are estimated in the coarsest resolution. Since the coarsest Gaussian image can improve resistance to noise, the computation in frequency domain for the initial estimation is more robust by removing noise that creates the high-frequency in the frequency domain, and the computation cost is smaller. The rotation or scale parameters may not be recovered in the coarsest level because the polar coordinate and log-polar coordinates represented in a small image resolution, which means too much quantization is applied to represent angle and magnification. But in the presence of large rotation or scale changes between consecutive scenes, rough initial estimations are still useful. Once we have initial parameters and the pyramid of image frames, we propagate features, increase translation parameters with respect to the finer image resolution, and refine the parameters.

6. Parameter Estimation

By using the initial translation parameters and correspondences, we minimize the error:

$$E = \sum \left(I_j(x,y) - I_j(M_{ij}(x',y')) \right)^2$$

Let us denote (x,y) and (x',y') two corresponding feature points. We obtain the following equations.

Affine

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ t_x \\ c \\ d \\ t_y \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

Projective

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -xx' & -yx' \\ 0 & 0 & 0 & x & y & 1 & -xy' & -yy' \end{pmatrix} \begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{31} \\ p_{32} \end{pmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

We use the linear least squares method to compute the parameters. The parameters obtained in coarser resolution are propagated to the finer resolution by changing the translation parameters with respect to the scale factor. Based on the new parameters, we adjust the locations of feature points and target points and repeat the error minimization steps.

In projective transformation, the correspondence is given by $x'=X/W$ and $y'=Y/W$ where

$$X = (p_{11}x+p_{12}y+p_{13})$$
$$Y = (p_{21}x+p_{22}y+p_{23})$$
$$W = (p_{31}x+p_{31}y+1)$$

Note that in the projective transformation case, this Linear Least Square method minimizes the error term

$$\varepsilon = (X,Y) - W(x,y)$$

But we wish to minimize

$$\varepsilon = (X/W, Y/W) - (x,y) = (x',y') - (x,y)$$

If the equation had been weighted by the factor $1/W$, the resulting error would have been what we want to minimize. Since W is dependent on (x,y), we cannot use a fixed weight, W, in the equation until we solve the equation. Therefore, we proceed iteratively to adapt W [Hartley97].

Let's denote the weight in the first step as W_0. In the next step, we can compute W_1 by finding P_{31} and P_{32}. We repeat this process at each step by multiplying the equation by $1/Wi$. If the number of repeated steps is n, the error measure by this process will be

$$\varepsilon = (Xn/Wn, Yn/Wn) - (x,y) = (x',y') - (x,y)$$

It approximates the error that we want to minimize. As an experiment, we tested the same four correspondences between images to recover parameters with and without this process. With this iterative refinement errors are reduced visibly as seen Figure 4. Without iterative refinement, the registration result shows misalignments on the table and large distortion on the right hand side.

(a)

(b)

Figure 4 Registration results (a) Without iterative refinement (b) With iterative refinement

7. RANSAC

Traditional least square algorithms use all the data to obtain the desired parameters. If there are a significant number of outliers, the estimated parameters are nowhere close to the real parameters. To prevent this situation, we use the RANdom SAmple Consensus (RANSAC) procedure in the parameter estimation stage. RANSAC is different in that it attempts to eliminate the invalid matches.

As started by Fisher and Bolles [Fisher81], RANSAC uses as small an initial data set as possible and enlarges this set with consistent data when possible. It partitions the data set into inliers and outliers based on a distance threshold, t. In this proposal, we use the symmetric transfer error, d, as the error metric.

$$d = \sum_k [d((x,y)_k, M_{ji}(x',y')_k) + d(M_{ij}(x,y)_k, (x',y')_k)]$$

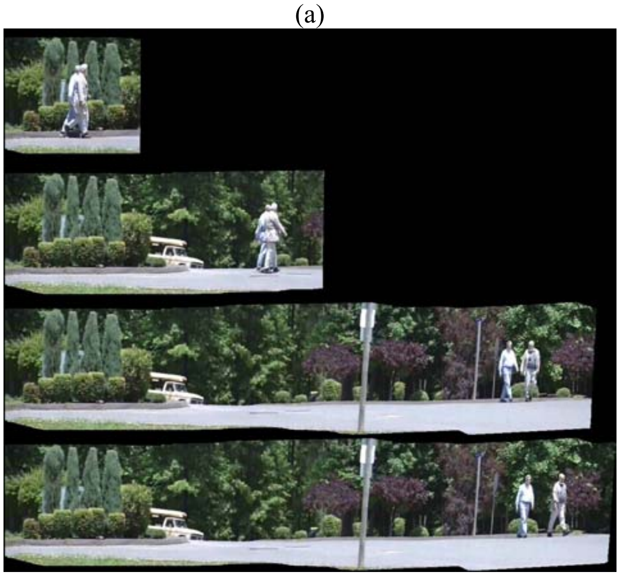
The idea is following: 3 or 4 points are selected randomly to estimate the affine or projective transformation. Then, the support for this transformation is measured by the number of points that match the transformation within the distance threshold, t. The random selection (called a sample) is repeated until the number of selection reaches the preset maximum iteration number or adaptively determined number. After trials, the largest consensus set is selected to estimate transformation parameters. As described, RANSAC performance is dependent upon the selection of points and the number of samples. In our work, for a sample in each iteration, we use bucket-based selection to enforce points to be evenly distributed over the entire image [Lacey00]. Also, the chosen sample is normalized before the estimation in order to increase stability of sample data [Hartley00][Zhang95].

8. Results

Video surveillance is one major application area that uses camera stabilization technique. Video surveillance performs the detection and analysis on targeted location and objects. However, when the camera moves, it is hard to perform surveillance on the target objects, for example for video sequences taken by a camera mounted on a moving vehicle, by a hand-held camera or by an unstable fixed camera. In Figure 5, we show that our method is a perfect solution for this kind of tasks.

As another application, we consider sports sequence. Sports sequences are usually taken by a pan-tilt-zoom camera whose motion can be perfectly stabilized by the 2D image registration. The challenging issue in sports sequences is to robustly remove mismatches caused by moving objects. To achieve this, we used RANSAC method in the parameter estimation stage. Figure 6 shows a sport sequence examples.

(a)



(b)

Figure 5 Registration Results (a) Three Input frames from 486 frames taken by a hand-held camera (b) Stabilized Sequences

(a)

(b)

Figure 6 Basketball sequence (a) Two frames from 99 input frames (b) Two Stabilized Sequences

9. Conclusions

In this report, we described a fast and robust pair-wise image registration method. Pair-wise image registration estimates transformation parameters for each pair of consecutive frames, and the transformation between arbitrary frames is computed by the concatenation of the pair-wise transformations. Pair-wise registration is based on 1) hierarchical parameter estimation and refinement 2) feature-matching 3) FFT(Fast Fourier Transformation)-based global matching 4) RANSAC-based parameter estimation. Our method registers 7~9 pairs of frames per second on 1GHZ Pentium III.

10. References

[Adelson91] E. Adelson, Layered Representation for Image Coding, *Technical Report 181, MIT Media Lab. Vision and Modeling Group*, December 1991.

[Ben-Ezra98] M. Ben-Ezra, S. Peleg and M. Werman, Efficient Computation of the Most Probable Motion from Fuzzy Correspondences, *IEEE Workshop on Applications of Computer Vision (WACV)*, 1998.

[Benedetti98] A. Benedetti and P. Perona, Real-time 2-D Feature Detection on a Reconfigurable Computer, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[Burt83] P. Burt and E. Adelson, The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. on Communication*, Vol. 31, No. 4, 1983.

[Chen94] Q. Chen, M. Defrise and F. Deconinck, Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transformation for Image Registration and Recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 16, No. 12, 1994.

[Cohen99] I. Cohen and G. Medioni, Detecting and Tracking Objects in Video Surveillance, *Proc. of the IEEE Computer Vision and Pattern Recognition (CVPR)*, Fort Collins, June 1999.

[Davis98] J. Davis, Mosaics of Scenes with Moving Objects, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1998.

[Faugeras93] O. Faugeras. Three Dimensional Computer Vision: A Geometric Viewpoint. *MIT Press*, 1993.

[Fisher81] M. A. Fisher and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. Assoc. Comp*. Vol. 24, No. 6, pp. 381-395, 1981.

[Hartley97] R. Hartley and P.Strum, Triangulation, *Computer Vision and Image Understanding (CVIU),* 1997.

[Hartley00] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, *Cambridge University Press*, 2000.

[Irani93] M. Irani and S. Peleg, Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency, *Journal of Visual Communication and Image Representation (VCIR)*, Vol. 4, No. 4, pp. 324-335, December 1993.

[Irani95] M. Irani, P. Anandan and S. Hsu, Mosaic Based Representation of Video Sequences and Their Applications, *IEEE International Conference on Computer Vision (ICCV)*, 1995.

[Irani96] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, Efficient Representations of Video Sequences and Their Applications. *Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application*, Vol. 8, No. 4, May 1996.

[Irani98] M. Irani, P. Anandan and D. Weinshall, From Reference Frames to Reference Planes: Multi-View Parallax Geometry and Applications, *European Conference on Computer Vision (ECCV)*, June, 1998.

[Kang00] E.Y. Kang, I. Cohen and G. Medioni, A Graph-based Global Registration for 2D Image Mosaic, *International Conference on Pattern Recognition (ICPR)*, 2000.

[Kumar95] R. Kumar, P. Anandan, M. Irani, J. Bergen and K. Hanna, Representation of Scenes from Collection of Images, *IEEE Workshop on Representations of Visual Scenes (WRVS)*, June 1995.

[Lacey00] A. Lacey, N. Pinitkarn and N. Thacker, An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration, *The British Machine Vision Conference (BMVC)*, 2000.

[McMillan95] L. McMillan and G. Bishop, Plenoptic Modeling: An Image-Based Rendering System, *Proceedings of ACM SIGGRAPH*, 1995.

[Morimoto97] C. Morimoto, R. Chellappa and S. Balakirsky, Fast Image Stabilization and Mosaicking, *DARPA Image Understanding Workshop (IUW),* New Orleans, May, 1997.

[Morimoto98] C. Morimoto and R. Chellapa, Evaluation of Image Stabilization Algorithms, *Proceedings of IEEE ICASSP*, May, 1998.

[Peleg97] S. Peleg and J. Herman, Panoramic Mosaics by Manifold Projection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* 1997.

[Reddy96] B. S. Reddy and B. N. Chatterji, An FFT-based Technique for Translation, Rotation and Scale-Invariant Image Registration, *IEEE Trans. on Image Processing (IP)*, Vol. 5, No. 8, 1996.

[Rousso97] B. Rousso, S. Peleg and I. Finci, Mosaicing with Generalized Strips, *DARPA Image Understanding Workshop (IUW)*, New Orleans, May, 1997.

[Rousso98] B. Rousso, S. Peleg, I. Finci and A. Rav-Acha, Universal Mosaicing Using Pipe Projection, *IEEE International Conference on Computer Vision (ICCV)*, 1998

[Sawheny95] H. Sawhney. S. Ayer and M. Gorkani, Model-based 2D&3D Dominant Motion Estimation for Mosaicing and Video Representation, *IEEE International Conference on Computer Vision (ICCV)*, 1995.

[Sawheny97] H. Sawhney and R. Kumar, True Multi-Image Alignment and its Application to Mosaicing and Lens Distortion Correction, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997.

[Sawheny98a] H. Sawhney, S. Hsu and R. Kumar, Robust Video Mosaicing through Topology Inference and Local to Global Alignment, *European Conference on Computer Vision (ECCV)*, 1998.

[Sawheny98b] H. Sawhney, R. Kumar, G. Gendel, J. Bergen, D. Dixon and V. Paragano, Video Brush: Experiences with Consumer Video Mosaicing, *IEEE Workshop on Applications of Computer Vision (WACV)*, 1998.

[Sawheny99] H. Sawhney and R. Kumar, True Multi-Image Alignment and Its Application to Mosaicing and Lens Distortion Correction, *IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 21, No. 3, 1999.

[Szeliski94] R. Szeliski, Image Mosaicing for Tele-Reality Applications, *IEEE Workshop on Applications of Computer Vision (WACV),* 1994.

[Szeliski97] R. Szeliski and H. Shum, Creating Full View Panoramic Image Mosaics and Environment Maps, *Proceedings of ACM SIGGRAPH*, pp. 251-258, 1997.

[Wren97] C. Wren, A. Azarbayejani, T. Darrel and A. Pentland, Pfinder: Real-time tracking of the human body *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 19, No. 7, pp. 780-785, 1997.

[Zoghlami97] I. Zoghlami, O. Faugeras and R. Deriche, Using Geometric Corners to Build a 2D Mosaic from a Set of Images, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1997.