Real-time Video Mosaicing with Adaptive Parameterized Warping

Xiangtian Dai Le Lu Gregory D. Hager

CIRL Lab, Computer Science Department, Johns Hopkins University, MD 21218, USA

Abstract

This paper briefly describes a system for real-time video mosaic construction using intensity-based registration. One novel aspects of this algorithm is a dynamic selection of warping models, including similarly transforms, affine models, projective models, and quadratic models based on the complexity of the incoming imagery. A second result is a new method for subsampling the images based on the sensitivity of image pixels in template matching. By exploiting the latter, it is possible to greatly accelerate the calculation. Some real image sequence mosaicing results for different parametric models are shown in this paper. These are all processed in real-time.

1 Introduction

Image registration or alignment for video mosaicing has many research and real applications. Our motivations come primarily from the medical field, and primarily seek to over come fundamental field of view and resolution tradeoffs that occur ubiquitously in endoscopic surgery.

There are two general approaches to computing the visual motions between successive images, the critical issue for registration. Direct approaches [3, 4, 6] use all the image pixels available to compute an image-based error, which is then optimized. Complementary approaches [1, 2] are to specifically detect certain image features first then to estimate the corresponding relations between image feature pairs in different camera views. The latter often has the advantage of a larger range of convergence, although at the cost of a prior feature-detection and correspondence stage.

Another dimension to registration is the type of deformations allowed in the image. Rigid, similarity, or affine motion models are often used. In [4], full 6 DOF face tracking are considered, thanks to a cylinder head model; the motion vectors are obtained by local disturbances of parameters. Joint View Triangulation (JVT) is a non-parametric representation proposed in [5] to interpolate the inter-frame visual motions of unmatched image parts by texture mapping from its semi-dense matched image feature neighbors. In [3] relative depth is used to create a full 3D tracking system.

We define our problem as building an image mosaic online from a sequence of video images of a geometrically constrained (planar or quadratic) static surface from a freely moving camera. The surface can be an approximation of a scene which has insignificant parallax by camera motion. We have constructed a hierarchical and adaptive framework for real-time video registration/mosaicing by extending from 4 parameters without out-of-plane rotation, to 6 or 8 parameter full 3D motion or even 12 parameter quadratic motion models. This adaptive warping model enables us to deal with more complex surfaces and more general camera motions. A second contribution of this paper is that we propose a simple but efficient rule to select part of image pixels that make the most significant contribution of optimization in Jacobian matrix.

2 Adaptive parameterized warping for video registration

The basic building block of the mosaicing process is the registration of image pairs. It consists of several steps, each refining the alignment result of the previous one by adding more parameters into the warping matrix. The number of steps is adaptive based on the warped image difference, or residue. If the residue is not sufficiently reduced with a low order warping model, higher order warping is activated for further compensation.

The steps of the algorithm are as follows. First of all, cross correlation is used for estimation of 2D translation d_x and d_y in the image coordinates.

After this initial registration, we optimize a warping function of the general form:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = W \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{1}$$

where (u, v) is the pixel coordinates of a physical point projected in the first image, and (x, y) is the pixel coordinates of its projection in the second image. λ is a scaling factor and W is an appropriately parameterized 3×3 matrix warping.

In the case of similarlity transformation, we only consider scale s and in-plane rotation θ in addition to d_x and d_y, yielding a 4 parameter warping matrix of the form:

$$W_a = \begin{pmatrix} s\cos\theta & -s\sin\theta & d_x \\ s\sin\theta & s\cos\theta & d_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

1063-6919/03 $17.00 © 2003 IEEE

More general affine warping can be approximated as:

$$W_p = \begin{pmatrix} s\cos\theta & -s\sin\theta & d_x \\ s\sin\theta & s\cos\theta & d_y \\ \alpha & \beta & 1 \end{pmatrix} \quad (3)$$

with two extra parameters to represent the image shearing deformation caused by out-of-plane rotations. We found that W_p is a good approximation of the standard homography transformation (8 parameters with 2 constraints) of W and often leads to more stable results.

For quadratic surfaces, we approximate the warping:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} w_{11}w_{12}w_{13}w_{14}w_{15}w_{16} \\ w_{21}w_{22}w_{23}w_{24}w_{25}w_{26} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \\ xy \\ x^2 \\ y^2 \end{pmatrix} \quad (4)$$

Optimization proceeds using standard gradient-descent techniques (robust IRLS is a straightforward extension [3]). Let p be the vector of warping parameters and let superscripts index the image points. Then we can write the registration error as

$$e = (I_1(U^{(1)}(p)) - I_2(X^{(1)}), \ldots, I_1(U^{(n)}(p)) - I_2(X^{(n)}))^T$$

The goal of minimizing $||e||$ can be achieved by $p^* = p - (J^TJ)^{-1}J^Te$ given an initial close estimate of p, where $J = (J^{(1)}, \ldots, J^{(n)})^T$ and

$$J^{(i)} = \frac{\partial I_1}{\partial p} = \frac{\partial I_1}{\partial U^{(i)}} \frac{\partial U^{(i)}}{\partial p} \quad (5)$$

The $\frac{\partial I_1}{\partial U}$ part of the Jacobian matrix J is the image gradient, and the $\frac{\partial U}{\partial p}$ part can be derived from W. In the 4 parameter case, for example,

$$\frac{\partial U}{\partial p}|_{d_x=0,d_y=0,s=1,\theta=0} = \begin{pmatrix} 1 & 0 & x & -y \\ 0 & 1 & y & x \end{pmatrix} \quad (6)$$

For 6 or 12 parameters, we can get Jacobian similarly.

In this form, image-based direct registration methods are usually computationally costly due to the large number of pixels involved; due to the changing imagery, the computation cannot be performed offline. Therefore, we try to only use a portion of the pixels to achieve acceptable registration error. Our is to minimize sum of squares of each J^TJ's element's change caused by not using a pixel. Observe that

$$J^TJ = \sum_{i=1}^{n} {J^{(i)}}^T J^{(i)} \quad (7)$$

and since $J^{(i)}$ is a row vector,

$$\sum_k \sum_l ({J^{(i)}}^T J^{(i)})^2_{k,l} = \sum_k \sum_l J^{(i)}_k J^{(i)}_l = (\sum_k J^{(i)}_k)^2 \quad (8)$$

So those $J^{(i)}$s that have the smallest magnitudes will be discarded.

For every incoming image frame, it is registered with the last frame, which has known warping parameters, then it is registered with the current global mosaic image to refine the result. Finally it is warped and added to the mosaic by a proper weight.

3 Experiments

One of the attached video sequence shows the image mosaicing process of a projected eye retina image. Quadratic warping is not used. The resolution of each image patch is 160 by 120. The other video sequence shows the image mosaicing of an indoor scene. The resolution of each image patch is 320 by 240. The experiments are carried on a Pentium4 1.7GHz computer with 5-10 Hz rate.

4 Summary and future work

Our work demonstrated a viable method for real-time image mosaicing by directly minimizing SSD. The method is suitable for a planar or quadratic surface with a free moving camera, or any scene with a rotation-only camera motion.

Future work will include a method to detect accumulated errors and dynamically re-distributed them, a better warping model for quadratic surface and other parametric or non-parametric constrained surfaces, and a layered or masked model when motion parallax is significant.

References

[1] A. Can, C.V. Stewart, B. Roysam, and H.L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina, *IEEE Trans. on PAMI*, **24:3** pp. 347-364, Mar. 2002.

[2] F. Dellaert and R. Collins Fast Image-Based Tracking by Selective Pixel Integration, ICCV 99 Workshop on Frame-Rate Vision, Sep. 1999.

[3] G. Hager and P. Belhumeur. Efficient Region Tracking With Parametric Models of Geometry and Illumination, *IEEE Trans. on PAMI*, **20:10**, pp. 1125-1139, 1998.

[4] M. La Cascia, S. Sclaroff, and V. Athitsos. Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Robust Registration of Texture-Mapped 3D Models. *IEEE Trans. PAMI*, **22:4**, Apr. 2000.

[5] Maxime Lhuillier and Long Quan. Image Interpolation by Joint View Triangulation, *CVPR'99*, pp. 139-145, Fort Collins, Colorado, USA, 1999.

[6] H.-Y. Shum and R. Szeliski. Construction of Panoramic Image Mosaics with Global and Local Alignment, *International Journal of Computer Vision* **36(2)**, pp. 101-130, 2000.