MULTIFRAME INTEGRATION VIA THE PROJECTIVE TRANSFORMATION WITH AUTOMATED BLOCK MATCHING FEATURE POINT SELECTION

Richard R. Schultz

Department of Electrical Engineering
University of North Dakota
Grand Forks, ND 58202-7165
rschultz@nyquist.ee.und.nodak.edu

Mark G. Alford

Rome Research Site
AFRL/IFEA
Rome, NY 13441-4114
alfordm@rl.af.mil

ABSTRACT

A subpixel-resolution image registration algorithm based on the nonlinear projective transformation model is proposed to account for camera translation, rotation, zoom, pan, and tilt. Typically, parameter estimation techniques for transformation models require the user to manually select feature points between the images undergoing registration. In this research, block matching is used to automatically select correlated feature point pairs between two images, and these features are used to calculate an iterative least squares solution for the projective transformation parameters. Since block matching is capable of estimating accurate translation motion vectors only in discontinuous edge regions, inaccurate feature point pairs are statistically eliminated prior to computing the least squares parameter estimate. Convergence of the projective transformation model estimation algorithm is generally achieved in several iterations. After subpixel-resolution image registration, a high-resolution video still may be computed by integrating the registered pixels from a short sequence of low-resolution image sequence frames.

1. INTRODUCTION

Many applications within the realm of sensor fusion require the accurate registration of multiple image channels. The challenge involves registering data acquired not only from the same sensor array after a camera transformation, but also registering images acquired by dissimilar sensors. In addition, subpixel accuracy may be necessary when the sensor arrays have different resolutions [8]. The goal is to perform multiframe integration [2][3][4][5] through the automated registration of image sequence frames. In this research, the image registration algorithm is based on the projective transformation [2][6][7], which takes into account camera translation, rotation, zoom, pan, and tilt. This transformation is accepted as the most accurate of the camera models. Since it is a nonlinear model, direct least squares estimation is not capable of estimating the parameters properly from the data sets. Typically, common feature points that appear within the images undergoing registration must be selected manually, and these features are used in the parameter estimation algorithm. In a directly related application to [7], the block matching algorithm [1] is used to automatically select correlated feature point pairs between two images, and these features are utilized in an iterative algorithm to estimate the nonlinear projective transformation parameters. Highly accurate subpixel-resolution registrations are possible using this technique. High-resolution video stills may be computed by integrating the pixels from a short sequence of registered low-resolution video frames. Simulations that perform enhancement on a short sequence of under-sampled intensity video frames show significant dealiasing when compared to the low-resolution reference frame.

This paper is organized as follows. In Section 2, the projective transformation model is introduced, and the automated parameter estimation algorithm is proposed. Section 3 describes simulations that verify the registration accuracy using the projective transformation, as well as experiments in multiframe integration. A brief summary along with future research directions is provided in Section 4.

This work was supported in part by the Air Force Office of Scientific Research Summer Faculty Research Program; the National Science Foundation Faculty Early Career Development (CAREER) Program, grant number MIP-9624849; and the U.S. Army Research Office under contract number DAAH04-96-1-0449.

2. AUTOMATED IMAGE REGISTRATION

This section discusses image registration in the context of transformation model parameter estimation, and presents the automated projective transformation estimation algorithm.

2.1 Image Registration Notation

Denote the point

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{1}$$

as a spatial location within image $\mathbf{y}^{(l)}$, and the corresponding point within the reference image $\mathbf{y}^{(k)}$ as

$$\mathbf{x}' = \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}. \tag{2}$$

These two spatial locations represent a pixel value that is correlated between the images; *i.e.*, the pixel at location $\mathbf{x}$ in $\mathbf{y}^{(l)}$ undergoes a transformation to position $\mathbf{x}'$ in $\mathbf{y}^{(k)}$. By estimating the parameters of this transformation and then warping image $\mathbf{y}^{(l)}$ accordingly, the pixels should be properly registered.

2.2 Projective Transformation Model

To accommodate for camera displacement, rotation, zoom, pan, and tilt, the eight-parameter projective model [2][6][7],

$$\mathbf{x}' = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^t \mathbf{x} + 1}, \tag{3}$$

has been selected for the registration algorithm. The numerator of the projective model is the linear six-parameter affine model, which accounts for translation, rotation, and zoom, while perspective transformations are handled in the denominator term. Explicitly, the rotation matrix, **A**, the displacement vector, **b**, and the perspective vector, **c**, may be expressed as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}. \tag{4}$$

Unfortunately, the nonlinearity of the model makes the direct estimation of these eight parameters difficult.

To estimate the projective transformation parameters, an iterative least squares approach will be taken. The least squares problem statement is defined as

$$\hat{\mathbf{A}}, \hat{\mathbf{b}}, \hat{\mathbf{c}} = \arg\min_{\mathbf{A},\mathbf{b},\mathbf{c}} \sum_{\mathbf{x}} \left\| \mathbf{x}' \mathbf{c}^t \mathbf{x} + \mathbf{x}' - \mathbf{A}\mathbf{x} - \mathbf{b} \right\|^2. \tag{5}$$

Block matching will be used to automatically select feature points $\mathbf{x}$ in $\mathbf{y}^{(l)}$ and their transformations $\mathbf{x}' = \mathbf{x} + \hat{\mathbf{b}}(\mathbf{x})$ in $\mathbf{y}^{(k)}$ during each iteration. The block matching estimate using the mean squared error (MSE) criterion [1] is defined as

$$\hat{\mathbf{b}}(\mathbf{x}) = \arg\min_{b_1,b_2} \frac{1}{(2p+1)^2} \sum_{m=x_1-p}^{x_1+p} \sum_{n=x_2-p}^{x_2+p} \left(y^{(l)}_{m,n} - y^{(k)}_{m+b_1,n+b_2} \right)^2 \tag{6}$$

for a $(2p + 1)$ x $(2p + 1)$ pixel block. During the i^{th} iteration of the algorithm, a least squares estimate of the transformation parameters, $(\mathbf{A}_i, \mathbf{b}_i, \mathbf{c}_i)$, is computed from a set of N feature points and their transformations, $(\mathbf{x}_k, \mathbf{x}'_k)$ for k=1,…,N, as image $\mathbf{y}^{(l)}$ is iteratively warped towards $\mathbf{y}^{(k)}$.

To select the feature points manually, the user must identify at least N=4 feature points $\mathbf{x}_k$ in $\mathbf{y}^{(l)}$ which correspond to pertinent objects within the scene [2]. These points should be spaced relatively far apart, and they must not lie on the same line. For every feature point $\mathbf{x}_k$ selected in $\mathbf{y}^{(l)}$, the corresponding transformed point $\mathbf{x}'_k$ must be located within $\mathbf{y}^{(k)}$. These feature point pairs, $(\mathbf{x}_k, \mathbf{x}'_k)$ for k=1,…,N, are then used to estimate the transformation parameters. The feature points may be selected automatically by sparsely sampling the data in a regular pattern and using block matching to estimate the transformed positions, with the knowledge that many of the feature points located in smooth regions will yield inaccurate feature point pairs [7]. By calculating a least squares solution for the projective transformation using only the "best" feature point pairs, the registration parameters can be estimated in most cases both automatically and efficiently.

Projective Transformation Estimation Algorithm:

1. Set $\mathbf{A}_0$=**I**, $\mathbf{b}_0$=**0**, $\mathbf{c}_0$=**0**, $\mathbf{y}_1^{(l)} = \mathbf{y}^{(l)}$, and iteration number i=1.

2. Select N feature points, $\mathbf{x}_k$ for k=1,…,N, by sparsely sampling a region within image $\mathbf{y}_i^{(l)}$ which contains a large number of edges. Every fourth point may be selected both horizontally and vertically within a spatially active region to achieve acceptable results.

3. Estimate block matching motion vectors at each of the N selected feature points. Denote the k^{th} transformed point as
$$\mathbf{x}'_k = \mathbf{x}_k + \hat{\mathbf{b}}(\mathbf{x}_k).$$

4. Estimate the projective transformation parameters, $(\mathbf{A}_i, \mathbf{b}_i, \mathbf{c}_i)$, using all N block matching feature point pairs, $(\mathbf{x}_k, \mathbf{x}'_k)$ for k=1,…,N, to calculate the least squares solution to the following problem statement:
$$\mathbf{A}_i, \mathbf{b}_i, \mathbf{c}_i = \arg\min_{\mathbf{A},\mathbf{b},\mathbf{c}} \sum_{k=1}^{N} \left\| \mathbf{x}'_k \mathbf{c}^t \mathbf{x}_k + \mathbf{x}'_k - \mathbf{A}\mathbf{x}_k - \mathbf{b} \right\|^2$$

5. Since block matching vectors located in smooth image regions and areas of object occlusion will be inaccurate, the M least accurate feature point pairs will be statistically eliminated from the parameter estimation problem. Denote the residual of the k^{th} transformed feature point, $\mathbf{x}'_k$, and the currently estimated projective transformation as
$$\mathbf{r}_k = \begin{bmatrix} r_{k_1} \\ r_{k_2} \end{bmatrix} = \begin{bmatrix} x'_{k_1} - (a_{i_{11}} x_{k_1} + a_{i_{12}} x_{k_2} + b_{i_1})/(\mathbf{c}_i^t \mathbf{x}_k + 1) \\ x'_{k_2} - (a_{i_{21}} x_{k_1} + a_{i_{22}} x_{k_2} + b_{i_2})/(\mathbf{c}_i^t \mathbf{x}_k + 1) \end{bmatrix}.$$
Calculate the sample mean, expressed as
$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{r}_k,$$
and the following residual sample variances:
$$\sigma_1^2 = \frac{1}{N-1} \sum_{k=1}^{N} (r_{k_1} - \mu_1)^2 \qquad \sigma_2^2 = \frac{1}{N-1} \sum_{k=1}^{N} (r_{k_2} - \mu_2)^2$$
If $|r_{k_1}| > \sigma_1$ or $|r_{k_2}| > \sigma_2$, eliminate the corresponding feature point pair $(\mathbf{x}_k, \mathbf{x}'_k)$ from the least squares problem.

6. Re-estimate the least squares projective transformation parameters, $(\mathbf{A}_i, \mathbf{b}_i, \mathbf{c}_i)$, using the $(N - M)$ most accurate block matching feature point pairs:
$$\mathbf{A}_i, \mathbf{b}_i, \mathbf{c}_i = \arg\min_{\mathbf{A},\mathbf{b},\mathbf{c}} \sum_{k=1}^{N-M} \left\| \mathbf{x}'_k \mathbf{c}^t \mathbf{x}_k + \mathbf{x}'_k - \mathbf{A}\mathbf{x}_k - \mathbf{b} \right\|^2$$

7. Warp the original image $\mathbf{y}^{(l)}$ by applying the overall projective transformation,
$$\mathbf{x}' = \frac{\left[\mathbf{A}_i \mathbf{A}_{i-1} + \mathbf{b}_i \mathbf{c}_{i-1}^t\right]\mathbf{x} + \left[\mathbf{b}_i + \mathbf{A}_i \mathbf{b}_{i-1}\right]}{\left[\mathbf{c}_i^t \mathbf{A}_{i-1} + \mathbf{c}_{i-1}^t\right]\mathbf{x} + 1},$$
to all pixels, and set $\mathbf{y}_{i+1}^{(l)}$ equal to the resulting image. This warped image is to be registered with the reference image $\mathbf{y}^{(k)}$ during the next iteration.

8. Calculate the change in the projective model parameters from iteration i-1 to i. If the change is small, the aggregate parameter estimates are given as follows:

$$\hat{\mathbf{A}} = \mathbf{A}_i \mathbf{A}_{i-1} + \mathbf{b}_i \mathbf{c}_{i-1}^t \qquad \hat{\mathbf{b}} = \mathbf{b}_i + \mathbf{A}_i \mathbf{b}_{i-1} \qquad \hat{\mathbf{c}} = \mathbf{A}_{i-1}^t \mathbf{c}_i + \mathbf{c}_{i-1}$$

Otherwise, set i=i+1, and return to Step 2. Convergence is generally achieved in two to three iterations.

For subpixel-resolution registration, the reference image, $\mathbf{y}^{(k)}$, and the image to be registered, $\mathbf{y}^{(l)}$, must first be up-sampled by a factor of q. These up-sampled images are used to estimate the 1/q-th pel resolution parameters. In the simulations, up-sampling is performed using cubic B-spline interpolation [4].

3. SIMULATIONS

Two frames from the *Film* image sequence were selected to show the efficacy of the automated registration algorithm. The image sequence is a film taken by a news crew from a helicopter, and it contains translational and rotational motion as well as a slight camera pan and tilt. Figure 1 shows the original high-resolution reference frame, and Figure 2 shows the results of integer- and subpixel-resolution registration. In the subpixel case, the original frames were down-sampled by a factor of 4. To show the quality of the registrations, the reference and the warped frames have been superimposed. Both the integer and subpixel data align extremely well, with the exception of the spire near the middle of the frames and the moving vehicles.

Multiframe enhancement involves the integration of a short sequence of low-resolution video frames to generate a high-resolution video still (HRVS) image [2][3][4][5]. By registering a set of low-resolution frames with respect to a reference image, the pixels should be aligned. However, the pixels may not be perfectly registered on the subpixel-resolution grid, and this subpixel overlap can be exploited to reduce aliasing artifacts. To perform multiframe enhancement, all candidate frames are first up-sampled by a factor of q using cubic B-spline interpolation. Using the interpolated data, subpixel-resolution projective transformation parameters are estimated for every frame with respect to the reference image. Each original low-resolution frame is then expanded by a factor of q using zero-order hold up-sampling, and these blocky images are warped using the estimated projective transformation parameters. Finally, a vector median filter is applied to the registered pixels at spatial location $\mathbf{x}$, with the filter output used as the value of the high-resolution image estimate at that point. Figure 3 depicts the results of the multiframe integration simulations on 3, 5, and 7 down-sampled frames from the *Film* image sequence. Visually, the estimate computed using 5 frames appears to be the closest to the original high-resolution reference image. An advantage of this multiframe enhancement technique over other super-resolution enhancement algorithms [3][4][5] is that the original video frames are not interpolated prior to their integration. Zero-order hold up-sampling simply replicates the pixels within a square block, and these large pixels are then warped using the estimated transformation parameters. No new image information is incorporated into the high-resolution video still estimate; the image sequence frames are utilized directly without alteration.

4. CONCLUSION

An automated image registration algorithm based on the projective transformation has been investigated which accounts for camera translation, rotation, zoom, pan, and tilt. Feature selection is performed by the block matching algorithm, in which the end-points of translation vectors serve as feature point pairs. Simulations were conducted to show the efficacy of the registration algorithm and the performance of multiframe integration applied to low-resolution video. Another application under investigation is image mosaicking, since the projective transformation may be used to match overlapping image regions acquired by a moving camera. Further research will also be conducted in dissimilar sensor image registration and fusion.

5. REFERENCES

[1] B. Furht, J. Greenberg, and R. Westwater, *Motion Estimation Algorithms for Video Compression.* Kluwer Academic Publishers, 1997.

[2] S. Mann and R. W. Picard, "Virtual Bellows: Constructing High Quality Stills from Video." In *Proc. IEEE Int. Conf. Image Processing*, (Austin, TX), pp. 363-367, 1994.

[3] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "High-Resolution Standards Conversion of Low-Resolution Video." In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Detroit, MI), pp. 2197-2200, May 1995.

[4] R. R. Schultz and R. L. Stevenson, "Extraction of High-Resolution Frames from Video Sequences," *IEEE Trans. Image Processing*, vol. 5, no. 6, pp. 996-1101, 1996.

[5] R. R. Schultz and R. L. Stevenson, "Improved Definition Video Frame Enhancement." In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, (Detroit, MI), pp. 2169-2172, May 1995.

[6] Y.-P. Tan, S. R. Kulkarni, and P. J. Ramadge, "A New Method for Camera Motion Parameter Estimation." In *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 406-409, October 1995.

[7] Y.-P. Tan, D. D. Saur, S. R. Kulkarni, and P. J. Ramadge, "Rapid Estimation of Camera Motion from Compressed Video With Application to Video Annotation," *IEEE Trans. Circuits Systems for Video Technology.* In press.

[8] Q. Tian and M. N. Huhns, "Algorithms for Subpixel Registration," *Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 220-233, 1986.

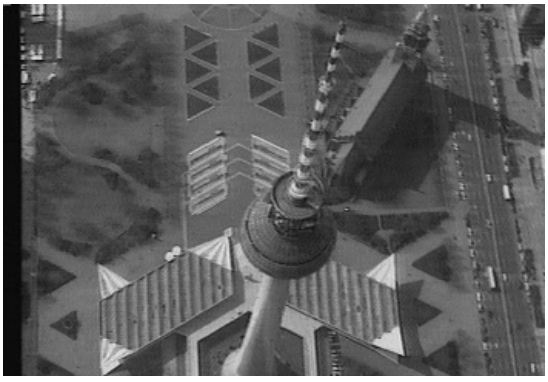

Figure 1. High-resolution *Film* reference frame.

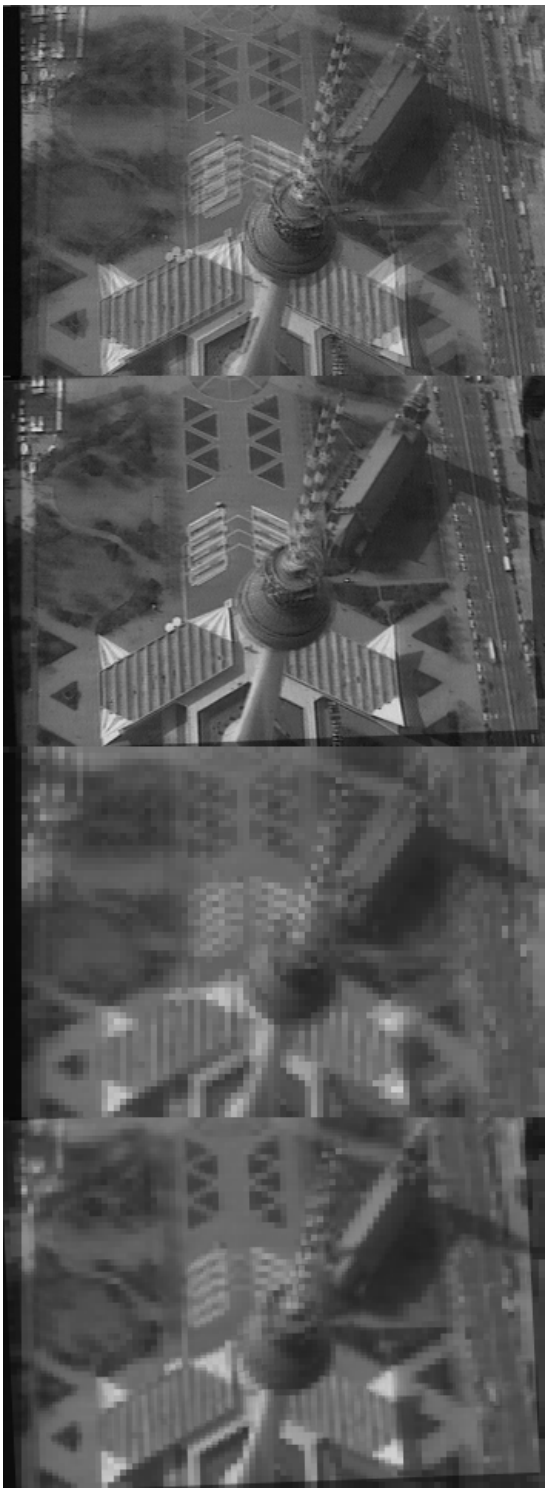

Figure 2. *Image 1:* Superimposed original integer-resolution frames. *Image 2:* Superimposed registered integer-resolution frames. *Image 3:* Superimposed original subpixel-resolution frames (1/4-th pel resolution). *Image 4:* Superimposed registered subpixel-resolution frames (1/4-th pel resolution).

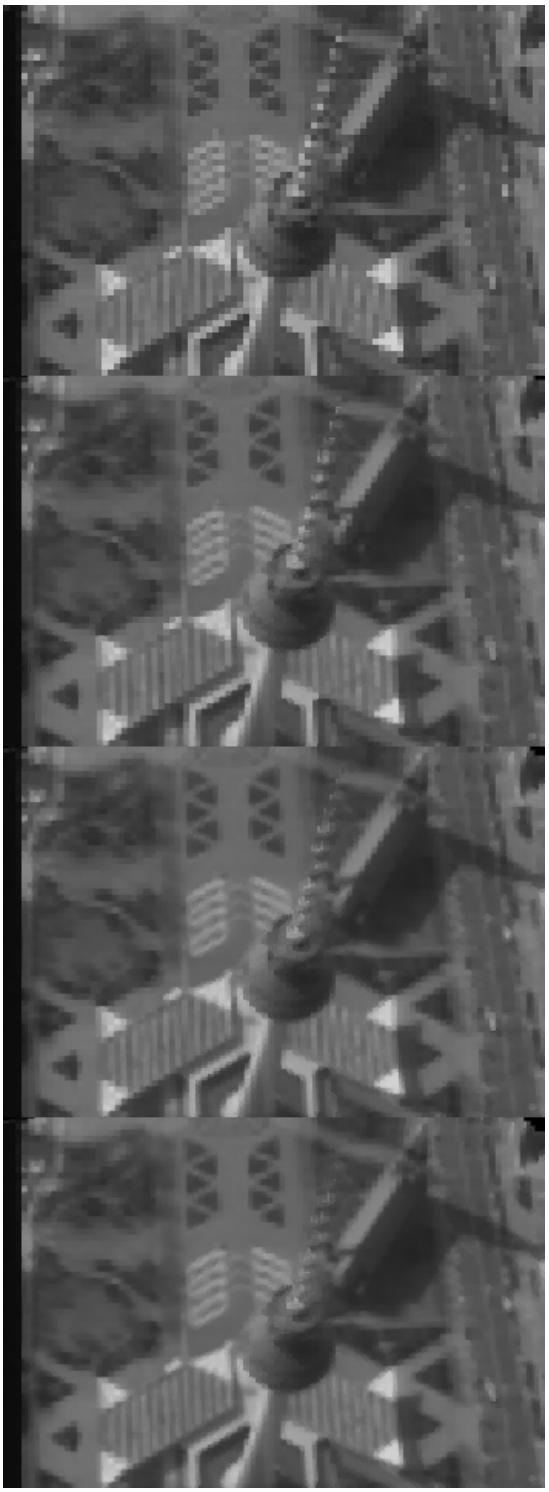

Figure 3. Multiframe integration of the *Film* image sequence. *Image 1:* Reference frame (PSNR=25.02 dB). *Image 2:* High-resolution video still computed using 3 frames (PSNR=25.49 dB). *Image 3:* High-resolution video still calculated using 5 frames (PSNR=25.52 dB). *Image 4:* High-resolution video still estimated using 7 frames (PSNR=25.43 dB).