ESTIMATING CORRESPONDENCE IN DIGITAL VIDEO

Richard Radke, Vitali Zagorodnov, Sanjeev Kulkarni, and Peter J. Ramadge

Department of Electrical Engineering
Princeton University
Princeton, NJ 08544
{rjradke, zvitali, kulkarni, ramadge}@ee.princeton.edu

1. INTRODUCTION

This paper addresses several estimation problems involving correspondence in digital video. We present three cases, in order of increasing complexity: affine transformations, projective transformations, and general correspondence. We review some recent research in the first two cases, and present a new framework for discussing the third case.

Many image and video processing problems hinge on possessing a dense (subpixel-level) estimate of correspondence between a set of still images. Generally the cameras which produced the images are uncalibrated, i.e., their locations and orientations are unknown. The classical correspondence problem is a fundamental and difficult problem in computer vision, as evidenced by more than 30 years of research. Notable approaches include feature-based parameter estimation [18, 23], interpolation of feature correspondences [4, 11, 20], optical flow [1, 3], layered motion [2, 9, 22], and correspondence along conjugate epipolar lines [5, 6, 10, 12, 21].

While in some applications, an unstructured optical flow field may be an adequate representation of correspondence between an image pair, there are many practical situations in which a parameterized or structured correspondence is induced by the geometry of the cameras. In addition to coupling the correspondence to a physical modeling assumption, parameterized correspondence is generally more consistent, easier to manipulate, and gives more information about the relationships between images.

Domains where parametric global motion models arise include orthographically projected images (6 affine parameters) or images taken by a rotating camera (8 projective parameters). In the first case, we compute the variance in estimation of affine parameters from noisy feature correspondences. We further discuss how multiple pairwise translation estimates can reduce the variance of the joint estimation of translation over an image sequence. In the second case, we demonstrate how the least-squares estimation of 8 projective parameters can be reduced analytically to a 2-parameter minimization that is computationally efficient and exposes new structure between the transformation parameters.

In general, the correspondence between an image pair has no simple global parameterization. In the difficult case of cameras separated by a non-negligible translation, traditional methods from stereo or optical flow are not applicable due to the large perspective difference between images. However, the set of correspondences which are physically realizable is not entirely unconstrained and has a structure described by *correspondence graphs*. We discuss how to use the formalism of correspondence graphs to ensure that any estimated correspondence is physically valid.

This research was partially supported by grants from the IBM Tokyo Research Laboratory and the New Jersey Center for Multimedia Research.

2. ESTIMATION OF AFFINE PARAMETERS

Affine transformations arise in the case of orthographically projected images or as approximations to nonlinear global transformations. The form of an affine transformation applied to a point $w = (x, y) \in \mathbb{R}^2$ is:

$$g_{A,b}(w) = Aw + b$$

where $A \in \mathbb{R}^{2\times 2}, b \in \mathbb{R}^2$. Affine transformations can be estimated from a noisy set of feature correspondences $\{w_j \mapsto w_j' \in \mathbb{R}^2, j = 1, \ldots, N\}$ by minimizing the least-squares functional

$$Q(A,b) = \tfrac{1}{2}\sum_{j=1}^{N} \left(w_j' - Aw_j + b\right)^T \left(w_j' - Aw_j + b\right) \quad (1)$$

The solution to this linear least squares problem is well known and is given by:

$$\hat{b} = \bar{w}' \quad (2)$$

$$\hat{A} = \Sigma_{ww'}(\Sigma_{ww})^{-1} \quad (3)$$

provided that we move the center of coordinates so that $\sum_{j=1}^{N} w_j = 0$. Here $\bar{w}' = \frac{1}{N}\sum_{j=1}^{N} w_j'$, $\Sigma_{ww'} = \sum_{j=1}^{N}(w_j' - \bar{w}')w_j^T$ and $\Sigma_{ww} = \sum_{j=1}^{N} w_j w_j^T$.

To evaluate the quality of the estimates we need to compute their mean and variance. If the errors in measurement of the feature correspondences, $e_j = w_j^{'} - Aw_j - b$, are zero-mean i.i.d. random variables with covariance matrix $\sigma^2 I$, we can determine the statistics of the estimates $(\hat{A}, \hat{b})$ by writing (2) and (3) as

$$\hat{b} = b + \frac{1}{N}\sum_{j=1}^{N} e_j \quad (4)$$

$$\hat{A} = A + (\sum_{j=1}^{N} e_j w_j^T)(\Sigma_{ww})^{-1} \quad (5)$$

It is clear that the estimates $(\hat{A}, \hat{b})$ are both unbiased. Straightforward computation gives the covariance matrix of $\hat{b}$ as $\Sigma_{\hat{b}} = \frac{\sigma^2}{N} I$. Setting $\hat{a} = (\hat{A}_{11}, \hat{A}_{12}, \hat{A}_{21}, \hat{A}_{22})$, the covariance matrix of $\hat{a}$ can be written as:

$$\Sigma_{\hat{a}} = \frac{\sigma^2}{N} \frac{1}{\delta_x^2 \delta_y^2 - \delta_{xy}^2} \begin{bmatrix} \delta_y^2 & -\delta_{xy}^2 & 0 & 0 \\ -\delta_{xy}^2 & \delta_x^2 & 0 & 0 \\ 0 & 0 & \delta_y^2 & -\delta_{xy}^2 \\ 0 & 0 & -\delta_{xy}^2 & \delta_x^2 \end{bmatrix}$$

where $\delta_x^2 = \frac{1}{N}\sum_{j=1}^{N} x_j^2$, $\delta_y^2 = \frac{1}{N}\sum_{j=1}^{N} y_j^2$ and $\delta_{xy}^2 = \frac{1}{N}\sum_{j=1}^{N} x_j y_j$. Lengthy but straightforward computation also shows that $\hat{A}$ and $\hat{b}$ are uncorrelated.

If the feature points are distributed uniformly over the image, the elements of $\hat{A}$ are uncorrelated and $var(\hat{A}_{11}) = var(\hat{A}_{21}) = \frac{\sigma^2}{N\delta_x^2}$, $var(\hat{A}_{12}) = var(\hat{A}_{22}) = \frac{\sigma^2}{N\delta_y^2}$.

Hence, the variance of the estimates of the affine parameters depends on the number and location of the feature points, as well as the precision of the feature matching algorithm which produced the correspondences. This information could be helpful in constructing an image registration algorithm.

This estimation problem can be extended to the joint registration of a sequence of images [7, 17]. Simply superimposing pairwise estimates is suboptimal for the joint estimation problem and leads to unstable error growth. Provided that the images are related by translation only and that every image can be registered to its M nearest neighbors, we can estimate the translation between all pairs of images and collect the measurements into a set of linear equations:

$$p_{ij} = X_j - X_i + e_{ij},\ 0 \le i, j \le N-1$$

where X_i is the unknown position of the i-th image, p_{ij} is the measurement of translation between the i-th and j-th images, and e_{ij} is the error in the estimation of translation. Rewriting this equation in the matrix form $p = AX + e$ and assuming $e \sim N(0, I)$ allows us to define the optimal maximum likelihood estimate of X as a linear function of the available measurements:

$$\hat{X} = (A^T A)^{-1} A^T p \quad (6)$$

The statistics of $\hat{X}$ were studied in [24], which proposed three different approaches to the analytical evaluation of the variance of (6) as a function of the sequence length N and the number of neighboring images M. It was shown that the three approaches give similar results (see Figure 1) and that compared to simple superposition (which gives a variance growth proportional to N), the worst variance in estimation for the optimal maximum likelihood solution is reduced by a factor of M^3.

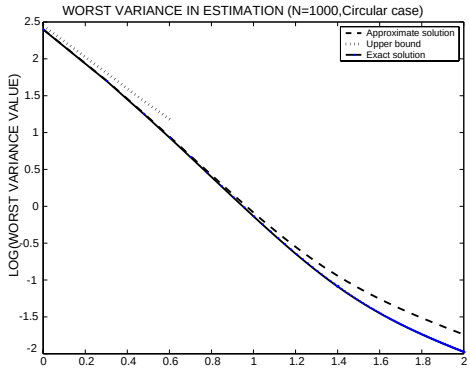

Fig. 1. The worst variance in estimation

Specifically, for linear and circular sequences of images:

$$\max_i var^{lin}(\hat{X}_i) \approx \frac{3N}{M^3} + \frac{1}{M}$$

$$\max_i var^{cir}(\hat{X}_i) \approx \frac{3N}{4M^3} + \frac{1}{M}$$

The optimal combination of additional measurements substantially reduces the worst variance in estimation, although for fixed M, the worst variance will be unbounded if we allow the length of the sequence to go to infinity. In this case, by allowing the maximum distance of measurements to grow as $\sqrt[3]{3N}$ (linear case) or $\sqrt[3]{3N/4}$ (circular case) we can ensure that the worst variance will not exceed the variance of an individual measurement; that is, the error growth can be stabilized.

3. ESTIMATION OF PROJECTIVE PARAMETERS

Projective transformations relate the coordinates of image pairs which are either (1) taken by a rotating camera without translation, or (2) images of a planar surface taken by a rotating and translating camera. The form of a projective transformation $M = (A, b, c)$ applied to a point $w \in \mathbb{R}^2$ is:

$$g_M(w) = \frac{Aw + b}{c^T w + 1} \quad (7)$$

where $A \in \mathbb{R}^{2\times 2}, b \in \mathbb{R}^2, c \in \mathbb{R}^2$. As above, we seek to minimize a least-squares functional of noisy feature corre-

spondences $\{w_j \mapsto w'_j \in \mathbb{R}^2, j = 1, \ldots, N\}$:

$$Q(M) = \frac{1}{2}\sum_{j=1}^{N}\left(w'_j - \frac{Aw_j+b}{c^T w_j+1}\right)^T\left(w'_j - \frac{Aw_j+b}{c^T w_j+1}\right) \quad (8)$$

As written, this is a nonlinear minimization over an 8-dimensional Euclidean space. However, it was shown in [13] that the optimal values of A and b are actually the solutions to a linear system involving the optimal value of c and the data, namely:

$$[\ A \quad b\ ]W(c) = V(c) \quad (9)$$

where $q_j(c) = c^T w_j + 1$ and

$$W(c) = \begin{bmatrix} \sum_{j=1}^{N}\frac{w_j w_j^T}{q_j^2(c)} & \sum_{j=1}^{N}\frac{w_j}{q_j^2(c)} \\ \sum_{j=1}^{N}\frac{w_j^T}{q_j^2(c)} & \sum_{j=1}^{N}\frac{1}{q_j^2(c)} \end{bmatrix} \quad (10)$$

$$V(c) = \begin{bmatrix} \sum_{j=1}^{N}\frac{w'_j w_j^T}{q_j(c)} & \sum_{j=1}^{N}\frac{w'_j}{q_j(c)} \end{bmatrix} \quad (11)$$

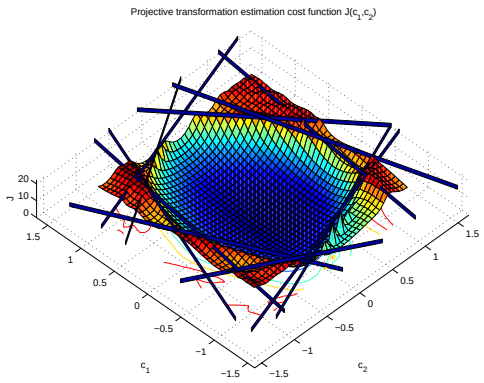

Fig. 2. The cost function $J(c)$.

We see that the problem of estimating the eight parameters of a projective transformation can be divided into a linear problem involving the six "affine" parameters (A, b) and a nonlinear problem involving the two "projective" parameters c. Therefore, we can reduce the projective transformation estimation to the 2-dimensional nonlinear minimization of

$$J(c) = \frac{1}{2}\sum_{j=1}^{N}\left(w'_j - \frac{A(c)w_j+b(c)}{c^T w_j+1}\right)^T\left(w'_j - \frac{A(c)w_j+b(c)}{c^T w_j+1}\right) \quad (12)$$

where $A(c)$ and $b(c)$ are defined through (9). This function can be efficiently minimized using an approximate Newton-Raphson scheme described in [13]. This result also gives us a way to visualize the shape of the cost function, as illustrated in Figure 2. In addition to providing a computational and visual advantage, casting the estimation problem in the two-dimensional setting allows us to more easily explore properties of the cost function such as continuity, convexity, and conditions for the existence of multiple local minima.

4. GENERAL CORRESPONDENCE

For an arbitrary image pair of the same scene, the only *a priori* constraint on correspondence is the well-known epipolar constraint [8]. The family of conjugate epipolar line pairs for an image pair may be generated by intersecting the set of all planes which contain the baseline with the two image planes. It follows that the correspondence of any point on an epipolar line in one image, if it exists, must lie on the conjugate epipolar line in the other image. In theory, this reduces the correspondence problem to a series of 1-D matching problems, each of which can be viewed as finding a path through a graph, as pictured in Figure 3.[1]

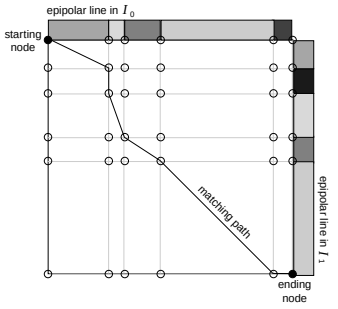

Fig. 3. Epipolar matching graph.

The monotonicity assumption that scene points are projected onto conjugate epipolar lines in the same order is a standard assumption of many correspondence algorithms, which typically use dynamic programming to efficiently obtain a solution. However, the monotonicity assumption is generally invalid for cameras whose centers of projection are widely separated with respect to their distance to points in the scene.

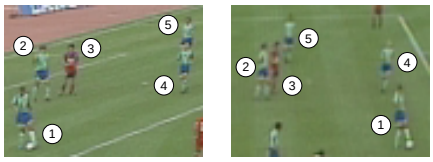

Fig. 4. Violations of monotonicity.

Figure 4 illustrates regions of two images of the same scene, rectified so that epipolar lines are horizontal. The numbered objects appear in different orders along conjugate epipolar lines due to the large perspective difference between the images. Each inconsistency in ordering generates a local violation of the monotonicity assumption in the affected conjugate epipolar lines. A monotonic path through a matching graph such as the one illustrated in Figure 3 cannot represent the correct matching.

[1] A projective transformation induces a linear path through such a matching graph.

Unfortunately, relaxing the monotonicity assumption to allow arbitrary matching of points between conjugate epipolar lines results in a problem of high combinatorial complexity, insoluble by a dynamic programming algorithm [15]. However, the set of correspondences which are physically realizable is not entirely unconstrained, and has a specific structure which we derive below.

4.1. The Correspondence Graph

In the following, we fix a pair of cameras $(\mathcal{C}_0, \mathcal{C}_1)$ whose centers of projection are O_0 and O_1, respectively. These cameras have associated image planes $\mathcal{P}_0$ and $\mathcal{P}_1$, which lie between the cameras' respective centers of projection and the scene $\mathcal{S}$, a collection of points in $\mathbb{R}^3$. Select a plane Φ containing the baseline, and view the intersection of Φ with the camera centers, the image planes, and the scene points as an imaging system with a 2-D scene $\mathbf{S} = \mathcal{S} \cap \Phi$ and 1-D image planes (the pair of conjugate epipolar lines (e_0, e_1)). We fix a coordinate system (x, y) on Φ by letting $O_0 = (0, 0)$ and placing O_1 at $(1, 0)$.[2] The epipolar lines e_0 and e_1 inherit natural one-dimensional coordinate systems (denoted i and j respectively), oriented so that increasing i and j correspond to increasing x. A *correspondence* is the realization of a point (x, y) in the scene as a pair $(i, j) \in e_0 \times e_1$. We will denote as $\mathbf{S}'$ the representation of the scene $\mathbf{S}$ in (i, j)-space.

We define two new coordinate systems, (r_0, θ_0), (r_1, θ_1) defined in terms of the coordinates (x, y) of a point p by:

$$(x, y) = (r_0 \cos\theta_0,\ r_0 \sin\theta_0) \quad (13)$$

$$(x, y) = (r_1 \cos\theta_1 + 1,\ r_1 \sin\theta_1) \quad (14)$$

These are just the "polar coordinates" of (x, y) with respect to O_0 and O_1, respectively. It is clear that the mappings between the four sets of coordinates are bijective, and hence the coordinate transforms $(i, j) = J_0(r_0, \theta_0)$ and $(i, j) = J_1(r_1, \theta_1)$ are well-defined. An important property of these mappings is:

Proposition 1 *For any fixed $\theta_0, \theta_1 \in (0, \pi)$,*

$$\frac{\partial i}{\partial r_0} = 0 \qquad \frac{\partial j}{\partial r_0} > 0 \quad (15)$$

$$\frac{\partial i}{\partial r_1} < 0 \qquad \frac{\partial j}{\partial r_1} = 0 \quad (16)$$

The proof is straightforward and can be seen from the diagram in Figure 5. The intuition is that any ray from O_0 maps to a line segment with fixed i in (i, j)-space, and provided that the ray is on the "right side" of the baseline, the segment is traversed in the direction of increasing j as we move away from O_0. Similarly, any ray from O_1 maps to a line segment with fixed j in (i, j)-space, which is traversed in the direction of decreasing i.

[2] Scene points are assumed to have positive y coordinates.

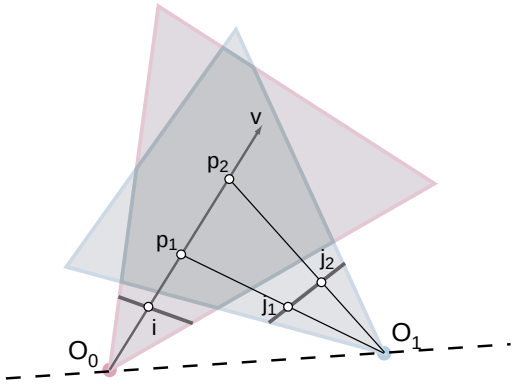

Fig. 5. Mapping from (x, y)-space into (i, j)-space.

Definition. The *correspondence graph* $C \subset e_0 \times e_1$ of a scene $\mathbf{S}$ with respect to the camera pair $(\mathcal{C}_0, \mathcal{C}_1)$ is the set of all points in $\mathbf{S}$ which are visible (i.e. unoccluded) in both e_0 and e_1, transformed into (i, j)-space.

Definition. A set A of points in (i, j)-space is a *Southeast set* if the subsets $\{(a, b) \in A \mid a = i\}$ and $\{(a, b) \in A \mid b = j\}$ have at most one element for all i, j.

Definition. The *Southeasting* operation produces a Southeast set A' from a set A as follows:

$$(i, j) \in A' \Leftrightarrow (i, j) \in A \text{ and } \{(a, j) \in A \mid a < i\} \text{ and } \{(i, b) \in A \mid b > j\} \text{ are empty}$$

Proposition 2 *The correspondence graph C for a scene $\mathbf{S}$ with respect to $(\mathcal{C}_0, \mathcal{C}_1)$ can be generated by Southeasting the transformed scene $\mathbf{S}'$.*

Proof. We know that the correspondence graph C is a subset of the transformed scene $\mathbf{S}'$. It remains to determine which points in $\mathbf{S}'$ actually appear in both images. Fix i and consider the set of points $\mathbf{S}'_\mathbf{i} = \{(a, b) \in \mathbf{S}' \mid a = i\}$. From Proposition 1, these points lie on the same ray from $\mathcal{C}_0$ in (x, y)-space. The point p' with the smallest j coordinate is closest to $\mathcal{C}_0$ and is hence the only point along the ray which is imaged by $\mathcal{C}_0$. Therefore, the points in $\mathbf{S}'_\mathbf{i}$ with larger j coordinates than p' are not retained in the correspondence graph. Similarly, for fixed j, consider the set $\mathbf{S}'_\mathbf{j} = \{(a, b) \in \mathbf{S}' \mid b = j\}$. These points lie on the same ray from $\mathcal{C}_1$ in (x, y)-space, and the only point which is retained in the correspondence graph is that point q' which has the largest i coordinate.

The operation described above is simply the Southeasting of the set $\mathbf{S}'$. By construction, the remaining elements in the Southeasted set are precisely those points which appear in both cameras and hence this Southeasted set is by definition the correspondence graph of $\mathbf{S}'$. □

The converse of the above theorem (i.e. any Southeast

set is the correspondence graph for some physical scene) is also true, but we omit the proof here.

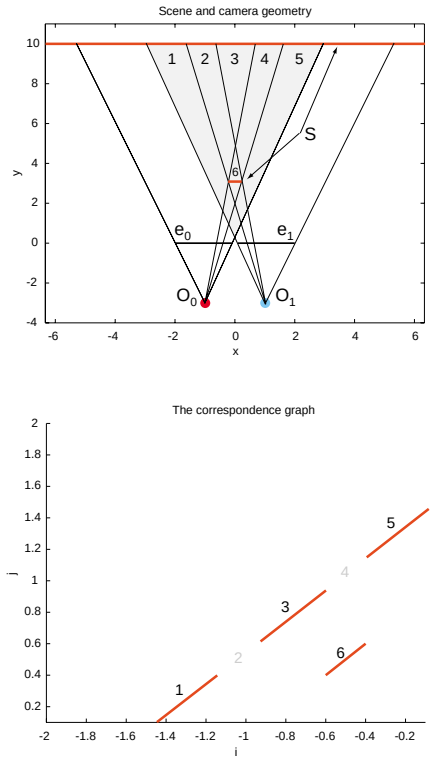

Fig. 6. Example correspondence graph.

A schematic example of a correspondence graph is shown in Figure 6. Note that segments 3 and 6 appear in different orders in the projections onto e_0 and e_1, a violation of monotonicity sometimes known as the *double-nail illusion*. While dismissed as relatively uncommon in stereo, this phenomenon is typical in widely-separated image pairs.

4.2. Estimation

Using a correspondence graph estimate, we can decide which regions appear in both images, which in only one image, and which in neither image. By ensuring that an estimated correspondence produces a valid (i.e. Southeast) correspondence graph, we are prevented from attempting to match pixels from regions that do not appear in both images. The "holes" in the correspondence graph can also be used to estimate the correct locations of regions not seen in both images. Our algorithm for estimating the correspondence between a general image pair proceeds as follows:

1. Segment and match the objects which violate monotonicity. Linearly interpolate between matched edges to obtain an initial correspondence within the segmented regions.

2. Estimate a global correspondence for the background regions. In the example from [14], this correspondence was initialized by estimating a projective transformation which related a dominant plane in the image pair, using the algorithm discussed in Section 3 above and in [13]. In a video sequence, this correspondence could be a propagated estimate from a previous frame pair.

3. Estimate the epipolar geometry [25].

4. For each pair of conjugate epipolar lines, generate the correspondence graph by Southeasting the set of background and foreground correspondences.

5. Refine the monotonic pieces of each correspondence graph using an interval matching algorithm (e.g. [12]).

Since by construction, each piece of the correspondence graph is monotonic, a correspondence algorithm which assumes monotonicity can be applied to each piece of the graph independently. If the estimate for the global background correspondence is sufficiently accurate, the refinement step may be constrained to select a matching path which lies in a nearby neighborhood of the initial matching path.

Fig. 7. Three images of a goal.

Operating on correspondence graphs allows us to obtain physically realistic transitions between object locations. Figure 7 illustrates a soccer goal seen from three perspectives: two real images (left and right) and a synthesized virtual image (center). The upright part of the goal and the goalie were segmented and used as input to the correspondence graph estimation algorithm. The floor of the goal is initially estimated to lie on the plane of the soccer field. The virtual image shows a perspective not seen in either of the real images: instead of being entirely within or outside of the goal mouth, the goalie passes in front of the upright.

4.3. Propagation

Though in some applications, tracking a small set of features through a video sequence [19] is sufficient, several modern computer graphics techniques require the accurate estimation of pixel-dense correspondence between frames of multicamera video.

Previously published work [14] addressed the creation of *virtual video* from a pair of synchronized video clips taken by widely separated, rotating cameras. The virtual camera can rotate and move along the baseline connecting the two camera centers, and the virtual video evolves at the same rate as the input video. The algorithm is based on view morphing [16], a computer graphics technique for creating a virtual image from a pair of stills given dense correspondence between them. Estimating and efficiently propagating this dense correspondence between image pairs was therefore the key problem to be solved. Furthermore, video in the database exhibited frequent monotonicity violations, which were dealt with using the formalism of the correspondence graph discussed above.

Space precludes its inclusion here, but Radke et al. [14] proposed a framework for the recursive propagation of correspondence graphs, exploiting the temporal regularity of video. The propagation consists of a time update step and a measurement update step. The time update depends only on the dynamics of the rotating source cameras, while the measurement update can be tailored to any member of a general class of image correspondence algorithms. Using these results, the correspondence graphs can be propagated and updated in a fraction of the time required to estimate them anew at every frame.

5. CONCLUSIONS

We addressed several aspects of estimating correspondence in digital video, moving from the assumption that global correspondence can be well-modeled by an affine or projective transformation to the general case when correspondence is nearly unconstrained. We have obtained reliable correspondence estimates in our research by applying a simple-to-complex approach. That is, a coarse-to-fine affine transformation estimation is used to obtain an initial registration of an image pair. From this registration, features are extracted and a projective transformation estimated. Finally, the projective transformation can be used as the initial background for estimating a set of correspondence graphs.

6. REFERENCES

[1] J.K. Aggarwal and N. Nandhakumar. On the Computation of Motion from Sequences of Images- A Review. *Proc. IEEE*, Vol. 76, No. 8, pp. 917-935, August 1988.

[2] S. Ayer and H.S. Sawhney. Layered Representation of Motion Video using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding. In *Proc. ICCV 1995*, June 1995.

[3] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, Vol. 12, No. 1, pp. 43-77, 1994.

[4] T. Beier and S. Neely. Feature-Based Image Metamorphosis. *Computer Graphics (SIGGRAPH '92)*, pp. 35-42, July 1992.

[5] P.N. Belhumeur. A Bayesian Approach to Binocular Stereopsis. *International Journal of Computer Vision*, Vol. 19, No. 3, pp 237-260, 1996.

[6] I.J. Cox, S.L. Hingorani, and S.B. Rao. A Maximum Likelihood Stereo Algorithm. *Computer Vision and Image Understanding*, Vol. 63, No. 3, pp. 542-567, May 1996.

[7] J. Davis. Mosaics of scenes with moving objects. In *Proc. CVPR 1998*, June 1998.

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

[9] S. Hsu, P. Anandan, and S. Peleg. Accurate computation of optical flow by using layered motion motion representations. In *Proc. ICPR 1994*, October 1994.

[10] H. Ishikawa and D. Geiger. Occlusions, Discontinuities, and Epipolar Lines in Stereo. In *Proc. ECCV '98*, Freiburg, Germany, 1998.

[11] D.T. Lee and B.J. Schachter. Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Inform. Sci.*, Vol. 9, pp. 219-242, 1980.

[12] Y. Ohta and T. Kanade. Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming. *IEEE PAMI*, Vol. 7, No. 2, pp. 139-154, March 1985.

[13] R. Radke, P. Ramadge, T. Echigo, and S. Iisaku. Efficiently Estimating Projective Transformations. In *Proc. ICIP 2000*, Vancouver, Canada, 2000.

[14] R. Radke, P. Ramadge, S. Kulkarni, T. Echigo, and S. Iisaku. Recursive Propagation of Correspondences with Applications to the Creation of Virtual Video. In *Proc. ICIP 2000*, Vancouver, Canada, 2000.

[15] D. Sankoff and J. Kruskal, eds. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.* Addison-Wesley, 1983.

[16] S.M. Seitz and C.R. Dyer. View Morphing. *Computer Graphics (SIGGRAPH '96)*, pp. 21-30, August, 1996.

[17] R. Szeliski and H.Y. Shum. Creating Full View Panoramic Mosaics and Environment Maps. *Computer Graphics (SIGGRAPH '97)*, pp. 251-258, August 1997

[18] Y.P. Tan, S. Kulkarni, and P. Ramadge. Extracting Good Features for Motion Estimation. *Proc. ICIP*, vol. 1, pp. 117-120, 1996.

[19] C. Tomasi and T. Kanade. Detection and Tracking of Point Features (Shape and Motion from Image Sequences: a Factorization Method – Part 3). Carnegie Mellon University Department of Computer Science Technical Report CMU-CS-91-132, April 1991.

[20] C. Tomasi and T. Kanade. Shape and Motion from Image Sequences under Orthography: a Factorization Method. *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137-154, 1992.

[21] C. Tomasi and R. Manduchi. Stereo Matching as a Nearest-Neighbor Problem. *IEEE PAMI*, Vol. 20, No. 3, pp. 333-340, March 1998.

[22] J.Y.A. Wang and E.H. Adelson. Representing Moving Objects with Layers. *IEEE Transactions on Image Processing Special Issue: Image Sequence Compression*, Vol. 3, No. 5, pp. 625-638, September 1994.

[23] J. Weng, N. Ahuja, and T.S. Huang. Matching Two Perspective Views. *IEEE PAMI*, Vol. 14, No. 8, pp. 806-825, 1992.

[24] V. Zagorodnov, P. Ramadge. Error stabilization in successive estimation of registration parameters. In *Proc. ICIP 2000*, September 2000.

[25] Z. Zhang. Determining the Epipolar Geometry and its Uncertainty: A Review. *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161-195, 1998.