1 Introduction

The general problem of achieving correspondence, or optical flow as it is known in the motion literature, is to recover the 2D displacement field between points across two images. Typical applications for which full correspondence (that is correspondence for all image points) is initially required include the measurement of motion, stereopsis, structure from motion, 3D reconstruction from point correspondences, and recently visual recognition, active vision and computer graphics animation.

In this paper we focus on two problems—one theoretical and the other more practical. On the practical side, we address the problem of establishing the full point-wise displacement field between two views (grey-level images) of a general 3D object. We achieve this by first considering a theoretical problem of establishing a quadric surface reference frame on the object. In other words, given any two views of some unknown textured opaque quadric surface in 3D projective space $\mathcal{P}^3$, is there a finite number of corresponding points across the two views that uniquely determine all other correspondences coming from points on the quadric? A constructive answer to this question readily suggests that we can associate a virtual quadric surface with any 3D object (not necessarily itself a quadric) and use it for describing shape, but more importantly, for achieving full correspondence between the two views.

On the conceptual level we propose combining geometric constraints, captured from knowledge of a small number of corresponding points (manually given, for example), and photometric constraints captured by the instantaneous spatio-temporal changes in image light intensities (conventional optical flow). The geometric constraints we propose, are related to the virtual quadric surface mentioned above. These constraints lead to a transformation (a nominal quadratic transformation) that is applied to one of the views with the result of bringing both views closer together. The remaining displacements (residuals) are recovered by using the spatial and temporal derivatives of image light intensity—either by correlation of image patches or by optical flow techniques.

2 The Quadric Reference Surface

We consider object space to be the 3D projective space $\mathcal{P}^3$, and image space to be the 2D projective space $\mathcal{P}^2$—both over the field $\mathcal{C}$ of complex numbers. Views are denoted by ψ_i, indexed by i. The epipoles are denoted by $v \in \psi_1$ and $v' \in \psi_2$, and we assume their location is known (for methods, see [6], for example). The symbol $\cong$ denotes equality up to a scale, GL_n stands for the group of $n \times n$ matrices, PGL_n is the group defined up to a scale, and $SPGL_n$ is the symmetric specialization of PGL_n.

Result 1. Given two arbitrary views $\psi_1, \psi_2 \subset \mathcal{P}^2$ of a quadric surface $Q \in \mathcal{P}^3(\mathcal{C})$ with centers of projection at $O, O' \in \mathcal{P}^3$, and $O, O' \in Q$, then five corresponding points across the two views uniquely determine all other correspondences.

Proof. Let (x_0, x_1, x_2) and (x'_0, x'_1, x'_2) be coordinates of ψ_1 and ψ_2, respectively, and $(z_0, \ldots, z_3)$ be coordinates of Q. Let $O = (0,0,0,1)$; then the quadric surface may be given as the locus $z_0 z_3 - z_1 z_2 = 0$, and ψ_1 as the projection from $O = (0,0,0,1)$ onto the plane $z_3 = 0$. In case where the centers of projection are on Q, the line through O meets Q in exactly one other point, and thus the mapping $\psi_1 \mapsto Q$ is generically one-to-one, and so has a rational inverse: $(x_0, x_1, x_2) \mapsto (x_0^2, x_0 x_1, x_0 x_2, x_1 x_2)$. Because all quadric surfaces of the same rank are projectively equivalent, we can perform a similar blow-up from ψ_2 with the result $(x_0'^2, x'_0 x'_1, x'_0 x'_2, x'_1 x'_2)$. The projective transformation $A \in PGL_4$ between the two representations of Q can then be recovered from five corresponding points between the two images. $\square$

The result does not hold when the centers of projection are not on the quadric surface. This is because the mapping between Q and $\mathcal{P}^2$ is not one-to-one (a ray through the center of projection meets Q in two points), and therefore, a rational inverse does not exist. We are interested in establishing a more general result that applies when the centers of projection are not on the quadric surface. One way to enforce a one-to-one mapping is by making "opacity" assumptions, defined below.

Definition 2 (Opacity Constraint). Given an object $Q = \{P_1, \ldots, P_n\}$, we assume there exists a plane through the camera center O that does not intersect any of the chords $P_i P_j$ (i.e., Q is observed from only one "side" of the camera). Secondly, we assume that the surface is opaque, which means that among all the surface points along a ray from O, the closest point to O is the point that also projects to the second view (ψ_2). The first constraint, therefore, is a camera opacity assumption, and the second constraint is a surface opacity assumption; together, we call them the opacity constraint.

With an appropriate re-parameterization of $\mathcal{P}^3$ we can obtain the following result:

Theorem 3. *Given two arbitrary views $\psi_1, \psi_2 \subset \mathcal{P}^2$ of an opaque quadric surface $Q \in \mathcal{P}^3$; then nine corresponding points across the two views uniquely determine all other correspondences.*

The following auxiliary propositions are used as part of the proof.

Lemma 4 (Relative Affine Parameterization). *Let p_o, p_1, p_2, p_3 and p'_o, p'_1, p'_2, p'_3 be four corresponding points coming from four non-coplanar points in space. Let A be a collineation of $\mathcal{P}^2$ determined by the equations $Ap_j \cong p'_j$, $j = 1, 2, 3$, and $Av \cong v'$. Finally let v' be scaled such that $p'_o \cong Ap_o + \boldsymbol{v}'$. Then, for any point $P \in \mathcal{P}^3$ projecting onto p and p', we have*

$$p' \cong Ap + k\boldsymbol{v}'. \tag{1}$$

The coefficient $k = k(p)$ is independent of ψ_2, i.e., is invariant to the choice of the second view, and the coordinates of P are $(x, y, 1, k)$.

The lemma, its proof and its theoretical and practical implications are discussed in detail in [10]. The scalar k is called a *relative affine invariant* and can be computed with the aid of a second arbitrary view ψ_2. For future reference, let π stand for the plane passing through P_1, P_2, P_3 in space.

Proof of Theorem: From Lemma 4, any point P can be represented by the coordinates $(x, y, 1, k)$ and k can be computed from Equation 1. Since Q is a quadric surface, then there exists $H \in SPGL_4$ such that $P^\top HP = 0$, for all points P of the quadric. Because H is symmetric and determined up to a scale, it contains only nine independent parameters. Therefore, given nine corresponding image points we can solve for H as a solution of a linear system; each corresponding pair p, p' provides one linear equation in H of the form $(x, y, 1, k)H(x, y, 1, k)^\top = 0$.

Given that we have solved for H, the mapping $\psi_1 \mapsto \psi_2$ due to the quadric Q can be determined uniquely (i.e., for every $p \in \psi_1$ we can find the corresponding $p' \in \psi_2$) as follows. The equation $P^\top HP = 0$ gives rise to a second order equation in k of the form $ak^2 + b(p)k + c(p) = 0$, where the coefficient a is constant (depends only on H) and the coefficients b, c depend also on the location of p. Therefore, we have two solutions for k, and by Equation 1, two solutions for p'. The two solutions for k are $k^1, k^2 = \frac{-b \pm r}{2a}$, where $r = \sqrt{b^2 - 4ac}$. The finding, shown in the next auxiliary lemma, is that if the surface Q is opaque, then the sign of r is fixed for all $p \in \psi_1$. Therefore, the sign of r for p_o that leads to a positive root (recall that $k_o = 1$) determines the sign of r for all other $p \in \psi_1$. $\square$

Lemma 5. *Given the opacity constraint, the sign of the term $r = \sqrt{b^2 - 4ac}$ is fixed for all points $p \in \psi_1$.*

Proof. Let P be a point on the quadric projecting onto p in the first image, and let the ray $\overline{OP}$ intersect the quadric at points P^1, P^2, and let k^1, k^2 be the roots of the quadratic equation $ak^2 + b(p)k + c(p) = 0$. The opacity assumption is that the intersection closer to O is the point projecting onto p and p'.

Recall that P_o is a point (on the quadric in this case) used for setting the scale of v' (in Equation 1), i.e., $k_o = 1$. Therefore, all points that are on the same side of π as P_o have positive k associated with them, and vice versa. There are two cases to be considered: either P_o is between O and π (i.e., $O < P_o < \pi$), or π is between O and P_o (i.e., $O < \pi < P_o$)—that is O and P_o are on opposite sides of π. In the first case, if $k^1 k^2 \leq 0$ then the non-negative root is closer to O, i.e., $k = \max(k^1, k^2)$. If both roots are negative, the one closer to zero is closer to O, again $k = \max(k^1, k^2)$. Finally, if both roots are positive, then the larger root is closer to O. Similarly, in the second case we have $k = \min(k^1, k^2)$ for all combinations. Because P_o can satisfy either of these two cases, the opacity assumption gives rise to a consistency requirement in picking the right root: either the maximum root should be uniformly chosen for all points, or the minimum root. $\square$

In the next section we will show that Theorem 3 can be used to surround an arbitrary 3D surface by a virtual quadric, i.e., to create quadric reference surfaces, which in turn can be used to facilitate the correspondence problem between two views of a general object. The remainder of this section takes Theorem 3 further to quantify certain useful relationships between the centers of two cameras and the family of quadrics that pass through arbitrary configurations of eight points whose projections on the two views are known.

Theorem 6. *Given a quadric surface $Q \subset \mathcal{P}^3$ projected onto views $\psi_1, \psi_2 \subset \mathcal{P}^2$, with centers of projection $O, O' \in \mathcal{P}^3$, there exists a parameterization of the image planes ψ_1, ψ_2 that yields a representation $H \in SPGL_4$ of Q such that $h_{44} = 0$ when $O \in Q$, and the sum of the elements of H vanishes when $O' \in Q$.*

Proof. The re-parameterization described here was originally introduced in [10] as part of the proof of Lemma 4. We first assign the standard coordinates in $\mathcal{P}^3$ to three points on Q and to the two camera centers O and O' as follows. We assign the coordinates $(1,0,0,0), (0,1,0,0), (0,0,1,0)$ to P_1, P_2, P_3, respectively, and the coordinates $(0,0,0,1)$, $(1,1,1,1)$ to O, O', respectively. By construction, the point of intersection of the line $\overline{OO'}$ with π has the coordinates $(1,1,1,0)$.

Let P be some point on Q projecting onto p, p'. The line $\overline{OP}$ intersects π at the point $(\alpha, \beta, \gamma, 0)$. The coordinates α, β, γ can be recovered (up to a scale) by the mapping $\psi_1 \mapsto \pi$, as follows. Given the epipoles v and v', we have by our choice of coordinates that p_1, p_2, p_3 and v are projectively (in $\mathcal{P}^2$) mapped onto $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$ and $e = (1,1,1)$, respectively. Therefore, there exists a unique element $A_1 \in PGL_3$ that satisfies $A_1 p_j \cong e_j$, $j = 1,2,3$, and $A_1 v \cong e$. Denote $A_1 p = (\alpha, \beta, \gamma)$. Similarly, the line $\overline{O'P}$ intersects π at $(\alpha', \beta', \gamma', 0)$. Let $A_2 \in PGL_3$ be defined by $A_2 p'_j \cong e_j$, $j = 1,2,3$, and $A_2 v' \cong e$. Further, let $A_2 p' = (\alpha', \beta', \gamma')$.

It is easy to see that $A \cong A_2^{-1}A_1$, where A is the collineation defined in Lemma 4. Likewise, the homogeneous coordinates of P are transformed into $(\alpha, \beta, \gamma, k)$. With this new coordinate representation the assumption $O \in Q$ translates to the constraint that $h_{44} = 0$ $((0,0,0,1)H(0,0,0,1)^{\top} = 0)$, and the assumption $O' \in Q$ translates to the constraint $(1,1,1,1)H(1,1,1,1)^{\top} = 0$. ∎

Corollary 7. *Theorem 6 provides a quantitative measure of proximity of a set of eight 3D points, projecting onto two views, to a quadric that contains both centers of projection.*

Proof. Given eight corresponding points we can solve for H with the constraint $(1,1,1,1)$ $H(1,1,1,1)^{\top} = 0$. This is possible since a unique quadric exists for any set of nine points in general position (the eight points and O'). The value of h_{44} is then indicative of how close the quadric is from the other center of projection O. ∎

Note that when the camera center O is on the quadric, then the leading term of $ak^2 + b(p)k+c(p) = 0$ vanishes ($a = h_{44} = 0$), and we are left with a linear function of k. We see that it is sufficient to have a bi-rational mapping between Q and only one of the views without employing the opacity constraint. This is because of the asymmetry introduced in our method: the parameters of Q are reconstructed with respect to the frame of reference of the first camera (i.e., relative affine reconstruction in the sense of [10]) rather than reconstructed projectively. Also note the importance of obtaining quantitative measures of proximity of an eight-point configuration of 3D points to a quadric that contains both centers of projection; this is a necessary condition for observing a "critical surface". A sufficient condition is that the quadric is a hyperboloid of one sheet [3, 7]. Theorem 6 provides, therefore, a tool for analyzing part of the question of how likely are typical imaging situations within a "critical volume".

3 Achieving Full Correspondence between Views of a General 3D Object

In this section we derive an application of Theorem 3 to the problem of achieving full correspondence between two grey-level images of a general 3D object. The basic idea, similar to [8, 9], is to treat the correspondence problem as composed of two parts: a nominal transformation with the aid of a small number of known correspondences, and a residual displacement field that is recovered using instantaneous spatio-temporal derivatives of image intensity values. along epipolar lines. This paradigm is general in the sense that it applies to any 3D object. However, it is useful if the nominal transformation brings the two views closer together.

Consider, for example, the case where the nominal transformation is a homography of $\mathcal{P}^2$ of some plane π. In that case, the residual field is simply the relative affine invariant k of Equation 1. In other words, if the object is relatively flat, then the k-field is small, and thereby the nominal transformation (which is Ap) brings the two views closer to each other. If the object is not flat, however, then the residual field may be large within regions in the image that correspond to object points that are far away from π. This situation is demonstrated in the second row display in Figure 1. Three points were chosen (two eyes and the right mouth corner) for the computation of the planar nominal transformation. The overlay of the second view and the transformed first view demonstrate that the central region of the face is brought closer at the expense of regions near the boundary (which correspond to object points that are far away from the virtual plane passing through both eyes and the mouth corner).

This example naturally suggests that a nominal transformation based on placing a virtual quadric reference surface on the object would give rise to a smaller residual field (note that the planar transformation is simply a particular case of a quadric transformation). The "nominal quadratic transformation" can be formalized as a corollary of Theorem 3 as follows:

Corollary 8 (of Theorem 3). *A virtual quadric surface can be fitted through any 3D surface, not necessarily a quadric surface, by observing nine corresponding points across two views of the object.*

Proof. First, it is known that there is a unique quadric surface through any nine points in general position. This follows from a *Veronese map* of degree two, $v_2 : \mathcal{P}^n \longrightarrow \mathcal{P}^{(n+1)(n+2)/2-1}$, defined by $(x_0, \ldots, x_n) \mapsto (\ldots, x^I, \ldots)$, where x^I ranges over all monomials of degree two in $x_0, \ldots, x_n$. For $n = 3$, this is a mapping from $\mathcal{P}^3$ to $\mathcal{P}^9$ taking hypersurfaces of degree two in $\mathcal{P}^3$ (i.e., quadric surfaces) into hyperplane sections of $\mathcal{P}^9$. Thus, the subset of quadric surfaces passing through a given point in $\mathcal{P}^3$ is a hyperplane in $\mathcal{P}^9$, and since any nine hyperplanes in $\mathcal{P}^9$ must have a common intersection, there exists a quadric surface through any given nine points. If the points are in general position this quadric is smooth (i.e., H is of full rank).

Therefore, by selecting any nine corresponding points across the two views we can apply the construction described in Theorem 3 and represent the displacement between corresponding points p and p' across the two views as follows:

$$p' \cong (Ap + k_q \boldsymbol{v}') + k_r \boldsymbol{v}', \tag{2}$$

where $k = k_q + k_r$, k_q is the relative affine structure of the virtual quadric and k_r is the remaining parallax which we call the residual. The term within parentheses is the nominal

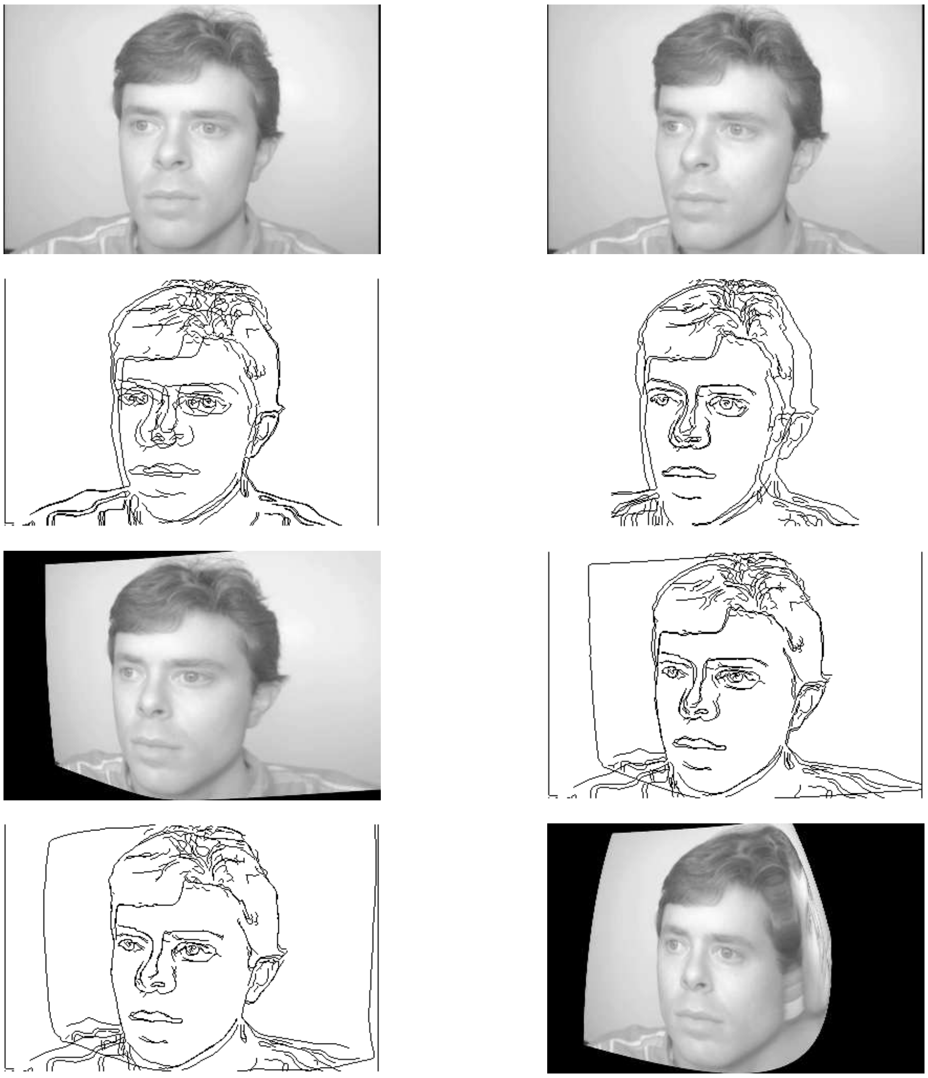

Figure 1. *Row 1*: Two views of a face, ψ_1 on the left and ψ_2 on the right. *Row 2*: The left-hand display shows the overlayed edges of the two views above. The right-hand display shows the overlay of the affine transformed ψ_1 and ψ_2. The three points chosen were the two eyes and the right mouth corner. Notice that the displacement across the center region of the face was reduced, at the expense of the peripheral regions that were taken farther apart. *Row 3*: The left-hand display shows ψ_1 followed by the nominal quadratic transformation. The right-hand display shows the overlay of edges of the display on the left and ψ_2. Notice that the original displacements between ψ_1 and ψ_2 are reduced to about 1–2 pixels. *Row 4*: The display on the left shows the overlay of the target view ψ_2 on the transformed view ψ_1 (nominal quadratic followed by residual flow). The right-hand display shows the effect of a nominal quadric transformation due to a hyperboloid of two sheets. This unintuitive solution due to an unsuccessful choice of sample points creates the mirror image on the right side.

quadratic transformation, and the remaining term $k_r \boldsymbol{v}'$ is the unknown displacement along the known direction of the epipolar line. Therefore, Equation 2 is the result of representing the relative affine structure of a 3D object with respect to some reference quadric surface, namely, k_r is a relative affine invariant (because k and k_q are both invariants by Theorem 3). $\square$

Note that the corollary is analogous to describing shape with respect to a reference plane [4, 8, 10]—instead of a plane we use a quadric and use the tools described in the previous section in order to establish a quadric reference surface. The overall algorithm for achieving full correspondence given nine corresponding points $p_o, p_1, \ldots, p_8$ is summarized below:

1. Determine the epipoles v, v' using eight of the corresponding points, and recover the collineation A from the equations $Ap_j \cong p'_j$, $j = 1, 2, 3$, and $Av \cong v'$. Scale v' to satisfy $p'_o \cong Ap_o + v'$.
2. Compute k_j, $j = 4, \ldots, 8$ from the equation $p'_j \cong Ap_j + k_j \boldsymbol{v}'$.
3. Compute the quadric parameters from the nine equations $(x_j, y_j, 1, k_j)H(x_j, y_j, 1, k_j)^\top = 0$ ($k_o = 1$, $j = 1, \ldots, 8$). Note that $k_1 = k_2 = k_3 = 0$.
4. For every point p compute k_q as the appropriate root of k of $ak^2 + b(p)k + c(p) = 0$, where the coefficients a, b, c follow from $(x_q, y_q, 1, k_q)H(x_q, y_q, 1, k_q)^\top = 0$, and the appropriate root follows from the sign of r for $ak_o^2 + b(p_o)k_o + c(p_o) = 0$ consistent with the root $k_o = 1$.
5. Warp ψ_1 according to the nominal transformation $\bar{p} \cong Ap + k_q \boldsymbol{v}'$.
6. The remaining displacement (residual) between p' and $\bar{p}$ consists of an unknown displacement k_r along the known epipolar line. The spatio-temporal derivatives of image light intensity can be used to recover k_r.

In case (and only then) the ray $\overline{OP}$ does not intersect the quadric, the solutions for the corresponding k are complex numbers (i.e., p cannot be reprojected onto $\bar{p}$ due to Q). This case can be largely avoided when the nine sample points are spread as far apart as possible on the view ψ_1 of the object (see [11] for analytic results and computer simulations on the existence and distribution of "complex pockets").

Also, a tight fit of a quadric surface onto the object can be obtained by using many corresponding points to obtain a least squares solution for H. Note that from a practical point of view we would like the quadric to lie as close as possible to the object; otherwise the algorithm, though correct, would not be useful, i.e., the residuals may be larger than the original displacement field between the two views.

4 Experimental Results on Real Images

We have implemented the method described in the previous section for purposes of computer simulation and for application in a real image situation. The computer simulations are shown in [11], and some of the real image experiments are shown here.

Figure 1, top row, shows two images of a face taken from two distinct viewpoints. Achieving full correspondence between two views of a face is extremely challenging for two reasons. First, a face is a complex object that is not easily parameterized. Second, the texture of a typical face does not contain enough image structure for obtaining point-to-point correspondence in a reliable manner. There are a few points (on the order of 10–20) that can be reliably matched, such as the corners of the eye, mouth and eyebrows. We rely on these few points to perform the quadratic nominal transformation and the epipolar geometry, and then apply optical flow techniques to "finish off" the correspondence everywhere else. The optical flow method we used was a modification on the technique described in [1, 8], which is a coarse-to-fine gradient-based method following [5].

The epipoles were recovered using the algorithm described in [2, 6] using a varying number of points. The results presented here used the minimal number of nine points, but similar performance was obtained using more than nine points with a least squares solution for H.

From Figure 1, second row, left-hand display, we see that typical displacements between corresponding points around the center region of the face vary around 20 pixels. Figure 1, third row, left-hand display, shows the quadric nominal transformation applied to the view ψ_1. The overlay of the edges of view ψ_2 and the transformed view are shown in the right-hand display. One clearly sees that both the center of the face and the boundaries are brought closer together. Typical displacements have been reduced to around 1–2 pixels. The optical flow algorithm restricted along epipolar lines was applied between the transformed view and the second view. The final displacement field (nominal transformation due the quadric plus the residuals recovered by optical flow) was applied to the first view to yield a synthetic image that, if successful, should look much like the second view. In order to test the similarity between the synthetic image and the second view, the overlay of the edges of the two images is shown in the fourth row of Figure 1, left-hand display.

Finally, to illustrate that a quadric surface may yield unintuitive results, we show in the fourth row of Figure 1, right-hand display, the result of having a hyperboloid of two sheets as a quadric reference surface. This is accidental, but evidently can happen with an unsuccessful choice of sample points.

5 Summary

Part of this paper addressed the theoretical question of establishing a one-to-one mapping between two views of an unknown quadric surface. We have shown that nine corresponding points are sufficient to obtain a unique map, provided we make the assumption that the surface is opaque. We have also shown that an appropriate parameterization of the image planes facilitates certain questions of interest such as the likelihood that eight corresponding points will be coming from a quadric lying in the vicinity of both centers of projection.

On the practical side, we have shown that the tools developed for quadrics can be applied to any 3D object by setting up a virtual quadric surface lying in the vicinity of the object. The quadric serves as a reference frame, but also to facilitate the correspondence problem. We have shown that one view can be transformed toward the second view, which is equivalent to first projecting the object onto the quadric and then projecting the quadric onto the second view. For example, in the implementation section we have shown that two views of a face with typical displacements of around 20 pixels are brought closer, to around a 1–2 pixel displacement, by the transformation. Most optical flow methods can deal with such small displacements quite effectively.

Acknowledgments

Thanks to David Beymer for providing the pair of images used for our experiments, and to Long Quan for providing the code necessary for recovering epipoles.

References

1. J.R. Bergen and R. Hingorani. Hierarchical motion-based frame rate conversion. Technical report, David Sarnoff Research Center, 1990.
2. O.D. Faugeras, Q.T. Luong, and S.J. Maybank. Camera self calibration: Theory and experiments. In *Proceedings of the European Conference on Computer Vision*, pages 321–334, Santa Margherita Ligure, Italy, June 1992.
3. B.K.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.
4. J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385, 1991.
5. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings IJCAI*, pages 674–679, Vancouver, 1981.
6. Q.T. Luong, R. Deriche, O.D. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report, INRIA, France, 1993.

7. S.J. Maybank. The projective geometry of ambiguous surfaces. *Proceedings of the Royal Society of London*, 332:1–47, 1990.
8. A. Shashua. Correspondence and affine shape from two orthographic views: Motion and recognition. AI Memo No. 1327, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, December 1991.
9. A. Shashua. *Geometry and photometry in 3D visual recognition.* PhD thesis, MIT Artificial Intelligence Laboratory, AI-TR-1401, November 1992.
10. A. Shashua. On geometric and algebraic aspects of 3D affine and projective structures from perspective 2D views. *2nd European Workshop on Invariants*, Azores Islands, Portugal, October 1993. Also AI Memo No. 1405, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, July 1993.
11. A. Shashua and S. Toelg. Quadric reference surfaces in 3D vision. A.I. Memo No. 1448, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1994.