Segmentation of Planar Objects and their Shadows in Motion Sequences

Yaser Yacoob and Larry Davis

Computer Vision Laboratory

University of Maryland

College Park, MD 20742

Abstract

We investigate segmentation of planar objects and their cast shadows in image sequences. Given two moving image regions in an image sequence we present an algorithm for determining if the two moving regions can be interpreted as a planar object and its cast shadow. Projective geometry and motion properties are employed to directly recover a homology that constrains point correspondences of the outlines of the image regions and determine if they obey an object/shadow relationship. This homology is derived directly from the motions of the regions and therefore is easier to accomplish than determining point-to-point correspondences between candidate object-shadow pairs. Several experiments under approximate point light source illumination are presented.

1 Background

Cast shadows are frequent in images of outdoor scenes due to direct sun light during the day and street lights at night. In indoor scenes, complex shadowing often takes place due to multiple light sources and their geometric and light emission properties. The primary objective of shadow segmentation is to eliminate distracting image regions and facilitate subsequent processing. Moreover, once detected, the cast shadow

can provide information about the shape of the object, light source location and the properties of the surface on which the shadow is cast.

The complexity in the image formation process of scenes with shadows stems from interactions between the light source, the object, the surface on which the shadow is cast and the camera's viewpoint. We describe here how to employ motion information to determine if two regions constitute a planar region and its cast shadow.

Almost all prior research on separation of shadows from shadow casting objects focused on a single image in which spatial or chromatic scene properties are exploited. Stauder et al. [11] proposed an approach for detection of moving shadows in MPEG sequences of object motion, as viewed by a stationary camera. They assumed that the light source intensity is high (so that the shadow points appear considerably darker than the foreground object), the camera and the planar surface on which the shadow is cast are stationary, and finally that the distance between the object and the shadow projection surface is not much smaller than the distance to the light source. Sonoda and Ogata [9] proposed an approach that is specialized for separation of people and their shadows moving in outdoor scenes. Assumptions on the camera position, human height and light source were employed. Noronha and Nevatia [8] employed a priori knowledge concerning the 3D structure of the scene and the position of the sun for shadow prediction and verification in outdoor scenes of aerial images of buildings. Several approaches that exploit differences in the brightness of colors between shaded areas and normal brightness have been proposed [3, 7, 5]. All of these approaches are applied to one frame at a time.

It is important to note that while it may be possible to determine object-to-shadow correspondence from a single image [8, 13]. less constrained than This could be done as follows: given two candidate regions, a triplet of points on the candidate object boundary should be matched against all triplets on the candidate shadow region to determine if a match exists and if it fully explains the mapping between the outlines of the given two regions. For example, any three points on the candidate object can be corresponded to any three points on the candidate shadow boundary and a planar homology can be computed [13]. While this was proposed in [13] as a manual process that can assist cartographers, its computational complexity is high.

This paper focuses on the following question: given two moving regions in an image do they constitute a pair of an object and its shadow? We show that if the object is approximately planar, its shadow is cast on a planar surface and the motion of the two regions is un-equal as viewed by the camera, then it is possible to compute a homology that answers the question. The fundamental observation is that an object and its cast shadow motions occur on different 3D surfaces and thus, the relationship between these two motions constitute an important visual cue. This cue is used to derive a closed-form solution to estimate the projective transformation (i.e., the homology) of the object region to the shadow region. This transformation can then be evaluated to determine if it explains the object and shadow region boundaries. Thus, the computational complexity of identifying the correct projective transformation is significantly reduced.

There are two prerequisites to the proposed approach. First, two regions have been located in an image sequence. Accomplishing this depends on the actual scene and is typically posed as a segmentation problem. The detection of two such regions may employ color, intensity, motion, texture, perceptual organization or model-based cues to infer candidate regions. In our experiments we employ straightforward intensity and color constraints to detect such regions and illustrate our algorithm, but clearly the extensive segmentation literature is pertinent. Note that the regions can be separate or adjacent without affecting the algorithm. The second prerequisite is that the two image regions are not conforming to a single motion model. Otherwise, motion cannot determine if they are an object/shadow pair.

The closest related work is that of Van Gool et al. [13]. The critical difference between the work reported here and [13] is that while computing the homology transformation required hand-picking object/shadow point-correspondences in [13], our approach computes the homology (up to a scalar) directly from the motion transformations of the regions. The approach of Van Gool et al. [13] is dependent on determining object-shadow point-pairs, automatic search for such pairs must be conducted over the whole image since there is no prior knowledge as to where the correspondent shadow point of an object point may lie in the image. In contrast, given that the motion of points between consecutive frames is small (a common assumption in motion estimation although it depends on the scene and imaging parameters) tracking points between consecutive frames is a well-established process since typically a search for the point is conducted in a

small neighborhood around its prior location. Furthermore, point-tracking may employ the assumption that the tracked point properties remain constant (instantaneously, between two frames), while the appearance relationship between a point and its corresponding shadow-point is not known in advance and is subject to the object, surface shadow and illumination properties.

In summary, our paper suggests that computing the homology in a closed-form using the motion of regions is computationally less complex and more constrained than computing the homology by evaluating all potential object-shadow point-pairs to determine which of these candidates lead to a plausible homology that can explain all point correspondences on the boundary of the object and shadwo region. Object-shadow point-correspondences were manually established before the computation of the homology in [13] and automating the process remains open for study. In contrast, our motion-based homology computation is automatic and employs existing techniques for point-tracking.

We present the assumptions and the basic approach in Section 2. In Section 3 we provide a detailed description of the geometric and computational processes. In Section 4 we describe the pre-processing steps that have been employed in the experiments, the results and a discussion of these experiments. Finally, Section 5 provides a summary and future research directions.

2 Assumptions and Approach

We make the following assumptions about the scene:

- The shadow casting object in the scene is approximately a 3D plane (or appears as a planar object under perspective projection).
- The surface on which the shadow is cast is also a 3D plane.
- The object is in motion.
- The light source is stationary and is approximately a point light source.

These assumptions, admittedly, rule out a large set of situations where objects and their cast shadows are created and may be imaged in motion sequences. Non-planar objects as well as non-planar projection surfaces are important cases we currently can not handle. Also, we do not consider articulated and deformable objects and these introduce further challenges to the problem. Nevertheless, near-planar objects, planar surfaces and point light sources frequently occur in indoor and outdoor environments. Moreover, the framework developed in this paper may be extended to piece-wise planar objects and cast surfaces at the cost of increased technical and computational complexity but using the basic projective/motion formulation in this paper.

Our approach is based on the following observations:

- Object motion and shadow motion, generally take place on different 3D planes and so produce different 2D motions. Specifically, while the object-projected image on the camera plane fits one motion model, the shadow, being cast on a different 3D plane, results in projected motion on the camera plane that is different from the object motion.

- While the dense interior optical flow generated by the moving object satisfies the well-known planar motion model and the brightness constancy assumption during the motion, the shadow region does not satisfy this since only the shadow boundary, in fact, appears to move as a planar region while the interior of the shadow exhibits no motion since it has fixed brightness (i.e., the brightness of a point in the interior of the shadow region does not change between time t and $t+1$ as long as the point remains in the interior of the cast shadow). Therefore, only the boundary of the shadow is useful in segmentation and for determining the correspondence between object/shadow regions.

- It is well known that a planar object and its cast shadow on a plane create a Desargues configuration and lead to a planar homology (to be reviewed below). We exploit the algebraic properties of the planar homology to estimate its value and validate the hypothesized object-shadow relationship.

Figure 1 describes the situation we model. A point light source L emits light rays that cross the planar object $O(t)$ and cast a shadow $S(t)$ on a 3D plane. Assuming that the planar object is in motion, the object and the cast shadow are shown at times t and $t+1$.

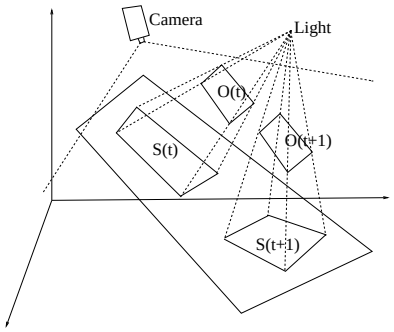


Figure 1: Planar object motion with cast shadow on a plane

3 Model and Approach

3.1 Planar Motion Transformation

We assume that each point in the regions of interest in the image is moving according to one of two motion models (the object motion or the shadow motion). If there is only a single motion model in the image (i.e., the object and the shadow motion models are approximately equal) then the region is moving as a single rigid region and the motion cue cannot assist in object/shadow segmentation. This may happen where the object and its shadow are close to each other with respect to their distance from the camera such as a vehicle moving on a road being observed by a distant camera.

Let the motion of the object region $O(t)$ as projected on the camera plane be represented by a 2D homography transformation matrix, [14]. Therefore, the point (x', y', w') at frame $t+1$ corresponds to the point (x, y, w) at t via

$$[x' \quad y' \quad w']^T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} [x \quad y \quad w]^T \tag{1}$$

The points are represented in homogeneous coordinates. Similarly, the motion of the boundary of the shadow region can be represented by a homography.

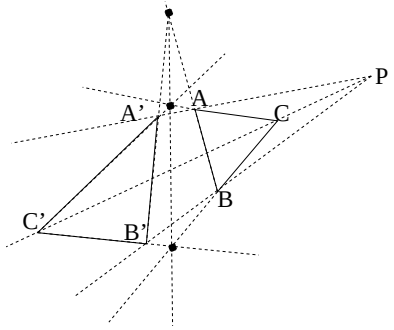


Figure 2: Two triangles in Desargues configuration

3.2 Desargues Configuration

Desargues Theorem is considered a cornerstone of projective geometry and was formulated by the French mathematician Desargues (1591-1661). It states that "Two triangles are in perspective from a point if and only if they are in perspective from a line" (see Figure 2):

Theorem:*Let A, B, C and A', B, C' be two triangles in projective 3D space. The lines AA', BB', CC' intersect in a single point if and only if the intersections of corresponding sides (AB, A'B'), (BC, B'C'), (CA, C'A') lie on a single line.*

The theorem has a clear self duality: given two triplets of lines defining two triangles, the intersections of the corresponding sides lie on a line if and only if the lines of intersection of the corresponding vertices intersect in a point.

If a point light source casts on a 3D plane a shadow A'B'C' of a non-coplanar triangle ABC then this constitutes a Desargues configuration in projective 3D space. A perspective image of this scene is also a Desargues configuration. The projected images of the two 3D triangles (the object and its shadow), the light source and the 3D line of intersection-points on the 2D plane of the camera constitute a Desargues configuration in this projective plane.

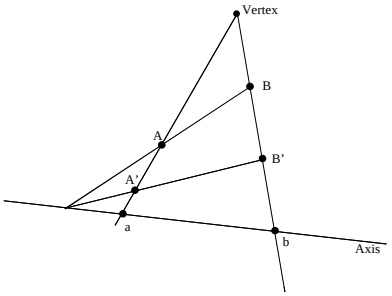


Figure 3: Planar homology, a vertex and an axis, A and B have corresponding points A' and B'. The intersection of the line through AB and A'B' lie on the axis while the intersection of AA' and BB' is at the vertex

3.3 Planar Homology

A planar homology is a plane projective transformation with five degrees of freedom, having a line of fixed points, called the *axis* and a distinct fixed point the *vertex*, (not on the axis). Figure 3 illustrates this projective relation. Under this transformation, a point A and its correspondent point A' are collinear with the vertex [10]. Moreover, the cross-ratio defined by the vertex, the corresponding points A, A' and the intersection of their join with the axis, is the characteristic invariant of the homology and the same for all corresponding points. Given three points and their correspondent points, the vertex of the homology is found by intersecting the three lines connecting the three pairs (two are sufficient) to find the vertex. The axis is found by connecting the sides of the point triplets in the same way as the Desargues configuration is defined.

The triangles of a Desargues configuration are related by a planar homology, with the common point of intersection being the vertex of the transformation, and the axis being the line containing the intersections of corresponding sides. A triplet of point pairs in correspondence under a homology are also triangle pairs in a Desargues Configuration.

Algebraically, a transformation, $\mathbf{H}$ (3x3), that is a planar homology has one distinct eigenvalue, λ_1 with corresponding eigenvector being the vertex, $\mathbf{v}$, and two equal eigenvalues, $\lambda_2 = \lambda_3$, whose corresponding eigenvectors span the axis, $\mathbf{a}$, of the transformation (i.e., $\mathbf{a} = ev_2 \times ev_3$ where $\times$ is the cross product of the eigenvectors ev_2 and ev_3) [10]. The ratio $\mu = \lambda_1/\lambda_2$ is the a characteristic invariant of the homology $\mathbf{H}$. The transformation $\mathbf{H}$ can be parameterized by a 3-vector representing the axis, $\mathbf{a}$, a 3-vector representing the

vertex, $\mathbf{v}$, and the characterstic cross-ratio, μ, to be

$$\mathbf{H} \equiv \mathbf{I} + (\mu - 1)\frac{\mathbf{v}\mathbf{a}^T}{\mathbf{v} \cdot \mathbf{a}} \tag{2}$$

where $\mathbf{I}$ is the identity matrix. This parameterization has only 5 degrees of freedom since the scales of $\mathbf{v}$ and $\mathbf{a}$ have no effect. Therefore, the parameterization is equivalent up to a scale factor and not a strict equality. Given three corresponding points, the 5 degrees of freedom become overdetermined and can be calculated up to a scalar.

3.4 Planar Object Motion and the Projected Shadow

The relation between a planar object and its shadow (on a plane) is a planar homology and satisfies Desargues configuration. Van Gool et. al. [13] employed this fact in determining shadows of planar sides of buildings in a cartographic tool that operators use by manually identifying corresponding points in images.

As observed by Van gool [13], the parameterization of a planar homology requires 3 pairs of corresponding points. Given two regions that are candidates for an object and its shadow, finding such pairs is challenging since the matching is combinatorial with a cost of $\mathrm{O}(n^3m^3)$ (n number of points on the object boundary and m the number of points on the shadow boundary). If we can reduce the matching to be between any point on the object and a constant number of points of the shadow then clearly the correspondence problem becomes much easier to solve. Then, for any triplet there will be one or at most a small constant number of triplets to consider and the computational complexity is reduced to $\mathrm{O}(n^3)$ if all point triplets of O are to be evaluated and in favorable conditions $\mathrm{O}(1)$ when only a single evaluation is sufficient to determine the object-shadow correspondence question.

Next, we show that the motion of the two regions O and S leads to a tight constraint on the matching between O and S if they are in fact an object and its shadow. Specifically, we show that a planar homology can be computed directly from the motion models of O and S; thus, enabling the matching of a point of O only to points, in S, that are on a single line determined by the homology lines and passing through a

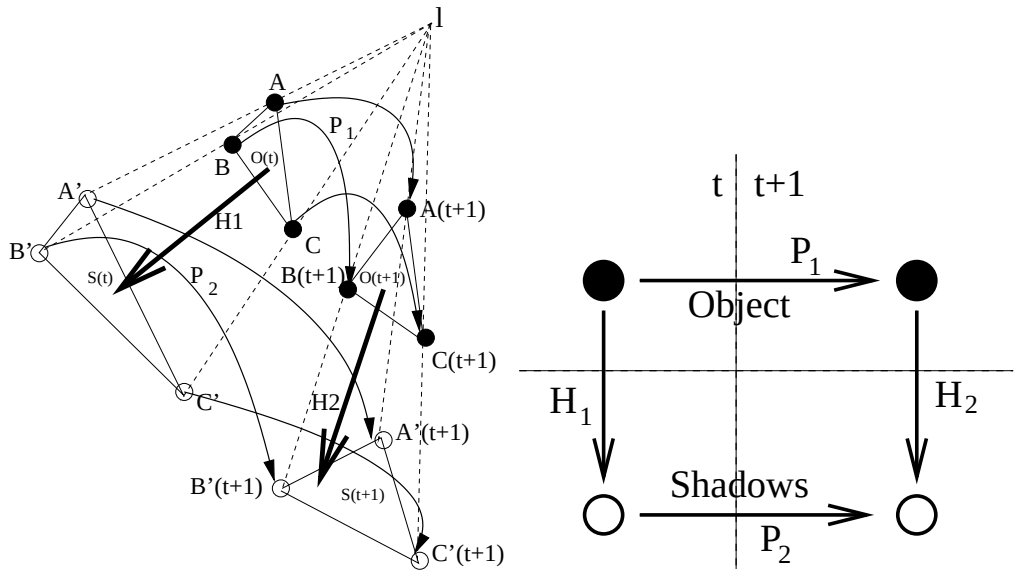


Figure 4: (Left) The stationary and motion constraints that are applicable to a triangle in motion. P_1 transforms $O(t)$ region to $O(t+1)$, P_2 transforms $S(t)$ region to $S(t+1)$, H_1 transforms $O(t)$ region to $S(t)$, H_2 transforms $O(t+1)$ region to $S(t+1)$, (Right) Functional description of the transformations, the time, region-shadow axes and the two paths that derive the equivalence equation.

candidate point O_i. It is critical to note that here lies an important difference with respect to [13] since instead of computing the homology from 3 corresponding object/shadow points we recover the homologies directly from the motion transformations. Typically, it is easier to establish motion point-correspondences over small space and time than determine object-shadow pairs.

Let P_1 be a (3x3) transformation matrix that describes the motion of the region O between time t and $t+1$, P_2 a transformation matrix that describes the motion of the region S between time t and $t+1$. Let us assume for now that these motions can be computed from the image sequence. Let H_1 be the planar homology that relates the region $O(t)$ to the region $S(t)$ at time t and H_2 the planar homology that relates the region $O(t+1)$ to the region $S(t+1)$ at time $t+1$ (see Figure 4 left). H_1 and H_2 are not known. For any point $\mathbf{x}$ of $O(t)$ (represented as a 3-vector in homogeneous coordinates), the following relation must hold (see Figure 4 right)

$$H_2 P_1 \mathbf{x} = P_2 H_1 \mathbf{x} \tag{3}$$

The left side of the equation describes how a point $\mathbf{x}$ in the region $O(t)$ moves to a point on $O(t+1)$ and then

how its shadow point is located in $S(t+1)$. The right side describes how $\mathbf{x}$'s shadow in $S(t)$ is determined and how it moves to the corresponding shadow point in $S(t+1)$. Since H_1 and H_2 have each 5 degrees of freedom the maximum degrees of freedom in Equation 3 are 10. However, since the light source location in the image does not move the vertices of H_1 and H_2 are at the same point. Therefore, the actual number of degrees of freedom is reduced to 8. Therefore, 4 corresponding points are sufficient to determine these 8 values. More importantly, since these transformation matrices are 3x3, there are 9 equations that can be written to overdetermine the solution to these 8 parameters. However, since the parameterization of H_1 and H_2 can be determined only up to a scalar, the relationship between the matrices is actually

$$H_2 P_1 \equiv P_2 H_1 \tag{4}$$

This equivalence can be turned into equality if we introduce a scalar s that implicitly captures how the scalars of H_1 and H_2 may be related

$$s H_2 P_1 = P_2 H_1 \tag{5}$$

where $s = \frac{s_2}{s_1}$ and s_1 and s_2 are the concrete scalars for the homologies H_1 and H_2. Therefore, we now have 9 equations with 9 degrees of freedom which leads to a direct closed-form solution. Given the parameterization in Equation 2, Equation 5 can be rewritten as

$$s(\mathbf{I} + \frac{\mu_2}{\mathbf{v} \cdot \mathbf{a}_2} \mathbf{v} {\mathbf{a}_2}^T) P_1 = P_2 (\mathbf{I} + \frac{\mu_1}{\mathbf{v} \cdot \mathbf{a}_1} \mathbf{v} {\mathbf{a}_1}^T) \tag{6}$$

where μ_1 and μ_2 are the cross ratios of the homologies H_1 and H_2 minus one, respectively (i.e., $\mu = (\lambda_1/\lambda_2 - 1)$ for each matrix), $\mathbf{v}$ is a 3-vector that defines the projected location of the light source in the image, and $\mathbf{a}_1$ and $\mathbf{a}_2$ are the 3-vector axes of homologies H_1 and H_2, respectively. Note that $\mathbf{v}, \mathbf{a}_1, \mathbf{a}_2$ have each 2 degrees of freedom since they are determined up to a scalar. The details of the solution of Equation 6 are given in Appendix A. Equation 6 is non-linear but it can be converted into a degree-eight polynomial in s and therefore results in up to 8 different solutions. While the 8 roots reflect at most 8 solutions, often 2, 4 or

6 are equal. For each root of s the other 8 parameters can be computed in closed form. Due to numerical considerations, once a solution is calculated, it is beneficial to refine it by minimizing the error between the right and left side of Equation 6 using a non-linear minimization such as Levenberg-Marquardt. These solutions are all considered as plausible planar homologies and are evaluated as discussed below.

Given the transformations P_1 and P_2, we can compute two homologies H_1 and H_2 that satisfy equation 6 and describe (a) the possible shadow location $S_i(t)$ of a point $O_i(t)$ and (b) the possible shadow location $S_i(t+1)$ of $O_i(t+1)$ (where $O_i(t+1)$ is location of $O_i(t)$ at $t+1$). Since the planar homologies are computed up to a scalar (recall that s captures the relationship between two implicit scalars s_1 for H_1 and s_2 for H_2), the transformation of $O_i(t)$ according to H_1 determines a line along which the true shadow of point $O_i(t)$ must lie and not just a single point.

The analysis so far does not take into account measurement inaccuracies. These inaccuracies may occur at several stages, (1) the computation of the transformations P_1 and P_2 (2) the minimization that determines the 9 degrees of freedom of Equation 6 (3) the fact that the light source may not be a true point light source or (4) the shadow has a penumbra and as a result shadow region boundary is inaccurate. These inaccuracies can be handled by expanding the matching of a point $O_i(t)$ to be along a line with a predetermined angle, α, with respect to the line computed by the recovered planar homologies H_1 and H_2 (see Figure 5). The choice of α should reflect what is known about the sources of inaccuracy in the image sequence. Note that a larger α increases the computational complexity which defeats the value of the direct homology recovery from the motions of the regions. Then, many more point-correspondences may have to be considered until a large enough α forces matching each point to all other points in the second region.

3.5 Determining Object/Shadow Pairs

The formulation in Equation 6 recovers the two homologies H_1 and H_2 that reflect possible object-shadow pairing at times t and $t+1$. In the rest of the paper we focus on H_1 (i.e., homology at time t) but the same treatment can be applied to H_2. The recovered planar homology allows us to determine if two region boundaries constitute an object and a shadow pair based on the degree to which point-pairs of the object

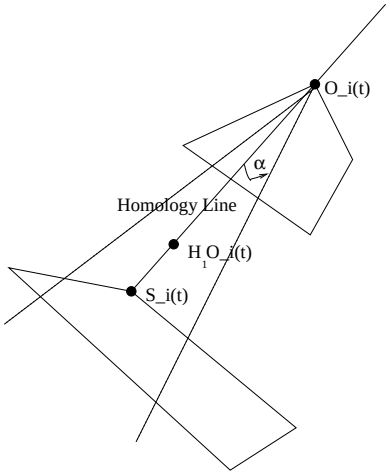


Figure 5: The point $O_i(t)$ corresponds to point $H_1O_i(t)$, the true shadow point of $O_i(t)$ is along the line $O_i(t)$ - $H_1O_i(t)$. Inaccuracy can be handled by allowing the search line to be α from the correct homology line.

and shadow boundaries are explained by this planar homology. If the region of the object and shadow are separate and apart in the image, then most points on the object should have a pairing in the shadow region (and vice versa). On the other hand, if the regions overlap then the pairing may not be complete (see examples in Figures 6 and 9).

To evaluate the estimated object-shadow planar homology, H_1, we

1. Select 4 points (any three must be non-colinear) on the boundary of the candidate object. It is advantageous to choose boundary points that occur at inflection points of the boundary as well as far apart as possible so in our implementation we choose from the vertices of the polygonal approximation of the boundary.

2. Determine all corresponding points on the candidate shadow region as determined by the computed planar homology H_1. As discussed earlier, the homology transformation will dictate a line that goes through the object point and after error measurements are taken into consideration all candidate shadow points within angle α (as measured with respect to the line connecting the candidate shadow point and the object point) of this line are considered candidates. Since vertices on the polygonal approximation of the object correspond to vertices on the polygonal approximation of the shadow

boundary we consider only vertices of the shadow (note, however, that angle-values of vertices are not preserved by the homology and there exist situations where vertices disappear by becoming part of edges but these are not critical to the algorithm since we need at least 4 vertices that are preserved). Compute an exact transformation matrix $\mathbf{T}$ from 4 points and their corresponding points according to H_1.

3. Use $\mathbf{T}$ to determine which points and how many on the object boundary have a correspondent points on the boundary of the shadow region.

Step (2) above may generate more than one set of 4 corresponding points due to the intersection of the lines with multiple boundary points (i.e., polygon vertices in practice), or due to several points being close to the line given the error angle α. All these candidates are subjected to step (3) and the best transformation $\mathbf{T}$ is chosen (the highest number of boundary to boundary matching via $\mathbf{T}$). Note also that since 8 homologies are computed some may not have any 4 pair correspondences and thus can be immediately rejected.

Finally, deciding whether two regions are an object-shadow pair is done by requiring that the percentage of the shadow boundary be explained by the best transformation $\mathbf{T}$ as projection of the object boundary under $\mathbf{T}$ exceed a certain threshold.

4 Experiments

4.1 Computation Issues

Locating image regions corresponding to the candidate object and its cast shadow in the first frame of the sequence is a challenging problem. Several approaches have been proposed in the literature for detecting foreground or independently moving regions in a scene. Specifically, for a stationary camera it is common to use background models to segment the scene into foreground/background regions [4, 5].

In the experiments described in the next subsection we compute a background model prior to the movement of objects into the field of view of the camera. Then, the foreground pixels are detected. In order to

detect two candidate regions for object/shadow pairing we assume in the first four experiments that if the two regions are separated spatially then they represent a potential pair, and if they overlap we classify pixels that are a shaded variant of the background into one region and the rest into another based on the color model proposed in [5]. In the fifth experiment we use a skin color model to detect the hand region and the rest of the foreground is considered the potential shadow region.

Once the regions have been detected, simple image processing steps such as filling holes, connecting components etc. are used to refine the detection. This segmentation need not be done perfectly.

Computing the motion of the regions can be done using a tracking algorithm such as [1] or by finding the best transformation that maps the boundary of a region in one frame to the boundary of the region in the next frame. Since the shadow region motion can only be computed based on its boundary we implemented a boundary-based motion estimation algorithm. Specifically, the boundaries of the regions are fit into a polygonal approximation and 4 point pairs (any 3 must be non-colinear) are selected from the region boundary at time t and $t+1$ to compute the motion transformation matrix directly [14]. The transformation that maps the majority of the points on the boundary between the two images is the best solution.

As discussed earlier the solution to the homology may involve up to 8 different solutions. These solutions are evaluated and the one that explains the largest percentage of boundary-to-boundary matches is considered the best. The largest percentage must also exceed the threshold of 50% for boundary-to-boundary matches.

4.2 Results

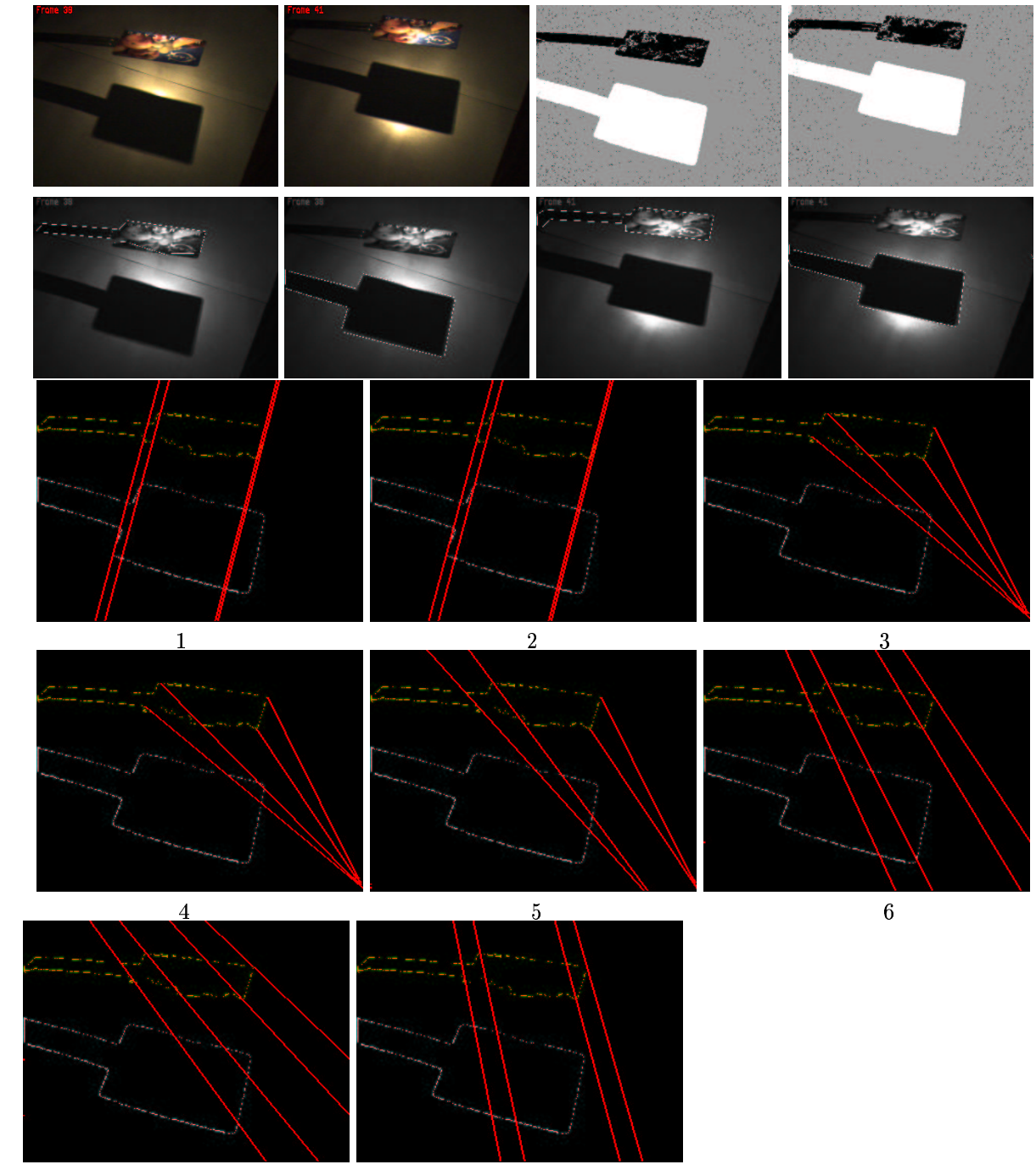
In the experiments below we assume that $\alpha = 17°$. This is slightly larger than the maximum angle that the light source spans over the object in these experiments and by choosing it we explicitly allow a point to move anywhere within the object (and its shadow).

Figure 6 shows two images from a sequence of a card moving and casting a shadow on a plane. The light source is approximately vertically above the object. The images show the two detected regions before image processing enhancements have been applied (white and black regions). The images on the second row show the polygonal approximation of the boundaries of these regions. Finally, the last three rows show the results

of the 8 solutions for the planar homology H_1. Four sample points on the object boundary are used to show the lines on which corresponding shadow points should lie according to each of the 8 computed homologies. As shown in the table, homologies 1 and 2 are the ones that provide the best mapping between object and shadow boundaries. Note that the light source is sufficiently far from the camera (about 6 times the size of the card in image pixels) to create the impression that the planar homology lines are parallel when viewed over this relatively small section. The best boundary match explains 73% of the boundary-to-boundary correspondences (see the table). Regarding the accuracy of the results, it should be noted that the angle of the light source and the object is about 6.6° (as we calculate manually from the first image). The best solution recovered an error-angle between the direction of the light and the farthest object-shadow pair that is 5.4° apart from the actual one. The processing time (rounded) for each homology is given in the table.

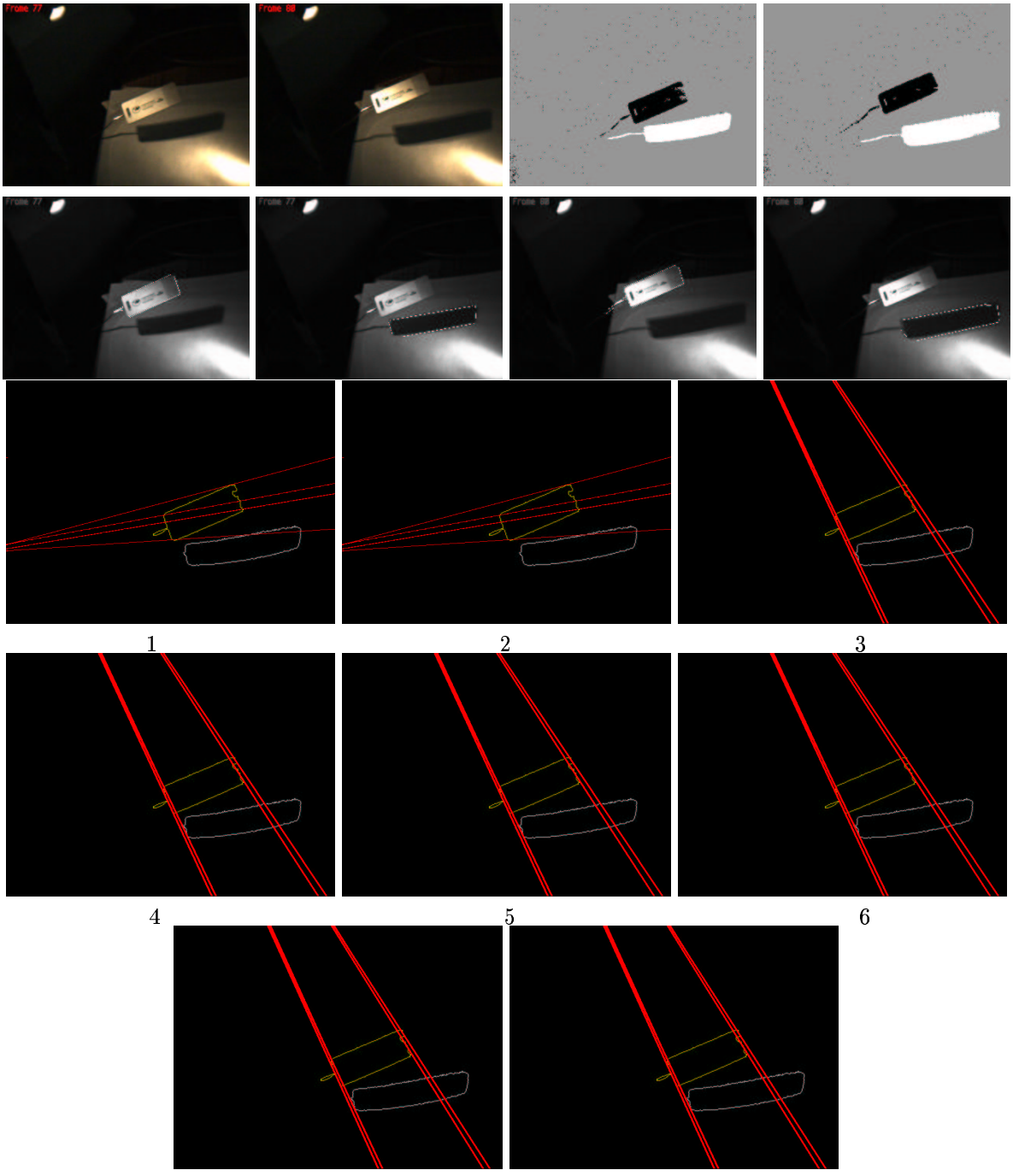
Figures 7 and 8 are similar to Figure 6 with the light source at different positions. In Figure 7 the light source is visible in the image (top left corner). As the figure shows, the 3-8 solutions provide equal and the good interpretation (1-2 are judged to be bad homologies), The angle that the light source spans with the object is 14.0°. The angular error between the true projection and the best computed projection for the farthest point is 11.4°. Figure 8 has the light source in a different position and the camera is closer to the scene. The angle that the light source spans with the object is 13.4°. The error in the solution is 5.1° between the true projection and the computed projection of the farthest point. The 3rd homology (and to lesser extent the 4rth and 5th) provides the best boundary-to-boundary matching results.

Figure 9 shows two images from a sequence of a cardboard cut in the shape of a hand and moving and casting a shadow on a plane. The light source is approximately vertically placed above the object. The images on the first row show the two detected regions before image processing enhancements have been applied. The regions are close and parts of the shadow are inaccurately associated with the object and vice versa. The images on the second row show the polygonal approximation of the boundaries of these regions. Homologies 4 and 5 and to a lesser extent 6 and 7 provide the best mapping between object and shadow boundaries. The best boundary-to-boundary match explains 60% of the correspondences. The angle of the light source and the object is about 6.0° (as we calculate manually from the first image). The best solution



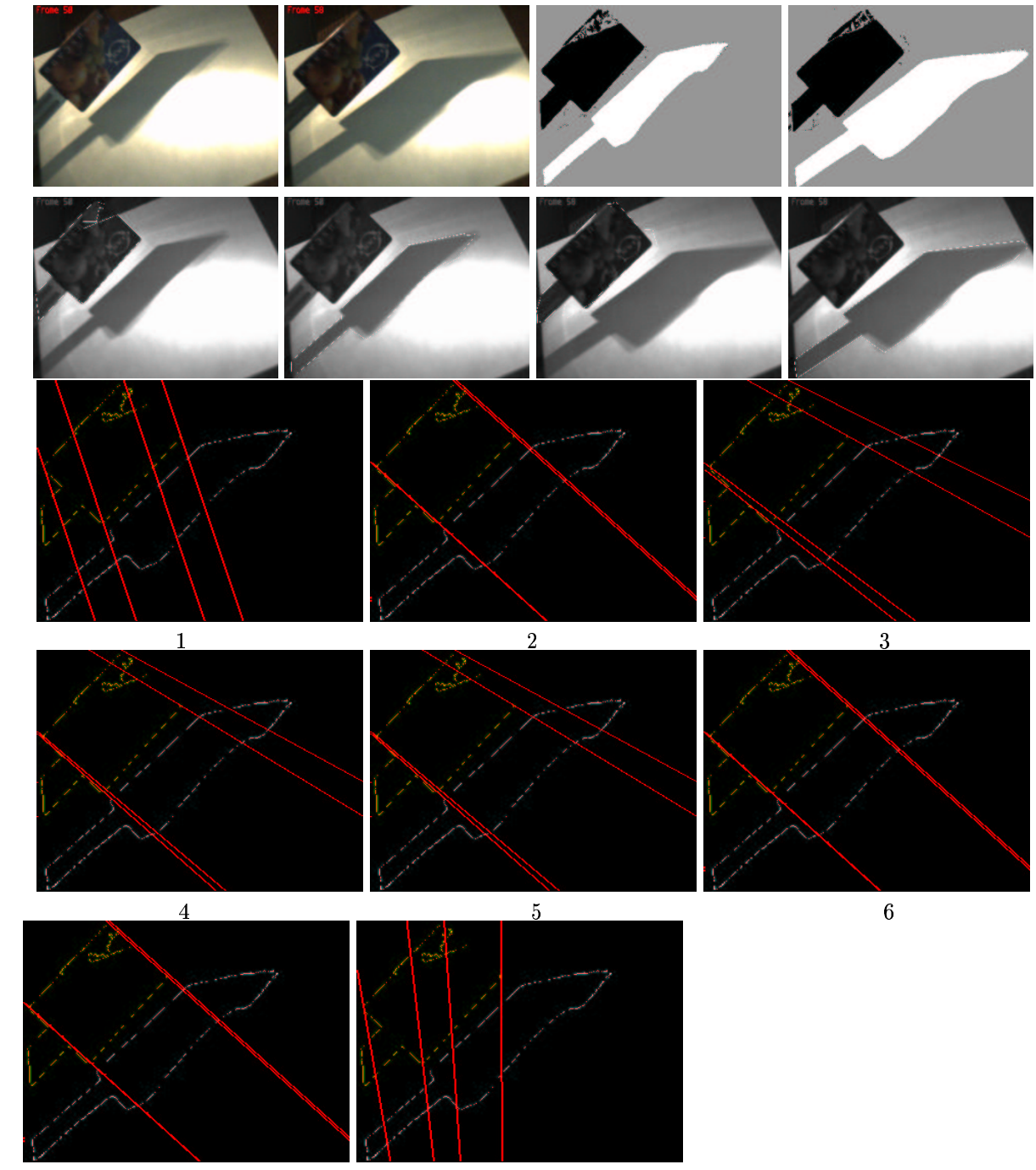
	H1	H2	H3	H4	H5	H6	H7	H8
Match	73%	73%	0%	0%	0%	0%	0%	39%
Time	0.5s	0.5s	0.0s	0.1s	0.0s	0.0	0.1s	0.4s
Error Angle	5.4°	5.4°	32.1°	32.1°	27.9°	17.4°	25.0°	11.2°

Figure 6: A card casting shadow on a plane in two frames (top row, left). Detection of two different regions using background subtraction (top row, right). Fitting the region boundary into a polyline (second row). The 8 different homologies that are computed. The first 2 explain the regions as an object and its shadow sine 73% of the points find correspondence between the boundaries according to the homology. The angle between the light and the object is 6.6°



	H1	H2	H3	H4	H5	H6	H7	H8
Match	0%	0%	83%	83%	83%	83%	83%	83%
Time	0.0s	0.0s	0.2s	0.2s	0.2s	0.2	0.2s	0.2s
Angle	24.9°	24.9°	11.4°	11.4°	11.4°	11.4°	11.4°	11.4°

Figure 7: A card casting shadow on a plane in two frames (top row, left). Detection of two different regions using background subtraction (top row, right). Fitting the region boundary into a polyline (second row). The 8 different homologies that are computed. Solutions 3-8 explain the regions as an object and its shadow.



	H1	H2	H3	H4	H5	H6	H7	H8
Match	34%	38%	50%	46%	46%	38%	38%	0%
Time	0.4s	2.3s	3.0s	3.0s	3.0s	2.1	2.1s	0.0s
Error Angle	27.7°	12.8°	5.1°	6.1°	6.1°	12.8°	12.8°	35.0°

Figure 8: A card casting shadow on a plane in two frames (top row, left). Detection of two different regions using background subtraction (top row, right). Fitting the region boundary into a polyline (second row). The 8 different homologies that are computed. The third and to a lesser extent 4rd and 5th solution explain the regions as an object and its shadow.

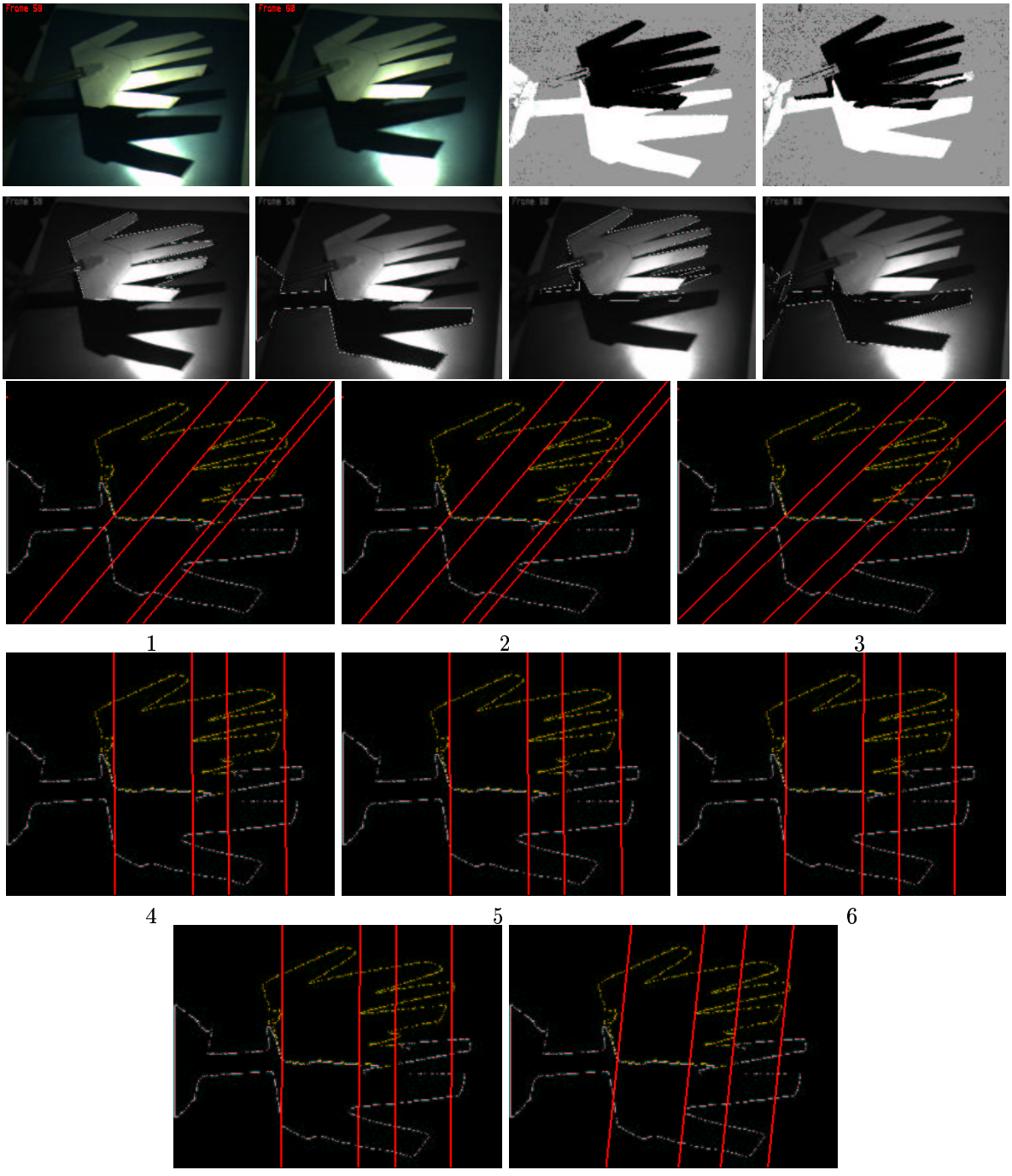
recovered an angle between the direction of the light and the farthest point pair from the correct one that is 5.5° apart from the actual one.

Figure 10 shows two images from a sequence of a hand moving and casting a shadow on a plane. The fingers are held in approximately a planar position. The light source is vertically placed above the hand. The images show the two detected skin regions before image processing enhancements have been applied. The regions are close and parts of the shadow are inaccurately associated with the hand and vice versa. The images on the second row show the boundaries of the two detected regions. Homologies 1-4 (6 and 7 are close as well) provide the best mapping between the hand and its and shadow boundaries. The angle of the light source and the hand is about 10.1° (as we calculate manually from the first image). The best solution recovered an angle between the direction of the light and the farthest point pair from the correct one that is 5.3° apart from the actual one. The best boundary-to-boundary match explains 73% of correspondences.

4.3 Discussion of Experimental results

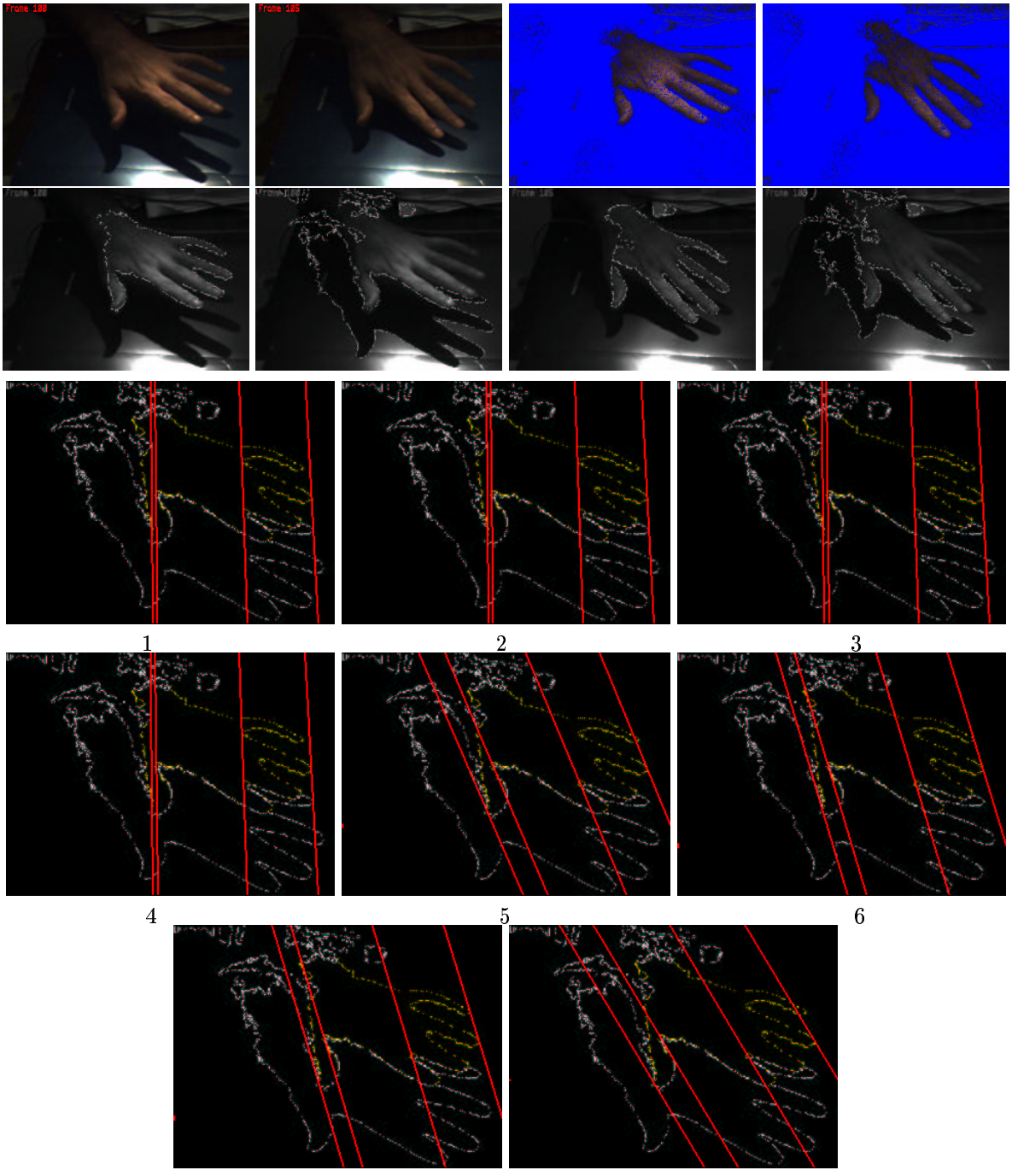
The experiments sought to evaluate the formulation of object-shadow correspondence in motion sequences. Theoretically, the motions of two regions can be used to compute a homology that constrains the only possible correspondences between points of the two regions if these obey an object/shadow pair. However, practically, we recover up to 8 different homologies that need to be evaluated. The results of the experiments confirmed the premise of the work but deviations inevitably occurred. Specifically,

- All but one experiment (Figure 10) involved a real planar object. The experiment with the hand demonstrated that the approximate planarity of the hand can be handled without changes in the algorithm. Coping with significant deviation from planarity without a priori knowledge of object shape may be possible if the object can be approximated by regions of piece-wise planarity. This scenario will require a more complex technical implementation and increased complexity to handle the region to region object/shadow correspondences. However, the formulation of projective/motion properties will remain unchanged.



	H1	H2	H3	H4	H5	H6	H7	H8
Match	45%	45%	44%	62%	62%	60%	60%	53%
Time	15.7s	15.5s	14.2s	10.5s	10.6s	10.8	11.5s	17.6s
Error Angle	20.4	22.3	5.6	5.6	5.6	5.6	5.6	8.3

Figure 9: A card casting shadow on a plane in two frames (top row, left). Detection of two different regions using background subtraction (top row, right). Fitting the region boundary into a polyline (second row). The 8 different homologies that are computed. Solution 4,5 and to a lesser extent 6-8 explain the regions as an object and its shadow.



	H1	H2	H3	H4	H5	H6	H7	H8
Match	73%	73%	73%	73%	54%	64%	64%	48%
Time	39.2s	39.3s	38.0s	38.6s	36.6s	34.8	35.0s	36.6s
Error Angle	5.3	5.3	5.3	5.3	16.3	13.5	13.5	20.4

Figure 10: A card casting shadow on a plane in two frames (top row, left). Detection of two different regions using background subtraction (top row, right). Fitting the region boundary into a polyline (second row). The 8 different homologies that are computed. The first solution (and to a lesser extent 2-4) explain the regions as the hand and its shadow.

- The assumptions of a point-light source did not hold in practice. This is clearly visible in Figure 7 as the light is visible in the image and it is sizeable. As a result, inaccuracy in the estimation of the 8 homologies was inevitable and consequently the large α accommodated this violation while increasing computational cost.

- The algorithm can cope with inaccuracies in the boundaries of the object and the shadow regions. For example, in most of the figures a penumbra can be observed in certain parts. Also, in Figures 7 the boundary of the shadow appears to be curved instead of being straight (either the cast surface was slightly warped or the light source proximity and size created this effect). In Figure 7 the light source and the camera properties make a straight line break into several piece-wise straight lines. Figures 9 and 10 show that local inaccuracies of the boundaries can be coped with. In Figure 9 part of the cardboard finger is considered a shadow and vice versa. While in Figure 10 the area close to the wrist and its corresponding shadow area are poorly segmented and their boundaries inaccurate. Automatic fitting of the boundaries to polylines reduced high frequency errors while evaluating the homologies at all plausible point-pairings enabled determining the homology that explains the maximal number of point correspondences between the object and shadow. It is worth noting that inaccuracies in boundary detection may not affect the homology estimation since it is based on motion transformations (as long as the motion is accurate). Inaccuracy may, however, affect the value of the threshold of boundary-to-boundary correspondences.

- The experiments used exactly the same parameters despite the variations in the imaging conditions (two different cameras were used, different objects, and different layouts of the scene). The variations in performance have been relatively small.

- Most but not all of the eight recovered homologies were near the correct one. This suggests that the solution to Equation 6 did not deviate too far despite our non-linear optimization near the calculated closed-form homologies. While theoretically there is a single solution, the measurement inaccuracies and the non-linearity of the equations lead to multiple solutions.

- The detection and motion estimation of the two regions in the image are prerequisites to the algorithm. We employed straightforward image processing techniques to accomplish this. This detection is essentially a region segmentation problem that continues to attract much research and as a result remained out of the scope of this paper.

5 Summary and Future Research

In this paper we developed a formulation under which the object/shadow segmentation and correspondence can be achieved in a motion sequence. Specifically 3D planar objects and their cast shadows projected on 3D planes offer geometric constraints that support a general solution of the problem using projective geometry and motion models. These basic principles eliminate the need to employ application specific assumptions for describing how objects and shadows appear in images.

Several sequences were captured to demonstrate the principles and the performance of the proposed algorithm. The algorithm is computationally efficient and can easily be integrated with various foreground/background detection techniques. However, we observed in our experiments that imaging and numeric factors play a role in the accuracy of the computation. The fact that the light source was never a true light source did not fulfill the assumptions of our model. Numerically, the estimation of the homology was improved by posing the problem as a non-linear minimization using the closed-form solution as a starting point.

Extending the principles employed in this paper to non-planar objects and cast shadow surfaces is possible if it is assumed that deviation from planarity translates into bounded deviation in the scene projected on the camera plane. Alternatively, piece-wise planarity requires determining the piece-wise homologies of permutations of object/shadow regions.

Determining candidate two regions in images is variation on the image segmentation problem and it remains a challenge. Scene and application constraints can be employed to compute these regions.

The treatment of general 3D objects and general cast shadow surfaces such as walls in an urban area

or curved surfaces remains a challenge. Also, the cases of camera motion and simultaneous camera object motion remain to be studied. Finally, indoor lighting may include multiple light sources and remains a serious challenge.

6 Appendix A

We solve the equation:

$$sH_2P_1 = P_2H_1 \tag{7}$$

We replace H_2 by variable Y_1 and H_1 by Y_2. The homologies can be reparameterized as

$$s(\mathbf{I} + \frac{\mu_1}{\mathbf{v}\cdot\mathbf{a}_1}\mathbf{v}\mathbf{a}_1{}^T)P_1 = P_2(\mathbf{I} + \frac{\mu_2}{\mathbf{v}\cdot\mathbf{a}_2}\mathbf{v}\mathbf{a}_2{}^T) \tag{8}$$

can be simplified to

$$sP_1 - P_2 + s\frac{\mu_1}{\alpha}\mathbf{v}\mathbf{a}_1{}^TP_1 = \frac{\mu_2}{\beta}P_2\mathbf{v}\mathbf{a}_2{}^T \tag{9}$$

where $\alpha = \mathbf{v}\cdot\mathbf{a}_1$ and $\beta = \mathbf{v}\cdot\mathbf{a}_2$ and

$$P_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{10}$$

$$-P_2 = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \tag{11}$$

Let the columns of P_1 be represented as $Q_1 = (a_{11}, a_{21}, a_{31})$, $Q_2 = (a_{12}, a_{22}, a_{32})$, $Q_3 = (a_{13}, a_{23}, a_{33})$, and the rows of P_2 be represented as $R_1 = (b_{11}, b_{12}, b_{13})$, $R_2 = (b_{21}, b_{22}, b_{23})$, $R_3 = (b_{31}, b_{32}, b_{33})$.

Let the vertex be $\mathbf{v} = (a, b, 1)$, axis of H_2 be $\mathbf{a}_1 = (c, d, 1)$ and axis of H_1 be $\mathbf{a}_2 = (e, f, 1)$. The following

values can be represented as

$$\mathbf{va}_1{}^T = \begin{bmatrix} ac & ad & a \\ bc & bd & b \\ c & d & 1 \end{bmatrix} \tag{12}$$

and

$$\mathbf{va}_2{}^T = \begin{bmatrix} ae & af & a \\ be & bf & b \\ e & f & 1 \end{bmatrix} \tag{13}$$

We can write 9 equations stemming from equation 9

$$sa_{11} + b_{11} + \frac{sa\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_1) = -\frac{\mu_2 e}{\beta} R_1 \cdot \mathbf{v} \tag{14}$$

$$sa_{12} + b_{12} + \frac{sa\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_2) = -\frac{\mu_2 f}{\beta} R_1 \cdot \mathbf{v} \tag{15}$$

$$sa_{13} + b_{13} + \frac{sa\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_3) = -\frac{\mu_2}{\beta} R_1 \cdot \mathbf{v} \tag{16}$$

$$sa_{21} + b_{21} + \frac{sb\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_1) = -\frac{\mu_2 e}{\beta} R_2 \cdot \mathbf{v} \tag{17}$$

$$sa_{22} + b_{22} + \frac{sb\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_2) = -\frac{\mu_2 f}{\beta} R_2 \cdot \mathbf{v} \tag{18}$$

$$sa_{23} + b_{23} + \frac{sb\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_3) = -\frac{\mu_2}{\beta} R_2 \cdot \mathbf{v} \tag{19}$$

$$sa_{31} + b_{31} + \frac{s\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_1) = -\frac{\mu_2 e}{\beta} R_3 \cdot \mathbf{v} \tag{20}$$

$$sa_{32} + b_{32} + \frac{s\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_2) = -\frac{\mu_2 f}{\beta} R_3 \cdot \mathbf{v} \tag{21}$$

$$sa_{33} + b_{33} + \frac{s\mu_1}{\alpha}(\mathbf{a}_1 \cdot Q_3) = -\frac{\mu_2}{\beta} R_3 \cdot \mathbf{v} \tag{22}$$

From 16 and 19 and Eqs 22 we can derive

$$(sa_{23} + b_{23} - sba_{33} - bb_{33})(R_1\mathbf{v} - aR_3\mathbf{v}) = (sa_{13} + b_{13} - saa_{33} - ab_{33})(R_2\mathbf{v} - bR_3\mathbf{v}) \tag{23}$$

From Eqs 15 18 and 21 we can derive

$$(sa_{12} + b_{12} - saa_{32} - ab_{32})(R_2\mathbf{v} - bR_3\mathbf{v}) = (sa_{23} + b_{22} - sba_{32} - bb_{32})(R_1\mathbf{v} - aR_3\mathbf{v}) \tag{24}$$

From Eqs 14 17 and 20 we can derive

$$(sa_{11} + b_{11} - saa_{31} - ab_{31})(R_2\mathbf{v} - bR_3\mathbf{v}) = (sa_{21} + b_{21} - sba_{31} - bb_{31})(R_1\mathbf{v} - aR_3\mathbf{v}) \tag{25}$$

Equations 23 24 and 25 have three variables, a, b, s (recall that $\mathbf{v} = (a, b, 1)$) and therefore they constrain a solution.

Dividing Eq 23 by 24 (with switching of the sides) and simplifying leads to

$$\begin{aligned} &a(s^2(a_{33}a_{22} - a_{32}a_{23}) + s(a_{33}b_{22} + a_{22}b_{33} - a_{23}b_{32} - a_{32}b_{23}) + (b_{33}b_{22} - b_{32}b_{23}))+ \\ &b(s^2(a_{32}a_{13} - a_{12}a_{33}) + s(a_{13}b_{32} + a_{32}b_{13} - a_{33}b_{12} - a_{12}b_{33}) + (b_{13}b_{32} - b_{12}b_{33}))+ \\ &s^2(a_{23}a_{12} - a_{13}a_{22}) + s(a_{23}b_{12} + a_{12}b_{23} - a_{13}b_{22} - a_{22}b_{13}) + (b_{23}b_{12} - b_{13}b_{22}) \quad = 0 \end{aligned} \tag{26}$$

Similarly, dividing Eq 23 by 25 (with switching of the sides) and simplifying leads to

$$\begin{aligned} &a(s^2(a_{33}a_{21} - a_{31}a_{23}) + s(a_{33}b_{21} + a_{21}b_{33} - a_{23}b_{31} - a_{31}b_{23}) + (b_{33}b_{21} - b_{31}b_{23}))+ \\ &b(s^2(a_{31}a_{13} - a_{11}a_{33}) + s(a_{13}b_{31} + a_{31}b_{13} - a_{33}b_{11} - a_{11}b_{33}) + (b_{13}b_{31} - b_{11}b_{33}))+ \\ &s^2(a_{23}a_{11} - a_{13}a_{21}) + s(a_{23}b_{11} + a_{11}b_{23} - a_{13}b_{21} - a_{21}b_{13}) + (b_{23}b_{11} - b_{13}b_{21}) \quad = 0 \end{aligned} \tag{27}$$

Similarly, dividing Eq 24 by 25 (with switching of the sides) and simplifying leads to

$$\begin{aligned} &a(s^2(a_{32}a_{21} - a_{31}a_{22}) + s(a_{32}b_{21} + a_{21}b_{32} - a_{22}b_{31} - a_{31}b_{22}) + (b_{32}b_{21} - b_{31}b_{22}))+ \\ &b(s^2(a_{31}a_{12} - a_{11}a_{32}) + s(a_{13}b_{12} + a_{12}b_{31} - a_{32}b_{11} - a_{11}b_{32}) + (b_{12}b_{31} - b_{11}b_{32}))+ \\ &s^2(a_{22}a_{11} - a_{12}a_{21}) + s(a_{22}b_{11} + a_{11}b_{22} - a_{12}b_{21} - a_{21}b_{12}) + (b_{22}b_{11} - b_{12}b_{21}) \quad = 0 \end{aligned} \tag{28}$$

The three equations 26, 27 and 28 can easily be combined into a polynomial of 8th degree in s. There are upto 8 roots for this polynomial (several may be the same). Once s is found, the rest of the parameters can be computed from the previous equations. Specifically, a and b are easily recoverable from Equations 26-28 and e, f, c, d, μ_1 and μ_2 from Equations 14-22.

Due to imaging and numerical considerations it is beneficial to use each computed solution as the starting point of a non-linear minimization of the error between the left and right wides of Equation 7. A Levenberg-Marquardt minimization is used.

References

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE PAMI*, 7(4), 1985, 384-401.

[2] M. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motions. *IJCV*, 25(1), 1997, 23-48.

[3] R. Cucchiara, C. Grana, M. Piccardi and A. Prati. Detecting Objects, Shadows and Gosts in Video Streams by Exploiting Color and Motion Information. Proceedings of 11th International Conference on Image Analysis and Processing (ICIAP 2001), September 2001.

[4] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis. Background and Foreground Modeling using Non-parametric Kernel Density Estimation for Visual Surveillance, Proceedings of the IEEE, July 2002, 1151 -1163.

[5] T. Horprasert, D. Harwood, and L.S. Davis, A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection, Proc. ICCV FRAME-RATE Workshop, Greece, September 1999.

[6] H.C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image. *Proc. Royal Society of London, B*, 208, 1980, 385-397.

[7] I. Mikic, P.C. Cosman, G.T. Kogut and M.M. Trivedi. Moving Shadow and Object Detection in Traffic Scenes. ICPR 2000.

[8] S. Noronah and R. Nevatia. Detection and Modeling of Buildings from Multiple Arial Images. IEEE PAMI, (23)5, 2001, 501-518.

[9] Y. Sonoda, T. Ogata. Separation of Moving Objects and their Shadows, and Applications to Tracking of Loci and Monitoring Images. Proceedings of ICIP, 1998, 1261-1264.

[10] C.E. Springer, *Geometry and Anlaysis of Projective Spaces*, W.H. Freeman and Company, 1964.

[11] J. Stauder, R. Mech and J. Ostermann. detection of Moving Cast Shadows for Object Segmentation. IEEE Trans.i on Multimedia, (1)1, 1999, 65-76.

[12] M. Subbarao and A.M. Waxman, Closed form solutions to image flow equations for planar surfaces in motion. *CVGIP*, 36, 1986, 208-288.

[13] L. Van Gool, M. Proesmans and A. Zisserman, Planar Homologies as a basis for Grouping and Recognition, Image and Vision Computing, 16, 1998, 21-26.

[14] G. Wolberg. *Digital Image Warping.* IEEE Computer Society Press, 1990.