On the Use of Shadows in Stance Recovery

Alfred M. Bruckstein,[1] Robert J. Holt,[1] Yves D. Jean,[2] Arun N. Netravali[1]

[1] Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974

[2] Avaya Communication, Murray Hill, NJ 07974

“ . . . the highest sum would be too little to pay for such a priceless shadow.”—A. von Chamisso, Peter Schlemiel: the man who sold his shadow.

ABSTRACT: The image of an object and of the shadow it casts on a planar surface provides important cues for three-dimensional (3D) stance recovery. We assume that the position of the plane on which the shadow lies with respect to a pinhole camera is known and that the position of the light source is unknown. If the light source is sufficiently far away that parallel projection may be assumed, then knowledge of two point correspondences between images of feature points and images of their shadows is enough to determine the position of the object and the direction of the light source. If the light source is close enough that the shadow points are obtained via perspective projection, then there is a one-parameter infinite family of solutions for the position of the object and the light source. Determining the stance of an object is highly sensitive to noise, so we provide algorithms for stance recovery that take into account known information about the object. In our experiments, the errors for the location of the 3D feature points obtained by these algorithms are generally less than 0.2% times the error in pixels in the image points and the errors for the 3D directions of the links is roughly 0.04° times the error in pixels, normalized by the distance to the object from the camera and the length of the link. © 2001 John Wiley & Sons, Inc. Int J Imaging Syst Technol, 11, 315–330, 2000

I. INTRODUCTION

Suppose an articulated object, like the human body, is seen in an image along with the shadow it casts on a planar surface, e.g., the ground or a wall. It is clear that the additional information provided by the shadow should make it easier for the viewer to assess the object’s three-dimensional (3D) location and shape. In this paper, we investigate the topic of shape and pose recovery from images of objects casting shadows from a strong single source of light such as the sun or a nearby omnidirectional light source.

There are few papers in the computer vision literature that report work on using shadows to recover scene geometry. The earliest work is perhaps due to Shafer (1985) and Shafer and Kanade (1983). They analyzed the role that shadows and silhouettes play in the automatic interpretation of images of solid objects under various viewpoints and illuminations. Shadows were also used to delineate and locate objects in images (Thompson et al., 1987) and to analyze scenes with electronic components (Tsuji et al., 1984). Researchers used shadows to analyze aerial imagery, mostly of urban or industrial areas (Nevatia, 1998; Shufelt, 1996, 1999).

A later development is the work of Kender and Smith (1987) and Yang and Kender (1996). They proposed an active vision method of illuminating a scene with a moving light source and learning about the scene geometry from the shadows that vary in time. Paralleling the work in computer vision and following the footsteps of artists like Leonardo da Vinci who realized the importance of shadows in rendering realistic scenes by providing a qualitative sense of depth, vision researchers assessed the importance of shadows for human image interpretation (for works on depth and motion perception, see Kersten et al., 1996, 1997; Yonas et al., 1978). Knill et al. (1997) summarized the geometric issues involved in shadow formation and analyzed in depth the ways shadows provide perception cues for scenes of objects with smooth boundaries. The use of shadows in photography was also investigated by Bouguet and Perona (1998, 1999).

As far as we know, the problem of viewing articulated thin objects in strong single source illumination, with a ground plane that is accurately located with respect to the camera, has not yet been discussed. Although these might be considered rather restrictive assumptions, they are realistic in a variety of practical applications, such as interpreting scenes at various sports events (e.g., tennis, soccer) where people tracking and stance recovery are needed for automated understanding and virtual replay.

This paper is organized as follows: the fundamentals of pose from shadows are discussed in Section II, least squares solutions for pose estimation are provided in Section III, and practical consideration together with experiments on synthetic and real images are presented in Sections IV and V.

II. FUNDAMENTALS OF POSE FROM SHADOWS

We will assume that an articulated 3D object is viewed under strong illumination by a perspective projection camera. The object casts shadows on a (background) plane and its shadows, by assumption, are also at least partially visible in the image. The illumination that

Correspondence to: Robert J. Holt. E-mail: rjh@research.bell-labs.com

Alfred M Bruckstein’s permanent address: Department of Computer Science, Technion—IIT 32000, Haifa, Israel.

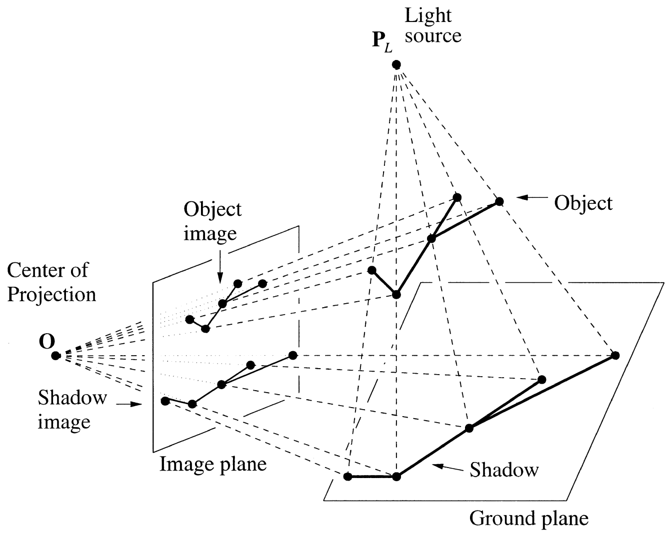

Figure 1. An object casts a shadow on the ground plane and the object and shadow are projected onto the image plane.

we assume is either a strong point source of light located not too far from the scene or a parallel illumination from a strong distant source like the sun (Fig. 1).

The questions we address are: Given the image of the object and its corresponding shadow, what are the constraints on the 3D world that provided the image? What additional items are necessary in order to recover fully the pose of the articulated object?

To begin our investigation, we address a simple case. It is the problem of recovery of pose for a rigid rod from its image and the image of its shadow on a plane. This simple instance of the (general) problem already provides most of the relevant insights and results for the general articulated object pose recovery.

In this paper, points in 3-space and their coordinates are denoted by capital letters, as in $\mathbf{P}_i = (X_i, Y_i, Z_i)$. The projection of points onto the image plane is denoted by the corresponding lower case letters, e.g., $\mathbf{p}_i = (x_i, y_i, F)$, where F is the focal length, the image plane is $Z = F$, and the camera center is the origin $\mathbf{O}$. The shadow point corresponding to $\mathbf{P}_i$ will be denoted by $\mathbf{P}_i^s = (X_i^s, Y_i^s, Z_i^s)$ and the image of the shadow point by $\mathbf{p}_i^s = (x_i^s, y_i^s, F)$.

Example: Rod Positioning Recovery from Image with Shadow

If the rod endpoints are $\mathbf{P}_1$ and $\mathbf{P}_2$ and their shadows are $\mathbf{P}_1^s$ and $\mathbf{P}_2^s$, then $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_1^s$, and $\mathbf{p}_2^s$ are their images as seen by the camera (Fig. 2). Assuming we can identify corresponding points and shadow pairs (i.e., $(\mathbf{p}_1, \mathbf{p}_1^s)$ and $(\mathbf{p}_2, \mathbf{p}_2^s)$), the following results are immediate consequences of the viewing and lighting projections:

1. The planes $(\mathbf{OP}_1\mathbf{P}_1^s)$ and $(\mathbf{OP}_2\mathbf{P}_2^s)$ intersect along the line $\overleftrightarrow{\mathbf{OP}_L}$ (if the light source is a point source at $\mathbf{P}_L$, as we assume for the time being).
2. The lines $\overleftrightarrow{\mathbf{p}_1\mathbf{p}_1^s}$ and $\overleftrightarrow{\mathbf{p}_2\mathbf{p}_2^s}$ meet at $\mathbf{p}_L$ (the projection into the image plane of $\mathbf{P}_L$). Hence, we can determine the line $\overleftrightarrow{\mathbf{Op}_L}$ on which $\mathbf{P}_L$ is located in space.
3. If we know the relative position of the ground plane with respect to the camera, then we can have readily the 3D locations of $\mathbf{p}_1^s$ and $\mathbf{p}_2^s$. For every possible $\tilde{\mathbf{P}}_L \in \overleftrightarrow{\mathbf{Op}_L}$, we shall have a rod $\overline{\tilde{\mathbf{P}}_1\tilde{\mathbf{P}}_2}$ that could have projected into the image seen. Hence, one more piece of information is required for complete recovery of the 3D geometry (i.e., light source position $\mathbf{P}_L$ and the rod $\overline{\mathbf{P}_1\mathbf{P}_2}$). This can be the length of the rod, the height of either $\mathbf{P}_1$ or $\mathbf{P}_2$, the angle the rod makes with the ground plane, the height of $\mathbf{P}_L$ above the ground plane, or the distance from $\mathbf{P}_L$ to the camera center ($\mathbf{O}$).

From the example above, we realize that any constellation of points $\{\mathbf{P}_1, \ldots, \mathbf{P}_K\}$ in 3D can be identified readily from their image when the corresponding shadows can be identified and it is known that a point source cast these shadows upon a known ground plane. The lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ in the image plane will form a pencil of lines passing through $\mathbf{p}_L$, as the planes $(\mathbf{Op}_i\mathbf{p}_i^s) \equiv (\mathbf{OP}_i\mathbf{P}_i^s)$ form a pencil of planes passing through $\overleftrightarrow{\mathbf{OP}_L}$ in 3D. One more item of information about the light source or the constellation of points (i.e., the height of one of them above the ground) will complete the whole 3D picture. The additional piece of information that enables complete 3D recovery can also be some prior knowledge about the configuration of points in space. For example, if we know that three points $\mathbf{P}_i$, $\mathbf{P}_j$, $\mathbf{P}_k$ are the vertices of an equilateral or isosceles triangle, i.e., we have $\mathbf{P}_i\mathbf{P}_j = \mathbf{P}_i\mathbf{P}_k$, this already disambiguates the recovery. If we know that four points are coplanar, this too provides enough constraints to achieve complete 3D recovery. If the points $\{\mathbf{P}_1, \ldots, \mathbf{P}_K\}$ are the vertices of an articulated 3D object made of interconnected rigid (linear) rods, the knowledge of the length ratio between two links will also suffice to identify completely the object from one image.

Therefore, when we have a 3D object that is simple enough to provide several feature points (like the well-defined joints of articulated bodies), its image and the image of the shadow it projects from a point source on a known plane (e.g., ground, wall) provide

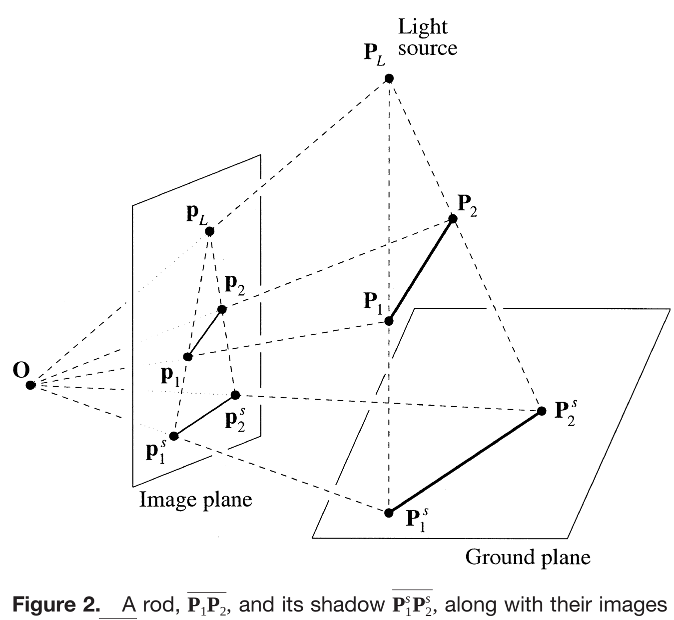

Figure 2. A rod, $\overline{\mathbf{P}_1\mathbf{P}_2}$, and its shadow $\overline{\mathbf{P}_1^s\mathbf{P}_2^s}$, along with their images $\overline{\mathbf{p}_1\mathbf{p}_2}$ and $\overline{\mathbf{p}_1^s\mathbf{p}_2^s}$.

enough information for 3D recovery up to a one-parameter uncertainty. An additional piece of information (e.g., the height of the light source) disambiguates completely the situation. In many cases of practical importance, the ground plane location and the height of the light source above ground are known. Therefore, unambiguous 3D recovery of even complex objects will be feasible, with the availability of images with shadows.

A. The Case of Parallel Illumination. When the light source is far away (like for example the sun, but also in a variety of other cases of interest), the lines in 3D connecting feature points with their shadows will be parallel (rather than forming a pencil through $\mathbf{P}_L$). The lines $\overleftrightarrow{\mathbf{P}_i\mathbf{P}_i^s}$ will have the form

$$\mathbf{P}(t) = \mathbf{P}_i + \mathbf{n}^L \cdot t$$

where $\mathbf{n}^L = (n_x^L, n_y^L, n_z^L)$ is the normalized direction of the light coming from the light source in the camera-centered coordinate system. $((n_x^L)^2 + (n_y^L)^2 + (n_z^L)^2 = 1.)$

In the image of the 3D configuration of points $\mathbf{P}_i$ and their shadows $\mathbf{P}_i^s$, the $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ are generally not parallel, but still form a pencil of lines meeting at a point $\mathbf{p}_L$, the image of the light source obtained through the parallel projection (Fig. 3). Note that $\mathbf{p}_i$ and $\mathbf{p}_i^s$ lie on the plane determined by the parallel lines $\overleftrightarrow{\mathbf{P}_i\mathbf{P}_i^s}$ and $\overleftrightarrow{\mathbf{O}\mathbf{p}_L}$ because they lie on the lines $\overleftrightarrow{\mathbf{O}\mathbf{P}_i}$ and $\overleftrightarrow{\mathbf{O}\mathbf{P}_i^s}$, respectively. Thus, $\mathbf{p}_i$, $\mathbf{p}_i^s$, and $\mathbf{p}_L$ all lie on both the image plane and the plane through $\mathbf{O}$, $\mathbf{P}_i$, and $\mathbf{P}_i^s$, and consequently lie on the line of intersection of these two planes and are collinear. A straightforward calculation shows that the coordinates of $\mathbf{p}_L$, the point of intersection of the image plane $Z = F$ and the line through the origin with direction $\mathbf{n}^L$, satisfy

$$\mathbf{p}_L = (x_L, y_L, z_L) = \left(\frac{Fn_x^L}{n_z^L}, \frac{Fn_y^L}{n_z^L}, F\right).$$

Hence, $\mathbf{p}_L$ determines $\mathbf{n}^L$ as

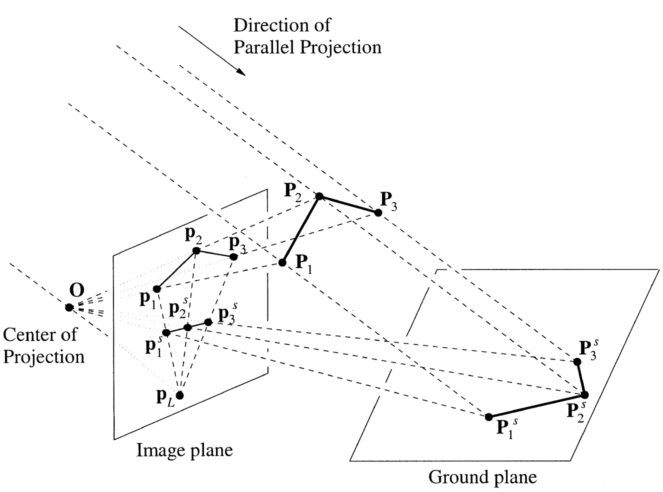

Figure 3. Parallel projection. For each *i*, lines $\overleftrightarrow{\mathbf{P}_i\mathbf{P}_i^s}$ and $\overleftrightarrow{\mathbf{O}\mathbf{p}_L}$ are parallel and the points $\mathbf{P}_i, \mathbf{P}_i^s$, $\mathbf{p}_i$, $\mathbf{p}_i^s$, $\mathbf{p}_L$, and $\mathbf{O}$ are all coplanar.

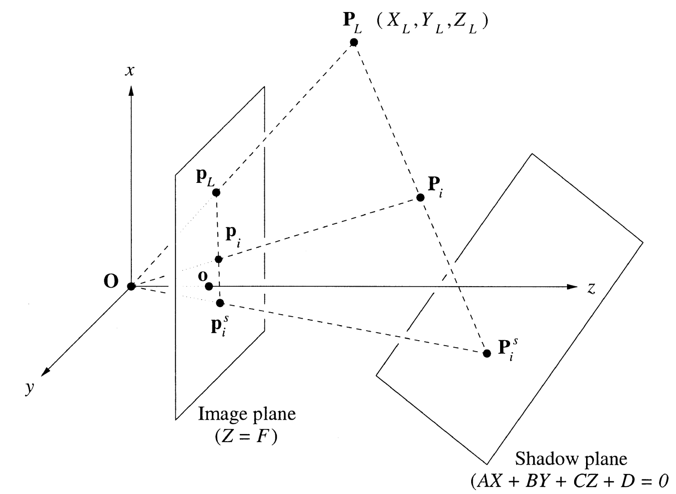

Figure 4. Image geometry for points and shadows. The image plane is taken to be $Z = F$. Shown is a feature point $\mathbf{P}_i$, its shadow $\mathbf{P}_i^s$, the light source $\mathbf{P}_L$, and their images $\mathbf{p}_i$, $\mathbf{p}_i^s$, and $\mathbf{p}_L$.

$$\mathbf{n}^L = \frac{\mathbf{p}_L}{\|\mathbf{p}_L\|} = \frac{(x_L, y_L, F)}{\sqrt{x_L^2 + y_L^2 + F^2}}.$$

(If the light source is in front of the camera, then $\mathbf{n}^L = -\mathbf{p}_L/\|\mathbf{p}_L\|$.) The image lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ are parallel only when the image plane $Z = F$ and $\mathbf{n}$ are parallel, and then $n_z^L = 0$. In this case, the lines parametrized as $(x, y) = (x_i + n_x^L t, y_i + n_y^L t)$ in the image plane uniquely specify (n_x^L, n_y^L) up to a sign.

If we regard $\mathbf{P}_i(t) = \mathbf{P}_i + \mathbf{n}^L t$ as a point on line $\overleftrightarrow{\mathbf{P}_i\mathbf{P}_i^s}$, as $t \to \infty$, we have

$$\begin{aligned}
\lim_{t\to\infty} \mathbf{p}_i(t) &= \lim_{t\to\infty} \frac{F\mathbf{P}_i(t)}{Z_i(t)} \\
&= \lim_{t\to\infty} F\left(\frac{X_i + n_x^L t}{Z_i + n_z^L t}, \frac{Y_i + n_y^L t}{Z_i + n_z^L t}, 1\right) \\
&= \left(\frac{Fn_x^L}{n_z^L}, \frac{Fn_y^L}{n_z^L}, F\right) \\
&= \mathbf{p}_L.
\end{aligned}$$

$\mathbf{p}_L$ is the vanishing point corresponding to lines parallel to the direction of illumination.

III. MATHEMATICAL ANALYSIS (AND STATISTICS) OF POSE RECOVERY FROM SHADOWS

The previous section discussed the problem of 3D recovery from images of feature points and their corresponding shadows cast on a known ground plane. We showed how to compute $\mathbf{n}_L$ and $\mathbf{p}_L$ from the image features under both parallel and perspective illumination. In this section, we derive the necessary equations to complete the recovery of pose as well as practical considerations. In typical situations, the points in the image plane will be specified with some

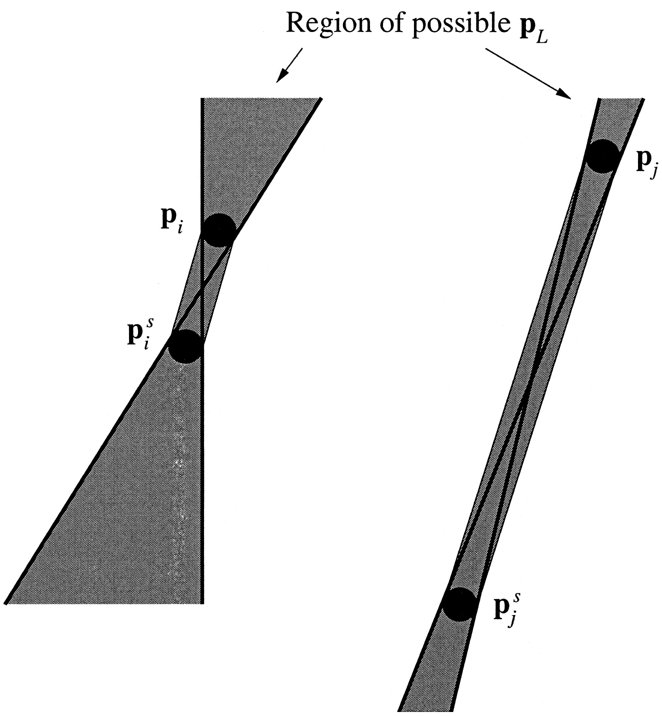

Figure 5. When the image and shadow image points are far apart, there is a narrower region in which the image of the light source may be located.

positional errors. We are interested in estimating the 3D structure from noisy data.

A. Computation of $\mathbf{p}_L$ under Noisy Conditions. Let us formulate the problem and write the relevant equations for 3D recovery from (possibly noisy) corresponding feature points in the image. We set the coordinate system so that the camera is located at the origin and looks toward the depth direction z and we assume that the image projection plane is $Z = F$ (Fig. 4). The shadow is projected onto a plane defined by

$$AX + BY + CZ + D = 0.$$

The light source ($\mathbf{P}_L$) coordinates (if finite) will be (X_L, Y_L, Z_L) if we assume a point source illumination. We could also assume a faraway source and define a normalized direction vector (n_x^L, n_y^L, n_z^L) specifying the direction of the parallel rays of light falling on the scene.

We assume that (A, B, C, D) is known a priori (i.e., we have a fully calibrated camera in the environment) and that the data we are gathering are pairs of feature point images and images of their shadows on the shadow plane. The data points are pairs of points in the image plane $(\mathbf{p}_i, \mathbf{p}_i^s)$, i.e.,

$$\{(x_i, y_i, F), (x_i^s, y_i^s, F)\}, \quad i = 1, \ldots, K,$$

and we must estimate from this data the location of $\mathbf{p}_L$, the image of $\mathbf{P}_L$ on the plane $Z = F$. The lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ should form a pencil of lines intersecting at $\mathbf{p}_L$. However, due to noise in the data, we have to estimate the location $\mathbf{p}_L$ that is most collinear with the data point pairs $(\mathbf{p}_i, \mathbf{p}_i^s)$ for $i = 1, \ldots, K$. To make this problem precise, we search for a point $\mathbf{p}_L = (x_L, y_L, F)$ so that a collinearity measure with all data pairs will be minimized. Various collinearity measures could be used. A particularly attractive one is:

$$M(\mathbf{p}_L, \{(\mathbf{p}_i, \mathbf{p}_i^s)\}_{i=1,\ldots,K}) = \sum_{i=1}^{K} [\text{area } \Delta \mathbf{p}_L \mathbf{p}_i \mathbf{p}_i^s]^2$$
$$= \frac{1}{4} \sum_{i=1}^{K} \left(\det \begin{vmatrix} x_L & x_i & x_i^s \\ y_L & y_i & y_i^s \\ F & F & F \end{vmatrix} \right)^2.$$

The problem is to determine the $\mathbf{p}_L$ that minimizes $M(\mathbf{p}_L, \{(\mathbf{p}_i, \mathbf{p}_i^s)\})$. Because the area of $\Delta \mathbf{p}_L \mathbf{p}_i \mathbf{p}_i^s$ is proportional to the product of the length $\mathbf{p}_i\mathbf{p}_i^s$ and the distance of $\mathbf{p}_L$ to the line $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$, we have an optimization problem that searches for the minimal weight point where each line $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ is surrounded by a weight field with profile given by

$$f(r) = \tfrac{1}{4}(\mathbf{p}_i\mathbf{p}_i^s)^2 \cdot r^2.$$

This function will have a proportionality gain factor that enhances proportionally the cost to the square of the length of the segment $\overline{\mathbf{p}_i\mathbf{p}_i^s}$. This makes sense. If the points are measured with some error ϵ in the image plane, longer segments will yield narrower cones of (uncertainty) possibility for the location of $\mathbf{p}_L$ (Fig. 5).

Other cost functions, either taking more precisely into consideration the geometry depicted in Figure 5, or simpler ones that penalize equally the distance from $\mathbf{p}_L$ to the line $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ can obviously be considered.

Assume we want to determine (x_L, y_L) by minimizing $M(\mathbf{p}_L, \{(\mathbf{p}_i, \mathbf{p}_i^s)\}_{i=1,\ldots,K})$. By differentiating M with respect to x_L and y_L, we get,

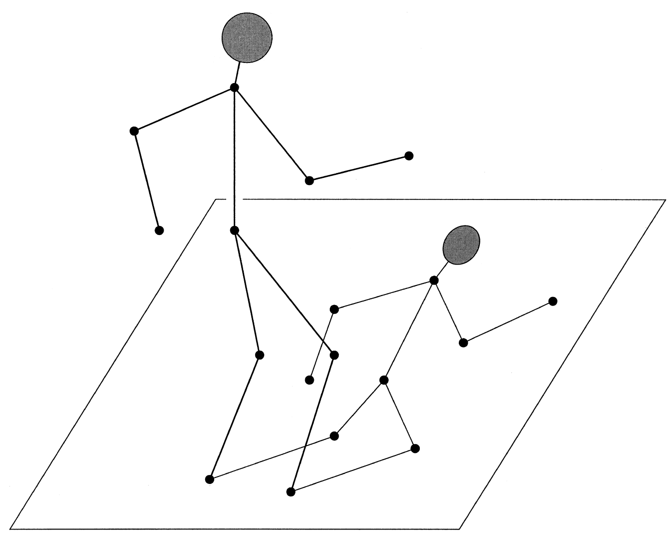

Figure 6. An articulated object with its shadow.

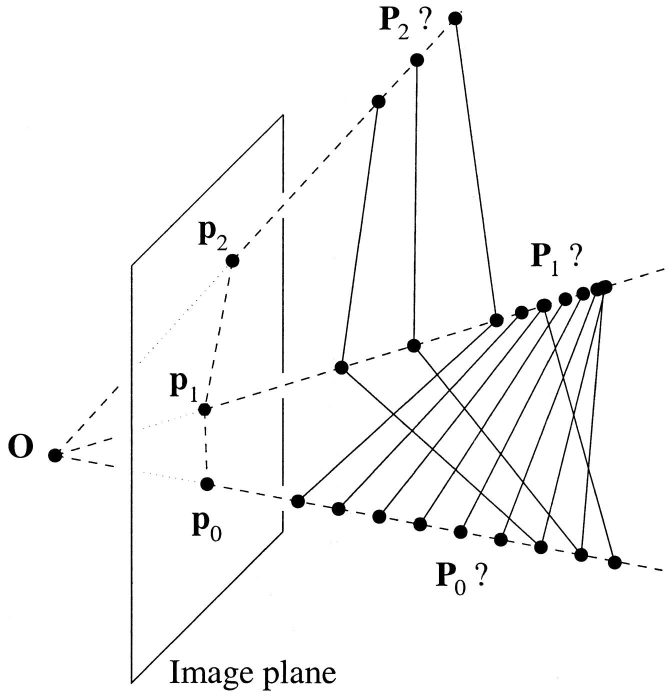

Figure 7. There are infinitely many possible positions for the feature points of an articulated object when given only their images and the lengths of the links.

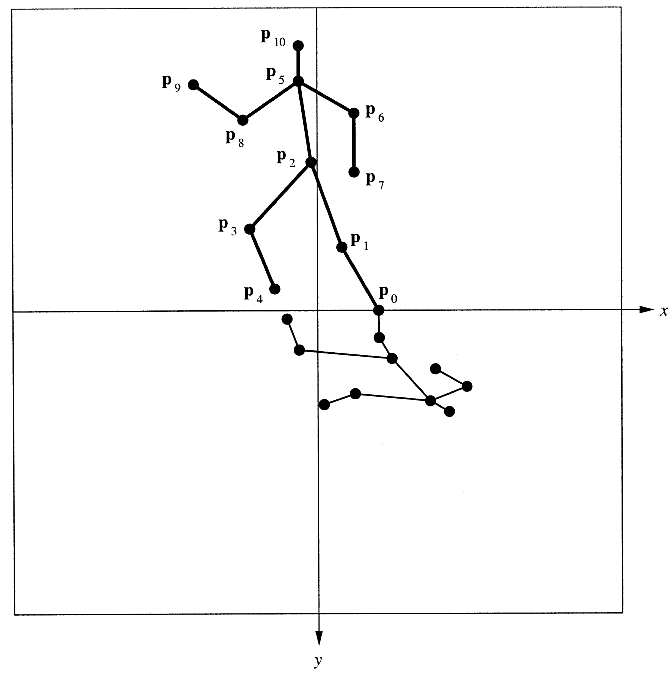

Figure 8. Synthetic example of an articulated object casting a shadow caused by parallel projection. Because the *z*-axis points into the picture, the *y*-axis points downward in order that this be a right-handed coordinate system.

$$\frac{\partial}{\partial x_L} M(\mathbf{p}_L, \{(\mathbf{p}_i, \mathbf{p}_i^s)\}) = \frac{F^2}{4} \sum_{i=1}^{K} 2 \begin{vmatrix} x_L & x_i & x_i^s \\ y_L & y_i & y_i^s \\ 1 & 1 & 1 \end{vmatrix} \cdot (y_i - y_i^s),$$

$$\frac{\partial}{\partial y_L} M(\mathbf{p}_L, \{(\mathbf{p}_i, \mathbf{p}_i^s)\}) = -\frac{F^2}{4} \sum_{i=1}^{K} 2 \begin{vmatrix} x_L & x_i & x_i^s \\ y_L & y_i & y_i^s \\ 1 & 1 & 1 \end{vmatrix} \cdot (x_i - x_i^s).$$

The optimal (x_L, y_L) should satisfy the system of equations below:

$$S_{yy}x_L - S_{xy}y_L + S_{xy,y} = 0$$

$$-S_{xy}x_L + S_{xx}y_L - S_{xy,x} = 0$$

where

$$S_{xx} = \sum_{i=1}^{K} (x_i - x_i^s)^2 \quad S_{yy} = \sum_{i=1}^{K} (y_i - y_i^s)^2$$

$$S_{xy} = \sum_{i=1}^{K} (x_i - x_i^s)(y_i - y_i^s)$$

$$S_{xy,x} = \sum_{i=1}^{K} (x_i y_i^s - y_i x_i^s)(x_i - x_i^s) \quad S_{xy,y} = \sum_{i=1}^{K} (x_i y_i^s - y_i x_i^s)(y_i - y_i^s)$$

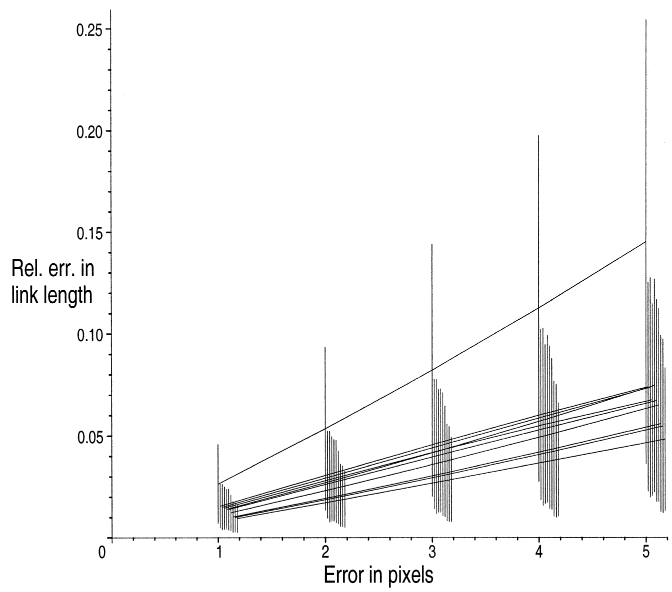

Figure 9. Average relative errors, with error bars of length 2σ, in the computed lengths of the links for Example 1. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_5\mathbf{P}_{10}$, $\mathbf{P}_3\mathbf{P}_4$, $\mathbf{P}_6\mathbf{P}_7$, $\mathbf{P}_0\mathbf{P}_1$, $\mathbf{P}_8\mathbf{P}_9$, $\mathbf{P}_5\mathbf{P}_6$, $\mathbf{P}_5\mathbf{P}_8$, $\mathbf{P}_2\mathbf{P}_3$, $\mathbf{P}_2\mathbf{P}_5$, and $\mathbf{P}_1\mathbf{P}_2$.

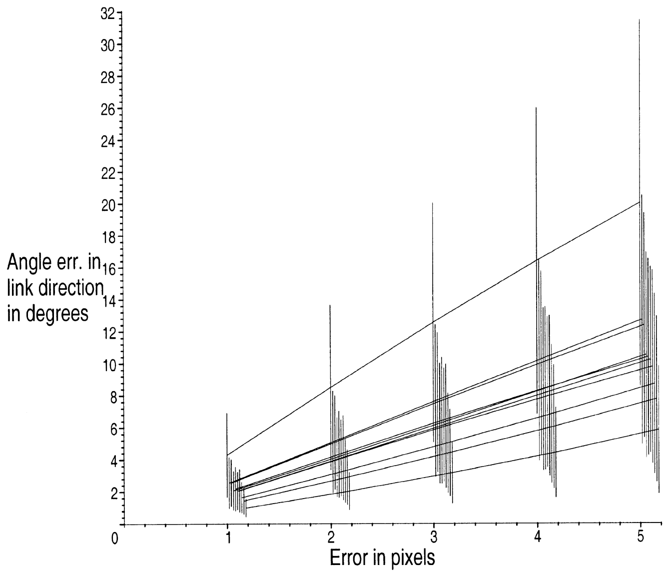

Figure 10. Average angular errors, with error bars of length 2σ, in the computed directions of the links for Example 1. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_5\mathbf{P}_{10}$, $\mathbf{P}_6\mathbf{P}_7$, $\mathbf{P}_8\mathbf{P}_9$, $\mathbf{P}_5\mathbf{P}_6$, $\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_5\mathbf{P}_8$, $\mathbf{P}_2\mathbf{P}_3$, $\mathbf{P}_0\mathbf{P}_1$, $\mathbf{P}_3\mathbf{P}_4$, and $\mathbf{P}_1\mathbf{P}_2$.

or

$$\begin{bmatrix} S_{yy} & S_{xy} \\ S_{xy} & S_{xx} \end{bmatrix} \begin{bmatrix} x_L \\ y_L \end{bmatrix} = \begin{bmatrix} -S_{xy,y} \\ S_{xy,x} \end{bmatrix}.$$

Therefore, by Kramer's rule:

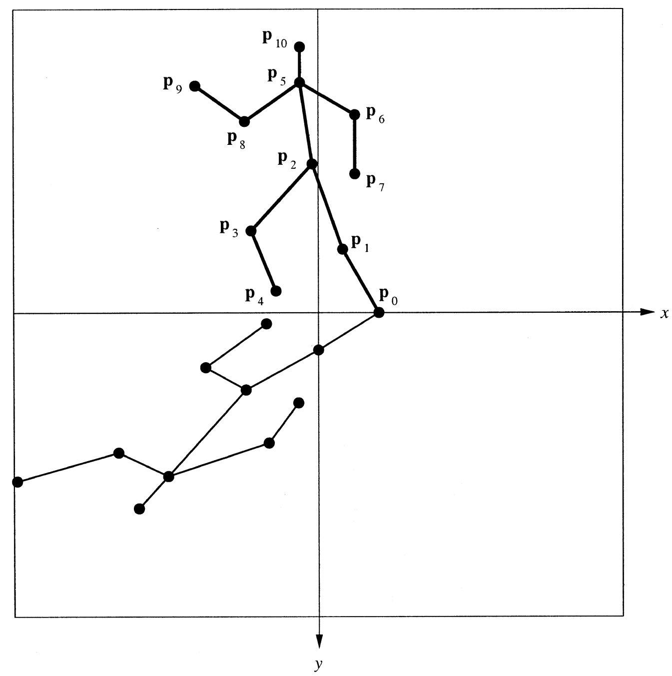

Figure 12. Synthetic example of an articulated object casting a shadow caused by perspective projection.

$$(x_L)_{\text{opt}} = \frac{-S_{xy,y}S_{xx} + S_{xy}S_{xy,x}}{S_{xx}S_{yy} - S_{xy}^2} \tag{1}$$

$$(y_L)_{\text{opt}} = \frac{S_{yy}S_{xy,x} + S_{xy}S_{xy,y}}{S_{xx}S_{yy} - S_{xy}^2} \tag{2}$$

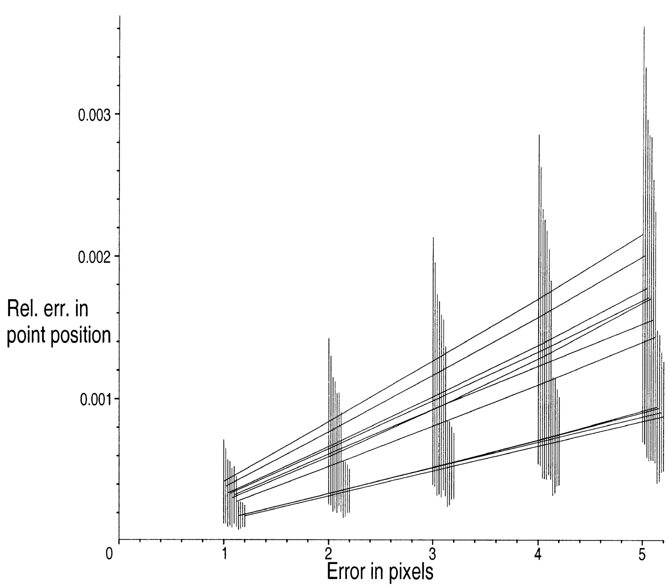

Figure 11. Average relative errors, with error bars of length 2σ, in the computed positions of the points for Example 1. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_{10}$, $\mathbf{P}_9$, $\mathbf{P}_8$, $\mathbf{P}_6$, $\mathbf{P}_5$, $\mathbf{P}_2$, $\mathbf{P}_7$, $\mathbf{P}_1$, $\mathbf{P}_3$, $\mathbf{P}_4$, and $\mathbf{P}_0$.

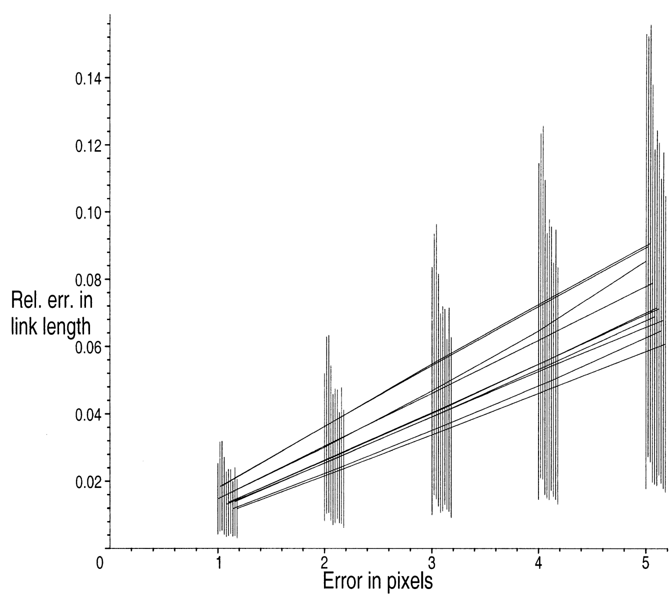

Figure 13. Average relative errors, with error bars of length 2σ, in the computed lengths of the links for Example 2. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_2\mathbf{P}_5$, $\mathbf{P}_0\mathbf{P}_1$, $\mathbf{P}_3\mathbf{P}_4$, $\mathbf{P}_2\mathbf{P}_3$, $\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_5\mathbf{P}_8$, $\mathbf{P}_8\mathbf{P}_9$, $\mathbf{P}_5\mathbf{P}_{10}$, $\mathbf{P}_5\mathbf{P}_6$, and $\mathbf{P}_6\mathbf{P}_7$.

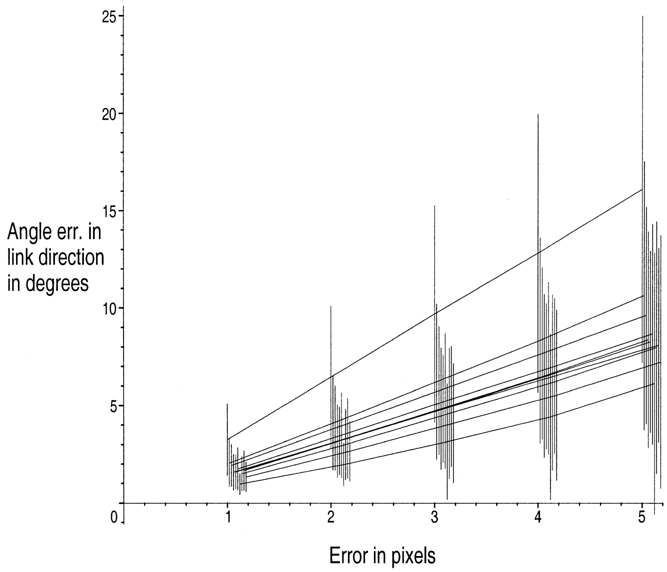

Figure 14. Average angular errors, with error bars of length 2σ, in the computed directions of the links for Example 2. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_5\mathbf{P}_{10}$, $\mathbf{P}_6\mathbf{P}_7$, $\mathbf{P}_8\mathbf{P}_9$, $\mathbf{P}_5\mathbf{P}_6$, $\mathbf{P}_5\mathbf{P}_8$, $\mathbf{P}_1\mathbf{P}_2$, $\mathbf{P}_2\mathbf{P}_5$, $\mathbf{P}_0\mathbf{P}_1$, $\mathbf{P}_2\mathbf{P}_3$, and $\mathbf{P}_3\mathbf{P}_4$.

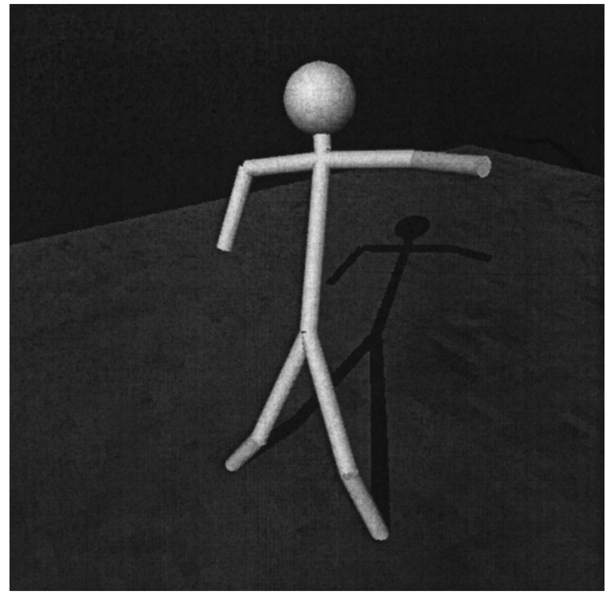

Figure 16. Example 3: Stickman and shadow created by parallel illumination in the direction (0.253, 0.055, 0.966). The ground plane is $-0.078x - 0.906y - 0.416z + 5.579 = 0$. The focal length is 2.365.

We have the optimal location for the light source projection (x_L, y_L) as an explicit formula involving the coordinates of the data points. (The denominators of Eqs. 1 and 2 are zero only if

$$S_{xx}S_{yy} = S_{xy}^2.$$

We have from Cauchy-Schwarz that for real a_i and b_i,

$$\left(\sum_i a_i^2\right)\left(\sum_i b_i^2\right) = \left(\sum_i a_i b_i\right)^2$$

if and only if $a_i/b_i = k$ for some constant k for all i; we need

$$(x_i - x_i^s)/(y_i - y_i^s) = k$$

for all i, i.e., when all the lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ are parallel. In this case, we assume that the illumination is from infinity and falls in a direction parallel to the image plane.)

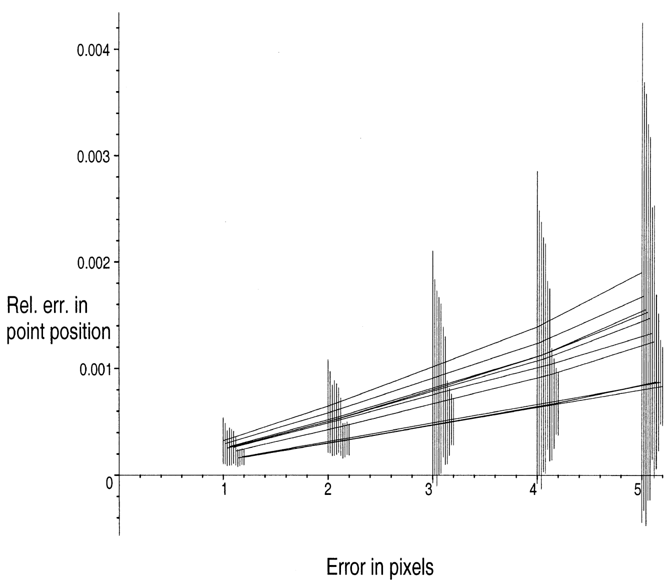

Figure 15. Average relative errors, with error bars of length 2σ, in the computed positions of the points for Example 2. From left to right at each integer number of pixels, the bars are for $\mathbf{P}_{10}$, $\mathbf{P}_5$, $\mathbf{P}_9$, $\mathbf{P}_8$, $\mathbf{P}_6$, $\mathbf{P}_7$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{P}_4$, $\mathbf{P}_1$, and $\mathbf{P}_0$.

B. Computation of 3D Feature Point Locations for Parallel Illumination. Once the coordinates (x_L, y_L) have been determined we proceed as follows:

If we know a priori that the illumination is parallel (i.e., a faraway source), then we have

Table I. Errors in the computed lengths and directions of the links for Example 3.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.046	2.03°
$\mathbf{P}_1\mathbf{P}_2$	0.010	0.50°
$\mathbf{P}_2\mathbf{P}_3$	0.062	2.15°
$\mathbf{P}_3\mathbf{P}_4$	0.001	0.71°
$\mathbf{P}_2\mathbf{P}_5$	0.033	0.39°
$\mathbf{P}_5\mathbf{P}_6$	0.005	2.51°
$\mathbf{P}_6\mathbf{P}_7$	0.087	5.94°
$\mathbf{P}_5\mathbf{P}_8$	0.030	3.64°
$\mathbf{P}_8\mathbf{P}_9$	0.032	2.48°
$\mathbf{P}_5\mathbf{P}_{10}$	0.080	1.50°

Table II. Errors in the computed positions of the points and the direction of parallel projection for Example 3.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(0.245, 0.802)	(0.245, 0.802)	(0.788, 2.563, 7.648)	0.004
1	(0.138, 0.595)	(0.242, 0.485)	(0.421, 1.776, 7.153)	0.010
2	(0.005, 0.115)	(0.232, 0.132)	(0.019, 0.375, 7.506)	0.011
3	(−0.165, 0.482)	(−0.042, 0.418)	(−0.550, 1.614, 8.129)	0.006
4	(−0.255, 0.568)	(−0.255, 0.568)	(−0.938, 2.108, 8.907)	0.005
5	(0.048, −0.465)	(0.332, −0.165)	(0.177, −1.557, 7.766)	0.005
6	(0.355, −0.462)	(0.498, −0.162)	(1.132, −1.465, 7.503)	0.013
7	(0.588, −0.438)	(0.608, −0.132)	(1.659, −1.232, 6.701)	0.017
8	(−0.202, −0.425)	(0.185, −0.165)	(−0.690, −1.516, 8.269)	0.002
9	(−0.282, −0.162)	(0.075, −0.045)	(−0.981, −0.582, 8.466)	0.005
10	(0.042, −0.675)	(0.355, −0.238)	(0.147, −2.169, 7.709)	0.008

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{n}$	Computed $\mathbf{n}$	Error in $\mathbf{n}$
(0.619, 0.136)	(0.621, 0.132)	(0.253, 0.055, 0.966)	(0.254, 0.054, 0.966)	0.10°

$$x_L = \frac{Fn_x}{n_z}, \quad y_L = \frac{Fn_y}{n_z},$$

which identifies uniquely the illumination direction (n_x, n_y, n_z). The shadow points $\mathbf{P}_i^s$ are determined readily from

$$x_i^s = \frac{FX_i^s}{Z_i^s} \quad y_i^s = \frac{FY_i^s}{Z_i^s}$$

and

$$AX_i^s + BY_i^s + CZ_i^s + D = 0$$

as

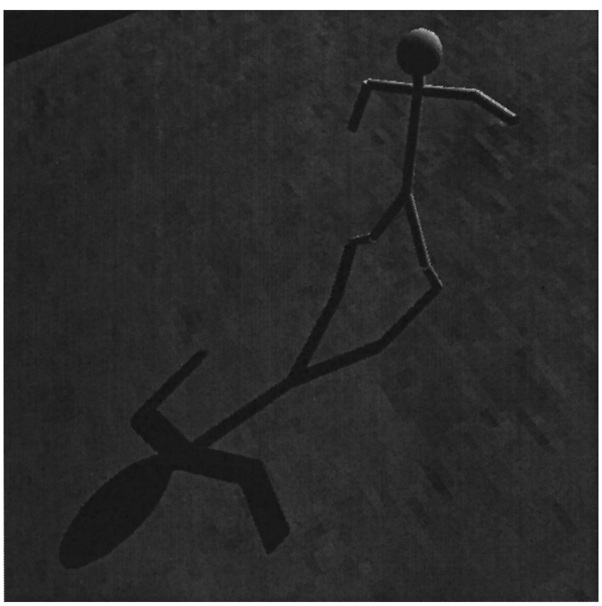

Figure 17. Example 4: Stickman and shadow created by a point light source located at (14.757, −23.354, 10.899). The ground plane is $-0.117x - 0.547y - 0.829z + 12.336 = 0$. The focal length is 2.363.

$$\mathbf{P}_i^s = (X_i^s, Y_i^s, Z_i^s) = -\frac{D}{Ax_i^s + By_i^s + CF}(x_i^s, y_i^s, F). \tag{3}$$

Next, we know that $\mathbf{P}_i$ lies on the lines $\mathbf{P}_i^s + t\mathbf{n} = \mathbf{P}_i^s + t(n_x, n_y, n_z)$ and $\overleftrightarrow{\mathbf{O}\mathbf{p}_i}$. From

$$x_i = \frac{FX_i}{Z_i} \quad y_i = \frac{FY_i}{Z_i} \tag{4}$$

and

$$X_i = X_i^s + tn_x \quad Y_i = Y_i^s + tn_y \quad Z_i = Z_i^s + tn_z \tag{5}$$

we can determine easily the unknown quantities X_i, Y_i, Z_i, and t as the best-fitting solution to the above system of five linear equations.

C. Computation of 3D Feature Point Locations for Point Illumination. If we assume point illumination from $\mathbf{P}_L = (X_L, Y_L, Z_L)$, then we have from

$$x_L = \frac{FX_L}{Z_L} \quad y_L = \frac{FY_L}{Z_L}$$

Table III. Errors in the computed lengths and directions of the links for Example 4.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.020	1.71°
$\mathbf{P}_1\mathbf{P}_2$	0.014	0.46°
$\mathbf{P}_2\mathbf{P}_3$	0.014	0.78°
$\mathbf{P}_3\mathbf{P}_4$	0.010	2.24°
$\mathbf{P}_2\mathbf{P}_5$	0.026	0.81°
$\mathbf{P}_5\mathbf{P}_6$	0.008	4.08°
$\mathbf{P}_6\mathbf{P}_7$	0.013	0.71°
$\mathbf{P}_5\mathbf{P}_8$	0.008	1.43°
$\mathbf{P}_8\mathbf{P}_9$	0.001	2.78°
$\mathbf{P}_5\mathbf{P}_{10}$	0.125	3.30°

Table IV. Errors in the computed positions of the points and the light source for Example 4.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(0.455, −0.055)	(0.455, −0.055)	(2.866, −0.377, 14.715)	0.048
1	(0.418, −0.128)	(0.258, 0.138)	(2.461, −0.779, 13.894)	0.028
2	(0.352, −0.375)	(−0.038, 0.258)	(2.026, −2.148, 13.469)	0.022
3	(0.242, −0.225)	(0.122, −0.032)	(1.513, −1.421, 14.674)	0.036
4	(0.172, −0.205)	(0.172, −0.205)	(1.161, −1.408, 15.611)	0.059
5	(0.392, −0.722)	(−0.398, 0.505)	(2.130, −3.928, 12.662)	0.069
6	(0.582, −0.708)	(−0.172, 0.565)	(3.081, −3.722, 12.450)	0.017
7	(0.715, −0.618)	(−0.075, 0.802)	(3.593, −3.107, 11.873)	0.002
8	(0.225, −0.738)	(−0.532, 0.368)	(1.277, −4.148, 13.144)	0.049
9	(0.172, −0.588)	(−0.332, 0.158)	(1.022, −3.456, 13.817)	0.059
10	(0.402, −0.852)	(−0.628, 0.695)	(2.079, −4.416, 12.292)	0.018

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{P}_L$	Computed $\mathbf{P}_L$	Relative Error in $\mathbf{P}_L$
(3.200, −5.064)	(3.119, −4.947)	(14.76, −23.35, 10.90)	(14.48, −22.97, 10.97)	0.016

that

$$(X_L, Y_L, Z_L) = (x_L Z_L/F, y_L Z_L/F, Z_L)$$

lies on a ray in space (parametrized by Z_L). The shadow points $\mathbf{P}_i^s$ are given by Eq. (3) as in Section IIIB. Now, each point $\mathbf{P}_i$ lies on the intersection of the lines $\overleftrightarrow{\mathbf{Op}_i}$ and $\overleftrightarrow{\mathbf{P}_L\mathbf{P}_i^s}$. In place of Eqs. (4) and (5), we use Eq. (4) and

$$X_i = X_i^s + t(X_L - X_i^s) = t\,\frac{x_L Z_L}{F} + (1-t)X_i^s$$

$$Y_i = Y_i^s + t(Y_L - Y_i^s) = t\,\frac{y_L Z_L}{F} + (1-t)Y_i^s$$

$$Z_i = Z_i^s + t(Z_L - Z_i^s) \tag{6}$$

to determine X_i, Y_i, Z_i, and t in terms of Z_L as the best-fitting solution to the system of five linear equations Eq. (4) and Eq. (6).

One more item of information about the scene beyond the locations of the images of the feature and shadow points on the image plane can lead to complete recovery of the 3D information. As discussed in the previous section, this can be the height of the light source (Z_L), the height of any point (Z_i), or data about the configuration of the points in 3D—like the equality of two links of an articulated object.

Consequently, parallel illumination is simpler than point source illumination. In the former situation, knowledge of the ground plane position in the camera coordinate system enables complete 3D recovery of the object from the images of itself and its shadow with two point correspondences. Knowing that the illumination is at infinity disambiguates the light source location.

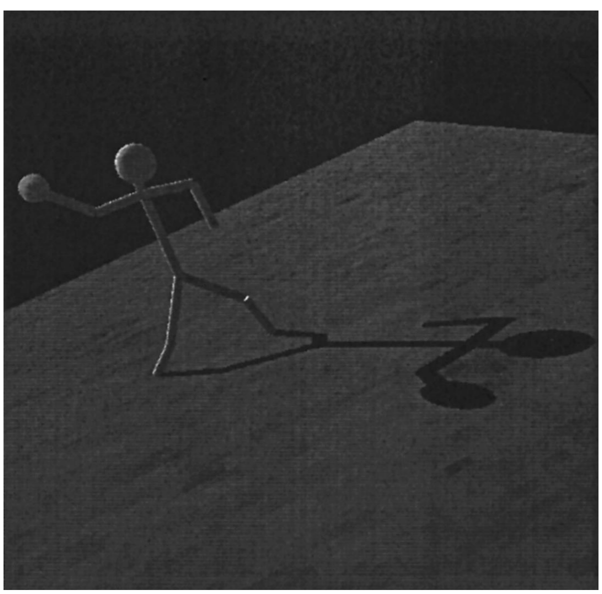

Figure 18. Example 5: Quarterback and shadow created by parallel illumination in the direction (−0.841, −0.412, 0.349). The ground plane is $-0.368x - 0.853y - 0.370z + 5.924 = 0$. The focal length is 2.367.

IV. APPLICATION: ARTICULATED OBJECT STANCE RECOVERY AND PRACTICAL CONSIDERATIONS

We discuss the above algorithms in the context of recovering the position of articulated objects (Fig. 6). For example, this could

Table V. Errors in the computed lengths and directions of the links for Example 5.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.006	2.51°
$\mathbf{P}_1\mathbf{P}_2$	0.039	0.68°
$\mathbf{P}_2\mathbf{P}_3$	0.050	1.48°
$\mathbf{P}_3\mathbf{P}_4$	0.005	1.93°
$\mathbf{P}_2\mathbf{P}_5$	0.034	1.37°
$\mathbf{P}_5\mathbf{P}_6$	0.044	4.39°
$\mathbf{P}_6\mathbf{P}_7$	0.030	8.31°
$\mathbf{P}_5\mathbf{P}_8$	0.027	2.44°
$\mathbf{P}_8\mathbf{P}_9$	0.089	11.97°
$\mathbf{P}_5\mathbf{P}_{10}$	0.055	2.12°

Table VI. Errors in the computed positions of the points and the direction of parallel projection for Example 5.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(−0.095, 0.108)	(−0.095, 0.108)	(−0.608, 0.692, 15.013)	0.002
1	(−0.185, −0.005)	(0.058, 0.118)	(−0.171, −0.047, 14.636)	0.005
2	(−0.408, −0.088)	(0.072, 0.148)	(−2.568, −0.551, 14.846)	0.007
3	(−0.438, 0.155)	(−0.272, 0.242)	(−2.724, 0.938, 14.721)	0.006
4	(−0.482, 0.252)	(−0.482, 0.252)	(−3.130, 1.631, 15.317)	0.005
5	(−0.525, −0.362)	(0.562, 0.152)	(−3.128, −2.196, 13.963)	0.006
6	(−0.365, −0.395)	(0.665, 0.075)	(−2.168, −2.400, 14.155)	0.004
7	(−0.288, −0.252)	(0.408, 0.082)	(−1.750, −1.514, 14.350)	0.012
8	(−0.688, −0.302)	(0.418, 0.245)	(−4.012, −1.751, 13.795)	0.003
9	(−0.845, −0.362)	(0.455, 0.285)	(−4.825, −2.136, 13.568)	0.017
10	(−0.552, −0.465)	(0.822, 0.148)	(−3.179, −2.691, 13.600)	0.003

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{n}$	Computed $\mathbf{n}$	Error in $\mathbf{n}$
(5.528, −11.869)	(5.649, −12.166)	(−0.841, −0.412, 0.349)	(−0.845, −0.401, 0.353)	0.72°

represent an athlete on a playing field. In theory, the knowledge that the light source is so far away that parallel projection can be assumed is sufficient for complete structure recovery. In practice, the problem is extremely sensitive to noise, and the procedure described above is not suitable for finding numerical solutions. We have to use known information about the object to recover its position. The main difficulty with the above algorithm is the precise determination of the image of the light source, the point $\mathbf{p}_L = (x_L, y_L)$, when it is far from the field of view. If $\mathbf{p}_L$ is actually directly visible in the image, then the above algorithm can be used starting with the known location of $\mathbf{p}_L$. If $\mathbf{p}_L$ is just a few screen widths away from the image, then the above algorithm may be used. However, if $\mathbf{p}_L$ is far from the visible screen, the lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ in the image plane are close to parallel and a small error in $\mathbf{p}_i$ and $\mathbf{p}_i^s$ can lead to large errors in the computed location of $\mathbf{p}_L$. Fortunately, the error in the computed position of $\mathbf{p}_L$ is not arbitrary. Because the lines $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$ are all close to parallel, the line through the origin $\mathbf{o}$ of the image plane and $\mathbf{p}_L$ is also close to parallel to all of these lines. It is the distance of $\mathbf{p}_L$ to the origin that may have considerable error. Another way of phrasing this is that the ratio x_L/y_L may be computed accurately, but the distance $(x_L^2 + y_L^2)^{1/2}$ often is not.

The numerical instability in this case can be explained by the fact that parallel projection is a limiting case of perspective projection as the light source tends to infinity and some piece of information about the object is required for pose recovery in the perspective case.

We propose an alternative algorithm. We find the direction (x_L, y_L) by averaging the directions of the line segments $\overleftrightarrow{\mathbf{p}_i\mathbf{p}_i^s}$, weighting each segment by some positive power of its length. (In our programs, we used the square of the length.) Longer segments naturally give more accurate direction information than shorter ones. Let the angle determined in this manner be θ and let $R = (x_L^2 + y_L^2)^{1/2}$, so that $x_L = R \cos \theta$ and $y_L = R \sin \theta$. Now θ is known with fairly high accuracy whereas R is not. The idea is to let R be a parameter and compute in terms of it the position of all

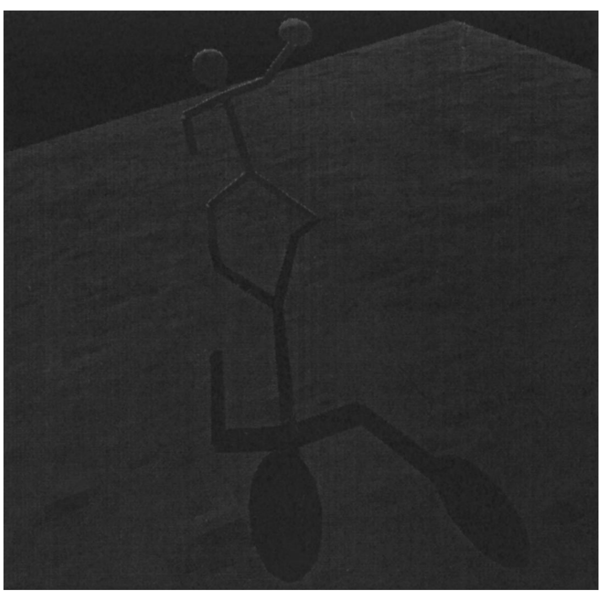

Figure 19. Example 6: Quarterback and shadow created by a point light source located at (−5.838, −24.987, 25.377). The ground plane is $-0.161x - 0.785y - 0.598z + 9.171 = 0$. The focal length is 2.365.

Table VII. Errors in the computed lengths and directions of the links for Example 6.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.041	3.08°
$\mathbf{P}_1\mathbf{P}_2$	0.024	2.65°
$\mathbf{P}_2\mathbf{P}_3$	0.025	2.97°
$\mathbf{P}_3\mathbf{P}_4$	0.045	2.03°
$\mathbf{P}_2\mathbf{P}_5$	0.041	3.71°
$\mathbf{P}_5\mathbf{P}_6$	0.016	3.88°
$\mathbf{P}_6\mathbf{P}_7$	0.008	1.28°
$\mathbf{P}_5\mathbf{P}_8$	0.015	3.52°
$\mathbf{P}_8\mathbf{P}_9$	0.043	8.17°
$\mathbf{P}_5\mathbf{P}_{10}$	0.110	10.55°

Table VIII. Errors in the computed positions of the points and of the light source for Example 6.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(−0.292, −0.205)	(−0.292, −0.205)	(−2.240, −1.579, 18.013)	0.002
1	(−0.298, −0.322)	(−0.268, −0.115)	(−2.186, −2.414, 17.461)	0.003
2	(−0.168, −0.442)	(−0.072, 0.018)	(−1.157, −3.057, 16.581)	0.003
3	(−0.048, −0.352)	(−0.018, −0.215)	(−0.399, −2.670, 17.818)	0.002
4	(0.052, −0.275)	(0.052, −0.275)	(0.415, −2.107, 17.968)	0.002
5	(−0.248, −0.695)	(−0.038, 0.465)	(−1.641, −4.643, 15.554)	0.006
6	(−0.382, −0.628)	(−0.272, 0.475)	(−2.501, −4.135, 15.536)	0.002
7	(−0.352, −0.502)	(−0.278, 0.182)	(−2.415, −3.457, 16.265)	0.001
8	(−0.112, −0.748)	(0.202, 0.388)	(−0.732, −5.007, 15.776)	0.006
9	(−0.025, −0.872)	(0.455, 0.578)	(−0.267, −5.700, 15.370)	0.007
10	(−0.295, −0.792)	(−0.048, 0.735)	(−1.912, −5.146, 15.327)	0.012

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{P}_L$	Computed $\mathbf{P}_L$	Relative Error in $\mathbf{P}_L$
(−0.544, −2.329)	(−0.572, −2.492)	(−5.84, −24.99, 25.38)	(−6.42, −27.97, 26.55)	0.090

3D points $\mathbf{P}_i$, then choose the value of R that leads to the 3D reconstruction that matches most closely the prior information on the articulated object. We use the bisection method to solve the resulting problem of determining the zero crossing of the numerical derivative thereby computed. One possible measure is to minimize the sum of the squares of the differences of the computed and actual lengths of some subset of the links of an articulated object. If it is not known whether the light source is in front of or behind the camera, then both positive and negative values of R must be considered. R is not going to be near zero, for that would mean the image of the light source would be visible on the image plane near its origin.

The images of the shadow points, in addition to the images of the feature points, are necessary for the recovery of the feature points of an articulated object with no closed loops. If only the images of the feature points are given, then there will be typically infinitely many possible stances for the collection of feature points (Fig. 7). There are infinitely many possible locations for the segment $\overline{\mathbf{P}_0\mathbf{P}_1}$ given that $\mathbf{P}_0$ lies on $\overrightarrow{\mathbf{Op}_0}$ and $\mathbf{P}_1$ on $\overrightarrow{\mathbf{Op}_1}$. For each location of $\mathbf{P}_1$, there can be one or two possible locations for $\mathbf{P}_2$ on $\overrightarrow{\mathbf{Op}_2}$ such that $\overline{\mathbf{P}_1\mathbf{P}_2}$ has the right length.

Once the best-fitting line for the direction of projection is found, the 3D feature points are determined as follows. If the data and all computations are exact, then the point $\mathbf{P}_i$ is the intersection of the lines $\overleftrightarrow{\mathbf{Op}_i}$ and $\mathbf{P}_i^s + \mathbf{n}t$. When these lines do not intersect, $\mathbf{P}_i$ is chosen to be the point on $\overleftrightarrow{\mathbf{Op}_i}$, which is nearest the line $\mathbf{P}_i^s + \mathbf{n}t$. This selection was slightly more accurate than taking the midpoint of the common perpendicular of the two lines. This may be expected because the line $\overleftrightarrow{\mathbf{Op}_i}$ is known to a higher accuracy than $\mathbf{P}_i^s + \mathbf{n}t$, due to the uncertainty in the computation of the direction $\mathbf{n}$.

For the case in which perspective projection of the light source may be assumed, it is possible to determine the position of $\mathbf{p}_L$, the image of the light source, with reasonably high accuracy using Eqs. (1) and (2). There is now a one-parameter family of possible locations for the light source $\mathbf{P}_L$. We can let Z_L sweep out values from $(0, \infty)$ or $(-\infty, \infty)$, just as we did for the quantity R in the parallel projection case. The value of Z_L, which comes closest to

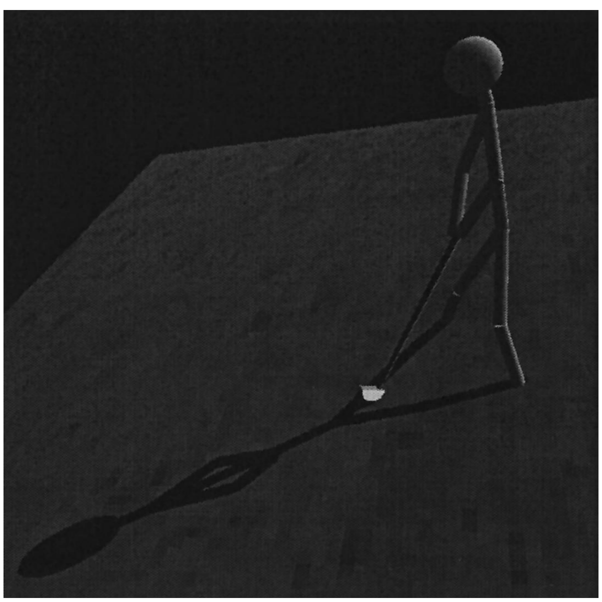

Figure 20. Example 7: Golfer and shadow created by parallel illumination in the direction (0.415, −0.892, 0.178). The ground plane is $-0.190x - 0.840y - 0.508z + 7.272 = 0$. The focal length is 2.365.

Table IX. Errors in the computed lengths and directions of the links for Example 7.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.059	0.86°
$\mathbf{P}_1\mathbf{P}_2$	0.040	1.04°
$\mathbf{P}_2\mathbf{P}_3$	0.114	3.58°
$\mathbf{P}_3\mathbf{P}_4$	0.210	6.92°
$\mathbf{P}_2\mathbf{P}_5$	0.043	3.21°
$\mathbf{P}_5\mathbf{P}_6$	0.039	3.94°
$\mathbf{P}_6\mathbf{P}_7$	0.041	4.81°
$\mathbf{P}_5\mathbf{P}_8$	0.106	0.87°
$\mathbf{P}_8\mathbf{P}_9$	0.010	0.95°
$\mathbf{P}_5\mathbf{P}_{10}$	0.141	2.10°

Table X. Errors in the computed positions of the points and the direction of parallel projection for Example 7.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(0.738, 0.248)	(0.738, 0.248)	(3.419, 1.232, 11.065)	0.012
1	(0.668, 0.065)	(0.455, 0.325)	(3.087, 0.294, 10.958)	0.013
2	(0.672, −0.258)	(0.118, 0.392)	(3.123, −1.204, 10.981)	0.011
3	(0.522, −0.038)	(0.362, 0.155)	(2.695, −0.268, 12.060)	0.021
4	(0.485, 0.048)	(0.485, 0.048)	(2.652, 0.251, 12.919)	0.012
5	(0.622, −0.652)	(−0.482, 0.668)	(2.648, −2.857, 10.050)	0.014
6	(0.648, −0.418)	(−0.218, 0.608)	(2.811, −1.834, 10.142)	0.009
7	(0.528, −0.232)	(−0.095, 0.508)	(2.235, −1.043, 10.508)	0.017
8	(0.538, −0.448)	(−0.275, 0.522)	(2.442, −2.050, 10.684)	0.012
9	(0.508, −0.235)	(−0.105, 0.495)	(2.310, −1.068, 10.638)	0.013
10	(0.572, −0.802)	(−0.782, 0.808)	(2.364, −3.296, 9.726)	0.008

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{n}$	Computed $\mathbf{n}$	Error in $\mathbf{n}$
(−64.47, 76.61)	(−103.13, 122.72)	(0.644, −0.765, −0.024)	(0.643, −0.765, −0.015)	0.51°

satisfying the known prior information, is chosen as the optimum position of the Z-coordinate of the light source.

V. EXPERIMENTAL RESULTS

Example 1: Parallel Projection

This example is illustrated in Figure 8. It is intended to model a tennis player on a court. A pinhole camera is located approximately 200 ft in front of and 50 ft above the person. The scene shown is 10 sq ft and the image is zoomed in to the point where the image is 6 in. square, although the units do not matter. The image is taken to be 500×500 pixels and 11 feature points are labeled $\mathbf{P}_0$ to $\mathbf{P}_{10}$. The direction of the light source is ($\frac{1}{3}$, $\frac{62}{75}$, $\frac{-34}{75}$).

Noise was added to the images of all of the feature and shadow points and the algorithm described in the previous section was used to determine the 3D positions of the feature points $\mathbf{P}_i$. Figure 9 shows the results of 1,000 trials of the reconstruction of the 3D position of the object. Added noise of k pixels means that the errors added to the x- and y- coordinates of the image points $\mathbf{p}_i$ and $\mathbf{p}_i^s$ were taken from a uniform distribution from $[-k, k]$ pixels. Figure 10 shows the errors in the computed directions of the links and Figure 11 shows the errors in the computed positions of the feature points.

In Figures 9–15, error bars of length two times the observed standard deviation are shown for each link and point. In order to display all of these, the values and error bars are staggered in the direction parallel to the horizontal axis. The horizontal coordinates are all integers, representing the error in pixels. The relative error in the computed length of a link is defined to be the difference in the computed and actual lengths, divided by its actual length, and the relative error in the computed position of a point is defined to be the absolute error in its computed position divided by its distance to the optic center of the camera.

Expressed as a fraction of the distance from the camera center to the object, the average relative errors in the lengths of the links range from 1% to 3% of the error in pixels in the image points and the maximum errors are about four times that. The greatest relative errors occur in the shortest links, as may be expected. The average error in the computed direction of the light is about 0.6° times the error in pixels, with the maximum error three to four times that. The relative error in the 3D positions of the feature points increases with their distance from the shadow plane; for the highest points, these

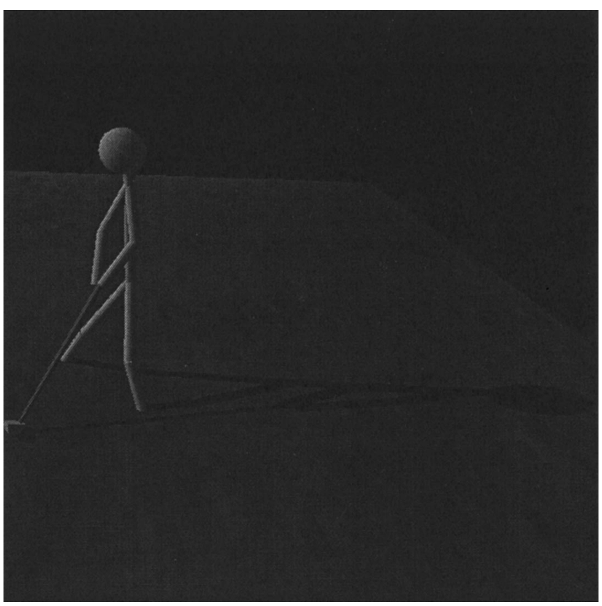

Figure 21. Example 8: Golfer and shadow created by a point light source located at (−20.955, −12.896, 3.277). The ground plane is $0.168x - 0.875y - 0.454z + 8.273 = 0$. The focal length is 2.365.

Table XI. Errors in the computed lengths and directions of the links for Example 8.

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.059	1.14°
$\mathbf{P}_1\mathbf{P}_2$	0.007	2.13°
$\mathbf{P}_2\mathbf{P}_3$	0.029	0.16°
$\mathbf{P}_3\mathbf{P}_4$	0.033	0.93°
$\mathbf{P}_2\mathbf{P}_5$	0.047	0.54°
$\mathbf{P}_5\mathbf{P}_6$	0.059	1.00°
$\mathbf{P}_6\mathbf{P}_7$	0.036	1.45°
$\mathbf{P}_5\mathbf{P}_8$	0.051	0.42°
$\mathbf{P}_8\mathbf{P}_9$	0.040	1.25°
$\mathbf{P}_5\mathbf{P}_{10}$	0.092	4.42°

Table XII. Errors in the computed positions of the points and of the light source for Example 8.

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(−0.535, 0.345)	(−0.535, 0.345)	(−2.979, 1.986, 13.373)	0.004
1	(−0.575, 0.188)	(−0.338, 0.325)	(−3.177, 1.032, 13.143)	0.002
2	(−0.568, −0.082)	(−0.022, 0.262)	(−3.182, −0.459, 13.287)	0.003
3	(−0.725, 0.095)	(−0.508, 0.222)	(−4.190, 0.525, 13.778)	0.003
4	(−0.785, 0.182)	(−0.785, 0.182)	(−4.765, 1.092, 14.374)	0.002
5	(−0.572, −0.402)	(0.568, 0.292)	(−2.974, −2.181, 12.382)	0.006
6	(−0.555, −0.215)	(0.322, 0.325)	(−2.894, −1.149, 12.487)	0.003
7	(−0.662, −0.068)	(−0.008, 0.348)	(−3.475, −0.354, 12.419)	0.001
8	(−0.658, −0.245)	(0.208, 0.275)	(−3.527, −1.345, 12.682)	0.002
9	(−0.682, −0.072)	(−0.055, 0.332)	(−3.588, −0.373, 12.496)	0.001
10	(−0.595, −0.522)	(0.795, 0.315)	(−2.991, −2.648, 11.983)	0.003

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{P}_L$	Computed $\mathbf{P}_L$	Relative Error in $\mathbf{P}_L$
(−15.12, −9.31)	(−9.21, −5.59)	(−20.96, −12.90, 3.28)	(−18.22, −11.06, 4.682)	0.144

errors are about 0.05% of the image pixel error on average and 0.2% in the worst case. The average error in the angles of the links is roughly 0.01° times the error in pixels, times the ratio of the distance to the object from the camera divided by the length of the link, with maximum errors three to four times as large.

Example 2: Perspective Projection

This example is illustrated in Figure 12. This is similar to Example 1, but now the light source is nearby, such as that on a tower at the edge of a playing field. In this example, the light source is located 15 units above, 20 units behind, and 5 units to the right of the subject, or at (5, −20, 215) in the camera coordinate system.

Noise was added to the images of all of the feature and shadow points as in Example 1. The algorithm described in the previous section was used to determine the 3D positions of the feature points $\mathbf{P}_i$. Figure 13 shows the results of 1,000 trials of the reconstruction of the 3D position of the object, Figure 14 shows the errors in the computed directions of the links, and Figure 15 shows the errors in the computed positions of the feature points.

For the most part, the errors in the computed lengths of the links and the 3D positions of the feature points are similar to those in the parallel projection case, although they are smaller. As a fraction of its distance to the camera center, the computed error in the position of the light source is about 0.3% times the error in pixels in the image points, with a maximum error of four to five times that value. This error is mostly in the Z-direction, along the optic axis of the camera. The angular error in the ray from the origin to the light source $\overrightarrow{\mathbf{OP}_L}$ is only about 0.1° times the image pixel error on average and 0.4° in the worst case. If the X- and Y-coordinates of P_L were on the same order of magnitude as Z_L, it would have been better to treat this example as one with parallel projection of the light rays.

Example 3: Parallel Projection of Stickman

This example is illustrated in Figure 16. In this and the following examples, point correspondences between the images of the feature and shadow points were obtained by locating the pixels representing the joints in the figure. The numbering of the joints is the same as in

Figure 22. Example 9: Tennis player and shadow created by the sun in the direction (−0.960, −0.279, −0.005). The ground plane is $0.995y + 0.099z - 5.125 = 0$. The focal length is 3.002.

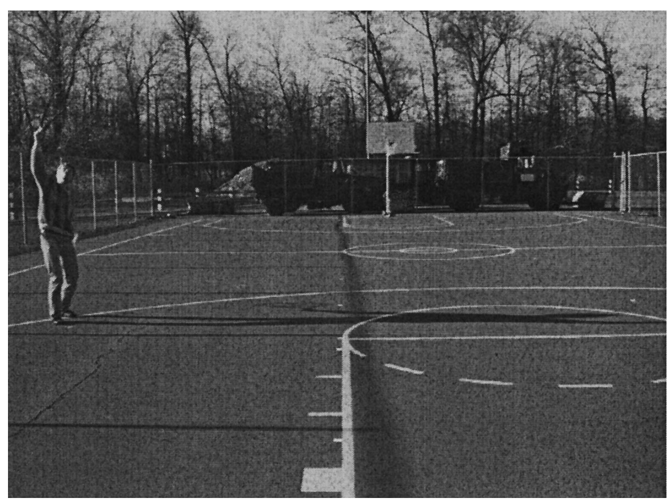

Figure 23. Example 9: Tennis player and shadow created by the sun in the direction (−0.962, −0.271, 0.019). The ground plane is $0.995y + 0.099z - 5.125 = 0$. The focal length is 3.002.

Table XIII. Errors in the computed lengths and directions of the links for Example 9(a).

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.055	2.84°
$\mathbf{P}_1\mathbf{P}_2$	0.176	14.56°
$\mathbf{P}_2\mathbf{P}_3$	0.087	6.91°
$\mathbf{P}_3\mathbf{P}_4$	0.022	2.48°
$\mathbf{P}_2\mathbf{P}_5$	0.077	17.14°
$\mathbf{P}_5\mathbf{P}_6$	0.216	7.13°
$\mathbf{P}_6\mathbf{P}_7$	0.456	13.87°
$\mathbf{P}_5\mathbf{P}_8$	0.255	12.43°
$\mathbf{P}_8\mathbf{P}_9$	0.204	2.59°
$\mathbf{P}_5\mathbf{P}_{10}$	0.049	2.20°

Table XV. Errors in the computed lengths and directions of the links for Example 9(b).

Link	Relative Error in Length	Error in Direction
$\mathbf{P}_0\mathbf{P}_1$	0.084	5.44°
$\mathbf{P}_1\mathbf{P}_2$	0.060	9.84°
$\mathbf{P}_2\mathbf{P}_3$	0.078	6.96°
$\mathbf{P}_3\mathbf{P}_4$	0.023	3.01°
$\mathbf{P}_2\mathbf{P}_5$	0.088	31.47°
$\mathbf{P}_5\mathbf{P}_6$	0.160	5.68°
$\mathbf{P}_6\mathbf{P}_7$	0.140	11.00°
$\mathbf{P}_5\mathbf{P}_8$	0.167	7.76°
$\mathbf{P}_8\mathbf{P}_9$	0.029	27.34°
$\mathbf{P}_5\mathbf{P}_{10}$	0.136	6.53°

the previous examples, with $\mathbf{P}_{10}$ denoting the center of the figure's head. Because the limbs are all several pixels wide, there will be some error in determining where the joints are in the image. The camera coordinate system is taken so that the images shown encompass the region $-1 \leq x \leq 1$, $-1 \leq y \leq 1$ and the resolution is 600×600 pixels.

In this example, $\mathbf{p}_L$ is located within the image. The algorithm described in Section III was used to determine the 3D positions of the feature points $\mathbf{P}_i$. Tables I and II show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

The errors in the computed positions of the 3D feature points are all less than 1.8% of their distance from the camera and the directions of the links are all computed to within 6°.

Example 4: Perspective Projection of Stickman

This example is illustrated in Figure 17. It is the same figure as in Example 3, but with a nearby point light source and a different camera position. Because this is a situation with a close light source, the algorithm of Section III is used to determine $\mathbf{p}_L$. Tables III and IV show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

The errors in the computed positions of the 3D feature points are all less than 0.6% of their distance from the camera and the directions of the links are all computed to within 5°.

Example 5: Parallel Projection of Quarterback

This example is illustrated in Figure 18. It is the same figure as in Example 5, but with a different light source direction and camera position. In this example, $\mathbf{p}_L$ is well outside the image and we use the algorithm described in Section IV. The algorithm of Section III could still be used, but provides less accurate results. Tables V and VI show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

The errors in the computed positions of the 3D feature points are all less than 1.7% of their distance from the camera and the directions of the links are all computed to within 12°.

Example 6: Perspective Projection of Quarterback

This example is illustrated in Figure 19. It is the same figure as in Example 5, but with a nearby point light source and a different camera position. Because this is a situation with a close light source, the algorithm of Section III is used to determine $\mathbf{p}_L$. Tables VII and VIII show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

The errors in the computed positions of the 3D feature points are all less than 1.2% of their distance from the camera and the directions of the links are all computed to within 11°.

Example 7: Parallel Projection of Golfer

This example is illustrated in Figure 20. It is the same figure as in Example 7, but with a different light source direction and camera

Table XIV. Errors in the computed positions of the points and the direction of parallel projection for Example 9(a).

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(−1.097, 0.238)	(−1.097, 0.238)	(−10.524, 2.287, 28.808)	0.002
1	(−1.060, 0.115)	(−0.473, 0.248)	(−9.771, 1.060, 27.673)	0.006
2	(−1.090, 0.002)	(−0.200, 0.262)	(−9.851, 0.015, 27.130)	0.008
3	(−1.123, 0.105)	(−0.447, 0.278)	(−9.865, 0.922, 26.364)	0.003
4	(−1.136, 0.268)	(−1.136, 0.268)	(−10.330, 2.439, 27.283)	0.001
5	(−1.120, −0.272)	(0.773, 0.255)	(−9.663, −2.344, 25.901)	0.020
6	(−1.063, −0.208)	(0.460, 0.228)	(−9.650, −1.891, 27.243)	0.032
7	(−1.023, −0.225)	(0.527, 0.215)	(−9.732, −2.140, 28.550)	0.010
8	(−1.170, −0.152)	(0.353, 0.292)	(−9.739, −1.262, 24.989)	0.002
9	(−1.137, −0.125)	(0.430, 0.312)	(−9.043, −0.994, 23.884)	0.013
10	(−1.120, −0.325)	(1.023, 0.262)	(−9.479, −2.750, 25.406)	0.019

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{n}$	Computed $\mathbf{n}$	Error in $\mathbf{n}$
(−567.38, 167.51)	(−33.64, 9.41)	(−0.960, −0.279, −0.005)	(−0.959, −0.268, −0.088)	4.66°

Table XVI. Errors in the computed positions of the points and the direction of parallel projection for Example 9(b).

i	$\mathbf{p}_i$	$\mathbf{p}_i^s$	Actual $\mathbf{P}_i$	Relative Error in $\mathbf{P}_i$
0	(−1.113, 0.255)	(−1.113, 0.255)	(−10.362, 2.373, 27.940)	0.000
1	(−1.077, 0.112)	(−0.510, 0.265)	(−9.663, 1.002, 26.943)	0.008
2	(−1.120, 0.008)	(−0.213, 0.272)	(−9.926, 0.074, 26.605)	0.001
3	(−1.137, 0.102)	(−0.490, 0.288)	(−9.897, 0.885, 26.137)	0.004
4	(−1.147, 0.275)	(−1.147, 0.275)	(−10.300, 2.470, 26.966)	0.001
5	(−1.113, −0.265)	(0.813, 0.258)	(−9.414, −2.241, 25.384)	0.053
6	(−1.087, −0.155)	(0.247, 0.238)	(−9.587, −1.368, 26.486)	0.060
7	(−1.070, −0.065)	(−0.123, 0.228)	(−9.815, −0.596, 27.535)	0.048
8	(−1.210, −0.342)	(1.133, 0.275)	(−9.711, −2.742, 24.093)	0.066
9	(−1.207, −0.498)	(1.400, 0.272)	(−9.441, −3.899, 23.488)	0.091
10	(−1.097, −0.332)	(0.983, 0.248)	(−9.381, −2.837, 25.680)	0.056

Actual $\mathbf{p}_L$	Computed $\mathbf{p}_L$	Actual $\mathbf{n}$	Computed $\mathbf{n}$	Error in $\mathbf{n}$
(−152.00, −42.82)	(−49.08, −13.82)	(−0.962, −0.271, 0.019)	(−0.961, −0.271, −0.059)	4.46°

position. In this example, $\mathbf{p}_L$ is well outside the image and we use the algorithm described in Section IV. Tables IX and X show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

The distance of $\mathbf{p}_L$ to the origin $\mathbf{o}$ of the image plane is not computed accurately, but that does not affect significantly the computation of the positions of the $\mathbf{P}_i$. This is because the direction of $\mathbf{n}$, which contains a very small component in the Z-direction, is still computed accurately. The errors in the computed positions of the 3D feature points are all less than 2.1% of their distance from the camera and the directions of the links are all computed to within 7°.

Example 8: Perspective Projection of Golfer

This example is illustrated in Figure 21. It is the same figure as in Example 7, but with a nearby light source and a different camera position. Because this is a situation with a close light source, the algorithm of Section III is used to determine $\mathbf{p}_L$. Tables XI and XII show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively.

In this example, the distance of $\mathbf{P}_L$ from the origin $\mathbf{O}$ is not computed accurately, but that does not affect greatly the computation of the positions of the $\mathbf{P}_i$. This is because the direction of $\overleftrightarrow{\mathbf{OP}}_L$, which contains a very small component in the Z-direction, is still computed accurately. The errors in the computed positions of the 3D feature points are all less than 0.7% of their distance from the camera and the directions of the links are all computed to within 5°.

Example 9: Real Images of Tennis Player

We conclude with two real outdoor images (Figs. 22 and 23). The subject is in two poses approximately 9 m in front of the camera. The photographs were taken a few minutes apart, just as the sun was crossing the camera plane. The positions of the 3D feature points were estimated by a combination of measured points on the court surface, the known link lengths, and their positions in the image. To make these measurements precise, we affixed bright stickers to the ground on a measured grid. Locating these fiducial points in the images was accomplished easily to high accuracy by straightforward thresholding. In some cases, particularly with the left arm in Figure 23, feature points were occluded and had to be estimated.

In these examples, $\mathbf{p}_L$ is well outside the image, so the algorithm of Section IV is used. Tables XIII and XIV show the errors in the computed directions and lengths of the links and the results of the reconstruction of the 3D position of the object, respectively, for the first image, Tables XV and XVI show the same errors and results (as shown in Tables XIII and XIV) for the second image.

In these examples, the distance of $\mathbf{p}_L$ to the origin $\mathbf{o}$ of the image plane is not computed accurately, but the direction of the sun is still computed to within 5°. As expected, the errors in the computed positions of the 3D feature points are greater than those in the previous examples in which the image feature and shadow points were computed with greater accuracy because the features were more distinct, and none were occluded. The algorithm still gives a reasonable idea where the feature points are in 3D-space, with all the feature points being computed to within 3.3% of their distance to the camera.

VI. CONCLUDING REMARKS

" . . . if you wish to live among your fellow man, learn to value your shadow more than gold."—A. von Chamisso, Peter Schlemiel: the man who sold his shadow.

This paper proposes the use of shadows for 3D recovery in scenes in which shadows are relatively easily associated with objects moving on a flat ground plane (e.g., tennis players on the court) under strong (natural or artificial) illumination. Shadows have been recognized as an important cue in 3D recovery. It seems to us that use of shadows for pose recovery in conjunction with articulated objects, which enable a straightforward determination of object-shadow correspondences, has not received its due attention in the literature. Several studies concerning 3D recovery from shadows under a variety of practical situations will be the subject of forthcoming reports.

REFERENCES

J.-Y. Bouguet and P. Perona, 3D photography on your desk, Sixth Int Conf on Computer Vision, Bombay, 1998, pp. 43–50.

J.-Y. Bouguet and P. Perona, 3D photography using shadows in dual-space geometry, Int J Comput Vision 35 (1999), 129–149.

J.R. Kender and E.M. Smith, Shape from darkness: Deriving surface information from dynamic shadows, Proc IEEE First Int Conf on Computer Vision, 1987, pp. 539–546.

D.K. Kersten, D.C. Knill, P. Mamassian, and I. Bulthoff, Illusory motion from shadows, Nature 379 (1996), 31.

D.K. Kersten, P. Mamassian, and D.C. Knill, Moving cast shadows induce apparent motion in depth, Perception 26 (1997) 171–192.

D.C. Knill, P. Mamassian, and D.K. Kersten, Geometry of shadows, J Optical Soc Am A, 14 (1997), 3216–3232.

R. Nevatia, Building detection and description from a single intensity image, Comput Vision Image Understand 72 (1998), 101–121.

S.A. Shafer, Shadows and silhouettes in computer vision, Kluwer, Boston, 1985.

S.A. Shafer and T. Kanade, Using shadows in finding surface orientations, Comput Vision Graphics Image Process 22 (1983), 145–176.

J.A. Shufelt, Performance evaluation and analysis of vanishing point detection techniques, ARPA96 Image Understanding Workshop, Palm Springs, CA, 1996, pp. 1113–1132.

J.A. Shufelt, Performance evaluation and analysis of monocular building extraction from aerial imagery, IEEE Trans Pattern Analysis Machine Intell 21 (1999), 311–326.

W.B. Thompson, M.T. Checky, and W.F. Kaemmerer, Shadow stereo-locating object boundaries using shadows, Proc AAAI-87 (Sixth Nat Conf on AI), Seattle, 1987, pp. 761–766.

S. Tsuji, M. Yachida, and H.-L. Guo, Wiresight: Robot vision for determining three-dimensional geometry of flexible wires from shadow information, 7th Int Conf on Pattern Recognition, Montreal, 1984, pp. 1358–1360.

D. Yang and J.R. Kender, Shape from darkness under error, ARPA96 Image Understanding Workshop, Palm Springs, CA, 1996, pp. 1141–1148.

A. Yonas, L.T. Goldsmith, and J.L. Hallstrom, Development of sensitivity to information provided by cast shadows in pictures, Perception 7 (1978), 333–341.