3D Trajectories from a Single Viewpoint using Shadows

I. Reid and A. North
Department of Engineering Science, Oxford University
Parks Road, Oxford, OX1 3PJ, UK
tel: +1865 273168
[ian|u94an1]@robots.ox.ac.uk

Abstract

We consider the problem of obtaining the 3D trajectory of a ball from a sequence of images taken with a camera which is possibly rotating and zooming (but not translating). Techniques are developed to compute the component of image motion of the ball due to camera rotation and zoom, using optic flow. The 3D location of the ball in each frame of the sequence is then determined using a novel geometric construction which makes use of shadows on the known ground plane in order to compute the vertical projection of the ball onto the ground, and the height of the ball above the ground.

1 Introduction

In the absence of any other constraints, the image projections of world points in a single view of a scene are insufficient to compute a 3D reconstruction of the scene. The most obvious way to obtain 3D structure is therefore to consider multiple views separated spatially (and possibly temporally). An alternative to using multiple viewpoints is to enforce physical and geometric constraints about the scene; for example, a particular illumination model, planarity, parallelism or symmetry.

In this paper we make use of shadows in order to compute 3D structure from a single viewpoint. We apply this to the problem of computing the 3D trajectory of a football from broadcast images of a game. The image location of a ball is tracked automatically using cross-correlation and a constant image velocity Kalman Filter, and a 3D reconstruction obtained using the locations of the ball's shadow, a (vertical) reference object and the known structure of the football pitch markings.

Our work is most closely related to [7] who computed 3D structure for bilaterally symmetric objects; they noted that a single view of a bilaterally symmetric object is equivalent to two views of half the object. In this paper we show that in some circumstances a point light source can be considered to be equivalent to a second view. We provide a constructive proof of how to determine the projection (in an arbitrary direction) of a point onto a distinguished plane, given the point's image location and the image location of its shadow from a point light source at infinity (e.g. the sun).

The other work most closely related is that of [2] who also analysed ball trajectories in football games, however we improve on their work in two main ways: (i) we compute the height of the ball using an elegant construction devised by [1] which is much cleaner than

the method proposed in [2]; and (ii) by using shadows we ultimately require much weaker assumptions about the physical motion model of the ball. [2] assumes the ball moves in a parabolic trajectory in a vertical plane, which is clearly false in many situations.

We are concerned in the present work with reconstructing the trajectory of the ball over a sequence of images taken with camera which rotates and zooms but does not translate. Since the camera is fixed, there is no baseline over which to triangulate so 3D reconstruction from a single view is essential. However the rotation of the camera, and changes in its intrinsic parameters, introduce an additional complication that the image motion of the ball consists of two parts: (i) the physical motion of the ball; (ii) that induced by the camera. We deal with this by observing the overall motion of the static parts of the scene induced by the camera and subtracting this component from the ball's image motion. To compute the image motion we introduce a new algorithm based on some tried and tested optic flow techniques from the literature.

In summary, we make the following contributions:

- We derive the equations which describe the image velocity field induced by a camera which rotates and zooms (section 2.1);
- We describe and implement an algorithm to compute the velocity field directly from spatial and temporal gradients (section 2.2);
- We devise a geometric construction to compute the projection (in an arbitrary direction) of a 3D point into a distinguished plane using the image locations of the point and its shadow in a single view (section 3.1);
- We determine a full 3D reconstruction of the point using this construction and one devised by [1] (section 3.2);
- We present an implementation of all our ideas with application to trajectory analysis of a moving football. Our implementation automatically tracks the image trajectory of the ball, and then a semi-automatic procedure (reference heights and shadows are picked manually) computes the 3D reconstruction of the trajectory (section 4).

In the remainder of the paper we adopt the following notation. Scalars are denoted by normal math-script Roman and Greek letters (e.g. λ, x, X). World locations are denoted by bold upper case letters, e.g. $\mathbf{X}$. Where required, these are assumed to be homogeneous 4-vectors whose components are $\mathbf{X} = [X\ Y\ Z\ W]^\top$. Image locations are denoted either by homogeneous 3-vectors $\mathbf{x} = [x\ y\ w]^\top$ or by inhomogeneous 2-vectors $\mathbf{u} = [u\ v]^\top$. Matrices are denoted in uppercase typewriter font, e.g. `M`, `H`. Derivatives with respect to time are denoted using dots, e.g. $\dot{\mathbf{x}}$ is the derivative of the vector $\mathbf{x}$ with respect to time.

2 Ball tracking

Since in typical broadcast images of a football match, the ball can appear rather small and irregular, we adopt a correlation (rather than a contour) based approach to the problem of measuring the ball's image position in each frame.

In order better to cope with noise, and with frames in the sequence when the ball is either occluded or indistinct, we have implemented a Kalman Filter which assumes a constant image velocity motion model upon the trajectory. The filter's predicted state and covariance provide a natural search region – a validation gate – in which to conduct the

(a) (b)

Figure 1: Two trajectories automatically computed: (a) Paul Gascoigne shooting (and scoring) against Scotland in Euro96; (b) Paul Ince shooting (and not scoring) against Italy in a World Cup qualifying match.

search for a template correspondence. The best correlating position within the validation gate, and above a similarity threshold, is assumed to be the correct match.

Much of the Kalman Filter implementation is straightforward and therefore details are omitted here. One aspect worthy of consideration is that of the motion of the camera observing the scene. Clearly any camera rotation or zooming will affect the image location of all features, including the ball. In the context of the standard Kalman Filter equations, the camera motion is a *control input*. If the camera's pan, tilt and zoom were known it would be a simple matter to determine the apparent image motion as was done in [6]. We derive the appropriate equations in section 2.1 below. In our case these data are not available directly, and so we make use of other computer vision techniques to determine the image motion. In particular we determine, using a gradient based scheme, the optic flow between frames, and use this to determine the apparent ball motion due to camera rotation and zoom. The details are given in section 2.2.

Figure 1 shows examples of the image trajectories of the ball as computed automatically by our system.

2.1 Equations of image motion

Here we derive the equations of image motion for a camera undergoing pure rotation and pure zoom.

We assume that the image axes are perpendicular, that the principal point is constant and located at (0,0), and that the pixels are square. Then the canonical projection equation can be written very simply in homogeneous coordinates as:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ w \end{bmatrix} = \mathtt{K} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{1}$$

The inhomogeneous representation is obtained as $\mathbf{u} = [u \;\; v] = [x/w \;\; y/w]^\top$.

If the camera's rotational velocity is given by $\mathbf{\Omega}$ then $\dot{\mathbf{X}} = \mathbf{\Omega} \times \mathbf{X}$ and combining this with the derivative of (1) yields

$$\begin{aligned} \dot{\mathbf{x}} &= \left[\dot{\mathtt{K}} + \mathtt{K}\mathbf{\Omega}_\times\right] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad \left(\text{where } \mathbf{\Omega}_\times = \begin{bmatrix} 0 & \omega_Z & -\omega_Y \\ -\omega_Z & 0 & \omega_X \\ \omega_Y & -\omega_X & 0 \end{bmatrix} \right) \\ &= \begin{bmatrix} \dot{f}/f & \omega_Z & -\omega_Y \\ -\omega_Z & \dot{f}/f & \omega_X \\ \omega_Y/f & -\omega_X/f & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} h_1 & h_2 & h_3 \\ -h_2 & h_1 & h_4 \\ h_5 & h_6 & 0 \end{bmatrix} \mathbf{x} = \dot{\mathtt{H}}\mathbf{x} \end{aligned} \tag{2}$$

Thus the image motion is related homographically to the image position. The homography has 5 degrees of freedom, but the equations can be solved linearly by allowing six degrees of freedom; coefficients $\mathbf{h} = [h_1 \dots h_6]$ in (2).

The optic flow, $\dot{\mathbf{u}}$ is given by

$$\dot{\mathbf{u}} = \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix} = \begin{bmatrix} \dot{x}/w - u\dot{w}/w \\ \dot{y}/w - v\dot{w}/w \end{bmatrix} \tag{3}$$

Dividing (2) by w and combining with (3) yields the usual equations of image motion [5], which can in turn be written in terms of $\mathbf{h}$ as

$$\dot{\mathbf{u}} = \begin{bmatrix} u & v & 1 & 0 & -u^2 & -uv \\ v & -u & 0 & 1 & -uv & -v^2 \end{bmatrix} \mathbf{h} \tag{4}$$

2.2 Computing optic flow

Since there exist few distinct features but much texture in a typical football broadcast image, we opt to use intensity gradient based flow rather than discrete feature matches to determine the overall image motion. By adopting the approach of Lucas and Kanade [3] to optic flow computation, we arrive at an elegant means of computing the 6 degree of freedom velocity field derived above.

The optic flow computation determines an approximation to the shape of the local sum-of-squared-differences function (SSD) at each point in the image, and finds the minimum of the function. Algebraically, at each point (u,v) in the image we wish to minimise

$$E_{\dot{\mathbf{u}}}(u,v) = \sum_{ij} (I(u+j+\dot{u}, v+i+\dot{v}, t+1) - I(u,v,t))^2 \tag{5}$$

where the indices i, j range over small patch centred on (u,v). Substituting a first order approximation for the term in brackets and expanding the square yields

$$E_{\dot{\mathbf{u}}}(u,v) = \sum_{ij} \dot{u}^2 (I_u)^2 + +2\dot{u}\dot{v} I_u I_v + \dot{v}^2 (I_v)^2 + 2\dot{u} I_u I_t + 2\dot{v} I_v I_t + (I_t)^2 \tag{6}$$

The sum over the patch can be replaced by a convolution (in our case Gaussian). We then differentiate with respect to $\mathbf{u}$ and set to zero, to find the minimum:

$$\begin{bmatrix} G \otimes I_u^2 & G \otimes I_u I_v \\ G \otimes I_u I_v & G \otimes I_v^2 \end{bmatrix} \dot{\mathbf{u}} + \begin{bmatrix} G \otimes I_u I_t \\ G \otimes I_v I_t \end{bmatrix} = \mathtt{M}\dot{\mathbf{u}} + \mathbf{b} = 0 \tag{7}$$

If there is no brightness gradient then $\mathtt{M}$ is singular and the flow cannot be computed, as one would expect. If there is a uniform edge at (u, v) then $\mathtt{M}$ has rank one, and only the component of flow in the kernel of $\mathtt{M}$ (in the direction of the brightness gradient) can be determined. This is the well known aperture problem. If $\mathtt{M}$ has full rank then the full flow can be obtained.

Thus each image location gives either zero, one or two constraints on the six degree of freedom optic flow field, resulting in an over-determined system. Algebraically this is derived by combining (4) and (7):

$$\mathtt{M} \begin{bmatrix} u & v & 1 & 0 & -u^2 & -uv \\ v & -u & 0 & 1 & -uv & -v^2 \end{bmatrix} \mathbf{h} + \mathbf{b} = 0 \tag{8}$$

We then solve for $\mathbf{h}$, the six coefficients of the flow field.

In typical sequences we have considered, $\mathtt{M}$ rarely has rank zero. As a result, the independently moving players and ball, which constitute a very small percentage of the overall scene, contribute little to the overall result. Outlier detection and removal (not implemented here) could improve matters further.

Since the camera undergoes pure rotation, image locations are related by a homography (known as the infinite homography). This is obtained by integrating the instantaneous homography $\dot{\mathtt{H}}$ over one time step:

$$\mathbf{x}' = (I + \dot{\mathtt{H}})\mathbf{x} = \mathtt{H}\mathbf{x} \tag{9}$$

The algorithm has been implemented as follows. For each pair of images:

- Normalise the brightness values and compute a Gaussian pyramid
- For each level of the pyramid, $i = n \ldots 1$ (coarse to fine):
  - Compute the flow field $\dot{\mathtt{H}}$ between images at level i as in section 2.2 above.
  - Compute the inter-image homography $\mathtt{H} = \mathtt{I} + \dot{\mathtt{H}}$ and warp the images at level $i - 1$ towards one another.
  - Accumulate the inter-image transformations $\mathtt{T}_i = \mathtt{H}\mathtt{T}_{i-1}$.

By way of demonstration we show an example of a mosaic which has been constructed from the inter-image homographies (figure 2). Because the inter-frame homographies are computed by integrating up from velocity, there is inevitable drift in the mosaic, however this could be addressed in the future by considering all inter-image transformations in a batch bundle adjustment, not just those between successive pairs.

3 Geometry

In this section we show how shadows from infinite point light sources can be employed to obtain affine and Euclidean structure from a single view. If one considers the light source to be analogous to a second viewpoint then the result is hardly surprising, but does involve some subtlety, as we show below. The ideas are discussed in more detail along with various different applications and further theoretical investigations in [1].

We concern ourselves with computing the structure of a point relative to a distinguished plane in the world. While this plane could be any world plane, for didactic

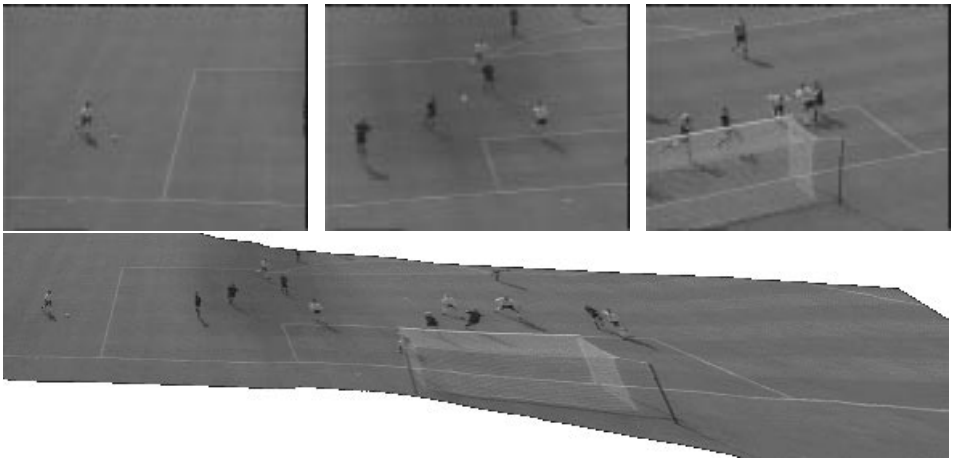

Figure 2: Above: the first, middle and last images in a 130 frame sequence of Alan Shearer scoring against Scotland in Euro96; below: the mosaic constructed using our flow technique.

purposes, here we will consider the ground plane. Likewise, the point could be any point in the scene, but here we will refer to the point as "the ball", for obvious reasons. Specifically, we show:

> Given an affine calibration of the ground plane (i.e. the image location of its vanishing line), the image locations of the top and bottom of a known reference height/direction, the image location of a second point, and the image locations of the shadow of the top of the reference height and shadow of the unknown point, we can determine:
>
> (i) the projection in the reference direction of the point onto the ground plane; i.e. its *affine coordinates* in the plane.
>
> (ii) the projection distance; i.e. the *affine height*.
>
> Hence we obtain the affine structure of the point. Furthermore, if the ground plane calibration is Euclidean and the reference direction is vertical, then we obtain Euclidean structure for the unknown point.

3.1 Computing the X and Y coordinates

We prove the results by construction, beginning with (i). The geometry is shown in figure 3. The shadow is assumed to derive from a light source at infinity (for example, the sun). The desired projection is obtained by construction as follows:

> • Line $\mathbf{l}_1$ is drawn through the reference shadow; Line $\mathbf{l}_2$ is drawn such that it passes through the shadow of the ball and is parallel in the world to $l1$ (i.e. in the image it intersects line $\mathbf{l}_1$ on the vanishing line); • The plane π_1 containing the light source, the ball, and top of the reference height intersects the

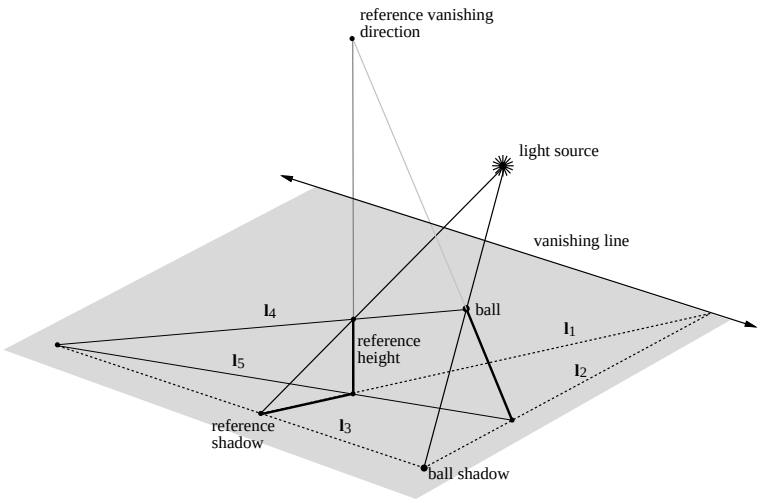

Figure 3: Obtaining the projection of a point onto the ground plane from a single view using shadows.

ground plane in line $\mathbf{l}_3$, obtained as the line which passes through the shadows of the ball and top reference; • The plane π_2 containing the reference direction and the ball intersects π_1 in line $\mathbf{l}_4$ which is obtained as the line passing through the ball and top reference; • The intersection of $\mathbf{l}_3$ and $\mathbf{l}_4$ is therefore the point of common intersection between the three planes π_1, π_2 and the ground plane; • The intersection of π_2 with the ground plane is then given by $\mathbf{l}_5$ which joins the common plane intersection point with the bottom reference; • The intersection of $\mathbf{l}_5$ with $\mathbf{l}_1$ is then the projection of the ball onto the ground plane (in the reference direction) as required.

Since we began with the assumption that the ground plane was affine calibrated, we therefore now know the affine coordinates of the projection of the ball on the ground plane and have proved part (i).

3.2 Computing the Z coordinate

In order to complete the proof of part (ii) we now show how to determine the projection distance (i.e. the affine height). Consider figure 4.

• The line $\mathbf{l}_5$ was obtained by construction in the previous step. It intersects the vanishing line uniquely; • Lines $\mathbf{l}_6$ and $\mathbf{l}_7$ are constructed such that they pass through the reference top and the ball (respectively) and are parallel in the world to $\mathbf{l}_5$, hence in the image they intersect $\mathbf{l}_5$ on the vanishing line of the ground plane• Lines $\mathbf{l}_8$ and $\mathbf{l}_9$ are parallel to the reference direction in the world. They are known since the top and bottom reference, and the ball and its projection are known. The intersection of $\mathbf{l}_8$ and $\mathbf{l}_9$ is the vanishing point in the reference direction, $\mathbf{v}$. We denote the intersection of $\mathbf{l}_7$ and $\mathbf{l}_8$ by $\mathbf{r}_i$ and the intersection of $\mathbf{l}_6$ and $\mathbf{l}_9$ by $\mathbf{p}_i$

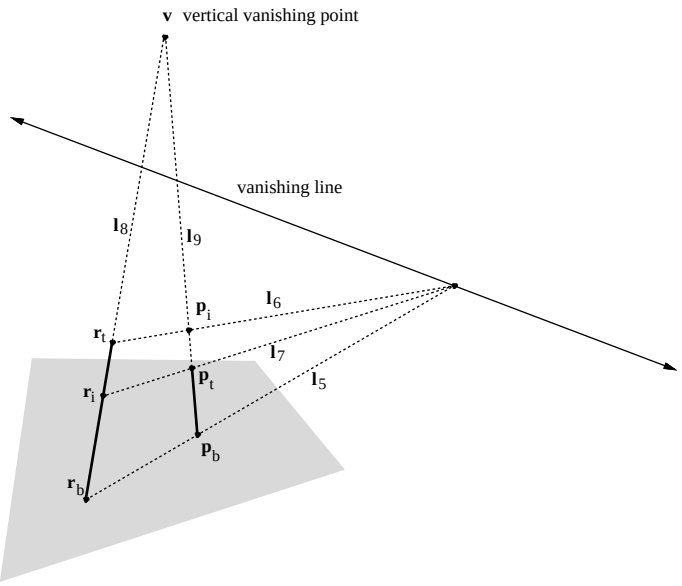

Figure 4: A construction for computing the height of a point from the ground.

The cross ratio of the four points on $\mathbf{l}_8$ must equal the cross ratio of the corresponding points on $\mathbf{l}_9$. Both cross ratios are equal to the cross ratio of the physical points on the world lines (i.e. with $\mathbf{v}$ at infinity), which is a simple ratio of involving only the reference and unknown height, viz $< \mathbf{r}_b, \mathbf{r}_i, \mathbf{r}_t, \mathbf{v} >=< \mathbf{p}_b, \mathbf{p}_t, \mathbf{p}_i, \mathbf{v} >= h_r/(h_r - h_p)$, where h_r is the reference height and h_p is the height of the ball. Hence

$$h_p = h_r - \frac{h_r}{< \mathbf{r}_b, \mathbf{r}_i, \mathbf{r}_t, \mathbf{v} >} \tag{10}$$

While these constructions are useful to understand intuitively the underlying geometry, in practice we have developed algebraic methods (see [1]) which are more robust and simpler to implement.

4 Results

In this section we present results of the trajectory reconstruction. Although we have analysed a number of sequences in this way with encouraging results, space permits only one set of results to be included here.

Although the ball tracking has been automated, the selection of reference heights throughout the sequence, and the localisation of the ball shadow are currently performed manually using a mouse.

The regular scene structure present on a football pitch – namely the pitch markings and goals – mean that affine calibration of the ground plane is straightforward. The known world locations of four pitch lines are used to compute the image to world homography which maps points in the image to their ground plane positions `H` [4]. The vanishing line of the ground plane in the image is then simply $\mathbf{u}^\top = [0\ 0\ 1]\,$`H`.

Figure 5: From left to right, the first middle and last images from a 40 frame sequence.

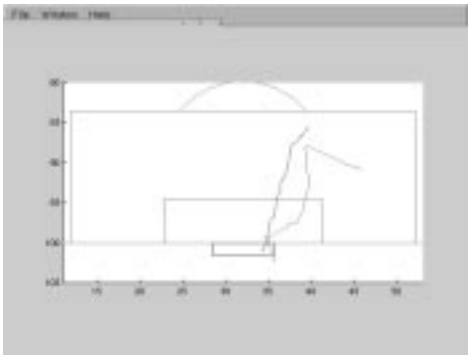

Figure 6: An overhead view of Gascoigne shot (see text for details)

Figure 5 shows the first, middle and last images from a sequence of 40 images of Paul Gascoigne scoring against Scotland in Euro96. One of the upright Scottish players (i.e. not Hendry who was floundering on the ground having been skinned) was chosen as vertical reference and his height estimated as 1.75m. An overhead view of the ball's trajectory is shown in figure 6. This view is achieved using the vertical reference direction and the shadow locations, as explained in section 3.1. The second (paler) trajectory, provided for comparison, is simply the projection of the ball onto the ground plane in the direction of sight (achieved by transforming the ball position via the image to ground plane homography).

The full 3d reconstruction of the scene, obtained using the theory set out in section 3.2 is shown from two different viewpoints in figure 7.

5 Discussion

We have presented a system for computing the 3D trajectory of a football in a sequence of images captured by a camera which can rotate and zoom (but not translate). The two main distinct contributions were (i) compensation for camera motion via a new method for computing the image velocity field, and (more importantly) (ii) a geometric construction for computing the 3D position of a point relative to a known ground plane and one vertical reference using its image location and the location of its shadow.

One idea we have not yet explored, but which seems is promising, is that of performing the filtering in three dimensional space, rather than in the image. This would have the

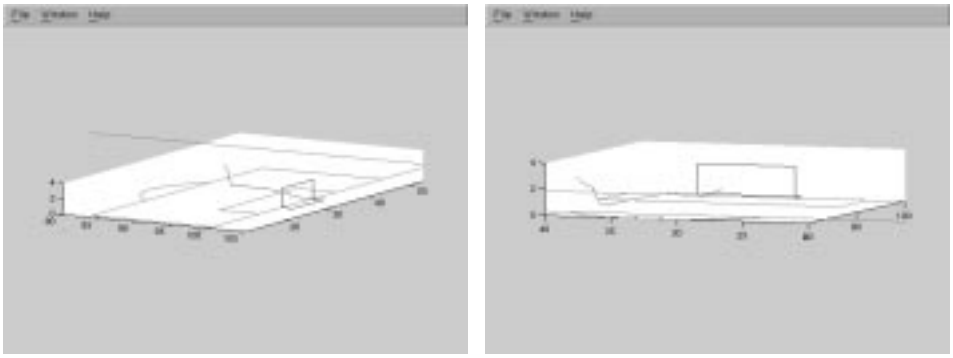

Figure 7: Two different viewpoints of the 3D reconstruction of Gascoigne's shot.

advantage that smoothness is imposed on the full trajectory, not just on its projection into the image, which would mitigate depth errors.

A more theoretical and complete discussion of 3D reconstruction from a single view can be found in [1]. We are grateful to Antonio Criminisi and Andrew Zisserman, our co-authors on [1], for many fruitful discussions.

References

[1] A. Criminisi, I. Reid, and A. Zisserman. Computing 3d euclidean distance from a single view. Technical Report OUEL report 2158/98, Dept. of Engineering Science, University of Oxford, 1998.

[2] Taeone Kim, Yongduek Seo, and Ki sang Hong. Physics-based 3D position analysis of a soccer ball from monocular image sequences. In *Proc. 6th Int'l Conf. on Computer Vision, Bombay*, pages 721–726, 1998.

[3] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *DARPA Image Understanding Workshop*, pages 121–130, 1981.

[4] J. L. Mundy and A. P. Zisserman, editors. *Geometric Invariance in Computer Vision*. MIT Press, Cambridge, MA, 1992.

[5] D. W. Murray and B. F. Buxton. *Experiments in the Machine Interpretation of Visual Motion*. MIT Press, Cambridge, MA, 1991.

[6] I. D. Reid and D. W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.

[7] C. Rothwell, D. Forsyth, A. Zisserman, and J. Mundy. Extracting projective structure from single perspective views of 3D point sets. In *Proc. 4th Int'l Conf. on Computer Vision, Berlin*, pages 573–582, 1993.