MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

Qualitative visual interpretation of 3D hand gestures using motion parallax

Roberto Cipolla * Yasukazu Okamoto and Yoshinori Kuno
Research and Development Center,
Toshiba Corporation
1, Komukai Toshiba-cho, Sawai-ku, Kawasaki 210

Abstract

We present an efficient and geometrically intuitive algorithm to reliably interpret the image velocities of moving objects in 3D. It is well known that in a small neighbourhood the image motion of points on a plane is characterised by an *affine* transformation. We show that the relative image motion of a nearby non-coplanar point and its projection on the plane is equivalent to *motion parallax* and that it is a reliable geometric cue to 3D shape and viewer/object motion. In particular we show how to interpret the motion parallax vector of a fourth point and the *curl*, *divergence* and *deformation* components of the affine transformation (defined by the three points of the plane) in order to recover the projection of the axis of rotation of a moving object; the change in relative position of the object; the rotation about the ray; the *tilt* of the surface and a one parameter family of solutions for the *slant* as a function of the magnitude of the rotation of the object. The latter is a manifestation of the *bas-relief* ambiguity. These measurements, although representing an incomplete solution to structure from motion, can be reliably extracted from 2 views even when perspective effects are small.

We present a real-time example in which the 3D visual interpretation of hand gestures or a hand-held object is used as part of a man-machine interface. This is an alternative to the Polhemus coil instrumented Dataglove commonly used in sensing manual gestures.

1 Introduction

Structure from motion

The way appearances change in the image due to relative motion between the viewer and the scene is a well known cue for the perception of 3D shape and motion. Computational attempts to quantify the perception of 3D shape have determined the number of points and the number of views needed to recover the spatial configuration of the points and the motion compatible with the views. Ullman, in his well-known structure from motion theorem [16], showed that a minimum of three distinct orthographic views of four non-planar points in a *rigid* configuration allow the structure and motion to be completely determined. If perspective projection is assumed two views are, in principle, sufficient. In fact two views of eight points allow the problem to be solved with linear methods [9] while five points from two views give a finite number of solutions [2].

Problems with this approach

Although structure from motion algorithms based on these formulations have been successfully applied in photogrammetry and some robotics systems [3] when a wide field of view, a large range in depths and a large number of accurately measured image data points are assured, these algorithms have been of little or no practical use in analysing imagery in which the object of interest occupies a small part of the field of view or is distant.

In this paper we summarise why structure from motion algorithms are often very sensitive to errors in the measured image velocities and then show how to efficiently and reliably extract an incomplete solution. We also show how to augment this into a complete solution if additional constraints or views are available.

The main problems with existing structure from motion formulations are:

1. **Perspective effects are often small**

   Structure from motion algorithms attempt to deliver a complete *quantitative* solution to the viewer or object motion (both 3D translation and 3D rotation) and then to reconstruct a euclidean 3D copy of the scene. Such *complete* quantitative solutions to the structure from motion problem, however, are not only often too difficult, but are numerically ill-conditioned, often failing in a graceless fashion [15] in the presence of image measurement noise. This is because they rely on the measurement of perspective effects which can be very small. In such cases the effects in the image of viewer translations parallel to the image plane are very difficult to discern from rotations about axes parallel to the image plane. Another ambiguity which often arises is the *bas-relief* ambiguity [4] which concerns the difficulty of distinguishing between a "shallow" structure close to the viewer and "deep" structures further away when perspective effects are small. Note that this concerns surface orientation and its effect – unlike the speed–scale ambiguity – is to distort the shape.

2. **Global rigidity and independent motion**

   Existing approaches place a lot of emphasis on global rigidity. Despite this it is well known that two (even orthographic) views give vivid 3D impressions even in the presence of a degree of non-

*University lecturer, Department of Engineering, University of Cambridge, CB2 1PZ, England.

rigidity such as the class of smooth transformations e.g. bending transformations which are locally rigid [7]. Many existing methods can not deal with multiple moving objects and they usually require the input image to be segmented into parts corresponding with the same rigid body motion. Segmentation based on image velocities alone is a non-trivial task if the image velocity data is noisy.

Our approach

In this paper we present an efficient and reliable solution to the structure from motion problem by avoiding small perspective effects or the constraint of global rigidity.

We assume *weak* perspective in a small neighbourhood and concentrate on shape and motion parameters which do not rely on perspective effects. The solution is however incomplete and motion and shape are expressed more qualitatively by spatial order (relative depths) and *affine structure* (Euclidean shape up to an arbitrary 3D affine transformation [7]).

The algorithm is based on a simple method of decomposing image velocities to remove the effect of viewer rotations to leave a component that depends simply on 3D shape and viewer translational motion. It is a development of the pioneering work of Longuet-Higgins and Prazdny [10] and Koenderink and Van Doorn [5, 7] (reviewed below).

2 Theoretical framework

2.1 Interpretation of image velocities under perspective projection

Consider an arbitrary co-ordinate system with the $x-y$ plane spanning the image plane and the z-axis aligned with the ray. Assume the viewer to have a translational velocity with components $\{U_1, U_2, U_3\}$ and an angular velocity with components $\{\Omega_1, \Omega_2, \Omega_3\}$. Let the image velocity field at a point (x, y) in the vicinity of $\mathbf{Q}$ be represented as a 2D vector field, $\vec{\mathbf{v}}(x, y)$ with x and y components (u, v). The two components of the image velocity of a point in space, (X, Y, Z) due to relative motion between the observer and the scene are given by [10]:

$$\begin{aligned} u &= \left[\frac{fU_1 - xU_3}{Z}\right] + f\Omega_2 - y\Omega_3 - \frac{xy}{f}\Omega_1 + \frac{x^2}{f}\Omega_2 \\ v &= \left[\frac{fU_2 - yU_3}{Z}\right] - f\Omega_1 + x\Omega_3 + \frac{xy}{f}\Omega_2 - \frac{y^2}{f}\Omega_1 \end{aligned} \quad (1)$$

The image velocity consists of two components. The first component is determined by relative translational velocity and encodes the structure of the scene, Z. The second component depends only on rotational motion about the viewer centre (eye movements). It gives no useful information about the depth of the point or the shape of the visible surface. It is this rotational component which complicates the interpretation of visual motion. The effects of rotation are hard to extricate however, although numerous solutions have been proposed [11]. As a consequence, point image velocities and disparities do not encode shape in a simple efficient way since the rotational component is often arbitrarily chosen to shift attention and gaze by camera rotations or eye movements. The rotational component can be removed if, instead of using raw image motion the difference of the image motions of a pair of points, is used. This is called *motion parallax*.

2.2 Motion Parallax

Consider two visual features at different depths whose projections on the image plane are instantaneously (x_i, y_i) $1, 2$ and which have image velocities given by (1). If these two features are instantaneously coincident in the image, $(x_1, y_1) = (x_2, y_2) = (x, y)$, their relative image velocity, (Δ_u, Δ_v) – motion parallax – depends only on their relative inverse-depths and on viewer translational velocity. It is independent of (and hence insensitive to errors in) the angular rotation $\mathbf{\Omega}$:

$$\begin{aligned} \Delta_u &= (fU_1 - xU_3)\left[\frac{1}{Z_1} - \frac{1}{Z_2}\right] \\ \Delta_v &= (fU_2 - yU_3)\left[\frac{1}{Z_1} - \frac{1}{Z_2}\right] \end{aligned} \quad (2)$$

Equations (2) can be used to a recover a linear constraint on the direction of translation. Namely:

$$\frac{\Delta_u}{\Delta_v} = \frac{(fU_1 - xU_3)}{(fU_2 - yU_3)} \quad (3)$$

The use of "motion parallax" for robust determination of the direction of translation $\mathbf{U}$ and relative depths from image velocities was described by Longuet-Higgins and Prazdny [10] and Rieger and Lawton [12]. The theory above relating relative depth to parallax however assumed that the two points were instantaneously coincident in the image. In practice, point pairs used as features will not coincide and this formulation can not be used in general. In the next section we will show how an effective motion parallax vector can be computed by considering the image velocities of points in a small neighbourhood. We first review the invariants of the image velocity field and how they relate to 3D shape and motion.

2.3 Affine transformation

For a sufficiently small field of view (defined precisely in [13]) and smooth change in viewpoint the image velocity field and the change in apparent image shape for a smooth surface is well approximated by a linear (*affine*) transformation [6, 5].

To first order the image velocity field at a point (x, y) in the neighbourhood of a given visual direction is given

by:

$$\begin{bmatrix} u \\ v \end{bmatrix} \approx \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (4)$$

The first term is a vector $[u_0, v_0]$ representing a pure translation (specifying the change in image position of the centroid of the shape) while the second term is a 2×2 tensor – the velocity gradient tensor – and represents the distortion of the image shape. The latter can be decomposed into independent components which have simple geometric interpretations. These are a 2D rigid rotation (vorticity), specifying the change in orientation, curl$\vec{\mathbf{v}}$; an isotropic expansion (divergence) specifying a change in scale, div$\vec{\mathbf{v}}$; and a pure shear or deformation which describes the distortion of the image shape (expansion in a specified direction with contraction in a perpendicular direction in such a way that area is unchanged) described by a magnitude, def$\vec{\mathbf{v}}$, and the orientation of the axis of expansion (maximum extension), μ. These quantities can be defined as combinations of the partial derivatives of the image velocity field, $\vec{\mathbf{v}} = (u, v)$, at an image point (x, y).

2.4 Differential invariants of image velocity field and their relation to 3D shape and motion

The differential invariants depend on the viewer motion and depth, Z and the relation between the viewing direction (ray, $\mathbf{Q}$) and the surface orientation in a simple and geometrically intuitive way. They are summarised below. We define two 2D vector quantities: $\mathbf{A}$, the component of translational velocity parallel to the image plane scaled by depth, Z, $((U_1/Z, U_2/Z))$ and $\mathbf{F}$ to represent the surface orientation:

$$|\mathbf{F}| = \tan\sigma \quad (5)$$
$$\angle\mathbf{F} = \tau \quad (6)$$

where σ and τ are the *slant* and *tilt* of the surface respectively.

$$\text{curl}\vec{\mathbf{v}} = -2\boldsymbol{\Omega}.\mathbf{Q} + f\mathbf{F}\wedge\mathbf{A} \quad (7)$$
$$\text{div}\vec{\mathbf{v}} = \frac{2\mathbf{U}.\mathbf{Q}}{\lambda} + f\mathbf{F}.\mathbf{A} \quad (8)$$
$$\text{def}\vec{\mathbf{v}} = f|\mathbf{F}||\mathbf{A}| \quad (9)$$

where μ (which specifies the axis of maximum extension) bisects $\mathbf{A}$ and $\mathbf{F}$:

$$\mu = \frac{\angle\mathbf{A} + \angle\mathbf{F}}{2}. \quad (10)$$

The geometric significance of these equations is easily seen with a few examples. For example, a translation towards the surface patch leads to a uniform expansion in the image, i.e. positive divergence. This encodes the distance to the object which due to the speed–scale ambiguity is more conveniently expressed as a time to contact, t_c. Translational motion perpendicular to the visual direction results in image deformation with a magnitude which is determined by the slant of the surface, σ and with an axis depending on the tilt of the surface, τ and the direction of the viewer translation. Divergence (due to foreshortening) and curl components may also be present.

Note that divergence and deformation are unaffected by (and hence insensitive to errors in) viewer rotations such as panning or tilting of the camera whereas these lead to considerable changes in point image velocities or disparities.

We note that measurement of the differential invariants in a single neighbourhood is insufficient to to completely solve for the structure and motion since (7,8,9,10) are four equations in the six unknowns of scene structure and motion. In a single neighbourhood a complete solution would require the computation of second order derivatives [10] to generate sufficient equations to solve for the unknowns. Even then the solution of the resulting set of non-linear equations is non-trivial.

In [1] we show how the 3D interpretation of the differential invariants of the image velocity field is especially suited to the domain of *active vision* in which the viewer makes deliberate (although sometimes imprecise) motions, or in stereo vision, where the relative positions of the two cameras (eyes) are constrained while the cameras (eyes) are free to make arbitrary rotations (eye movements). Estimates of the divergence and deformation of the image velocity field, augmented with constraints on the direction of translation, are then sufficient to efficiently determine the object surface orientation and time to contact. In this sequel we show how to use the differential invariants measured from a minimum of three points and the relative motion of a fourth point to efficiently and reliably estimate the certain attributes of the scene structure and the 3D motion.

3 Parallax-based SFM

3.1 Pseudo-parallax

We now describe the main theoretical contribution of this paper. We present a method that computes an effective *parallax motion* even when image features do not coincide.

Consider the image motion of a point P in the image plane. In a small neighbourhood of P consider the image motion of a triplet of points A, B, C (figure 1). As shown above for a small enough neighbourhood the image velocities in the plane defined by the three points can be approximated by an affine transformation. The velocity of a virtual point, P^*, which is coincident with P but lies on the plane can thus be determined as a linear sum of the image velocities of the other three points. The difference between the motion of the virtual point, P^*, and the real point, P, is equivalent to the motion parallax between P and a point coincident in the image but at a different depth. As shown above this pseudo–parallax vector constrains the direction of translation and allows us to effectively cancel the effects of viewer rotations.

We now show below that the analysis of structure from motion based on pseudo-parallax instead of raw image velocities is considerably simplified.

3.2 3D qualitative interpretation

We now show how to recover reliable, although incomplete shape and motion properties from the image velocity of points relative to a triplet of features in a small neighbourhood.

The main result follows directly from the parallax result described above. Namely that the direction of the parallax velocity can determine a constraint on the the projection of the direction of translation, $\angle\mathbf{A}$, when we consider the image velocities in its neighbourhood. Note that we have not determined the magnitude of $\mathbf{A}$. This would, in fact, be equivalent to knowing the direction of translation. We have simply determined a line in the image in which the direction of translation must pierce the image plane. Without loss of generality assume that position of the fourth point is aligned with the optical axis at $(0,0)$. This can always be achieved by rotating the camera about its optical centre. A solution can be obtained in the following way.

1. Determine the projection of the direction of translation, $\angle\mathbf{A}$, from the relative image motion of a fourth point relative to the image motion of a neighbourhood triplet. Note that if the visual motion arises from the rotation of a rigid object in-

   front of a stationary camera, the projection of the axis of rotation will be perpendicular to $\mathbf{A}$.

2. Compute the curl, divergence and deformation (axes and magnitude) from the image velocities of the 3 points from (7,8,9,10).

3. The axis of expansion (μ) of the deformation component and the projection in the image of the direction of translation ($\angle\mathbf{A}$) allow the recovery of the *tilt*, τ, of the planar triangle from (10).

4. The *slant* of the surface can not be fixed but is constrained depending on the magnitude of $\mathbf{A}$ by (9). This is an exposition of the *bas-relief* ambiguity (explained below). Knowing the "turn" of the object allows us to fix the orientation of the surface and vice versa. However, in general, from 2 views with no perspective effects surface orientation is recovered as a one-parameter family of solutions.

5. Having determined the tilt of the surface and the slant as a function of $|\mathbf{A}|$ it is possible to recover the important relative motion parameters such as change in overall scale and rotation about the image axis from the equations relating image divergence and curl to the motion and structure parameters. We can then subtract the contribution due to the surface orientation and viewer translation parallel to the image axis from the image divergence (8). This is equal to $|\mathrm{def}\vec{\mathbf{v}}|\cos(\tau - \angle\mathbf{A})$. The remaining component of divergence is due to movement towards or away from the object. This can be used to recover the time to contact, t_c or to express the change in overall scale due to a change in the distance between the object and viewer, U_3/Z. This can be recovered despite the fact that the viewer translation may not be parallel to the visual direction.

6. Similarly we can then subtract the contribution due to the surface orientation and viewer translation parallel to the image axis from the image curl (7). This is equal to $|\mathrm{def}\vec{\mathbf{v}}|\sin(\tau - \angle\mathbf{A})$. The remaining component of curl is due to a rotation of the object/camera about the direction of the ray (the cyclotorsion), Ω_3.

The advantage of this formulation is that camera rotations do not affect the estimation of shape and distance. The effects of errors in the direction of translation are clearly evident as scalings in depth or by a 3D affine transformation [7]. The quantities listed above are the only parameters which can be reliably extracted from the image velocities in a small field of view.

The *bas-relief* ambiguity manifests itself in the appearance of surface orientation, $\mathbf{F}$, with $\mathbf{A}$. Increasing the slant of the surface $\mathbf{F}$ while scaling the movement by the same amount will leave the local image velocity field unchanged. Thus, from two weak perspective views and with no knowledge of the viewer translation, it is impossible to determine whether the deformation in the image is due to a large $|\mathbf{A}|$ (equivalent to a large "turn" of the object or "vergence angle") and a small slant or a large slant and a small rotation around the object. Equivalently a nearby "shallow" object will produce the same effect as a far away "deep" structure. We can only recover the depth gradient $\mathbf{F}$ up to an unknown scale. These ambiguities are clearly exposed with this analysis whereas this insight is sometimes lost in the purely algorithmic approaches to solving the equations of motion from the observed point image velocities. A consequence of the latter is the numerically ill-conditioned nature of structure from motion solutions when perspective effects are small. In this analysis we have avoided attempting to recover absolute surface orientations. The resulting 3D shape and motion is however qualitative since we have not been able to recover the direction of translation.

4 Implementation and Applications

We have shown that the image motion of a minimum of four arbitrary points on a moving rigid object can be used to describe qualitatively the translation and rotation of a rigid object. In particular for a rotating object in front of a stationary camera image translations can be interpreted as *small* object translations parallel to the

image plane; changes in scale (computed from the divergence after subtracting the effects of foreshortening) are interpreted as movement along the optical axis; motion parallax is interpreted as resulting from the component of rotation of a rigid object about an axis parallel to the image plane; and 2D image rotations (computed from curl component after subtracting the component due to surface orientation) are interpreted as a rotation about the optical axis. This solution is not complete since we are not able to determine the exact ratios of the components of translation and rotation parallel to the image plane to those along the optical axis. The information extracted is however insensitive to small perspective effects and can be used in many tasks requiring 3D inferences of shape and motion.

We now describe a simple implementation in which this information is used to interpret hand and head gestures for a man–machine interface by tracking appropriate features. We present a simple real-time example in which the 3D hand gestures are used as the interface to a graphics system to generate changes in the viewing position and orientation of an object displayed on a computer graphics system.

The 3D motions of the hand are automatically interpreted as either small translations parallel to the image plane (image translations with zero parallax motion and zero deformations); changes in scale (zero parallax motion with non-zero divergence); rotations of the object about an axis specified by the parallax motion vector (non-zero parallax, deformation, curl and divergence).

In the present implementation 4 colour markers placed on a hand-held object (figure 2) or a glove (figure 3) are tracked in real-time (50Hz) using purpose built image processing system for detecting and tracking image features [8]. The interpretation of the visual motion is carried out on a host workstation and its results are communicated to a graphics workstation which responds by changing the position and orientation of a computer graphics model (see figure 3). Since the algorithm does not produce quantitative values of rotation it must be used with visual feedback – the user must continues to rotate or translation his hand until the object has rotated/translated by the desired amount. Real-time tests at the Tokyo Data Show 1992 have successfully demonstrated the usefulness and reliability of this partial solution to structure from motion.

5 Conclusions

We have presented an efficient and geometrically intuitive algorithm to reliably interpret the image velocities of a minimum of four points on a moving objects under weak perspective using motion parallax. Preliminary implementation based on tracking coloured markers has proved the power and reliability of this algorithm even in the presence of small perspective effects and non-rigidity. The solution is however incomplete. In principle it can be augmented into a complete solution to structure from motion with additional constraints. Knowledge of the *slant* of the plane containing 3 of the reference points from monocular cues, for example, allows us to determine the exact direction of translation or angle of rotation of the object. Adding additional views will also allow a complete solution but this may, in general, be ill-conditioned unless a large number of views and image velocities are processed [14]. We believe, however, that the qualitative partial solution is preferable in many visual tasks which require shape and motion cues since it can be computed reliably and efficiently.

We are aiming to develop a more practical glove free real-time system to detect and track arbitrary grey-level image features using *cross-correlation*. We are also investigating algorithms to group image velocities into independently moving rigid body motions based on their parallax velocities. We are also improving the reliability of the solution by integrating information from more points and views.

Acknowledgements

Roberto Cipolla acknowledges discussions with Christopher Longuet-Higgins and Steve Maybank and the support of the Toshiba (visiting researcher) fellowship.

References

[1] R. Cipolla and A. Blake. Surface orientation and time to contact from image divergence and deformation. In *Proc. 2nd European Conference on Computer Vision*. Springer–Verlag, 1992.

[2] O.D. Faugeras and S.J. Maybank. Motion from point matches: multiplicity of solutions. In *IEEE Workshop on Motion*, pages 248–255, Irvine,1989.

[3] C.G. Harris. Determination of ego-motion from matched points. In *3rd Alvey Vision Conference*, pages 189–192, 1987.

[4] C.G. Harris. Structure from motion under orthographic projection. In *Proc. 1st European Conference on Computer Vision*, pages 118–123. Springer–Verlag, 1990.

[5] J.J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986.

[6] J.J. Koenderink and A.J. Van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22(9):773–791, 1975.

[7] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. *J. Opt. Soc. America*, pages 377–385, 1991.

[8] H. Kubota, Y. Okamoto, H. Mizoguchi, and Y. Kuno. Vision processor for moving object analysis. In B Zavigovique and P.L. Wendel, editors, *Computer Architecture for Machine Perception*, pages 461–470. 1992.

[9] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[10] H.C. Longuet-Higgins and K. Pradzny. The interpretation of a moving retinal image. *Proc. R. Soc. Lond.*, B208:385–397, 1980.

[11] S.J. Maybank. *A theoretical study of optical flow.* PhD thesis, Birbeck College, University of London, 1987.

[12] J.H. Rieger and D.L. Lawton. Processing differential image motion. *J. Opt. Soc. America*, A2(2):354–360, 1985.

[13] D.W. Thompson and J.L. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *Proceedings of IEEE Conference on Robotics and Automation*, 1987.

[14] C Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *Int. Journal of Computer Vision*, 9(2), 1992.

[15] R.Y. Tsai and T.S. Huang. Uniqueness and estimation of three-dimensional motion parameters of a rigid objects with curved surfaces. *IEEE Trans. PAMI*, 6(1):13–26, 1984.

[16] S. Ullman. *The interpretation of visual motion.* MIT Press, Cambridge,USA, 1979.

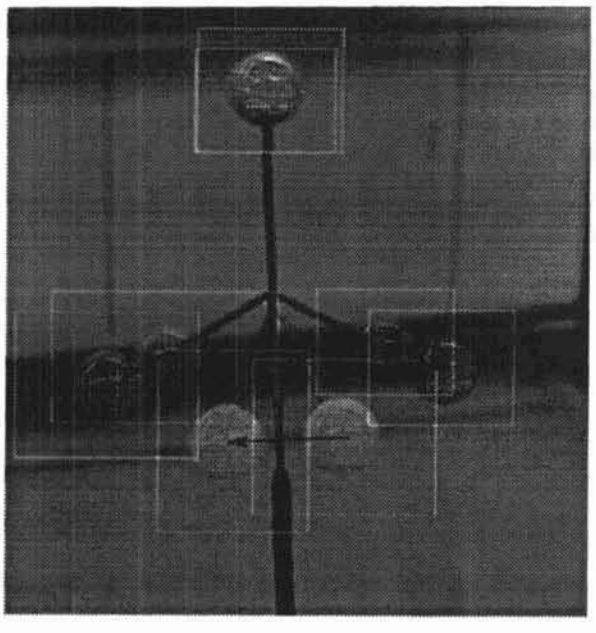

Figure 2: Real-time tracking of coloured markers and measurement of visual motion.

Colour marker detection and tracking is performed on a purpose built image processor. Colour markers are detected by comparing the pixels intensities in a validation window generated by the tracker to the colour of the feature being tracked (taught by showing in the beginning of each session). If a colour blob is detected its convex hull co-ordinates are passed onto a tracker which controls the position of the search/validation window in the next image. If a colour blob is not found the validation window size is doubled until it reaches its maximum of 128 × 128. Detected pixels and windows for each feature are shown superimposed on the image of 4 colour attached to a glove. Each window is controlled by a separate processor. The spatial positions of the balls is unknown. The image motion of these 4 points is sufficient to determine the 3D motion qualitatively. In this case the motion is correctly interpreted as arising from a rotation about a vertical axis.

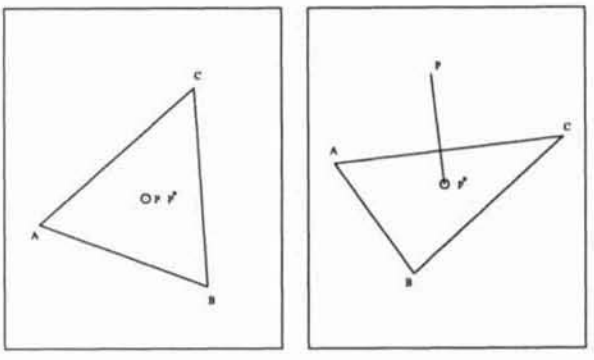

Figure 1: Motion parallax from the image motion of a point P relative to a triangle of 3 points.

Motion parallax is defined as the relative image motion of 2 points which are instantaneously coincident in the image but at different depths. This can be computed in practice from the relative motion of a 4th point, P, relative to a small neighbourhood defined by a triangle of three image points, A, B, C. The image positions of A, B, C can be used to predict the image position of a virtual point lying on the same plane, P^* and instantaneously co-incident with P in the first view. The two points will not, however, coincide in the second view (unless P lies in the same plane as A, B, C) and their relative velocity, P^*P is equivalent to the motion parallax.

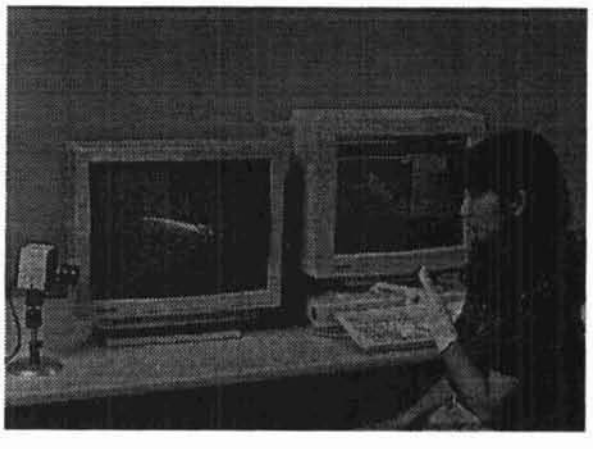

Figure 3: Vision interface using 3D hand gestures and a wireless glove.

Movement of the hand results in relative image motion between the images of the coloured markers. The parallax motion vector and the divergence, curl and deformation components of the affine transformation of an arbitrary triangle of points are sufficient to determine the projection of the axis of rotation, change in scale (zoom) and the cyclotorsion. This information is sent to a computer graphics workstation and the image of a model is changed accordingly (translation, rotation and change in scale).