Image Divergence and Deformation from Closed Curves

Roberto Cipolla[*] and Andrew Blake[†]

March 19, 1997

Abstract

This paper describes a novel method to measure the differential invariants of the image velocity field from the integral of normal image velocities around image contours. This is equivalent to measuring the temporal changes in the area of a closed contour. This avoids having to recover a dense image velocity field and taking partial derivatives. It also does not require point or line correspondences. Moreover integration provides some immunity to image measurement noise.

It is shown how an *active* observer making small, deliberate motions can use the estimates of the divergence and deformation of the image velocity field to determine the object surface orientation and time to contact. The results of real-time experiments are presented in which arbitrary image shapes are tracked using B-spline snakes and the invariants are computed efficiently as closed-form functions of the B-spline control points. This information is used to guide a robot manipulator in obstacle collision avoidance, object manipulation and navigation.

1 Introduction

Relative motion between an observer and a scene induces deformation in image detail and shape. If these changes are smooth they can be economically described locally by the first order differential invariants of the image velocity field [23] – the curl (vorticity), divergence (dilation), and deformation (pure shear) components. These invariants have simple geometrical interpretations which do not depend on the particular choice of co-ordinate system. Moreover they are related to the three dimensional structure of the scene and the viewer's motion – in particular the surface orientation and the time to contact [1] – in a simple geometrically intuitive way. Better still, the divergence and deformation components of the image velocity field (measured on the image sphere or at the optical axis) are unaffected by arbitrary viewer rotations about the projection centre. They therefore provide an efficient, reliable way of recovering these parameters.

[*]Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, England.
[†]Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, England.

[1]The time duration before the observer and object collide if they continue with the same relative translational motion [19, 15]

Although the analysis of the differential invariants of the image velocity field has attracted considerable attention [23, 21] their application to real tasks requiring visual inferences has been disappointingly limited [32, 14]. This is because existing methods have failed to deliver reliable estimates of the differential invariants when applied to real images. They have attempted the recovery of dense image velocity fields [6] or the accurate extraction of points or corner features [21]. Both methods have attendant problems concerning accuracy and numerical stability. An additional problem concerns the domain of applications to which estimates of differential invariants can be usefully applied. First order invariants of the image velocity field at a single point in the image cannot be used to provide a *complete* description of shape and motion as attempted in numerous structure from motion algorithms [36]. This in fact requires second order spatial derivatives of the image velocity field [29, 37]. The advantage of the first order invariants lies in their ability to efficiently recover reliable but incomplete solutions to the structure from motion problem which can be augmented with other information to accomplish useful visual tasks.

The reliable, real-time extraction of these invariants from image data and their application to visual tasks will be addressed in this paper. First we present a novel method to measure the differential invariants of the image velocity field by computing average values from the integral of simple functions of the normal image velocities around image contours. This is equivalent to measuring the temporal changes in the area of a closed contour and avoids having to recover a dense image velocity field and taking partial derivatives. It only requires the gross correspondence of closed contours and does not require point or line correspondences. Moreover integration provides some immunity to image measurement noise.

Second we show that the 3D interpretation of the differential invariants of the image velocity field is especially suited to the domain of *active vision* [7] 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.

The results of preliminary real-time experiments in which arbitrary image shapes are tracked using B-spline snakes [8, 9, 10] are presented. The invariants are computed as closed-form functions of the B-spline snake control points. This information is used to guide a robot manipulator in obstacle collision avoidance, object manipulation and navigation.

2 Differential Invariants of the Image Velocity Field

2.1 Review

Consider the viewer to be moving relative to a stationary scene. Let $\mathbf{v}(x, y)$ represent the image velocity field – a 2D vector field defined on the image plane where each vector assigned to an image point (x, y) represents the image motion (u, v) of the scene point that projects to that image point. For a sufficiently small field of view and smooth change in viewpoint the image velocity field is well approximated by a linear (*affine*) transformation [23] (see Appendix A for derivation):

$$\begin{bmatrix} u \\ v \end{bmatrix} \approx \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{1}$$

where (u_0, v_0) is an image translation (specifying the change in image position of the centroid of the shape) and where the first order partial derivatives, (u_x, u_y, v_x, v_y) (subscripts denote differentiation with respect to the subscript parameter) define the velocity gradient tensor which describes the change in apparent image shape.

The velocity tensor gradient can be decomposed into independent components which have simple geometric interpretations. These are an isotropic expansion (divergence) specifying a change in scale, $\mathrm{div}\mathbf{v}$; a 2D rigid rotation (vorticity), specifying the change in orientation, $\mathrm{curl}\mathbf{v}$; 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, $\mathrm{def}\mathbf{v}$, and the orientation of the axis of expansion (maximum extension), μ (Figure 1 and Appendix B).

$$\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \frac{\mathrm{div}\mathbf{v}}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\mathrm{curl}\mathbf{v}}{2} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \frac{\mathrm{def}\mathbf{v}}{2} \begin{bmatrix} \cos 2\mu & \sin 2\mu \\ \sin 2\mu & -\cos 2\mu \end{bmatrix} \tag{2}$$

These quantities can be defined [2] as combinations of the partial derivatives of the image velocity field, $\mathbf{v} = (u, v)$, at an image point (x, y):

$$\mathrm{div}\mathbf{v} = (u_x + v_y) \tag{3}$$

$$\mathrm{curl}\mathbf{v} = -(u_y - v_x) \tag{4}$$

$$(\mathrm{def}\mathbf{v}) \cos 2\mu = (u_x - v_y) \tag{5}$$

$$(\mathrm{def}\mathbf{v}) \sin 2\mu = (u_y + v_x) \tag{6}$$

The curl, divergence and the magnitude of the deformation are scalar invariants and do not depend on the particular choice of image co-ordinate system [23, 21]. The axes of maximum extension and contraction change with rotations of the image plane axes.

2.2 Relation to 3D Shape and Viewer Motion

The differential invariants depend on the viewer motion (translational velocity, $\mathbf{U}$, and rotational velocity, $\mathbf{\Omega}$), scene structure (depth), Z, and the relation between the viewing direction (ray direction, unit vector $\mathbf{Q}$) and the surface orientation in a simple and geometrically intuitive way. Before summarising these relationships let us define two 2D vector quantities:

- $\mathbf{A}$ – The component of translational velocity parallel to the image plane scaled by depth, Z, where:

$$\mathbf{A} = \frac{\mathbf{U} - (\mathbf{U} \cdot \mathbf{Q})\mathbf{Q}}{Z} \tag{7}$$

[2] The divergence of a vector field and the magnitude of the curl component are defined by $\mathrm{div}\mathbf{v} = \nabla \cdot \mathbf{v}$ and $\mathrm{curl}\mathbf{v} = |\nabla \times \mathbf{v}|$ respectively.

Note that the direction of $\mathbf{A}$ is determined by the intersection of the epipolar plane (defined by direction of translation and ray) and the image plane. The magnitude is a measure of the angle between the two views in an infinitesimal sense. For an object which is rotating in front of a stationary camera it is a measure of the "turn" or rotation of the object.

- $\mathbf{F}$ – the *depth gradient* scaled by depth[3], is used represent the surface orientation and can be defined in terms of the 2D vector gradient:

$$\mathbf{F} = f\nabla(\log Z) \quad = \quad \frac{f\nabla Z}{Z} \ . \tag{8}$$

  where f is the distance between the image plane and optical centre. The magnitude of the depth gradient, $|\mathbf{F}|$, determines the tangent of the *slant* of the surface (angle between the surface normal and the visual direction). It vanishes for a frontal view and is infinite when the viewer is in the tangent plane of the surface. Its direction, $\angle\mathbf{F}$, specifies the direction in the image of increasing distance. This is equal to the *tilt* of the surface tangent plane. The exact relationship between the magnitude and direction of $\mathbf{F}$ and the slant and tilt of the surface (σ, τ) is derived in Appendix C and is given by:

$$|\mathbf{F}| \quad = \quad \tan\sigma \tag{9}$$
$$\angle\mathbf{F} \quad = \quad \tau \ . \tag{10}$$

With this new notation the relations between the differential invariants, the motion parameters and the surface position and orientation are given by [22] (see Appendix C for derivation):

$$\mathrm{div}\mathbf{v} \quad = \quad \frac{2\mathbf{U}\cdot\mathbf{Q}}{Z} + \mathbf{F}\cdot\mathbf{A} \tag{11}$$
$$\mathrm{curl}\mathbf{v} \quad = \quad -2\mathbf{\Omega}\cdot\mathbf{Q} + |\mathbf{F}\times\mathbf{A}| \tag{12}$$
$$\mathrm{def}\mathbf{v} \quad = \quad |\mathbf{F}||\mathbf{A}| \ . \tag{13}$$

The axis of maximum extension of the deformation component, μ, bisects $\mathbf{A}$ and $\mathbf{F}$:

$$\mu = \frac{\angle\mathbf{A} + \angle\mathbf{F}}{2} \ . \tag{14}$$

The geometric significance of these equations is easily seen with a few examples. For example, a translation along the ray towards the surface patch (i.e. $|\mathbf{A}| = 0$) leads to a uniform expansion in the image, i.e. positive divergence (figure 2). This encodes the distance to the object which, due to the speed–scale ambiguity[4], is more conveniently expressed as a time to contact, t_c:

$$t_c = \frac{Z}{\mathbf{U}\cdot\mathbf{Q}} \ . \tag{15}$$

Translational motion perpendicular to the visual direction results in image deformation with a magnitude which is determined by the slant of the surface, σ,

[3] Koenderink [22] defines $\mathbf{F}$ as a "nearness gradient", $\nabla(\log(1/Z))$. In this paper $\mathbf{F}$ is defined as a scaled depth gradient. These two quantities differ by a sign.

[4] Translational velocities appear scaled by depth making it impossible to determine whether the effects are due to a nearby object moving slowly or a far-away object moving quickly.

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 (figures 2 and 3).

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[5]. As a consequence the deformation component efficiently encodes the orientation of the surface while the divergence component can be used to provide an estimate of the time to contact or collision.

This formulation clearly exposes both the speed–scale ambiguity and the *bas–relief* ambiguity [16]. The latter manifests itself in the appearance of surface orientation, $\mathbf{F}$, with $\mathbf{A}$ in the equations relating visual motion to surface position and orientation. 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 $|\mathbf{A}|$ (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.

3 Extraction of Differential Invariants

The analysis above treated the differential invariants as observables of the image. There are a number of ways of extracting the differential invariants from the image. These are summarised below before presenting a novel method based on the temporal derivatives of the moments of the area enclosed by a closed curve.

3.1 Summary of Existing Methods

1. **Partial derivatives of image velocity field**

   This is the most common approach. It is based on recovering a dense field of image velocities and computing the partial derivatives using discrete approximation to derivatives [25] or a least squares estimation of the affine transformation parameters from the image velocities (a minimum of six *normal* velocities) estimated by spatio-temporal methods [32, 1]. The

[5] This is somewhat related to the reliable estimation of relative depth from the relative image velocities of two nearby points – motion parallax [29, 33, 10]. Both motion parallax and the deformation of the image velocity field relate local measurements of relative image velocities to scene structure in a simple way which is uncorrupted by the rotational image velocity component. In the case of parallax, the depths are discontinuous and differences of discrete velocities are related to the difference of inverse depths. Equation (13) on the otherhand assumes a smooth and continuous surface and derivatives of image velocities are related to derivatives of inverse depth.

recovery of the image velocity field is usually computationally expensive and ill-conditioned [18].

2. **Point velocities in a small neighbourhood**

   The image velocities of a minimum of three points in a small neighbourhood are sufficient, in principle, to estimate the components of the affine transformation and hence the differential invariants [21, 26]. In fact it is only necessary to measure the change in area of the triangle formed by the three points and the orientations of its sides. There is, however, no redundancy in the data and hence this method requires very accurate image positions and velocities. In [11] this is attempted by tracking large numbers of "corner" features [17] and using Delaunay triangulation [5] in the image to approximate the physical world by planar facets. Preliminary results showed that the localisation of "corner" features was insufficient for reliable estimation of the differential invariants.

3. **Relative orientation of line segments**

   Koenderink [22] showed how temporal texture density changes can yield estimates of the divergence.

   From equation(2) it is easy to show that the change in orientation (clockwise), $\Delta\phi$ of an element with orientation ϕ is given to first order by [24]

   $$\Delta\phi = -\frac{\mathrm{curl}\mathbf{v}}{2} + \frac{1}{2}\mathrm{def}\mathbf{v}\sin 2(\phi - \mu). \tag{16}$$

   Orientations are not affected by the divergence term. They are only affected by the curl and deformation components. In particular the curl component changes all the orientations by the same amount. It does not affect the angles between the image edges. These are only affected by the deformation component. The relative changes in orientation can be used to recover deformation in a simple way since the effects of the curl component are cancelled out. By taking the difference of (16) for two orientations, ϕ_1 and ϕ_2, it is easy to show (using simple trigonometric relations) that the relative change in orientation specifies both the magnitude, $\mathrm{def}\mathbf{v}$, and axis of expansion, μ, of the shear as shown below.

   $$\Delta\phi_2 - \Delta\phi_1 = \mathrm{def}\mathbf{v}\left[\sin(\phi_2 - \phi_1)\cos 2\left(\frac{\phi_2 + \phi_1}{2} - \mu\right)\right]. \tag{17}$$

   Measurement at three oriented line segments is sufficient to completely specify the deformation components. Note that the recovery of deformation can be done without any explicit co-ordinate system and even without a reference orientation. The main advantage is that point velocities or partial derivatives are not required.

4. **Curves and closed contours**

   The methods described above require point and line correspondences. Sometimes these are not available or are poorly localised. Often we can only reliably extract portions of curves (although we can not always rely on the end points) or closed contours.

Image shapes or contours "sample" the image velocity field. At contour edges it is only possible to measure the normal component of image velocity. This information can in certain cases be used to recover the image velocity field. Waxman and Wohn [38] showed how to recover the full velocity field from the normal components at image contours. In principle, measurement of eight normal velocities around a contour allow the characterisation of the full velocity field for a planar surface. Kanatani [20] related line integrals of image velocities around closed contours to the motion and orientation parameters of a planar contour. In the following we will not attempt to solve for these structure and motion parameters directly but only to recover the divergence and deformation.

3.2 Recovery of Invariants from Area Moments of Closed Contours

It has been shown that the differential invariants of the image velocity field conveniently characterise the changes in apparent shape due to relative motion between the viewer and scene. Contours in the image sample this image velocity field. It is usually only possible, however, to recover the normal image velocity component from local measurements at a curve [36, 18]. It is now shown that this information is often sufficient to estimate the differential invariants within closed curves.

Our approach is based on relating the temporal derivative of the area of a closed contour and its moments to the invariants of the image velocity field (figure 4). This is a generalisation of the result derived by Maybank [30] in which the rate of change of area scaled by area is used to estimate the divergence of the image velocity field. The advantage is that point or line correspondences are not used. Only the correspondence between shapes is required. The computationally difficult, ill-conditioned and poorly defined process of making explicit the full image velocity field [18] is avoided. Moreover, since taking temporal derivatives of area (and its moments) is equivalent to the integration of normal image velocities (scaled by simple functions) around closed contours our approach is effectively computing average values of the differential invariants (not point properties) and has better immunity to image noise leading to more reliable estimates. Areas can also be estimated accurately, even when the full set of first order derivatives cannot be obtained.

The moments of area of a contour, I_g, are defined in terms of an area integral with boundaries defined by the contour in the image plane:

$$I_g = \iint_{a(t)} g(x,y)\,dx\,dy \tag{18}$$

where $a(t)$ is the area of a contour of interest at time t and g is a scalar function of image position (x, y) that defines the moment of interest. For instance setting $g = 1$ gives us area. Setting $g = x$ or $g = y$ gives the first-order moments about the image x and y axes respectively.

The moments of area can be measured directly from the image (see below for a novel method involving the control points of a B-spline snake). Better still, their temporal derivatives can also be measured. Differentiating (18) with respect to time and using a result from calculus[6] we can relate the temporal derivative of the moment of area to an integral around the contour, $c(t)$ of the normal component of image velocities $\mathbf{v}.\mathbf{n}$, weighted by a scalar $g(x,y)$.

$$\frac{d}{dt}(I_g) = \oint_{c(t)} [g\mathbf{v}.\mathbf{n}]ds \tag{19}$$

By Green's theorem the contour integral can be re-expressed as an integral over the area enclosed by the contour, $a(t)$, of the divergence of the vector field:

$$\frac{d}{dt}(I_g) = \iint_{a(t)} [\nabla.(g\mathbf{v})]dxdy \tag{20}$$

If we assume that the velocity field is linear in the area of interest (i.e. partial derivatives of $\mathbf{v}$ are constant) and using simple vector algebra:

$$\nabla.(g\mathbf{v}) = g\nabla.\mathbf{v} + \nabla g.\mathbf{v} \tag{21}$$

to substitute into (20) we obtain a linear equation between the temporal derivative of the moment of area I_g (which can be measured, see below) and the velocity field together with its partial derivatives. Each term is a parameter of the affine transformation. The coefficients of each term are moments of area (also directly measurable).

$$\frac{d}{dt}(I_g) = \iint_{a(t)} [g\nabla.\mathbf{v} + \nabla g.\mathbf{v}]dxdy \tag{22}$$

In summary, the image velocity field deforms the shape of contours in the image (figure 4). Shape can be described by moments of area. Hence measuring the change in the moments of area is an alternative way characterising the transformation. In this way the change in the moments of area have been expressed in terms of the parameters of the affine transformation.

If we initially set up the x, y co-ordinate system at the centroid of the image contour of interest so that the first moments are zero, (22) with $g = x$ and $g = y$ shows that the centroid of the deformed shape specifies the mean translation $[u_0, v_0]$. Setting $g = 1$ leads to the simple and useful result that the divergence of the image velocity field can be estimated as the derivative of area scaled by area:

$$\frac{da(t)}{dt} = \iint \nabla.\mathbf{v}dxdy \quad \approx \quad a(t)\mathrm{div}\mathbf{v} \ . \tag{23}$$

Increasing the order of the moments, i.e. different values of $g(x,y)$, generates new equations and additional constraints. Some examples which can be used to estimate the affine transformation are listed below.

$$\frac{d}{dt}\begin{bmatrix} a \\ I_x \\ I_y \\ I_{x^2} \\ I_{y^2} \\ I_{x^3y} \end{bmatrix} = \begin{bmatrix} 0 & 0 & a & 0 & 0 & a \\ a & 0 & 2I_x & I_y & 0 & I_x \\ 0 & a & I_y & 0 & I_x & 2I_y \\ 2I_x & 0 & 3I_{x^2} & 2I_{xy} & 0 & I_{x^2} \\ 0 & 2I_y & I_{y^2} & 0 & 2I_{xy} & 3I_{y^2} \\ 3I_{x^2y} & I_{x^3} & 4I_{x^3y} & 3I_{x^2y^2} & I_{x^4} & 2I_{x^3y} \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ u_x \\ u_y \\ v_x \\ v_y \end{bmatrix} \ . \tag{24}$$

[6] This equation can be derived by considering the *flux* linking the area of the contour. This changes with time since the contour is carried by the velocity field. The *flux* field, g, in our example does not change with time. Similar integrals appear in fluid mechanics, e.g. the *flux transport theorem* [13].

(Note that in this equation the subscripts of I are used to label the moments of area. The left-hand side represents the temporal derivative of the moments in the column vector.)

In principle, if it is possible to find six linearly independent equations, we can solve for the affine transformation parameters and combine the coefficients to recover the differential invariants. The validity of the affine approximation can be checked by looking at the error between the transformed and observed image contours.

Some contours may lead to equations which are not independent and their solution is ill-conditioned. The interpretation of this is that the normal components of image velocity are insufficient to recover the true image velocity field globally, e.g. a fronto-parallel circle rotating about the optical axis. This was termed the "aperture problem in the large" by Waxman and Wohn [38] and investigated by Berghom and Carlsson [4]. Note however, that it is always possible to recover the divergence directly from the area of a closed contour. In practice higher order moments are more sensitive to image noise.

4 Recovery of Surface Orientation and Time to Contact

Applications of the estimates of the image divergence and deformation of the image velocity field are summarised below. It has already been noted that measurement of the differential invariants in a single neighbourhood is insufficient to completely solve for the structure and motion since (11–14) 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 [29, 37] 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 the following, the information available from the first-order differential invariants alone is investigated. It will be seen that the differential invariants are sufficient to constrain surface position and orientation and that this partial solution can be used to perform useful visual tasks when augmented with additional information. Useful applications include providing information which is used by pilots when landing aircraft [15], estimating time to contact in braking reactions [28] and in the recovery of 3D shape up to a 3D affine (relief) transformation [26, 27]. We now show how surface orientation and position (expressed as a time to contact) can be recovered from the estimates of image divergence and the magnitude and axis of the deformation component.

1. **With knowledge of translation but arbitrary rotation**

   An estimate of the direction of translation is usually available when the viewer is making deliberate movements (in the case of active vision) or in the case of binocular vision (where the camera or eye positions are constrained). It can also be estimated from image measurements by motion parallax [29, 33, 12].

   If the viewer translation is known, (11),(13) and (14) are sufficient to unambiguously recover the surface orientation and the distance to the object. Due to the speed–scale ambiguity the latter is expressed as a time to contact. A solution can be obtained in the following way.

(a) 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 surface from (14).

$$\tau = 2\mu - \angle\mathbf{A} \tag{25}$$

(b) We can then subtract from the image divergence component the contribution due to foreshortening. This arises from the component of viewer translation parallel to the image axis and depends on the surface orientation.

From (11) it is easy to show that this component is equal to $\mathrm{def}\mathbf{v}\cos(\tau - \angle\mathbf{A})$. The remaining component of divergence is due to movement towards or away from the object. Rearranging (11) allows us to recover the time to contact, t_c.

$$\frac{2}{t_c} = \mathrm{div}\mathbf{v} - \mathrm{def}\mathbf{v}\cos(\tau - \angle\mathbf{A}) \tag{26}$$

This can be recovered despite the fact that the viewer translation may not be parallel to the visual direction.

(c) The time to contact fixes the viewer translation in temporal units. It allows the specification of the magnitude of the translation parallel to the image plane (up to the same speed–scale ambiguity), $\mathbf{A}$. The magnitude of the deformation can then be used to recover the slant, σ, of the surface from (13).

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 (relief) transformation [22].

2. **With fixation**

If the cameras or eyes rotate to keep the object of interest in the middle of the image (null the effect of image translation) the magnitude of the rotations needed to bring the object back to the centre of the image determines $\mathbf{A}$ (see Appendix A (40– 41)) and hence allows us to solve for surface orientation, as above. Again the major effect of any error in the estimate of rotation is to scale depth and orientations.

3. **With no additional information – constraints on motion**

Even without any additional assumptions it is still possible to obtain useful information from the first-order differential invariants. The information obtained is best expressed as bounds. For example inspection of (11) and (13) shows that the time to contact must lie in an interval given by:

$$\frac{2}{(\mathrm{div}\mathbf{v} + \mathrm{def}\mathbf{v})} \leq t_c \leq \frac{2}{(\mathrm{div}\mathbf{v} - \mathrm{def}\mathbf{v})} \quad . \tag{27}$$

The lower bound on time to contact occurs when the component of viewer translation parallel to the image plane is in the opposite direction to the depth gradient. The upper bound occurs when the translation is parallel to the depth gradient. The upper and lower estimates of time to contact are equal when there is no deformation component. This is the case in which the viewer translation is along the ray or when viewing a fronto-parallel surface (zero depth gradient locally). The estimate of time to contact is then exact. A similar equation was recently described by Subbarao [35].

4. **With no additional information – the constraints on 3D shape**

   Koenderink and Van Doorn [26] showed that surface shape information can be obtained by considering the variation of the deformation component alone in a small field of view when weak perspective is a valid approximation. This allows the recovery of 3D shape up to a scale and 3D affine (relief) transformation. That is they effectively recover the axis of rotation of the object but not the magnitude of the turn. This yields a family of solutions depending on the magnitude of the turn. Fixing the latter determines the slants and tilts of the surface. This has recently been extended in the affine structure from motion theorem [27] and is equivalent to recovering 3D shape in the presence of a single motion parallax constraint [12].

The solutions presented above use knowledge from a single viewer translation and measurement of the divergence and deformation of the image velocity field only. An alternative solution exists if the observer is free to translate along the ray and also in two orthogonal directions parallel to the image plane. In this case measurement of divergence alone is sufficient to recover the surface orientation and the time to contact. This has the advantage that divergence can be estimated efficiently and reliably from the area enclosed by a single contour.

5 Implementation and Experimental Results

We now describe the implementation and application of the methods proposed in (3.2) and (4.1) to recover time to contact and surface orientation from image divergence and deformation estimated from area moments of closed image curves.

5.1 Tracking Closed Contours

Multi-span closed loop B-spline snakes [8, 7] are used to localise and track closed image contours. The B-spline is a curve in the image plane

$$\mathbf{x}(s) = \sum_i b_i(s)\mathbf{q}_i \tag{28}$$

where b_i are the spline basis functions with coefficients $\mathbf{q}_i$ (control points of the curve) and s is a curve parameter (not necessarily arc length) [3]. The snakes are initialised as points in the centre of the image and are forced to expand radially outwards until they are in the vicinity of an edge where image "forces" make the snake stabilise close to a high contrast closed contour. Subsequent image motion is automatically tracked by the snake [7].

B-spline snakes have useful properties such as local control and continuity. They also compactly represent image curves. In our applications they have the additional advantage that the area enclosed is a simple function of the control points. This also applies to the other area moments. From Green's theorem in the plane it is easy to show that the area enclosed by a curve with parameterisation $x(s)$ and $y(s)$ is given by [7]:

$$a = \int_{s_0}^{s_N} x(s) y'(s) ds \tag{29}$$

where $x(s)$ and $y(s)$ are the two components of the image curve and $y'(s)$ is the derivative with respect to the curve parameter s. For a B-spline, substituting (28) and its derivative:

$$\begin{aligned} a(t) &= \int_{s_0}^{s_N} \sum_i \sum_j (q_{x_i} q_{y_j}) b_i b'_j ds & (30) \\ &= \sum_i \sum_j (q_{x_i} q_{y_j}) \int_{s_0}^{s_N} b_i b'_j ds. & (31) \end{aligned}$$

Note that for each span of the B-spline and at each time instant the basis functions remain unchanged. The integrals can thus be computed off-line in closed form. (At most 16 coefficients need be stored. In fact due to symmetry there are only 10 possible values for a cubic B-spline). At each time instant multiplication with the control point positions gives the area enclosed by the contour. This is extremely efficient, giving the exact area enclosed by the contour. The same method can also be used for higher moments of area. The temporal derivatives of the area and its moments are then used to estimate image divergence and deformation.

5.2 Applications

Here we present the results of a preliminary implementation of the theory. The examples are based on a camera mounted on a robot arm whose translations are deliberate while the rotations around the camera centre are performed to keep the target of interest in the centre of its field of view. The camera intrinsic parameters (image centre, scaling factors and focal length) and orientation are unknown. The direction of translation is assumed known and expressed with bounds due to uncertainty.

5.2.1 Braking

Figure 5 shows four samples from a sequence of images taken by a moving observer approaching the rear windscreen of a stationary car in front. In the first frame (time $t = 0$) the relative distance between the two cars is approximately 7m. The velocity of approach is uniform and approximately 1m/time unit.

A B-spline snake is initialised in the centre of the windscreen, and expands out until it localises the closed contour of the edge of the windscreen. The snake can then automatically track the windscreen over the sequence. Figure 6 plots the apparent area, $a(t)$ (relative to the initial area, $a(0)$) as a function of time, t. For uniform translation along the optical axis the relationship between area

[7] Green's theorem in the plane can be written as: $\int_c F_1 dx + F_2 dy = \int\int_a (\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}) dx dy$. Substituting $F_2 = x$ and a change of variable gives one form of the well-known area relation.

and time can be derived from (11) and (23) by solving the first-order differential equation:

$$\frac{da}{dt} = \left(\frac{2}{t_c}\right) a(t) \ . \tag{32}$$

For uniform motion the time to contact decreases linearly with time:

$$t_c(t) = t_c(0) - t \tag{33}$$

and the solution of the differential equation is given by:

$$a(t) = \frac{a(0)}{\left[1 - \frac{t}{t_c(0)}\right]^2} \tag{34}$$

where $t_c(0)$ is the time to contact at time $t = 0$. This is in close agreement with the data (Fig. 6). This is more easily seen if we look at the variation of the time to contact with time. For uniform motion this should decrease linearly. The experimental results are plotted in Fig. 7. These are obtained by dividing the area of the contour at a given time by its temporal derivative (estimated by finite differences). The variation is linear, as predicted. These results are of useful accuracy, predicting the collision time to the nearest half time unit (corresponding to 50cm in this example).

For non-uniform motion the profile of the time to contact as a function of time is a very important cue for braking and landing reactions [28].

5.2.2 Collision avoidance

It is well known that image divergence can be used in obstacle collision avoidance. Nelson and Aloimonos [32] demonstrated a robotics system which computed divergence by spatio–temporal techniques applied to the images of highly textured visible surfaces. We describe a real-time implementation based on image contours and "act" on the visually derived information.

Figure 8 shows the results of a camera mounted on an Adept robot manipulator and pointing in the direction of a target contour. (We hope to extend this so that the robot initially searches by rotation for a contour of interest. In the present implementation, however, the target object is placed in the centre of the field of view.) The closed contour is then localised automatically by initialising a closed loop B-spline snake in the centre of the image. The snake "explodes" outwards and deforms under the influence of image forces which cause it to be attracted to high contrast edges.

The robot manipulator then makes a deliberate motion towards the target. Tracking the area of the contour and computing its rate of change allows us to estimate the divergence. For motion along the visual ray this is sufficient information to estimate the time to contact or impact. The estimate of time to contact – decreased by the uncertainty in the measurement and any image deformation (27) – can be used to guide the manipulator so that it stops just before collision (Fig. 8d). The manipulator in fact, travels "blindly" after its sensing actions (above) and at a uniform speed for the time remaining until contact. In repeated trials (see Fig. 9a and b) image divergences measured at distances of 0.5m to 1.0m were estimated accurately to the nearest half of a time unit. This corresponds to a positional accuracy of 20mm for a manipulator translational velocity of 40mm/s.

The affine transformation approximation breaks down at close proximity to the target. This may lead to a degradation in the estimate of time to contact when very close to the target.

5.2.3 Landing reactions and object manipulation

If the translational motion has a component parallel to the image plane, the image divergence is composed of two components. The first is the component which determines immediacy or time to contact. The other term is due to image foreshortening when the surface has a non-zero slant. The two effects can be computed separately by measuring the deformation. The deformation also allows us to recover the surface orientation.

Note that unlike stereo vision, the magnitude of the translation is not needed. The camera parameters (focal length and aspect ratio are not needed for divergence) are also not known. The magnitudes and directions of the camera rotations needed to keep the target in the field of view are also not needed. Simple measurements of area and its moments – obtained in closed form as a function of the B-spline snake control points – are used to estimate divergence and deformation. The only assumption is of uniform motion and known direction of translation.

Figure 9 shows an example in which a robot manipulator uses these estimates of time to contact and surface orientation to approach the object surface perpendicularly so as to position a suction gripper for manipulation. The image contours are shown in Fig. 2 highlighting the effect of deformation due to the sideways component of translation. The successful execution is shown in Fig. 9 c and d.

5.2.4 Qualitative visual navigation

Existing techniques for visual navigation have typically used stereo or the analysis of image sequences to determine the camera ego-motion and then the 3D positions of feature points. The 3D data are then analysed to determine, for example, navigable regions, obstacles or doors. An example of an alternative approach is presented. This computes qualitative information about the orientation of surfaces and times to contact from estimates of image divergence and deformation. The only requirement is that the viewer can make deliberate movements or has stereoscopic vision. Figure 10a shows the image of a door and an object of interest, a pallet. Movement towards the door and pallet produce a distortion in the image. This is seen as an expansion in the apparent area of the door and pallet in Fig. 10b. This can be used to determine the distance to these objects, expressed as a time to contact – the time needed for the viewer to reach the object if the viewer continued with the same speed. The image deformation is not significant. Any component of deformation can, anyhow, be absorbed by (27) as a bound on the time to contact. A movement to the left (Fig. 10c) produces image deformation, divergence and rotation. This is immediately evident from both the door (positive deformation and a shear with a horizontal axis of expansion) and the pallet (clockwise rotation with shear with diagonal axis of expansion). These effects with the knowledge of the direction of translation between the images taken at figure 10a and 10c are consistent with the door having zero tilt, i.e. horizontal direction of increasing depth, while the pallet has a tilt of approximately 90^o, i.e. vertical direction of increasing depth. These are the effects predicted by (11, 12, 13 and 14) even though there are also strong perspective effects in the images. They are sufficient to determine the orientation of the surface qualitatively (Fig. 10d). This has been done without knowledge of the intrinsic properties of the cameras (camera calibration), the orientations of the cameras, their rotations or translational velocities. No knowledge of epipolar geometry is used to determine exact image velocities or disparities. The solution is incomplete. It can, however, be easily augmented into a complete solution by adding additional information. Knowing the magnitude of the sideways translational velocity, for example, can determine the exact quantitative orientations of the visible surfaces.

6 Conclusions

We have presented a simple and efficient method for estimating image divergence and deformation by tracking closed image contours with B-spline snakes. This information has been successfully used to estimate surface orientation and time to contact and exploited in the visual guidance of a robot manipulator.

The implemtation described in this paper required a high contrast image contour on a flat surface and approximate knowledge of the direction of camera translation. We have recently extended this work to highly textured images [34] and when camera translation is unknown [12].

Acknowledgements

The authors acknowledge discussions with Mike Brady, Kenichi Kanatani, Christopher Longuet-Higgins, Steve Maybank and Andrew Zisserman.

Appendix A

Consider an arbitrary co-ordinate system (the final results will be invariant to this choice) 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 the ray be represented as a 2D vector field, $\mathbf{v}(x, y)$ 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 [29, 37]:

$$u = \left[\frac{-fU_1 + xU_3}{Z}\right] - f\Omega_2 + y\Omega_3 + \frac{xy}{f}\Omega_1 - \frac{x^2}{f}\Omega_2 \qquad (35)$$

$$v = \left[\frac{-fU_2 + yU_3}{Z}\right] + f\Omega_1 - x\Omega_3 - \frac{xy}{f}\Omega_2 + \frac{y^2}{f}\Omega_1 \qquad (36)$$

The image velocity consists of two components. The first component is determined by relative translational velocity and encodes the structure of the scene, $Z(x, y)$. 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 [31]. 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.

For a sufficiently small field of view and in a small neighbourhood of a smooth surface such that

$$\frac{\Delta Z}{Z_0} \ll 1 \tag{37}$$

$$\frac{x^2+y^2}{f^2} \ll 1 \tag{38}$$

where ΔZ is the variation in scene depth (relief) and $Z_0 = Z(0,0)$, it is possible to ignore the quadratic components of image motion. The image velocity field is well approximated by a translation in the image (u_0, v_0) and by the first-order partial derivatives of the image velocity (u_x, u_y, v_x, v_y). Namely to first order the image velocity field at a point (x, y) in the neighbourhood of the origin can be approximated by an affine transformation:

$$\begin{bmatrix} u \\ v \end{bmatrix} \approx \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{39}$$

where:

$$u_0 = -\frac{fU_1}{Z_0} - f\Omega_2 \tag{40}$$

$$v_0 = -\frac{fU_2}{Z_0} + f\Omega_1 \tag{41}$$

$$u_x = \frac{U_3}{Z_0} + \frac{fU_1 Z_x}{Z_0^2} \tag{42}$$

$$u_y = +\Omega_3 + \frac{fU_1 Z_y}{Z_0^2} \tag{43}$$

$$v_x = -\Omega_3 + \frac{fU_2 Z_x}{Z_0^2} \tag{44}$$

$$v_y = \frac{U_3}{Z_0} + \frac{fU_2 Z_y}{Z_0^2} \tag{45}$$

and where the x and y subscripts represent differentiation with respect to these spatial parameters. Note that the average image translation (u_0, v_0) can always be cancelled out by appropriate camera rotations (eye movements) (Ω_1, Ω_2).

Consider points that lie on a plane in space. The depth of each point is then given by:

$$Z = Z_0 + pX + qY \tag{46}$$

and the inverse-depth by:

$$\frac{1}{Z} = \frac{1}{Z_0}(1 - px/f - qy/f) \tag{47}$$

where p and q represent the orientation of the plane. Note that p and q are related to the scaled depth gradients by:

$$p = \frac{fZ_x}{Z_0} \tag{48}$$

$$q = \frac{fZ_y}{Z_0} \tag{49}$$

The image velocity gradients can now be expressed in terms of the surface orientation by:

$$u_x = \frac{U_3}{Z_0} + \frac{pU_1}{Z_0} \tag{50}$$

$$u_y = \Omega_3 + \frac{qU_1}{Z_0} \tag{51}$$

$$v_x = -\Omega_3 + \frac{pU_2}{Z_0} \tag{52}$$

$$v_y = \frac{U_3}{Z_0} + \frac{qU_2}{Z_0} \tag{53}$$

Appendix B

The velocity gradient tensor represents the distortion of the image shape. It can be decomposed into three components which are invariant under a transformation of the image co-ordinate system. [8] These components are the first-order differential invariants of the image velocity field – the vorticity (curl) , dilation (divergence) and pure shear (deformation) components.

$$\begin{aligned}
\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} &= \frac{\mathrm{curl}\mathbf{v}}{2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \frac{\mathrm{div}\mathbf{v}}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \\
&\quad \frac{\mathrm{def}\mathbf{v}}{2}\begin{bmatrix} \cos\mu & -\sin\mu \\ \sin\mu & \cos\mu \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{bmatrix} && (54) \\
&= \frac{\mathrm{curl}\mathbf{v}}{2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \frac{\mathrm{div}\mathbf{v}}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\mathrm{def}\mathbf{v}}{2}\begin{bmatrix} \cos 2\mu & \sin 2\mu \\ \sin 2\mu & -\cos 2\mu \end{bmatrix} && (55)
\end{aligned}$$

where $\mathrm{curl}\mathbf{v}$, $\mathrm{div}\mathbf{v}$ and $\mathrm{def}\mathbf{v}$ represent the curl, divergence and deformation components and where μ specifies the orientation of the axis of expansion (maximum extension). [9] $\mathrm{curl}\mathbf{v}$, $\mathrm{div}\mathbf{v}$ and $\mathrm{def}\mathbf{v}$ are defined in equations (3-6).

Appendix C

The relationships between the observed differential invariants, the three-dimensional configuration and the viewer motion are given below. We first express the differential invariants in terms of the viewer translation $(U_1/Z_0, U_2/Z_0, U_3/Z_0)$ and

[8] The decomposition is known in applied mechanics as the Cauchy–Stokes decomposition theorem [2].

[9] $(\cos\mu, \sin\mu)$ is the eigenvector of the traceless and symmetric component of the velocity gradient tensor. It corresponds the positive eigenvalue with magnitude $\mathrm{def}\mathbf{v}$. The other eigenvector specifies the axis of contraction and is orthogonal. It corresponds to the negative eigenvalue with magnitude $-\mathrm{def}\mathbf{v}$.

the surface orientation (p, q). From (3-6) and (50 -53) we have:

$$\begin{aligned}
\mathrm{div}\mathbf{v} &= \frac{2U_3}{Z_0} + \frac{(pU_1 + qU_2)}{Z_0} & (56)\\
\mathrm{curl}\mathbf{v} &= -2\Omega_3 + \frac{(-qU_1 + pU_2)}{Z_0} & (57)\\
(\mathrm{def}\mathbf{v})\cos 2\mu &= \frac{(pU_1 - qU_2)}{Z_0} & (58)\\
(\mathrm{def}\mathbf{v})\sin 2\mu &= \frac{(qU_1 + pU_2)}{Z_0}. & (59)
\end{aligned}$$

The differential invariants depend on the viewer motion, depth and surface orientation. We can express them in a co-ordinate free manner by introducing two 2D vector quantities: the component of translational velocity parallel to the image plane scaled by depth, Z_0, $\mathbf{A}$ where:

$$\begin{aligned}
\mathbf{A} &= \left(\frac{U_1}{Z_0}, \frac{U_2}{Z_0}\right)\\
&= \frac{\mathbf{U} - (\mathbf{U}.\mathbf{Q})\mathbf{Q}}{Z_0} & (60)
\end{aligned}$$

and the *depth gradient* scaled by depth, $\mathbf{F}$, to represent the surface orientation and which we define in terms of the 2D vector gradient[10]:

$$\begin{aligned}
\mathbf{F} &= (p, q) & (61)\\
&= \frac{f\nabla Z}{Z}. & (62)
\end{aligned}$$

With this new notation equations (56, 57, 58 and 59) can be re-written to show the relation between the differential invariants, the motion parameters and the surface position and orientation – equations (11,12 and 13).

The relationship between μ (which specifies the axis of maximum extension) and $\mathbf{A}$ and $\mathbf{F}$ (14):

$$\mu = \frac{\angle\mathbf{A} + \angle\mathbf{F}}{2}.$$

follows from the trigonometric relation:

$$\tan(\angle\mathbf{A} + \angle\mathbf{F}) = \frac{\tan(\angle\mathbf{A}) + \tan(\angle\mathbf{F})}{1 - \tan(\angle\mathbf{A})\tan(\angle\mathbf{F})} \tag{63}$$

[10]There are three simple ways to represent surface orientation: components of a unit vector, $\mathbf{n}$; gradient space representation (p, q) and the spherical co-ordinates (σ, τ). Changing from one representation to another is trivial and is listed here for completeness.

$$\begin{aligned}
\tan\sigma &= \sqrt{(p^2 + q^2)}\\
\tan\tau &= \frac{q}{p}\\
\mathbf{n} &= (\sin\sigma\cos\tau, \sin\sigma\sin\tau, \cos\sigma).
\end{aligned}$$

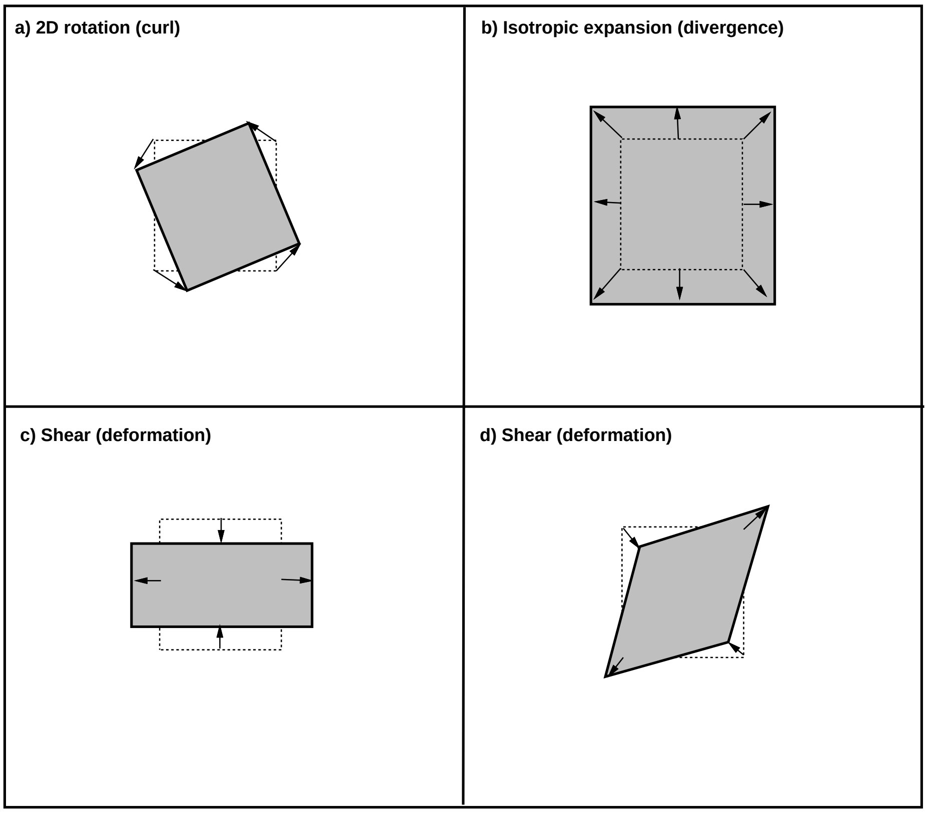

Figure 1: Differential invariants of the image velocity field.

To first order the image velocity field can be decomposed into curl (vorticity), divergence (dilation) and pure shear (deformation) components. The curl, divergence and the magnitude of the deformation are differential invariants and do not depend on the choice of image co-ordinate system. Their effect on apparent image shape can be described by four independent components of an affine transformation. These are: (a) a 2D rotation; (b) an isotropic expansion (scaling); (c) and (d) two deformation components. The latter two are both pure shears about different axes. Any deformation can be conveniently decomposed into these two components. Each component is dependent on an arbitrary choice of co-ordinate system and is not a differential invariant.

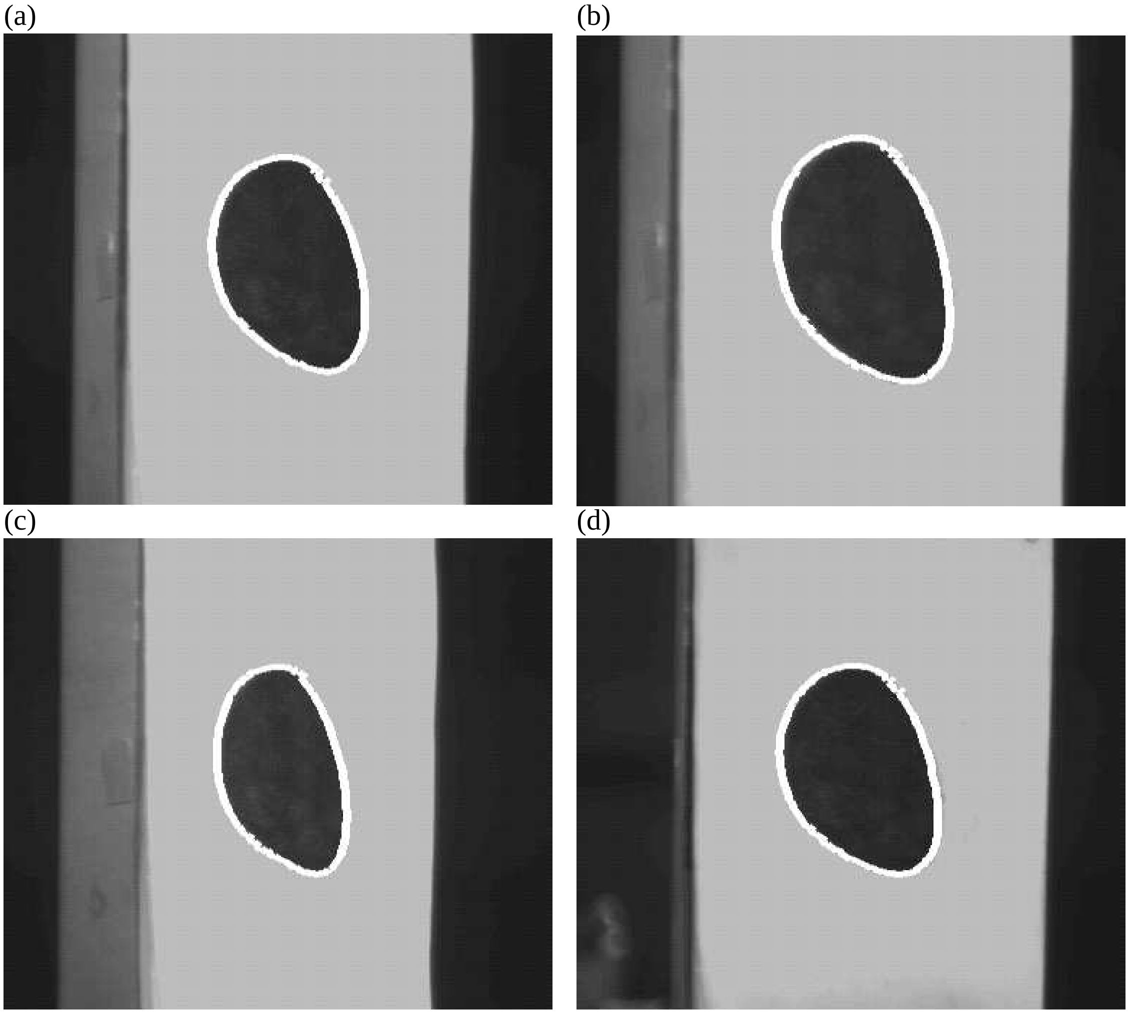

Figure 2: Distortions in apparent shape due to viewer motion.

(a) The image of a planar contour (zero tilt and positive slant, i.e. the direction of increasing depth, **F***, is horizontal and from left to right). The image contour is localised automatically by a B-spline snake initialised in the centre of the field of view. (b) The effect on apparent shape of a viewer translation towards the target. The shape undergoes an isotropic expansion (positive divergence). (c) The effect on apparent shape when the viewer translates to the left while fixating on the target (i.e.* **A** *is horizontal, right to left). The apparent shape undergoes an isotropic contraction (negative divergence which reduces the area) and a deformation in which the axis of expansion is vertical. These effects are predicted by equations (11,12,13 and 14) since the bisector of the direction of translation and the depth gradient is the vertical. (d) The opposite effect when the viewer translates to the right. The axes of contraction and expansion are reversed. The divergence is positive. Again the curl component vanishes.*

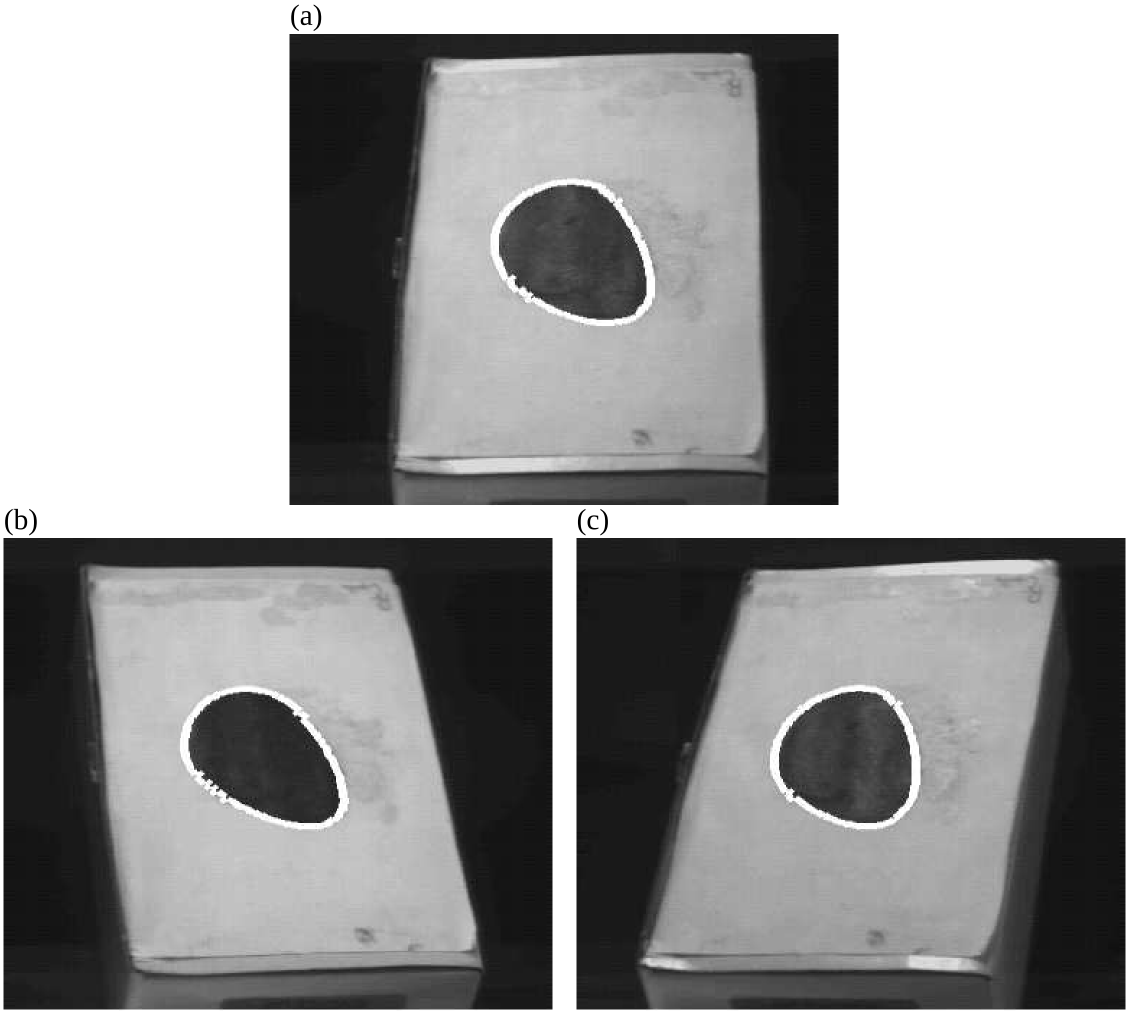

Figure 3: Image deformations and rotations due to viewer motion.

(a) The image of a planar contour (90^o tilt and positive slant – i.e. the direction of increasing depth, $\mathbf{F}$*, is vertical, bottom to top). (b) The effect on apparent shape of a viewer translation to the left. The contour undergoes a deformation with the axis of expansion at 135^o to the horizontal. The area of the contour is conserved (vanishing divergence). The net rotation is however non-zero. This is difficult to see from the contour alone. It is obvious, however, by inspection of the sides of the box, that there has been a net anticlockwise rotation. (c) These effects are reversed when the viewer translates to the right.*

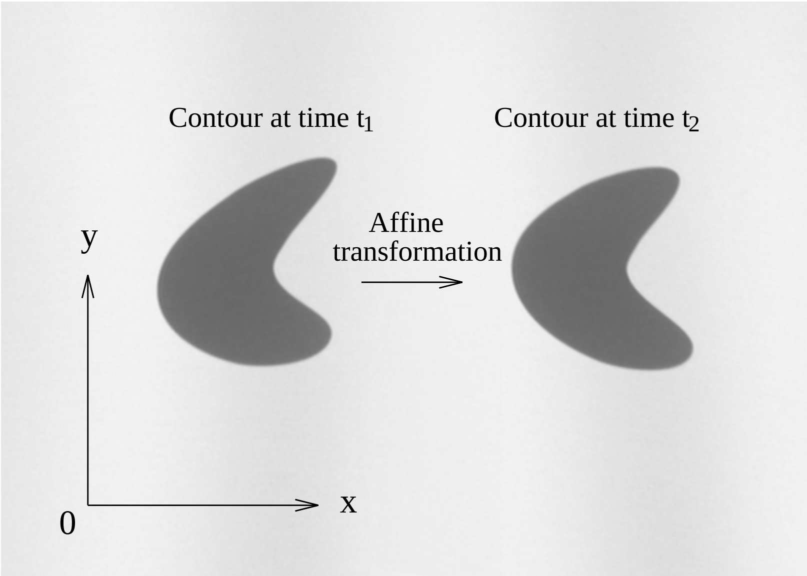

Figure 4: The temporal evolution of image contours.

For small fields of view the distortion in image shape can be described locally by an affine transformation. The components of the affine transformation can be expressed in terms of contour integrals of normal image velocities. More conveniently the temporal derivatives of the area and its moments can be used to characterise the distortion in apparent shape and the affine transformation.

Figure 5: Using image divergence to estimate time to contact.

Four samples of a video sequence taken from a moving observer approaching a stationary car at a uniform velocity (approximately 1m per time unit). A B-spline snake automatically tracks the area of the rear windscreen. The image divergence (figure 6) is used to estimate the time to contact (figure 7). The next image in the sequence corresponds to collision!

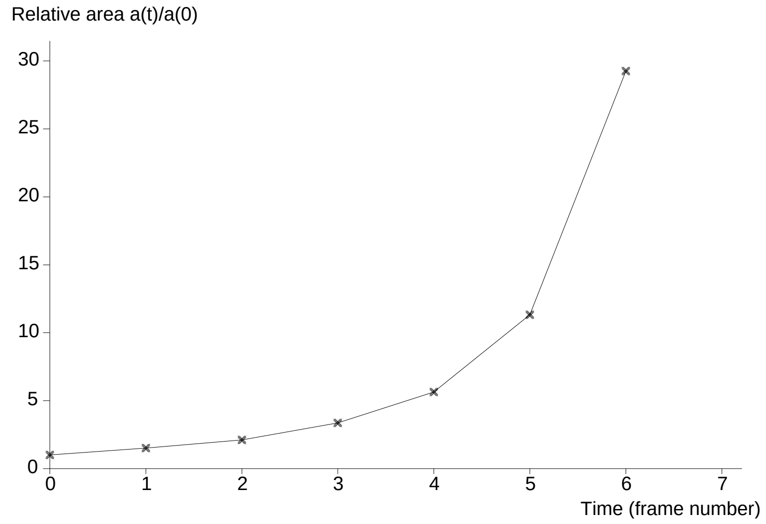

Figure 6: Apparent area of windscreen for approaching observer.

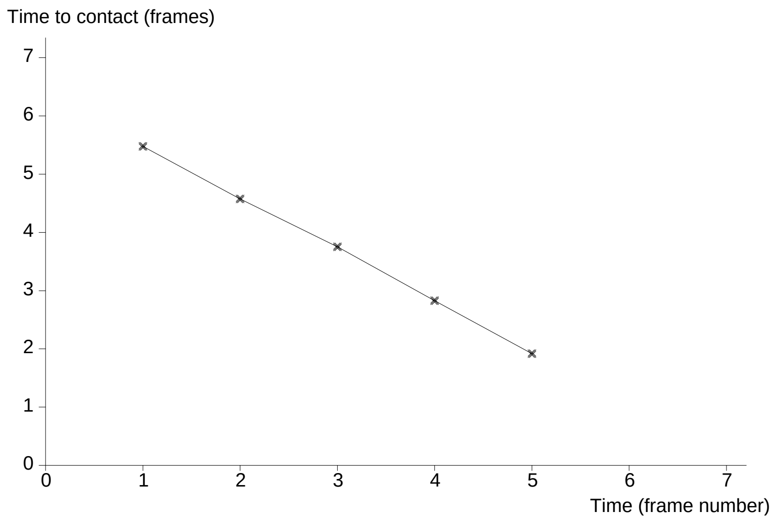

Figure 7: Estimated time to contact for approaching observer.

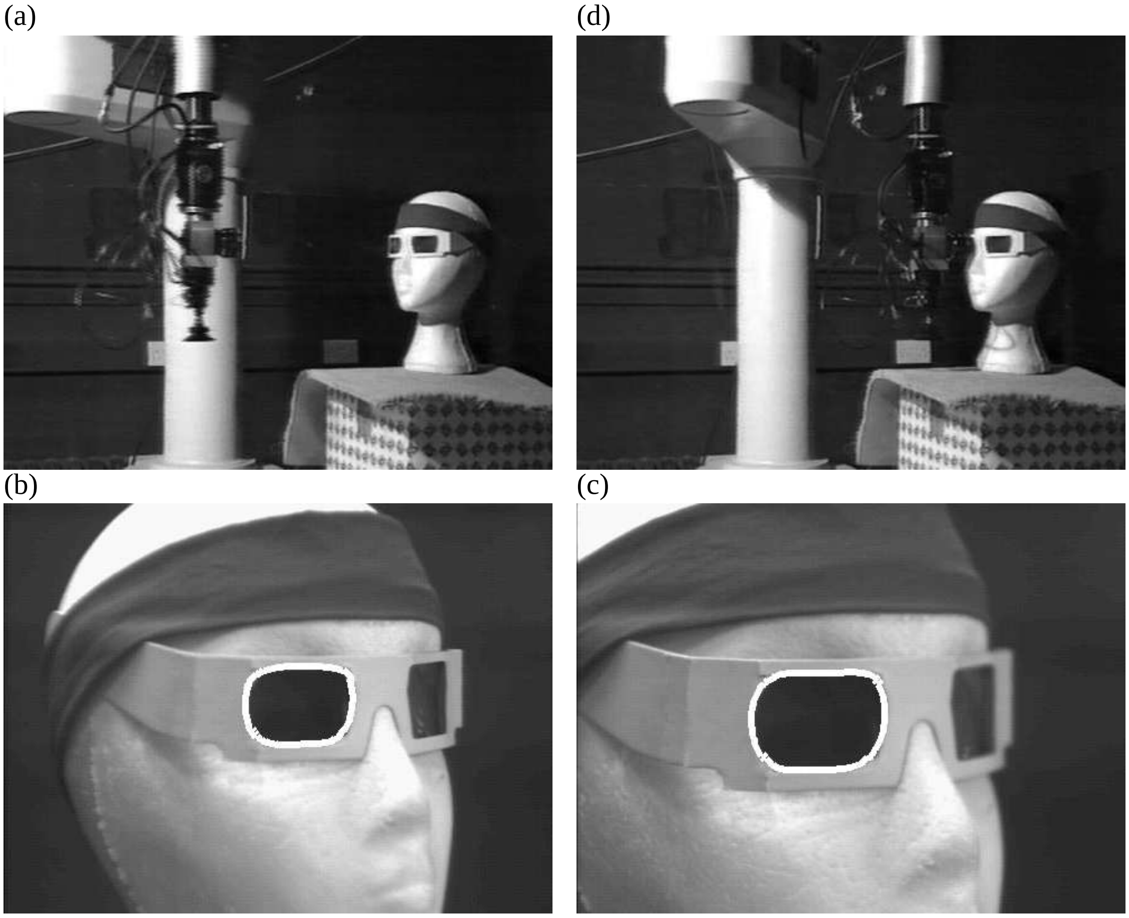

Figure 8: Using image divergence for collision avoidance.

A CCD camera mounted on a robot manipulator (a) fixates on the lens of a pair of glasses worn by a mannequin (b). The contour is localised by a B-spline snake which "expands" out from a point in the centre of the image and deforms to the shape of a high contrast, closed contour (the rim of the lens). The robot then executes a deliberate motion towards the target. The image undergoes an isotropic expansion (divergence)(c) which can be estimated by tracking the closed loop snake and monitoring the rate of change of the area of the image contour. This determines the time to contact *– a measure of the distance to the target in units of time. This is used to guide the manipulator safely to the target so that it stops before collision (d).*

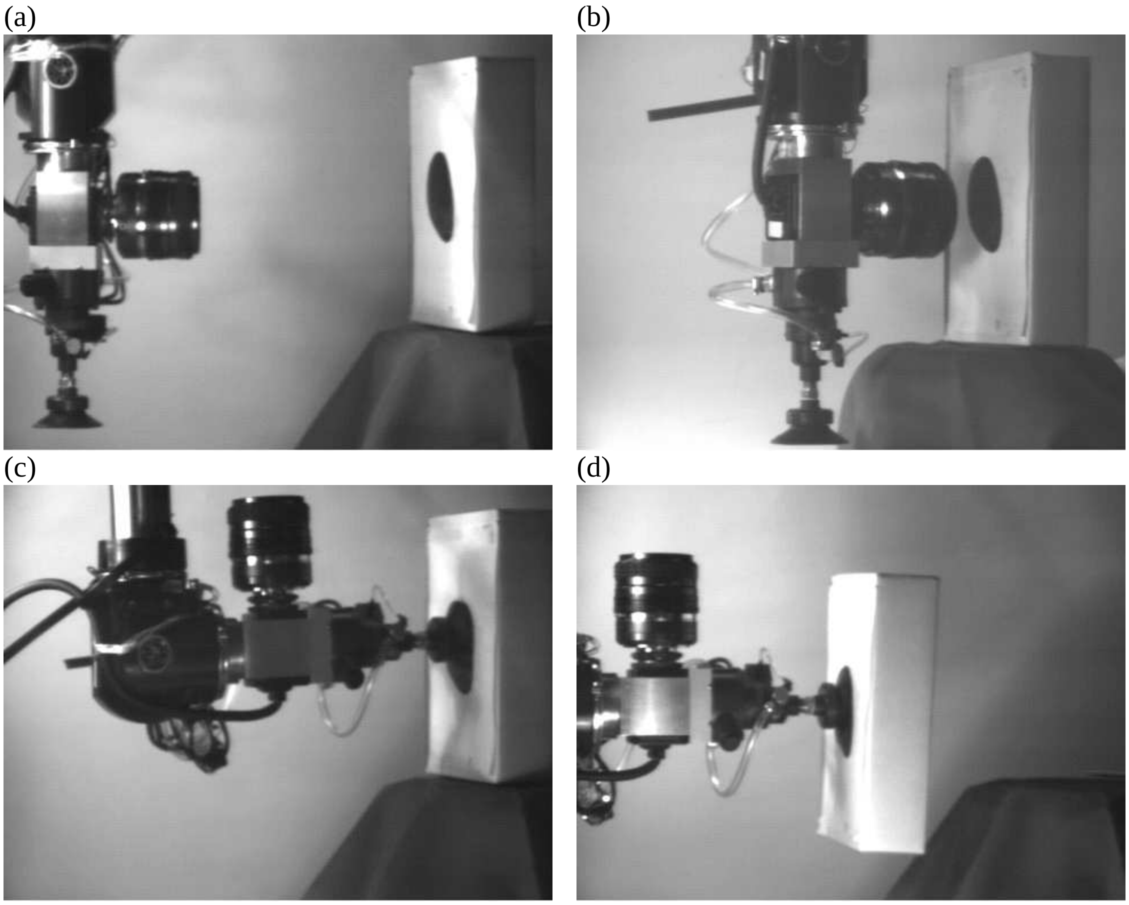

Figure 9: Visually guided landing and object manipulation.

Two examples in which a robot manipulator uses the estimates of time to contact and surface orientation to land on a target surface (approaching perpendicular to object surface) and to pick it up. The tracked image contours used to estimate image divergence and deformation are shown in figure 2. In (a) and (b) the estimate of the time to contact and surface orientation is used to guide the manipulator so that it comes to rest perpendicular to the surface with a pre-determined clearance. Estimates of divergence and deformation made approximately 1m away were sufficient to estimate the target object position and orientation to the nearest 2cm in position and 1^o in orientation.
In the second example, figures (c) and (d), this information is used to position a suction gripper in the vicinity of the surface. A contact sensor and small probing motions can then be used to refine the estimate of position and guide the suction gripper before manipulation. An accurate estimate of the surface orientation is essential. The successful execution is shown in (c) and (d).

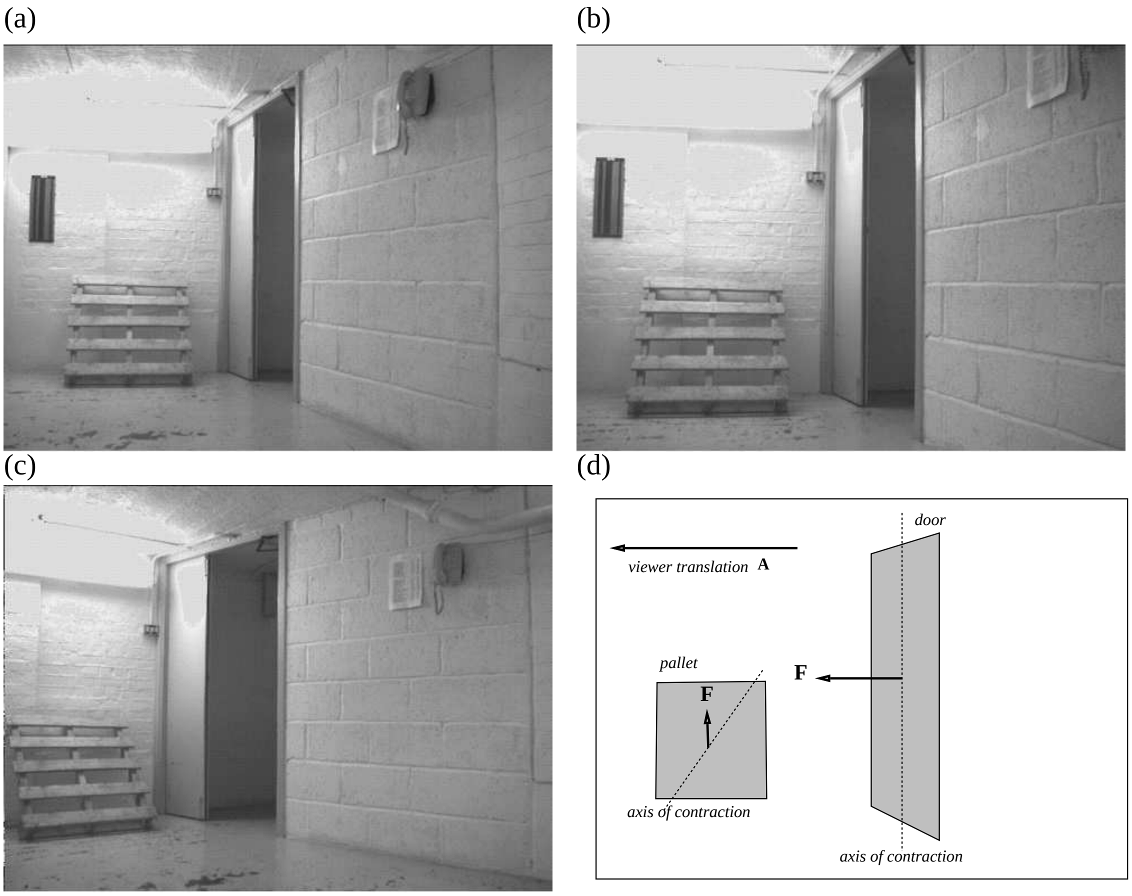

Figure 10: Qualitative visual navigation using image divergence and deformation.

(a) The image of a door and an object of interest, a pallet. (b) Movement towards the door and pallet produces a deformation in the image seen as an expansion in the apparent area of the door and pallet. This can be used to determine the distance to these objects, expressed as a time to contact – the time needed for the viewer to reach the object if it continued with the same speed. (c) A movement to the left produces combinations of image deformation, divergence and rotation. This is immediately evident from both the door (positive deformation and a shear with a horizontal axis of expansion) and the pallet (clockwise rotation and shear with diagonal axis of expansion). These effects, combined with the knowledge that the movement between the images, are consistent with the door having zero tilt, i.e. horizontal direction of increasing depth, while the pallet has a tilt of approximately 90^{o}*, i.e. vertical direction of increasing depth. They are sufficient to determine the orientation of the surface qualitatively (d). This has been done with no knowledge of the intrinsic properties of the camera (camera calibration), its orientations or the translational velocities. Estimation of divergence and deformation can also be recovered by comparison of apparent areas and the orientation of edge segments.*

References

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *Int. Journal of Computer Vision*, pages 283–310, 1989.

[2] R. Aris. *Vectors, Tensors, and the basic equations of fluid mechanics*. Prentice-Hall, Englewood, N.J., 1962.

[3] R.H. Bartels, J.C. Beatty, and B.A. Barsky. *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. Morgan Kaufmann, 1987.

[4] F. Bergholm. Motion from flow along contours: a note on robustness and ambiguous case. *Int. Journal of Computer Vision*, 3:395–415, 1989.

[5] J.D. Boissonat. Representing solids with the delaunay triangulation. In *Proc. ICPR*, pages 745–748, 1984.

[6] M. Campani and A. Verri. Computing optical flow from an overconstrained system of linear algebraic equations. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 22–26, 1990.

[7] R. Cipolla. *Active Visual Inference of Surface Shape*. PhD thesis, University of Oxford, 1991. (To appear in Springer-Verlag Lecture Notes in Computer Science).

[8] R. Cipolla and A. Blake. The dynamic analysis of apparent contours. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 616–623, Osaka, Japan, 1990.

[9] R. Cipolla and A. Blake. Surface orientation and time to contact from image divergence and deformation. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision*, pages 187–202. Springer–Verlag, 1992.

[10] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112, 1992.

[11] R. Cipolla and P. Kovesi. Determining object surface orientation and time to impact from image divergence and deformation. (University of Oxford (Memo)), 1991.

[12] R. Cipolla, Y. Okamoto, and Y. Kuno. Robust structure from motion using motion parallax. In *Proc. 4th Int. Conf. on Computer Vision*, pages 374–382, 1993.

[13] H.F. Davis and A.D. Snider. *Introduction to vector analysis*. Allyn and Bacon, 1979.

[14] E. Francois and P. Bouthemy. Derivation of qualitative information in motion analysis. *Image and Vision Computing*, 8(4):279–288, 1990.

[15] J.J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979.

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

[17] C. Harris. Geometry from visual motion. In A. Blake and A. Yuille, editors, *Active Vision*. MIT Press, 1992.

[18] E.C. Hildreth. *The measurement of visual motion*. The MIT press, Cambridge Massachusetts, 1984.

[19] F. Hoyle. *The Black Cloud*. Heinemann, London, 1957.

[20] K. Kanatani. Detecting the motion of a planar surface by line and surface integrals. *Computer Vision, Graphics and Image Processing*, 29:13–22, 1985.

[21] K. Kanatani. Structure and motion from optical flow under orthographic projection. *Computer Vision, Graphics and Image Processing*, 35:181–199, 1986.

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

[23] 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.

[24] J.J. Koenderink and A.J. Van Doorn. Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, 21:29–35, 1976.

[25] J.J. Koenderink and A.J. Van Doorn. How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In G. Hauske and E. Butenandt, editors, *Kybernetik*. Oldenburg, Munchen, 1978.

[26] J.J. Koenderink and A.J. Van Doorn. Depth and shape from differential perspective in the presence of bending deformations. *J. Opt. Soc. America*, 3(2):242–249, 1986.

[27] J.J. Koenderink and A.J. van Doorn. Affine structure from motion. *J. Opt. Soc. America*, 8(2):377–385, 1991.

[28] D.N. Lee. The optic flow field: the foundation of vision. *Phil. Trans. R. Soc. Lond.*, 290, 1980.

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

[30] S. J. Maybank. Apparent area of a rigid moving body. *Image and Vision Computing*, 5(2):111–113, 1987.

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

[32] R.C. Nelson and J. Aloimonos. Using flow field divergence for obstacle avoidance: towards qualitative vision. In *Proc. 2nd Int. Conf. on Computer Vision*, pages 188–196, 1988.

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

[34] J. Sato and R. Cipolla. Image registration using multi-scale texture moments. *Image and Vision Computing*, 13(5):341–353, 1995.

[35] M. Subbarao. Bounds on time–to–collision and rotational component from first-order derivatives of image flow. *Computer Vision, Graphics and Image Processing*, 50:329–341, 1990.

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

[37] A.M. Waxman and S. Ullman. Surface structure and three-dimensional motion from image flow kinematics. *Int. Journal of Robotics Research*, 4(3):72–94, 1985.

[38] A.M. Waxman and K. Wohn. Contour evolution, neighbourhood deformation and global image flow: planar surfaces in motion. *Int. Journal of Robotics Research*, 4(3):95–108, 1985.