FORTH-ICS / TR-156 December 1995

Non-metric depth representations: Preliminary results †

Manolis I. A. Lourakis

ABSTRACT: The ability of a system to construct reliable representations of its surroundings is crucial for the purposes of autonomous navigation. Based on the assumption that representations suitable for navigation can be of qualitative nature, this paper presents a method for computing an ordering of the points in a scene with regard to their depth. We show that this ordering does not require the structure from motion problem to be solved. The proposed method makes use of a pair of images that depict a scene from slightly different viewpoints, in such a way that the center of both images corresponds to the same world point (*fixated image pair*). Such a pair of images can be acquired from the two cameras of a stereo system and can be regarded as resulting from the displacement of one camera to the position of the other. This hypothetical motion combined with the approximation of the image formation process by the weak perspective model, leads to simple motion equations for the image points, which enable the ordering of a pair of points based on depth. The only information about the hypothetical motion that is extracted from the images is the normal flow field; no attempt is made to correspond points in the stereo pair, either in the form of optical flow or in the form of feature correspondence. To demonstrate a possible use of the resulting representation, a method is presented for qualitatively characterizing surface shape based on the sign of the Gaussian curvature.

† This report summarizes work that was carried out in partial fulfillment of the requirements for the degree of Master of Science in Computer Science at the University of Crete

Non-metric depth representations: Preliminary results

Manolis I. A. Lourakis †

Computer Vision and Robotics Laboratory
Institute of Computer Science (ICS)
Foundation for Research and Technology --- Hellas (FORTH)
Science and Technology Park, Heraklio, Crete
POBox 1385, GR-711-10 Greece
E-mail: lourakis@ics.forth.gr, Tel: +30 (81) 391705, Fax: +30 (81) 391601

Technical Report FORTH-ICS / TR-156 --- December 1995

©Copyright 1995 by FORTH

ABSTRACT: The ability of a system to construct reliable representations of its surroundings is crucial for the purposes of autonomous navigation. Based on the assumption that representations suitable for navigation can be of qualitative nature, this paper presents a method for computing an ordering of the points in a scene with regard to their depth. We show that this ordering does not require the structure from motion problem to be solved. The proposed method makes use of a pair of images that depict a scene from slightly different viewpoints, in such a way that the center of both images corresponds to the same world point (*fixated image pair*). Such a pair of images can be acquired from the two cameras of a stereo system and can be regarded as resulting from the displacement of one camera to the position of the other. This hypothetical motion combined with the approximation of the image formation process by the weak perspective model, leads to simple motion equations for the image points, which enable the ordering of a pair of points based on depth. The only information about the hypothetical motion that is extracted from the images is the normal flow field; no attempt is made to correspond points in the stereo pair, either in the form of optical flow or in the form of feature correspondence. To demonstrate a possible use of the resulting representation, a method is presented for qualitatively characterizing surface shape based on the sign of the Gaussian curvature.

---

† The author is a Ph.D. candidate at the Department of Computer Science, University of Crete

1 Introduction

One of the most important capabilities of an intelligent system is that of navigation, that is the capability of autonomous motion in the environment, based on measurements provided by various sensors. Of particular interest is the study of *visual* navigation mechanisms, which are based on the processing of image sequences acquired by optical sensors (e.g. cameras). Such sensors are very common in biological organisms and supply data from which a plethora of information can be extracted. Given that an intelligent system is often required to operate in dynamically changing and possibly unknown environments, the capability of space perception is a prerequisite for autonomous navigation. Undoubtedly, an important component of space perception is the perception of depth, the dimension that is lost during the projection of world parts on two dimensional surfaces.

Up to now, the vast majority of the studies related to visual navigation have been influenced by Marr's reconstructive theory [18], according to which the ultimate goal of vision is to construct detailed metric representations of the world, capable of providing input to any function that needs information extracted from visual data. Since, according to this theory, the perceptual functions are distinct and separated from those of action, the problem of navigation is reduced to the following two problems: Considering an observer that is moving relative to the environment, the first problem is to compute both a geometric representation of the environment and an estimate of the observer's 3D velocity. The second problem concerns the use of such a representation for computing a sequence of movements that will permit the observer to maneuver safely in space. The first problem is best known as the *structure from motion (SFM)* problem [26, 19], while the second is known as the *motion planning* problem [13, 15].

Computing the depth of a scene within the framework of the reconstructive theory, amounts to a simple application of methods dealing with the SFM problem. Such methods aim at computing metric (quantitative) depth, i.e. they aim at accurately estimating the distance of every point from the observer. The enormous and long-term effort towards solving the SFM problem, indicates that a general solution is very difficult to obtain. Although systems capable of solving the problem in restricted cases have appeared, their extension to general, real-world environments has not been successful. Besides, a detailed metric representation can be computationally expensive to exploit, while it can not easily deal with uncertainty. It is possible, however, that the understanding of some aspects of space does not necessarily require the computation of metric depth. Representations of depth that are less informative compared to a metric one, but are easier to compute, are often enough for the purposes of navigation. Results of psychophysical experiments, performed using human subjects, raise doubts as to the capability of the human visual system to compute metric depth [14, 21]. On the contrary, the human visual system is capable of computing an ordering of the points observed, based on depth [24]. Relying on such observations, this paper deals with the computation of *ordinal* and not metric depth. By the term ''ordinal'', we mean the ordering of points in a scene with respect to their depth, that is their distance from the observer measured along the optical axis. In order for such an ordering to be possible, the relation of the depths of the points in any pair has to be identified. More specifically, the point which is closer to the observer has to be found. Unquestionably, ordinal depth provides less information compared to metric

depth. We believe, however, that it can prove very useful in the perception of space [17][1]. The motivation for developing the method that will be described, is to show that useful information regarding the structure of a scene can be recovered without resorting to metric depth. Functions of depth - not necessarily known exactly - can be computed instead, without having to solve the SFM problem. This is in agreement with the recent work by Huang and Aloimonos [12], which circumvents the SFM problem and permits an active observer to compute the ratio of the depths for pairs of object points for which the intensity gradient is perpendicular to the vector that defines the parallel translational motion of the object with respect to the focal plane.

Our approach differs from that of classical reconstructive methods, in that it aims at computing qualitative information, suitable for space understanding. It is in agreement with the spirit of the theory of *purposive and qualitative active vision* [2, 3]. In brief, this theory dictates the computation of quantities that are well suited to the goals of a particular system, by exercising suitable control to the image acquisition process so that the recorded images will facilitate the computations that have to be carried out.

The remainder of the paper is organized as follows. Section 2 defines the normal flow field, which will serve as input to the proposed algorithms. Section 3 describes the coordinate system and the experimental setup used for image acquisition. The equations for ordinal depth estimation are developed in section 4. Based on these, section 5 presents some results from differential geometry and develops a method for qualitatively characterizing the local shape of surfaces. In section 6, results from the implementation of the proposed methods are presented and discussed. Conclusions are contained in section 7.

2 Normal flow

The motion of an observer relative to its environment causes intensity changes for the points in the image sequence that is recorded. These changes result in the apparent motion of some brightness patterns, which define a 2D vector field describing the observed displacements. This vector field is known as the *optical flow field*. According to the reconstructive theory, the first step of motion analysis is the estimation of a dense optical flow field. Unfortunately, the problem of estimating the optical flow field is ill-posed [1]. Regularization methods, which impose constraints on the solution, are often used to get around this problem. Such constraints are related to certain assumptions about the structure of the viewed scene. Typically, these assumptions hold only for selected environments and lead to restrictive methods for estimating optical flow. Another problem has to do with the nature of the optical flow field. Intensity changes in an image sequence are not exclusively due to the observer's motion relative to the environment [10]. Consequently, the optical flow field differs from the *motion field*, which is a purely geometrical concept, defined by the projection of the 3D velocity of every world point on the focal plane. Verri and Poggio have shown that the two fields are equal in very restricted cases [27]. Finally, the process of estimating the 3D velocity of the observer and the structure of a scene is very sensitive to the presence of

[1] Imagine for example a visually guided manipulator that must grasp an object. Grasping can be achieved by looking at (i.e. fixating) the object in question and then adjusting the position of the manipulator according to whether it is closer to the observer than the object to be grasped.

noise in the optical flow field [25, 22]. For the reasons stated above, we will avoid estimating the optical flow field and we will restrict ourselves to the estimation of the *normal flow field* which is defined below. The normal flow field has been used in the past for estimating 3D velocity [8].

Assuming that a sequence of images can be regarded as a continuous function $I(x, y, t)$ of two spatial (x, y) and one temporal (t) coordinates and that the intensity of corresponding points is conserved between successive frames, the well known *optical flow constraint equation* can easily be derived as [11]

$$I_x u + I_y v + I_t = 0, \tag{1}$$

where (u, v) is the optical flow vector and I_x, I_y, I_t are the partial derivatives of I with respect to x, y, t [2]. If eq. (1) is written in the form of a dot product,

$$(I_x, I_y) \cdot (u, v) = -I_t \tag{2}$$

it shows that it is possible to compute the projection of the optical flow along the intensity gradient vector, which is perpendicular to the edge at the corresponding point. Equation (2) is the mathematical equivalent of the so called *aperture problem*, which states that based on local information only, the estimation of a component of the optical flow field along a direction different from that of the gradient is impossible. The projection of the optical flow along the gradient direction, is known as *normal flow* and eq. (2) gives its algebraic value as $-I_t/\sqrt{I_x^2 + I_y^2}$.

In [27] it is shown that the difference between the algebraic values of the normal flow and the normal motion field (the projection of the motion field along the gradient) is equal to $\frac{1}{||\nabla I||}\frac{dI}{dt}$, where $||\nabla I||$ is the gradient magnitude and $\frac{dI}{dt}$ the total derivative of I with respect to time. Consequently, normal flow is a good approximation to the normal motion field at points with large gradient magnitudes.

3 Preliminaries

3.1 The coordinate system

Consider the $OXYZ$ coordinate system, which is fitted to the nodal point of a camera, so that the OZ axis coincides with the optical axis. Assume also that the camera is moving relative to a stationary scene with translational velocity $\vec{t} = (U, V, W)$ and rotational velocity $\vec{\omega} = (\alpha, \beta, \gamma)$, as indicated in figure 1.

This case is equivalent to the camera being stationary and the scene moving relative to it with translational velocity $-\vec{t}$ and rotational velocity $-\vec{\omega}$. The combined velocity of a point $P = (X, Y, Z)$ with regard to the $OXYZ$ coordinate system equals $-\vec{\omega} \times P - \vec{t}$, where $\times$ denotes the cross product. Thus, the velocity of P is given by

$$(dX/dt, dY/dt, dZ/dt) \quad = \quad (-U - \beta Z + \gamma Y, -V - \gamma X + \alpha Z, -W - \alpha Y + \beta X) \tag{3}$$

Equation (3) will be used in a following section to relate the motion field with the motion parameters $\vec{t}$ and $\vec{\omega}$.

[2] All quantities in eq. (1) are defined with respect to a particular image point (x, y).

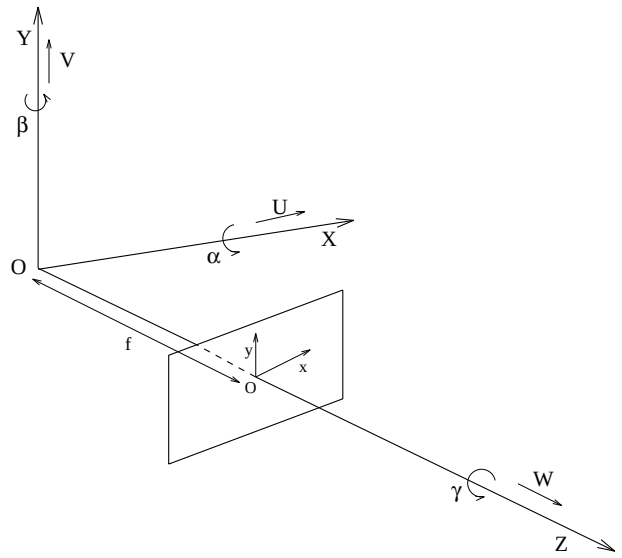


Figure 1: The coordinate system

3.2 The visual input

The method that will be presented makes use of a pair of images that depict a scene from slightly different viewpoints. The images are such that their centers correspond to the same point F of a scene. Image pairs of this kind are known as *fixated image pairs* and can be captured by a binocular system, whose cameras are pointed in a way such that their optical axes intersect at the *fixation point* F. The experimental setup is shown in figure 2, which depicts the relative positions of the camera coordinate systems. It is a stereo system, where the optical axes are converging instead of being parallel as it it commonly assumed for stereo systems. Such a configuration can easily be attained with a head-eye system, such as the one described in [20]. Fixation is very common among the biological organisms that possess stereoscopic vision.

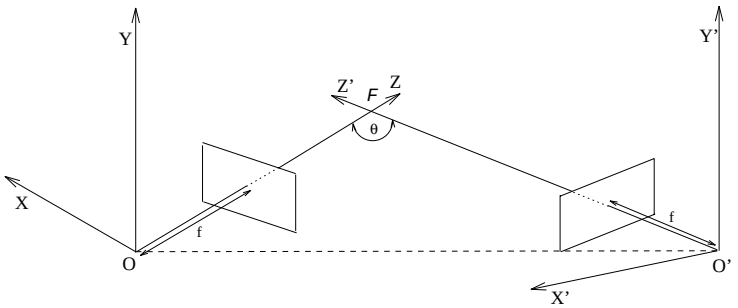


Figure 2: Binocular system geometry.

The information relevant to depth that can be extracted from a stereo image pair manifests itself in the form of *disparities* defined by the displacements of points between images. Typically, processing of a stereo pair aims at computing the disparities by corresponding image points and then extracting depth through triangulation. The hardest step in this procedure is to solve the correspondence problem. The proposed method follows a different approach, which does not

require this problem to be solved.

A pair of images that is captured by a stereo system can be thought of as an image sequence and processed using motion analysis techniques. Indeed, such a pair is identical to the one that would be captured by a moving camera which is at the position of the left (right) camera of the stereo system at time t and at the position of the right (left) camera at time $t+1$. This hypothetical camera is moving in such a way that its optical axis goes through the same world point at both positions and it is not rotating around the OX and OZ axes. This observation can reduce each problem of stereo analysis to a motion analysis problem and is valid independently of the relative position of the stereo system cameras. The estimation of ordinal depth will be based on the analysis of this hypothetical motion. It is worth noting that, during the hypothetical motion, the environment is stationary, since both images of a fixated pair are acquired at the same time.

The hypothetical motion is defined by less parameters compared to the case of a general, unrestricted motion and is described by simpler equations. The fact that the disparities within a small area, around the image center, are small enables the computation of the spatiotemporal derivatives of the intensity, which define the normal flow. In addition, approximating the image projection process by the model of *weak perspective projection* is also valid in the same area. According to that model, a point P with coordinates (X, Y, Z) with respect to the camera reference system is projected to point (x, y) on the image plane, defined by

$$x = sX \quad , \quad y = sY, \tag{4}$$

where s is a positive constant, usually equal to f/Z_0 and f, Z_0 are the focal length and the average depth of the observed scene respectively. Weak perspective projection leads to simple motion equations since eqs. (4) are linear in X, Y. A study of the error introduced by the weak perspective projection, has shown that the latter is a good approximation of the perspective projection for points lying near the optical axis, which are projected around the image center [17].

4 Ordinal depth estimation

4.1 The equations

We begin with the equations relating normal flow with the motion parameters. Let (X, Y, Z) be a world point which is projected to image point (x, y) by the weak perspective model. By definition, the motion field at (x, y) is given by

$$u = \frac{dx}{dt} = s\frac{dX}{dt} \quad , \quad v = \frac{dy}{dt} = s\frac{dY}{dt} \tag{5}$$

By substituting $\frac{dX}{dt}$ and $\frac{dY}{dt}$ from eq. (3), into (5) we obtain

$$u = s(-\beta Z + \gamma Y - U) \quad , \quad v = s(-\gamma X + \alpha Z - V) \tag{6}$$

We will follow the convention that the direction of the hypothetical motion of the camera is from the left camera of the stereo system to the right. This motion is described by less parameters

compared to the five appearing in eq. (6). Specifically, with reference to figure 3, the $OXYZ$ coordinate system can be moved to the position of $O'X'Y'Z'$ by a rotation about the OY axis, a translation along the OX and an other translation along OZ. These elementary motions are described by the parameters β, U, W respectively. The rotation around OY has a counter-clockwise direction, so $\beta > 0$. Both translations are negative, since they have opposite directions from the positive OX, OZ axes, so $U < 0, W < 0$. Parameters α, γ, V are all equal to zero. Given the above,

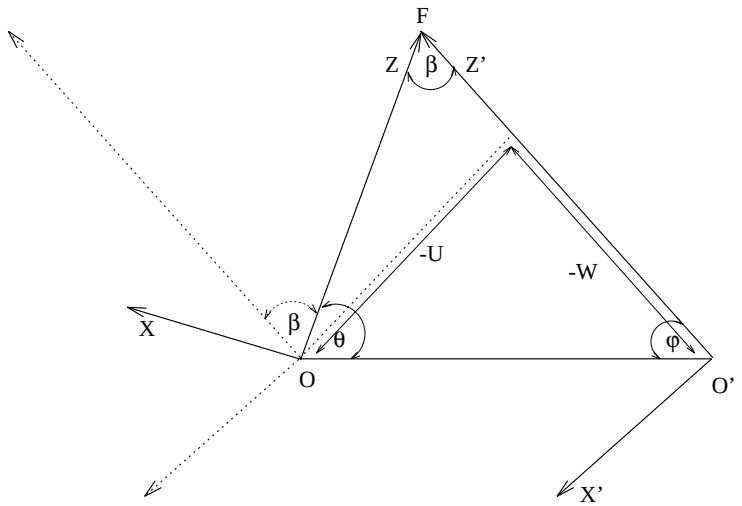


Figure 3: Top view of the coordinate systems for both cameras. The OY and $O'Y'$ axes are normal to the page, facing upwards.

equations (6) reduce to

$$u = s(-\beta Z - U) \quad , \quad v = 0 \tag{7}$$

The algebraic value of the normal flow field is given by

$$u^n = \quad (u, v)(\tfrac{I_x}{||\nabla I||}, \tfrac{I_y}{||\nabla I||}) = \tfrac{I_x}{||\nabla I||} s(-\beta Z - U) \tag{8}$$

Assuming that $I_x \neq 0$, eq. (8) can be written as

$$u^n \frac{||\nabla I||}{I_x} = \quad -s\beta Z - sU \tag{9}$$

Approximating the normal motion field with the normal flow field, we can substitute u^n with the algebraic value of the normal flow field, which is equal to $-\frac{I_t}{||\nabla I||}$. Consequently, equation (9) becomes

$$-\frac{I_t}{I_x} = \quad -s\beta Z - sU \tag{10}$$

At this point, it is observed that a function of depth has been obtained, which is of the form

$$f(Z) = AZ + B, \tag{11}$$

where $A \equiv -s\beta$ and $B \equiv -sU$ are unknown constants.

Let $P_1 and P_2$ be a pair of points having coordinates (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) and corresponding intensities I^1, I^2. Equation (10) gives for these points

$$-\frac{I_t^1}{I_x^1} = -s\beta Z_1 - sU \quad and \quad -\frac{I_t^2}{I_x^2} = -s\beta Z_2 - sU \tag{12}$$

Subtracting the above equations yields

$$B_{1,2} \equiv -\frac{I_t^1}{I_x^1} + \frac{I_t^2}{I_x^2} = \quad s\beta(Z_2 - Z_1) \tag{13}$$

Since $s, \beta > 0$, the sign of the difference $Z_2 - Z_1$ equals that of $B_{1,2}$. The fact that we employ just the sign of β and not its arithmetic value provides for some tolerance for errors during the image acquisition process, where the fixation point is not projected to the center of the images but is displaced by a few pixels in the horizontal dimension.

Equation (13) permits a comparison of the difference between the depths of two points with the maximum difference in the depths of points in a scene. Suppose that m is the point with minimum depth Z_m in the scene and M the point with the maximum depth Z_M. These points are identified as the first and the last elements of the sequence of depths sorted in ascending order. Let i and j be a pair of points with depths Z_i and Z_j respectively. Equation (13) gives

$$B_{M,m} \equiv -\frac{I_t^M}{I_x^M} + \frac{I_t^m}{I_x^m} = s\beta(Z_m - Z_M) \quad and \quad B_{i,j} \equiv -\frac{I_t^i}{I_x^i} + \frac{I_t^j}{I_x^j} = s\beta(Z_j - Z_i) \tag{14}$$

Dividing the two expressions in eq. (14) in parts, we obtain

$$\frac{Z_i - Z_j}{Z_M - Z_m} = \frac{B_{i,j}}{B_{M,m}} \tag{15}$$

Equation (15) yields the *relative difference of depths* for points i and j, since it defines the ratio of their depth difference with respect to a constant quantity.

4.2 Log-polar transform

The method described in the previous section is based on the computation of normal flow. Given that this computation assumes that the image sequence can be described by a continuous function of three variables, the motion between two successive frames must be small. This means that the disparities should be on the order of a few pixels. In the case of a fixated pair, this requirement is satisfied at points that are close to the center of the images. Moving towards the periphery, however, the disparities become larger, making the computation of the temporal partial derivative of the intensity impossible. To deal with this problem, the *log-polar transform* is employed.

This transform maps a point (x, y) to the point (ξ, η), where

$$\xi = ln(\sqrt{x^2 + y^2}) \quad , \quad \eta = arctan(\frac{y}{x}) \tag{16}$$

Since the natural logarithm of zero is not defined, the log-polar transform is not applied to the points of an image that lie within a small disk centered on the image center. The mapping defined

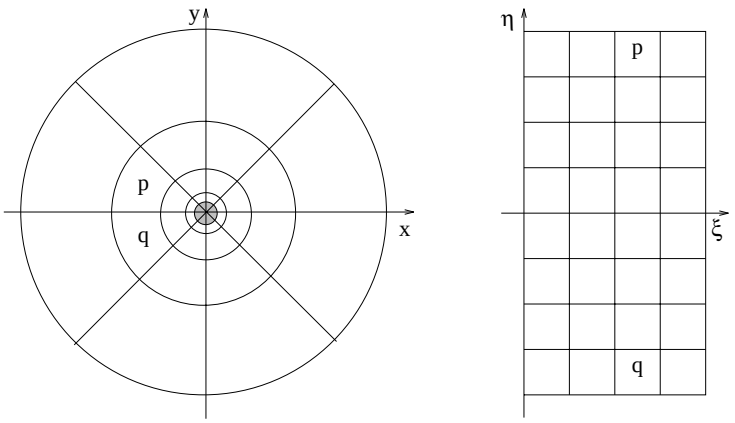


Figure 4: The log-polar transform.

by equations (16) can be understood with the aid of figure 4. The log-polar transform is a popular model for space variant sensing and CCD cameras implementing this mapping in hardware have recently been developed [6].

The number of points (x, y) which are mapped to the same point (ξ_0, η_0) in the log-polar domain depends on the distance of this point from the origin of the ξ, η axes. This is shown in fig. 4, where it is clear that the area of log-polar pixels increases with their distance from the center of the image. The basic reason for this is that the radii of the concentric circles in fig. 4 increase exponentially. The net result is that the disparities measured in the (ξ, η) domain are much smaller compared to these measured in the cartesian (x, y) domain, enabling the computation of the temporal partial derivative.

Based on the above, the estimation of normal flow makes use of the log-polar transform. In [17] it is shown that the partial derivatives of intensity with regard to x and y as a function of the partial derivatives with regard to ξ and η, are given by

$$\frac{\partial I}{\partial x} = \frac{\partial I}{\partial \xi}\frac{1}{\rho}cos(\eta) - \frac{\partial I}{\partial \eta}\frac{1}{\rho}sin(\eta) \quad and \quad \frac{\partial I}{\partial y} = \frac{\partial I}{\partial \xi}\frac{1}{\rho}sin(\eta) + \frac{\partial I}{\partial \eta}\frac{1}{\rho}cos(\eta) \tag{17}$$

Equations (17) facilitate the computation of the cartesian partial derivatives in the log-polar domain.

5 Local shape characterization

5.1 Definitions

The following subsections will describe a method for computing the sign of the Gaussian curvature of the surfaces in a scene, using the depth function that was computed in section 4. We begin by summarizing a few basic results that are needed in the development. For a more rigorous treatment, the reader is referred to a book on differential geometry, e.g. [7]. [23] can also serve as an introduction to mathematical descriptions of surface shape.

A *space curve* C is defined as a function $\vec{r}(s)$ from an interval of $\mathcal{R}$ into $\mathcal{R}^3$. The variable s is assumed to represent the arclength of C. Let $\vec{t} = \frac{d\vec{r}}{ds}$ be the tangent vector to C at some point. According to the well-known Frenet formulas, $\frac{d\vec{t}}{ds} = \kappa\vec{n}$, where $\vec{n}$ is the unit vector in the direction of $\vec{t}$ and κ is the *curvature* of C. The definition of the curvature of a surface patch is based on the notion of the curvature of space curves. Specifically, a surface patch S is defined as a function $\vec{r}(x, y)$ from an open set of $\mathcal{R}^2$ into $\mathcal{R}^3$. The shape of a patch is intrinsically characterized by a pair of quadratic forms, known as the *fundamental forms*, which are defined at every point of the patch. The partial derivatives $\vec{r}_x$ and $\vec{r}_y$ of S with respect to x and y at a point P, define the tangent plane $T_P(S)$ to S at P. The first fundamental form defines the length of a vector in the tangent plane $T_P(S)$. The cross-product of $\vec{r}_x$ and $\vec{r}_y$ defines a direction $\vec{\zeta}$, perpendicular to the tangent plane $T_P(S)$ and to S. Every direction $\vec{\xi}$ on the plane $T_P(S)$ defines a plane which includes $\vec{\xi}$ and $\vec{\zeta}$, and is perpendicular to S. The intersection of this plane with S defines a space curve. The second fundamental form defines the curvature of this curve at P. The second fundamental form has a unique maximum κ_1 and a unique minimum κ_2, along two normal directions. These directions are known as the *principal directions*, while κ_1 and κ_2 are known as *principal curvatures*. Knowledge of the principal curvatures completely describes the local shape of S, since the curvature of a normal intersection along any other direction can be computed as a linear combination of κ_1 and κ_2. The principal curvatures are invariants of S, that is they do not depend on the coordinate system that was used for defining them.

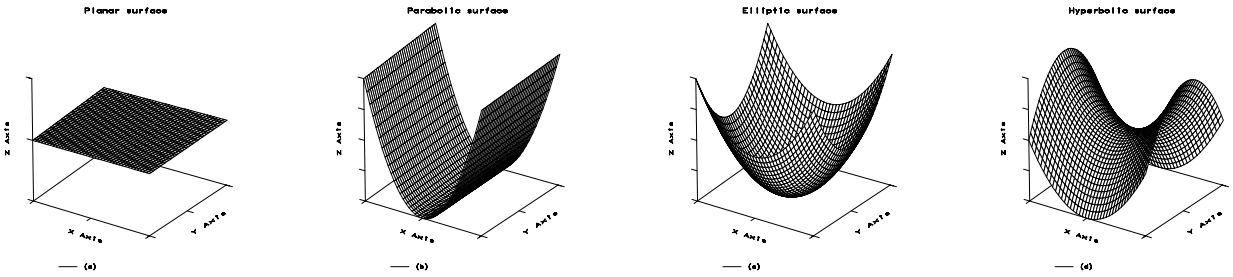


Figure 5: Surface patches with Gaussian curvatures of various signs
(a) $K = 0,\ \kappa_1^2 + \kappa_2^2 = 0$: planar
(b) $K = 0,\ \kappa_1^2 + \kappa_2^2 \neq 0$: parabolic
(c) $K > 0$: elliptic
(d) $K < 0$: hyperbolic

A qualitative characterization of shape is given by the product of the principal curvatures, which is known as the *Gaussian curvature* K. When K is positive, then the surface under consideration curves in the same manner along the two principal directions. In this case, the surface is called *elliptic* and its shape is locally convex or concave, depending on the point of observation. When K is negative, then the surface curves differently along the principal directions. The surface is then called *hyperbolic* and its shape resembles that of a saddle. Finally, when K is zero the surface is called *developable* and its shape is either planar (both principal curvatures are zero) or *parabolic* (exactly one principal curvature is zero). Figure 5 presents some examples of surfaces with different signs of Gaussian curvatures.

5.2 Computation of the Gaussian curvature sign

Let (x, y) be a point in an image and $Z(x, y)$ be its depth. $Z(x, y)$ defines the surface $r : \mathcal{R}^2 \rightarrow \mathcal{R}^3$ with $r(x, y) = (x, y, Z(x, y))$. According to [16], the Gaussian curvature of r is given by

$$K = \frac{DD'' - (D')^2}{EG - F^2}, \tag{18}$$

where

$$\begin{aligned} D &= -N \cdot \frac{\partial^2 r}{\partial x^2} \quad , \quad D' = -N \cdot \frac{\partial^2 r}{\partial x \partial y} \\ D'' &= -N \cdot \frac{\partial^2 r}{\partial y^2} \quad , \quad E = (\frac{\partial r}{\partial x})^2 \\ F &= \frac{\partial r}{\partial x} \cdot \frac{\partial r}{\partial y} \quad , \quad G = (\frac{\partial r}{\partial y})^2 \end{aligned} \tag{19}$$

and $N = \frac{\frac{\partial r}{\partial x} \times \frac{\partial r}{\partial y}}{\sqrt{EG - F^2}}$.

A little algebraic manipulation reveals that $EG - F^2 = |\frac{\partial r}{\partial x} \times \frac{\partial r}{\partial y}|^2 > 0$. This means that the sign of the Gaussian curvature depends only on the numerator in eq. (18). It can also be shown (see [17]) that the sign of the numerator equals that of L, where[3]

$$L = \frac{\partial^2 Z}{\partial x^2} \frac{\partial^2 Z}{\partial y^2} - (\frac{\partial^2 Z}{\partial x \partial y})^2 \tag{20}$$

The above equation cannot be used directly for computing the sign of the Gaussian curvature, since it requires the exact depth Z for each point to be known. However, it can easily be shown that the function of depth that is defined by (11) can be used in Z's place. Indeed,

$$G \equiv \frac{\partial^2 f(Z)}{\partial x^2} \frac{\partial^2 f(Z)}{\partial y^2} - (\frac{\partial^2 f(Z)}{\partial x \partial y})^2 = A^2 (\frac{\partial^2 Z}{\partial x^2} \frac{\partial^2 Z}{\partial y^2} - (\frac{\partial^2 Z}{\partial x \partial y})^2) = A^2 L \tag{21}$$

Equation (21) leads to the conclusion that the sign of the Gaussian curvature is equal to the sign of the quantity G, which can be computed from the function of depth that was recovered in section 4. Thus, we have shown that a qualitative characterization of surface shape is possible without having an accurate depth map. A map of the sign of the Gaussian curvature can be computed directly from the normal flow field. Similar results are reported in [28], with the major difference being that that work assumes the availability of a dense optical flow field.

6 Implementation and results

6.1 Computation of first order partial derivatives

The first order partial derivatives I_x, I_y that are needed in eq. (10) are computed through I_ξ, I_η and equations (17). The discrete log-polar transform is computed using bilinear interpolation in the

[3] In the following equations, the x and y arguments of the function $Z(.,.)$ have been omitted for brevity.

cartesian domain. The transform is not computed for points that lie within a circular disk with a five pixel radius, centered on the image center. The distance of every sample on the ξ axis from the image center is 1.05 times larger than that of the previous sample. The sample step for the η axis equals 0.5°. I_ξ and I_η are computed as a byproduct of Canny's edge detector [4]. The reason for not computing I_x and I_y directly in the cartesian domain is that an image in the log-polar domain is much smaller than the corresponding cartesian image, a fact that results in less computations. For example, the images that were used are 650×574, while their log-polar transforms are 93×720, that is 5.5 times smaller with regard to the number of pixels.

It can be shown that the temporal partial derivative is the same, independently of the domain (cartesian or log-polar) in which it is computed. I_t is computed as the average difference of pixels in the log-polar domain, using 3×3 neighborhoods.

6.2 Ordinal representation of depth

As shown in previous sections, in order to determine the relation of a point to other points with regard to depth, the following must hold:

- The point in question is within a fairly small circular disk, centered on the image center, so that the weak perspective projection is a good approximation to the perspective projection.
- The gradient magnitude is greater than some threshold, so that the normal motion field is well approximated by the normal flow field.
- I_x is nonzero, so that the algebraic value of the normal motion field is also nonzero.

For each pair of points satisfying the constraints stated above, equation (13) can be used to relate their depths. These points are sorted in ascending order based on their depths. Figures 6(a) and 6(b) show a fixated image pair, which was used as input in our experiments[4]. Figure 6(c) shows ordinal depth computed from 6(a) and 6(b) in the form of an image. Regions that are red denote points where the computation of ordinal depth is not possible. For the rest of the points, equation (15) for $j = m$ gives the relation of the difference of their depths from that of the point closest to the camera, with the largest depth difference in the scene. Painting the closest point white and the one furthest away black, the intermediate grey levels encode the depths for the remaining points. To facilitate the interpretation of figure 6(c), fig. 6(d) shows the circular disk for which ordinal depth is computed. The number of pixels for which reliable estimates of normal flow, and thus ordinal depth, can be computed, is about 20% of the total number of points contained within the disk.

6.3 Surface classification

Based on the results of section 5, local surface shape can be classified in three classes depending on the sign of the Gaussian curvature. The classification employs the depth function computed from a fixated pair in computing the sign of G in eq. (21). The second order derivatives that appear

[4]This pair has been captured by the vision head of the Royal Institute of Technology (KTH) in Sweden.

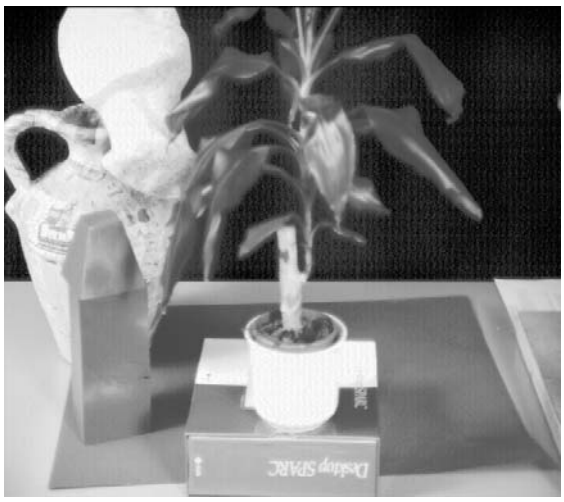
in (21) are computed with the aid of 3×3 masks [5]. To cope with the presence of noice, a fixed threshold of 0.05 is used for G and absolute values that are below it are assumed to be zero. This scheme can misclassify surfaces with low curvature as developable.

As already pointed out, the depth function cannot be computed at each point in the image. Since we do not attempt to interpolate for the depth function at points where it is unknown, the Gaussian curvature cannot be computed for every point where the depth function is known. Specifically, the Gaussian curvature can be computed only at points where the depth function is known for all of their 3×3 neighbors. The sign of the Gaussian curvature that is computed from images 6(a) and 6(b) is shown in fig. 6(e). Local surface shape is developable for red points, elliptic for green points and hyperbolic for blue points. For the rest of the points, the information available for computing the Gaussian curvature sign, is insufficient.

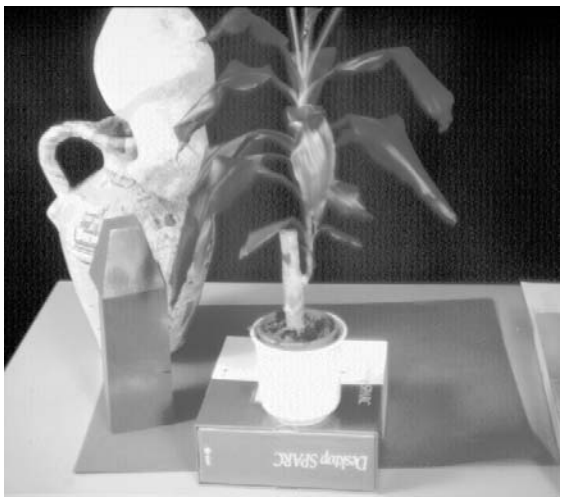
7 Conclusions

This paper presented a scheme for computing an ordering of the points in a scene according to their depths and an application of the resulting representation to characterizing the local shape of surfaces. The proposed method requires a partial solution to the correspondence problem, in the sense that it computes just one out of the two components of the optical flow (i.e. the normal flow). In addition, the method does not require a calibrated stereo system, but it is based on qualitative knowledge regarding its geometry. No special assumptions about the environment are made. The only requirement is that the environment has enough texture so that the normal flow can be computed, a condition that usually holds in practice. The proposed algorithms are computationally efficient and can greatly benefit from a parallel implementation, since most computations are data independent.

The ordinal depth computation that was presented is local, meaning that it is with regard to a small area around the fixation point. Therefore, an interesting problem would be to study methods for combining many such local representations that arise from different fixations, aiming at more global ordinal depth maps. Another interesting problem has to do with the identification of ordinal depth properties that are invariant to the position of the observer relative to the viewed scene. Such properties could be used for describing and later recognizing parts of the world [9]. A third interesting problem arises when the computation of ordinal depth is considered in the case of the more realistic perspective projection model. Some results towards this direction are already available and will be reported in a future publication. Finally, additional qualitative representations of depth need to be investigated, so that one can make a suitable choice among different representations, for specific navigation tasks.



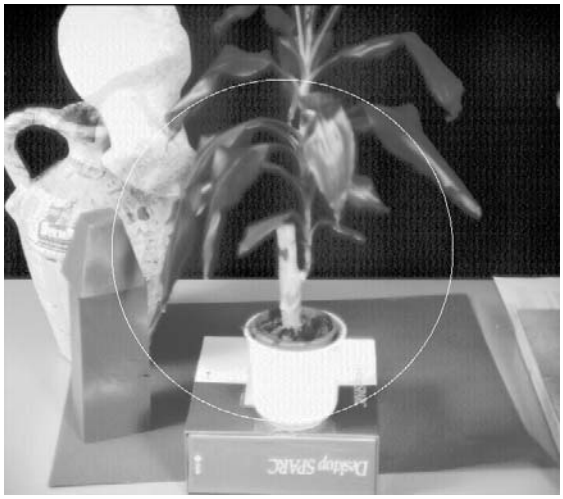
(a)



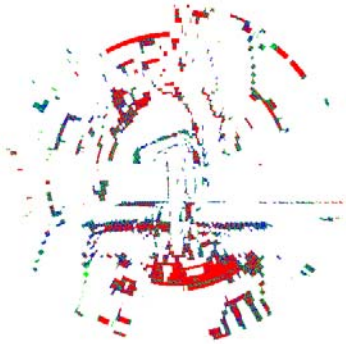
(b)



(c)



(d)



(e)

Figure 6: A fixated image pair (a, b), ordinal depth representation (c) and surface classification (e).

References

[1] J. Aloimonos, I. Weiss, and A. Bandopadhay. Active vision. *International Journal of Computer Vision*, 2:333--356, 1988.

[2] Y. Aloimonos. Purposive and qualitative active vision. In *DARPA Image Understanding Workshop*, pages 816--828, 1990.

[3] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):996--1005, Aug. 1988.

[4] J. Canny. A computational approach to edge detection. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-8:679--697, 1986.

[5] A. Cumani. Edge detection in multispectral images. *CVGIP: Graphical models and image processing*, 53(1):40--51, 1991.

[6] J. Van der Speigel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and G. Soncini. A foveated retinal-line sensor using CCD technology. In C. Mead and M. Ismail, editors, *Analog VLSI and Neural Network Implementations*, Boston, 1989. DeKluwer Pubs.

[7] M. P. do Carmo. *Differential geometry of curves and surfaces.* Prentice Hall, Englewood Cliffs, NJ, 1976.

[8] C. Fermuller. *Basic Visual Capabilities.* PhD dissertation, Center for Automation Research, University of Maryland, 1993.

[9] D. Forsyth, J. Mundy, A. Zisserman, C. Coelho, A. Heller, and C. Rothwell. Invariant descriptors for 3-D object recognition and pose. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-13(10):971--991, Oct. 1991.

[10] B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[11] B. K. P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185--203, 1981.

[12] L. Huang and Y. Aloimonos. How normal flow constrains relative depth for an active observer. *Image and Vision Computing*, 12(7):435--445, Sept. 1994.

[13] Y. K. Hwang and N. Ahuja. Gross motion planning -- a survey. *ACM Computing Surveys*, 24(3):221--291, Sept. 1992.

[14] E. Johnston. Systematic distortions of shape from stereopsis. *Vision Research*, 31:1351--1360, 1991.

[15] J. C. Latombe. *Robot Motion Planning.* Kluver Academic Publishers, Boston, MA, 1991.

[16] E. A. Lord and C. B. Wilson. *The mathematical description of shape and form.* Ellis Horwood Ltd., London, 1984.

[17] M. Lourakis. An alternative approach for studying space perception: Use of ordinal instead of metric depth. Master's thesis, University of Crete, Computer Science Department, Jan. 1995. (In Greek).

[18] D. Marr. *Vision.* W. H. Freeman, San Francisco, 1982.

[19] V. S. Nalwa. *A Guided tour of Computer Vision*, chapter 8. Addison-Wesley, 1993.

[20] K. Pahlavan and J. O. Eklundh. A head-eye system -- analysis and design. *CVGIP: Image Understanding, Special Issue on Purposive, Qualitative Active Vision*, 56:41--56, 1992.

[21] H. Pick, A. Yonas, D. Gentile, P. Melendez, D. Wagner, and D. Wegesin. Perceptual aspects of navigation. In *DARPA Image Understanding Workshop*, pages 481--483, 1993.

[22] M. E. Spetsakis and J. Aloimonos. Optimal motion estimation. In *IEEE Workshop on visual motion*, pages 229--237, 1989.

[23] K. A. Stevens. The visual interpretation of surface contours. *Artificial Intelligence*, 17:47--73, 1981.

[24] J. Todd and Reichel. Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychology Review*, 96:643--657, 1989.

[25] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-6:13--27, 1984.

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

[27] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-11:490--498, 1989.

[28] D. Weinshall. Direct computation of qualitative 3-D shape and motion invariants. *IEEE Trans. on Pat. Anal. and Mach. Int.*, PAMI-13(12):1236--1240, 1991.