Stereo

Uncalibrated Relief Reconstruction and Model Alignment from Binocular Disparities

Jonas Gårding[1], John Porrill[2], John P Frisby[2], John E W Mayhew[2]

[1] Computational Vision and Active Perception Laboratory (CVAP)
Dept. of Numerical Analysis and Computing Science
KTH (Royal Institute of Technology), S-100 44 Stockholm, Sweden
Email: Jonas.Garding@bion.kth.se

[2] AI Vision Research Unit, Sheffield University
P.O. Box 603, Psychology Building
Sheffield S10 2UR, United Kingdom
Email: J.P.Frisby@sheffield.ac.uk

Abstract. We propose a computational scheme for uncalibrated reconstruction of scene structure up to a relief transformation from binocular disparities. This scheme, which we call *regional disparity correction* (RDC), is motivated both by computational considerations and by psychophysical observations regarding human stereoscopic depth perception. We describe an implementation of RDC, and demonstrate its performance experimentally. As an example of applications of RDC, we show how it can be used to align a three-dimensional object model with an uncalibrated disparity field.

1 Introduction

The process of stereoscopic depth perception in human and machine vision comprises two major computational steps. First, the correspondence between the left and right images of points in three-dimensional space must be established, resulting in a disparity map which may be sparse or dense. Then, the disparity map must somehow be interpreted in terms of the depth structure of the scene. In this paper we consider the second of these two steps, i.e., the problem of disparity interpretation.

The classical approach to interpretation of stereo images is to perform a careful camera calibration, which is then used to reconstruct the three-dimensional structure of the scene by intersecting the left and right visual rays of each point. This technique was originally developed and refined by photogrammetrists, and it continues to be an important and powerful tool in circumstances that allow an accurate calibration to be performed; see e.g. [18].

However, in many situations a separate calibration stage is impractical, and there has recently been a great interest in approaches that require no calibration at all [6, 11], and there now exists a rich literature on the subject. Typically, these methods produce structure up to an arbitrary projective or affine transformation.

Another useful approach is possible in applications in which it is known that there exists a ground plane (or other surface) in the scene. By computing disparities relative to this surface, many computational simplifications are achieved. Early applications of the approach were intended for obstacle detection [4], but more recently this idea has been put in a more general framework [3, 17].

It has also been pointed out that disparities can be useful for segmentation or control of attention [10, 23] without attempting to reconstruct anything at all. These methods use raw disparity measurements to filter out regions for which the depth differs significantly from that of the region of interest, and for this purpose uncalibrated horizontal disparities are typically good enough.

In this paper we propose an alternative approach to disparity interpretation, which we call *regional disparity correction* (RDC). This approach originates from our work on modelling certain aspects of human stereoscopic depth perception, and it is a generalization of the model proposed in [9]. Like the raw disparity approach, RDC avoids explicit estimation of physical viewing parameters such as fixation distance. In contrast, however, RDC allows recovery of both general projective structure and of the metric depth ordering of the scene, and it is therefore potentially useful as a basis e.g. for object recognition and other tasks that can be aided by a characterization of the three-dimensional shape of objects in the scene. A more detailed account of the present work is given in [8].

2 Regional Disparity Correction (RDC)

In short, RDC comprises the following three steps:

1. Approximate the *vertical* component of disparity by a quadratic polynomial $\hat{v}(x, y)$.
2. Compute a "correction polynomial" $g(x, y)$ by a certain reshuffling of the coefficients of $\hat{v}(x, y)$.
3. Compute *affine nearness* $\rho(x, y) = h(x, y) + g(x, y)$.

It will be shown that $\rho(x, y)$ is itself a projective reconstruction of the scene, but more importantly, it allows the structure of the scene to be recovered up to a *relief transformation*, which preserves the depth ordering of the scene. Before describing the method in detail, we shall briefly discuss the psychophysical observations and computational considerations that motivate the approach.

2.1 Human Stereoscopic Depth Perception

Two aspects of human stereoscopic depth perception are particularly relevant for the RDC model; the role of vertical disparities, and the geometric nature of the depth percept.

The fact that the vertical component of disparity plays a significant role in human stereoscopic depth perception was first pointed out by Helmholtz [12].

He found that an array of vertical threads arranged in the fronto-parallel plane appears significantly curved, but that the perceived curvature can be eliminated by attaching beads to the threads. Presumably the beads allow the visual system to estimate vertical disparities, which allows the horizontal disparities of the threads to be interpreted correctly. Another well-known demonstration of the influence of vertical disparities is due to Ogle [21], who showed that unilateral vertical magnification (achieved by inserting a horizontal cylindrical lens in front of one eye) induces an apparent slant of the fronto-parallel plane. More recently, a number of researchers have investigated the effect of vertical disparity manipulations in a variety of circumstances, and there now exists a substantial body of empirical data; see e.g. [7] for a review.

Concerning the geometric aspects of stereoscopic depth perception, several studies (e.g. [14, 27]) have found that human performance in tasks involving estimation of metric structure from binocular disparities is remarkably poor, even in the presence of a richly structured disparity field. We interpret this as an indication that metric reconstruction is not the primary purpose of human stereopsis.

2.2 Computational Motivation

The RDC model takes an intermediate position between fully calibrated metric recovery and weakly calibrated recovery of projective structure.

We believe that it is beneficial to use as much knowledge about the viewing geometry as is available, without relying on unrealistic assumptions or computationally demanding calibration procedures. In this sense, the calibrated approach uses too much information, whereas the weakly calibrated approach uses too little. The RDC model is based on fixating binocular vision, typical of both human vision and anthropomorphic robot vision systems [23, 24]. In such systems the viewing geometry is constrained in a number of ways. The extrinsic geometry has essentially only three degrees of freedom if Donder's law is assumed, and it seems reasonable to assume that most of the intrinsic parameters are relatively stable or change only slowly over time (perhaps with the exception of the focal length). Nevertheless, it would be unrealistic to assume perfect fixation with zero cyclovergence etc., so we explicitly incorporate small-angle deviations from the idealized geometry in the RDC model.

Concerning the end result of the disparity interpretation process, there are good computational reasons for considering alternatives to metric reconstruction. It has been shown [22, 10] that the quality of metric reconstruction depends crucially on an accurate estimate of the fixation distance d (or equivalently the vergence angle between the two optical axes). RDC avoids this difficulty by not attempting to recover metric depth, and hence not having to estimate d. Instead, RDC recovers depth up to a relief transformation by computations performed only in the disparity domain, i.e., by adding small corrections to the horizontal disparities. This means that errors in the estimated disparity vectors are never magnified e.g. by division by small scale factors or by non-linear operations, in contrast to the case of recovery of metric structure.

3 Description and Analysis of RDC

The RDC method described here extends the model proposed in [9] in several ways, notably by allowing unknown fixation errors, unknown cyclovergence and unknown focal length.

A schematic representation of the binocular viewing geometry is shown in Figure 1. We represent visual space with respect to a virtual cyclopean eye, constructed such that the cyclopean visual axis (the Z axis) bisects the left and right visual axes. We explicitly model deviations from this idealized model by including (small) rotation angles (ω_x, ω_z), where ω_x is a relative rotation around the X axis and hence represents a vertical fixation error, and ω_z is a relative rotation around the Z axis, representing cyclovergence.

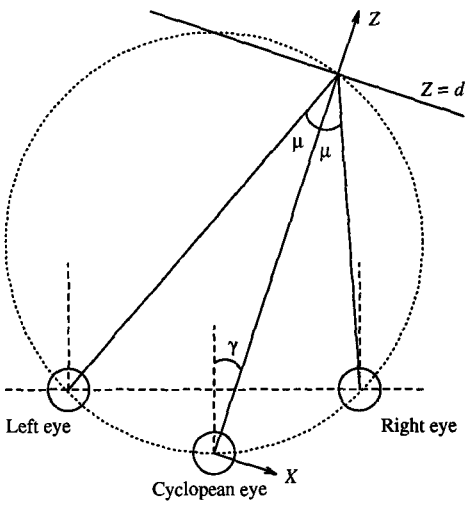

Fig. 1. Idealized representation of the viewing geometry. The plane of the drawing is referred to as the fixation plane. The Vieth-Müller circle (dotted) through the fixation point and the eyes indicates a part of the point horopter, i.e., the locus of points that yield zero disparity. μ is the vergence half-angle, and γ is the angle of asymmetric gaze.

We use a standard pinhole projection model, i.e., for each camera

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{X}, \tag{1}$$

where (x, y) are image coordinates, f is the focal length[3] and (X, Y, Z) are (left-handed) camera coordinates with the Z axis pointing toward the scene.

We use a flow approximation of disparity. This is equivalent to assuming that the angles ω_x and ω_z as well as μ (vergence) are small enough to be represented

[3] We assume the f is unknown but equal in the left and right images.

by a first-order approximation. As shown in [9], the error in this approximation is negligible for realistic viewing distances.

To simplify the notation, we define the inverse distance function $\lambda(x,y) = 1/Z(x,y)$. We relate the translation and rotation parameters $(t_x, t_y, t_z; \omega_x, \omega_y, \omega_z)$ to the viewing geometry by

$$\omega_y = t_x/d = t_x\lambda_0.$$
$$(t_x, t_y, t_z) = -I(\cos\gamma, 0, \sin\gamma).$$

This generalized definition of the fixation distance $d = 1/\lambda_0$ in terms of t_x and ω_y makes sense even if the optical axes do not cross. The only assumption made here is thus that $t_y = 0$, i.e., that there is no vertical displacement between the cameras. Applying these definitions to the standard image flow equations [19], we obtain the disparity as

$$\begin{aligned} h &= x_r - x_l \approx \dot{x} = \rho + (I\sin\gamma)\lambda\, x + (I\cos\gamma)\lambda_0\, x^2/f + \omega_z\, y + \omega_x\, xy/f, \\ v &= y_r - y_l \approx \dot{y} = (I\sin\gamma)\lambda\, y + (I\cos\gamma)\lambda_0\, xy/f - \omega_z\, x + \omega_x\, (f + y^2/f), \end{aligned}$$

where we have defined the *affine nearness*

$$\rho(x,y) = fI\cos\gamma\,(\lambda_0 - \lambda(x,y)). \tag{2}$$

For reasons which will be elaborated in Section 4, the purpose of RDC is to estimate $\rho(x,y)$. The key to achieving this aim is the observation that the vertical disparity depends very weakly on $\lambda(x,y)$, i.e., the depth structure of the scene. In fact, for symmetric vergence this small dependency vanishes completely, so that $v(x,y)$ encodes only the camera geometry. The idea is therefore to use $v(x,y)$ to "calibrate" $h(x,y)$ in order to estimate $\rho(x,y)$. We do this by approximating $v(x,y)$ in some region by the vertical disparity field corresponding to the average plane $\pi : Z = PX + QY + R$ in that region. Due to the weak dependency on depth, the error in this approximation can be expected to be significantly smaller than what would be the case if we had instead attempted to approximate $h(x,y)$. In this way RDC is different from methods based on subtracting the disparity of a reference surface [3, 17].

In terms of the image coordinates, we have

$$\lambda_\pi(x,y) = \frac{1}{Z(x,y)} = \frac{f - Px - Qy}{fR}. \tag{3}$$

The corresponding disparity field can be expressed as

$$h_\pi(x,y) = fI\cos\gamma\,(\lambda_0 - \lambda_\pi(x,y)) - g_\pi(x,y), \tag{4}$$
$$v_\pi(x,y) = A + Bx + Cy + Exy + Fy^2, \tag{5}$$

where

$$g_\pi = -Cx + By - Ex^2 - Fxy,$$

and (A, B, C, E, F) are constants depending on the viewing geometry and the parameters (P, Q, R) of the plane. They are estimated directly from $v(x, y)$ by linear least squares. Affine nearness can now be estimated by computing

$$\hat{\rho}(x, y) = h(x, y) + g_\pi(x, y).$$

By substituting (4), we can express this estimate as

$$\hat{\rho}(x, y) = \rho(x, y) + \epsilon(x, y), \tag{6}$$

where

$$\epsilon(x, y) = I \sin\gamma \, (\lambda(x, y) - \lambda_\pi(x, y))x \tag{7}$$

is the estimation error. This error is the product of three factors which are generally small: the deviation from symmetric vergence, the deviation from the average plane, and the horizontal image eccentricity. Consequently, $\hat{\rho}(x, y)$ is typically a good estimate of $\rho(x, y)$.

4 Affine Nearness and Relief Transformations

The affine nearness $\rho(x, y)$ contains important and useful information about the structure of the world. From (2) and (1) we can solve for the relation between $(x, y, \rho(x, y))$ and 3D-points (X, Y, Z). In homogeneous coordinates, we obtain the invertible relation

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \cong \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/L & f\lambda_0 \end{pmatrix} \begin{pmatrix} x \\ y \\ \rho \\ 1 \end{pmatrix}, \tag{8}$$

where $L = I \cos\gamma$, and the symbol $\cong$ denotes equality up to an arbitrary scale factor. This relation demonstrates that the surface in $\mathbb{P}^3$ with homogeneous coordinates $(x, y, \rho(x, y), 1)$ is in fact a projective reconstruction of the scene. Hence, if for example $\rho(x, y)$ is flat, then so is the corresponding physical surface in the scene.

However, $\rho(x, y)$ is distinctively different from the physical surface in other ways, and there is no choice of parameters (λ_0, L, f) for which they become identical. It is therefore more useful to consider the equivalence class of surfaces $(X(x, y), Y(x, y), Z(x, y))$ that can be reconstructed from $\rho(x, y)$ by "guessing" the parameters (λ_0, L, f). From (8) we see that two such surfaces, corresponding to (λ_0, L, f) and (λ_0', L', f') respectively, are related by the transformation

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} \cong \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & b & a \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \tag{9}$$

where

$$a = L/L' > 0, \qquad b = \frac{f'\lambda_0' L' - f\lambda_0 L}{fL'}, \qquad c = f/f' > 0.$$

In the following, the matrix in (9) will be denoted by $T_{a,b,c}$. It is easily verified that $T_{a,b,c}$ defines a *transformation group*, which in turn defines an *equivalence class* of 3-D shapes compatible with a given affine nearness $\rho(x, y)$.

We shall refer to $T_{a,b,c}$ as a generalized[4] *relief transformation.* Transformations of this type have a long history in vision, and have been considered e.g. in [12, 15, 16]. Perhaps the most important property of a relief transformation is that it preserves the *depth ordering* of the scene.[5] Hence, knowing the structure of the scene up to $T_{a,b,c}$ entails knowing what is in front of what, in addition to the general projective properties such as coplanarity and collinearity.

5 Experimental Evaluation

In this section we shall give some quantitative examples of the performance of RDC, using the following procedure:

1. Choose the binocular camera geometry $(f, d, I, \gamma, \omega_x, \omega_z)$
2. Generate a random cloud of points in space around the fixation point
3. For each 3-D point, generate its left and right projection using the pinhole camera model (i.e., *not* the flow approximation)
4. Perform RDC, i.e., fit a quadratic function to the vertical disparity field, and use this function to compute the estimate $\hat{\rho}(x, y)$ of affine nearness
5. Reconstruct the 3-D points from $\hat{\rho}(x, y)$ by resolving the relief ambiguity, i.e., by applying (8) using the true values of (d, L, f)
6. Compute the difference between the reconstructed points and the true 3-D points.

The reason for performing the metric 3-D reconstruction is to obtain a quantitative and geometrically intuitive measure of the quality of the estimate $\hat{\rho}(x, y)$.

As shown in Table 1, in the noise-free case the reconstruction errors from RDC, which originate from the flow approximation of disparity and the error term $\epsilon(x, y)$, are on the order of a millimeter or less with a viewing geometry representative of human vision. It is perhaps more interesting to consider what happens under the more realistic assumption that the disparity vectors are noisy. The columns labeled $\sigma = 1$ show the corresponding data obtained after adding Gaussian noise to each disparity vector. It is evident that RDC handles this noise level relatively gracefully.

[4] The definition of a relief transformation given in [9] corresponds to $T_{a,b,1}$, i.e., the case of known focal length.

[5] Strictly speaking, the depth order is preserved on either side of the singularity $a + bZ = 0$, but not across it. However, any physically valid relief reconstruction (X', Y', Z') must satisfy $Z' > 0$ for all points, which implies $a + bZ > 0$.

Viewing geometry	n	*RDC error*		*Raw disparity error*	
		$\sigma = 0$	$\sigma = 1$	$\sigma = 0$	$\sigma = 1$
Symmetric vergence;	5	0.037	2.681	2.437	2.412
$\gamma = \omega_x = \omega_z = 0$	10	0.041	1.002	2.416	2.726
	100	0.043	0.929	2.812	2.848
Asymmetric vergence	5	0.385	1.682	5.096	5.240
with cyclovergence;	10	0.400	1.249	9.033	9.796
$\gamma = 25°$, $\omega_z = 5°$	100	0.464	1.257	10.879	10.875

Table 1. Average distances (in cm) between reconstructed and true 3-D points, for two different viewing geometries and varying number n of points. The fixation distance was 50cm, the interocular baseline was 6cm, and the 3-D points were randomly distributed in a box 40cm wide and 20cm deep centered around the fixation point. The columns labelled *RDC error* indicate the result when depth was reconstructed from estimated affine nearness $\hat{\rho}(x, y)$. As a comparison, the columns labelled *Raw disparity error* show the result of neglecting the vertical disparities and using $\hat{\rho}(x, y) \approx h(x, y)$. σ indicates the standard deviation (in pixels) of Gaussian noise, which was added independently to the horizontal and vertical components of each disparity vector (in a 512×512 image).

6 Model Alignment

In applications such as object recognition, the need often arises to align an object model with image or 3-D data using hypothesized correspondences between model and data features [13, 28]. When only a monocular image is available, the alignment is by necessity done from 3-D model features to 2-D image features. A binocular image pair, however, opens the possibility of performing the alignment in 3-D [20], if the binocular system can be calibrated with high enough accuracy.

In this section we propose an intermediate approach which combines the strengths of both 2-D and 3-D methods, namely model alignment based on affine nearness. This approach provides more information and hence better disambiguation than monocular data, while avoiding the need for accurate camera calibration.

We assume that the model consists of a list of 3-D points, and that there exists a hypothesized correspondence between each model point (X_i, Y_i, Z_i) and a point on the affine nearness surface $(\hat{x}, \hat{y}, \hat{\rho})$ which has been estimated by RDC. In order to align the object we need to estimate (i) the pose $(R, \mathbf{t})$ of the object relative to the cyclopean coordinate system, and (ii) the parameters (d, L, f) needed for resolving the relief ambiguity (Section 4).

Simple equation counting tells us that three point correspondences may suffice, but we shall generally use more points and perform a least squares estima-

tion. More precisely, we wish to minimize the goal function

$$\delta^2(R,\mathbf{t},d,L,f)) = \sum \left(\frac{1}{\sigma_x^2}(\hat{x}_i - x_i)^2 + \frac{1}{\sigma_y^2}(\hat{y}_i - y_i)^2 + \frac{1}{\sigma_\rho^2}(\hat{\rho}_i - \rho_i)^2 \right), \tag{10}$$

where $(\hat{x}_i, \hat{y}_i, \hat{\rho}_i)$ are the estimated data features, $(\sigma_x, \sigma_y, \sigma_\rho)$ their associated covariances, and

$$x_i = f\frac{X_i}{Z_i + d}, \qquad y_i = f\frac{Y_i}{Z_i + d}, \qquad \rho_i = \frac{fL}{d}\frac{Z_i}{Z_i + d}.$$

(X_i, Y_i, Z_i) are the coordinates of a model point rotated to be parallel to the cyclopean system and translated such that the model point $(0,0,0)$ coincides with the fixation point. These coordinates are related to the given model points by a rotation R and a translation $\mathbf{t}$.

To simplify this minimization problem, we use an iterative procedure in which we in each iteration first minimize δ^2 with respect to the viewing parameters (d, L, f) only, and then use these estimated parameters to reconstruct the 3D points and to perform a 3D-to-3D pose estimation. To obtain an initial pose estimate, we use a simple monocular method based on a scaled orthography approximation [5].

The computations performed in each iteration are relatively simple. For the pose estimation the optimal estimate can be computed by singular value decomposition of a 3×3 matrix [2]. For the viewing parameter estimation it is easily shown (by differentiating δ^2 with respect to L and f) that for a given estimate of d, the optimal estimates of f and L are given by

$$f = \frac{\frac{1}{\sigma_x^2}\sum \frac{\hat{x}_i X_i}{Z_i + d} + \frac{1}{\sigma_y^2}\sum \frac{\hat{y}_i Y_i}{Z_i + d}}{\frac{1}{\sigma_x^2}\sum \frac{X_i^2}{(Z_i + d)^2} + \frac{1}{\sigma_y^2}\sum \frac{Y_i^2}{(Z_i + d)^2}}, \tag{11}$$

and

$$L = \frac{d}{f}\frac{\sum \frac{\hat{\rho}_i Z_i}{Z_i + d}}{\sum \frac{Z_i^2}{(Z_i + d)^2}}. \tag{12}$$

Hence, the problem of estimating the viewing parameters for a given pose estimate can be reduced to a 1-D minimization problem in d, which can be solved numerically with a very limited computational effort.

We have found that the scheme typically converges in 2–5 iterations, after which the estimates change very slowly. Numerical values obtained with the data set corresponding to the second viewing geometry and $n = 10$, $\sigma = 1$ in Table 1 are shown in Table 2.

Since this alignment method produces explicit estimates of the parameters (d, L, f) needed to resolve the relief ambiguity, it is of course possible to apply

Iteration	d	L	f	δ^2
Initial	100	10.59	2.21	$3.8 \cdot 10^{-3}$
1	55.9	6.02	1.19	$3.0 \cdot 10^{-4}$
2	51.8	5.65	1.06	$1.2 \cdot 10^{-4}$
3	49.4	5.45	1.00	$8.4 \cdot 10^{-5}$
5	47.5	5.36	0.95	$7.0 \cdot 10^{-5}$
10	47.1	5.43	0.94	$6.4 \cdot 10^{-5}$
50	47.2	5.48	0.93	$6.2 \cdot 10^{-5}$
True	49.8	5.44	1.00	

Table 2. Estimation of viewing parameters by first applying RDC to a noisy disparity field and then aligning five model points. The cyclovergence is 5°, and the remaining true parameter values are shown in the bottom row. The other rows show the estimates after each iteration.

the resulting transformation to the entire disparity field. Figure 2 shows a stereo image pair of a calibration cube, in which the intersections of the white rulings were matched to produce a sparse disparity field. A model containing the seven visible points of the top corner of the cube was used in the procedure described above, and the estimated parameters were then used to reconstruct the entire cube. The result is shown in Figure 3.

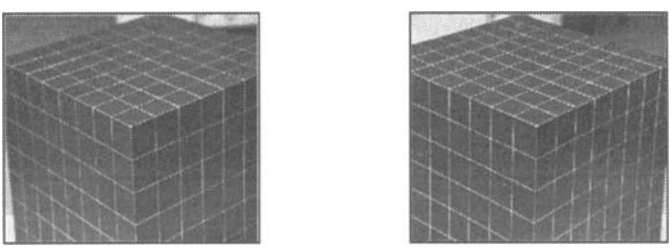

Fig. 2. Stereo pair showing a calibration cube (arranged for cross-eyed fusion).

7 Future Work

Future work includes integration of RDC with the stereo matching stage; the matching and the construction of the parametric representation $\hat{v}(x, y)$ of the vertical disparity field should be performed in parallel, so that early estimates of $\hat{v}(x, y)$ can be used to constrain the epipolar geometry which reduces the matching ambiguity and hence allows more matches to be found, which in turn can be used to improve the estimate $\hat{v}(x, y)$, and so on.

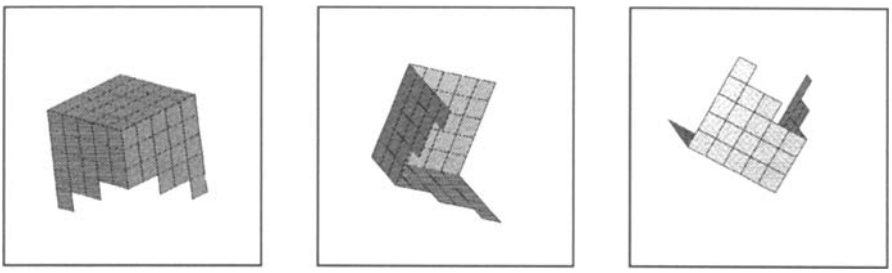

Fig. 3. Three different views of a cube reconstructed from affine nearness by aligning seven model points around the central corner of the cube.

Moreover, in this paper we have not explored the "R" in RDC, i.e., we have always used the vertical disparity field in the entire image as a reference. There are, however, indications [26, 25, 1] that the human visual system may use more localized representations, which would also make sense from a computational point of view.

References

1. W. Adams, J.P. Frisby, D. Buckley, H. Grant, J. Gårding, S.D. Hippisley-Cox, and J. Porrill, "Pooling of vertical disparities by the human visual system", *Perception*, 1996. (To appear).
2. K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-squares fitting of two 3-D point sets", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 9, pp. 698–700, 1987.
3. J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation", in *Proc. 2nd European Conf. on Computer Vision*, vol. 508 of *Lecture Notes in Computer Science*, pp. 237–252, Springer Verlag, Berlin, 1992.
4. S. Carlsson and J.-O. Eklundh, "Object detection using model based prediction and motion parallax", in *Proc. 1st European Conf. on Computer Vision* (O. Faugeras, ed.), vol. 427 of *Lecture Notes in Computer Science*, pp. 297–306, Springer Verlag, Berlin, Apr. 1990. (Antibes, France).
5. D.F. DeMenthon and L.S. Davis, "Model-based object pose in 25 lines of code", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 335–343, Springer-Verlag, May 1992.
6. O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, pp. 563–578, Springer-Verlag, May 1992.
7. J.P. Frisby, H. Grant, D. Buckley, J. Gårding, J.M. Horsman, S.D. Hippisley-Cox, and J. Porrill, "The effects of scaling vertical disparities on the perceived amplitudes of three-dimensional ridges", May 1995. (Submitted for publication).
8. J. Gårding, J. Porrill, J.P. Frisby, and J.E.W. Mayhew, "Uncalibrated relief reconstruction and model alignment from binocular disparities", Tech. Rep. ISRN

KTH/NA/P--96/02--SE, Dept. of Numerical Analysis and Computing Science, KTH (Royal Institute of Technology), Jan. 1996.

9. J. Gårding, J. Porrill, J.E.W. Mayhew, and J.P. Frisby, "Stereopsis, vertical disparity and relief transformations", *Vision Research*, vol. 35, pp. 703–722, Mar. 1995.
10. W.E.L. Grimson, A.L. Ratan, P.A. O'Donnell, and G. Klanderman, "An active visual attention system to play "Where's Waldo?"", in *Proc. Workshop on Visual Behaviors*, (Seattle, Washington), pp. 85–90, IEEE Computer Society Press, June 1994.
11. R. Hartley, R. Gupta, and T. Chang, "Stereo from uncalibrated cameras", in *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, (Champaign, Illinois), pp. 761–764, June 1992.
12. H.L.F. von Helmholtz, *Treatise on Physiological Optics*, vol. 3. (trans. J.P.C Southall, Dover, New York 1962), 1910.
13. D.P. Huttenlocher and S. Ullman, "Recognizing solid objects by alignment with an image", *Int. J. of Computer Vision*, vol. 5, pp. 195–212, 1990.
14. E.B. Johnston, "Systematic distortions of shape from stereopsis", *Vision Research*, vol. 31, pp. 1351–1360, 1991.
15. J.J. Koenderink and A.J. van Doorn, "Geometry of binocular vision and a model for stereopsis", *Biological Cybernetics*, vol. 21, pp. 29–35, 1976.
16. J.J. Koenderink and A.J. van Doorn, "Affine structure from motion", *J. of the Optical Society of America A*, vol. 8, pp. 377–385, 1991.
17. R. Kumar, P. Anandan, and K. Hanna, "Shape recovery from multiple views: a parallax based approach", in *Proc. Image Understanding Workshop*, (Monterey, CA), 1994.
18. M. Li and D. Betsis, "Head-eye calibration", in *Proc. 5th International Conference on Computer Vision*, (Cambridge, MA), pp. 40–45, June 1995.
19. H.C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image", *Proc. Royal Society London B*, vol. 208, pp. 385–397, 1980.
20. J.E.W. Mayhew and J.P. Frisby, eds., *3D Model Recognition from Stereoscopic Cues*. MIT Press, 1991.
21. K.N. Ogle, *Researches in Binocular Vision*. Saunders, Philadelphia, 1950.
22. T.J. Olson, "Stereopsis for verging systems", in *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, (New York), pp. 55–66, 1993.
23. T.J. Olson and D.J. Coombs, "Real-time vergence control for binocular robots", *Int. J. of Computer Vision*, vol. 7, no. 1, pp. 67–89, 1991.
24. K. Pahlavan, T. Uhlin, and J.-O. Eklundh, "Dynamic fixation", in *Proc. 4th Int. Conf. on Computer Vision*, (Berlin, Germany), pp. 412–419, May 1993.
25. B.J. Rogers and J.J. Koenderink, "Monocular aniseikona: a motion parallax analogue of the disparity-induced effect", *Nature*, vol. 322, pp. 62–63, 1986.
26. S.P. Stenton, J.P. Frisby, and J.E.W. Mayhew, "Vertical disparity pooling and the induced effect", *Nature*, vol. 309, pp. 622–623, June 1984.
27. J.S. Tittle, J.T. Todd, V.J. Perotti, and J.F. Norman, "Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis", *J. of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 3, pp. 663–678, 1995.
28. S. Ullman and R. Basri, "Recognition by linear combinations of models", *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 13, pp. 992–1006, 1991.