Pairwise comparison technique: a simple solution for depth reconstruction

Leonid L. Kontsevich

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115

Received August 7, 1992; revised manuscript received January 4, 1993; accepted January 19, 1993

A new technique dramatically simplifies the analysis of matching and depth reconstruction by extracting three-dimensional rigid depth interpretation from pairwise comparisons of weak perspective projections. This method provides a simple linear criterion for testing the correctness of correspondence for a pair of images; the method also provides a description of a one-parameter family of interpretations for each pair of images that satisfies this criterion. We show that if at least three projections of a volumetric object are known, then a three-dimensional (3D) rigid interpretation can be inferred from pairwise comparisons between any one of these images and other images in the set. The 3D interpretation is derived from the intersection of corresponding one-parameter families. The method provides a common computational basis for different processes of depth perception, for example, depth-from-stereo and depth-from-motion. Thus, a single mechanism for these processes in the human visual system would be sufficient. The proposed method does not require information about relative positions of eye(s) or camera(s) for different projections, but this information can be easily incorporated. The method can be applied for pairwise comparison within a single image. If any nontrivial correspondence is found, then several views of the same object are present in the same image. This happens, for example, in views of volumetrically symmetric objects. Symmetry facilitates depth reconstruction; if an object possesses two or more symmetries, its depth can be reconstructed from a single image.

1. INTRODUCTION

The derivation of a three-dimensional (3D) structure of a scene from several projections is accomplished by various mechanisms in the visual system, the most important of which are based on motion and stereo parallax. For this reason the depth-from-motion problem has been the focus of a computational approach[1] to visual perception. Several approaches to this problem have been outlined, and some special cases of it have been fully described. Differences among the computational approaches arise from the necessity of restricting possible interpretations for a given sequence of images. These restrictions can be applied to motion itself as well as to the objects in the scene. Many cases have been analyzed in which some parameters of motion or shape are known (e.g., translatory motion of the observer,[2] fixed axis of rotation,[3-6] rotation with translation perpendicular to the rotation axis,[7] and arbitrary motion of a plane surface[8-10]). The problem with such particular solutions is that in any given situation it is necessary to decide which particular solution is applicable. A preferable approach is to use general constraints that will be applicable to the largest number of possible situations. One such constraint is smoothness of the motion of massive objects. This constraint is generally applied to correspondence problems and estimation of motion parameters.[11-13] A more general constraint is based on the intrinsic property of shape invariance with respect to motion. This rigidity constraint can provide a unique interpretation of 3D shape[14-18] without the need for additional information about motion parameters and object shape. Nevertheless, in the domain of nonrigid motion the human visual system works surprisingly well. For computation in that domain a more general, maximal rigidity constraint has been proposed.[19]

Ullman[14] has described the minimal conditions for which depth reconstruction with a rigidity constraint is possible. For central projection there should be at least two projections with at least five corresponding points. For the case of orthographic projection at least three projections with at least four points are required. The same requirements hold for weak perspective projections.[17] These computational minimality conditions have been tested psychophysically. Interestingly, the human visual system creates a rigid 3D interpretation even in those cases when, theoretically, insufficient information is available. In particular, depth perception is possible from a sequence of two projections[20] and under finite point lifetime conditions.[21,22]

The latter observation prompted several research groups to study the recovery of structure from two orthographic projections. Bennett *et al.*[23] investigated this question in detail and provided a complete account of possible solutions and simple closed-form solutions for the depth. A similar problem for the case of weak perspective projections was considered by Lee and Huang.[24] They showed that, given points on one image, the sets of all possible corresponding points on the other image constitute parallel lines. This fact was used to solve the correspondence task between the two images. Both of these studies nevertheless overcomplicate the problem. To clarify the logic underlying these studies, I reproduce them in simpler terms in the sections that follow.

In this paper I propose a new method for depth reconstruction that is based on pairwise comparison of weak perspective projections. Our algorithm for this method has important merits over that for other methods. First, unlike all other algorithms it is truly linear. For this reason it is simple and fast. It is also much more stable than algorithms based on nonlinear equations. Second, it is ro-

0740-3232/93/061129-07$06.00

bust because it integrates information over any number of points in the image and over any number of images. Most other methods impose stringent limitations on the number of points and projections used in analysis. Third, the proposed method is highly parallel, because pairwise comparisons can be performed independently. And, finally, it is general because it can integrate information from different sources, such as motion parallax, stereo parallax, and others to be described.

2. SHAPE-CONSTANCY CONSTRAINT FOR ORTHOGRAPHIC VIEWS

A rigidity constraint is appropriate when one infers 3D structure from perspective views. It is, however, too strong when applied to orthographic projection. An orthographic image provides a good approximation to the perspective case when the size of the object is significantly less than the distance between object and observer and this distance is approximately constant in the course of observation. However, in reality the distance can vary on a large scale. To consider this variable in terms of orthographic projection, one can use the concept of weak perspective projection. The weak perspective projection is the orthographic projection that can subsequently be contracted or expanded with an arbitrary scale factor. This scaling simulates the motion of an object toward or away from an observer.

In this paper I use an equivalent method to achieve the same goal through the introduction of a shape-constancy constraint. This constraint allows us to work with the standard orthographic projection and not with the two-stage weak perspective one. Assume that scaling is part of the motion and that objects are not rigid in the conventional sense. Note that by this definition, an object undergoing motion can change size (but not form). In this case displacement in depth does not affect the object's orthographic projection. However, displacement's effect on perspective projection can be simulated by the object's change of size.

It must be emphasized that from this point onward the term motion refers not only to objects' rotations and translations in space but also to their dilations that leave their shapes invariant. In other words, motion assumes angle invariance, as opposed to the standard rigidity constraint. The images that we consider are orthographic projections of objects moving in this manner.

3. LINEAR CRITERION FOR THE EXISTENCE OF AN INTERPRETATION WITH CONSTANT SHAPE

Consider a projection plane that is stable in space (i.e., a camera-centered world coordinate system) and a moving object that consists of a limited number of isolated points $\{a_i\}$. Suppose that at two instants two orthographic images of this object are obtained and denoted P and P'. Suppose also that the correspondence of points is known such that point a_i corresponds to points b_i and b_i' in the images.

An object's motion can be represented as a composition of translation, rotation, and scaling. We can immediately exclude translation by considering the transformation of the edge vectors that connect the points of the object rather than the points themselves. We denote the vector with the origin at point a_i and the end at point a_j by $\mathbf{r}_{i,j}$ at the instant of the first projection and by $\mathbf{r}'_{i,j}$ of the second projection. The correspondence of edges is defined by the correspondence of points. Denote by $\mathbf{p}_{i,j}$ and $\mathbf{p}'_{i,j}$ the orthographic projections of the edges $\mathbf{r}_{i,j}$ and $\mathbf{r}'_{i,j}$ on images P and P'. Rotation and scaling of an object, i.e., the linear components of motion, can be considered as rotation and scaling of edges in vector space.

Consider pure rotation in space. If some plane is fixed (in our case, the image plane), then an arbitrary rotation can be represented as the superposition of rotation around an axis V, lying in this plane and rotation around the axis Z orthogonal to the fixed plane (Fig. 1). Denote V' the image of the V axis after completion of the rotation.

This decomposition, like Euler's decomposition, can be defined by three independent parameters: orientation of axis V, rotation around V, and the rotation that transfers V to V'.

From the proposed decomposition, it immediately follows that the orthogonal projection of the edge $\mathbf{r}_{i,j}$ onto axis V and its projection onto axis V' after the complete rotation are equal. Consequently, the norms of the projections of vector $\mathbf{p}_{i,j}$ onto the V axis and vector $\mathbf{p}'_{i,j}$ onto the V' axis are also equal (Fig. 2).

According to the motion-decomposition method proposed at the beginning of this section, an object changes its scale after a rotation. Let s denote the scale factor. Then $s = |\mathbf{r}'_{i,j}|/\mathbf{r}_{i,j}|$. Suppose that axes V and V' are determined by nonzero vectors $\mathbf{v}$ and $\mathbf{v}'$, adjusted so that

$$|\mathbf{v}'| = |\mathbf{v}|/s. \tag{3.1}$$

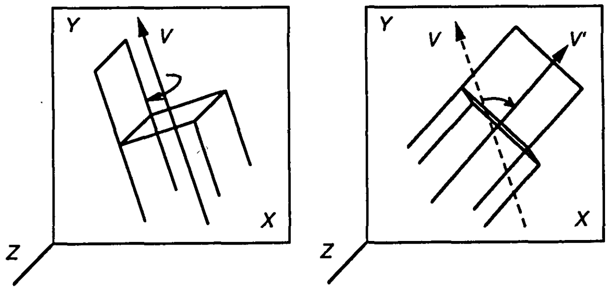


Fig. 1. Example of decomposition of an arbitrary rotation. Any rotation can be decomposed to rotation around some axis V in the image plane and rotation around axis Z. The edges are added only for purposes of illustration.

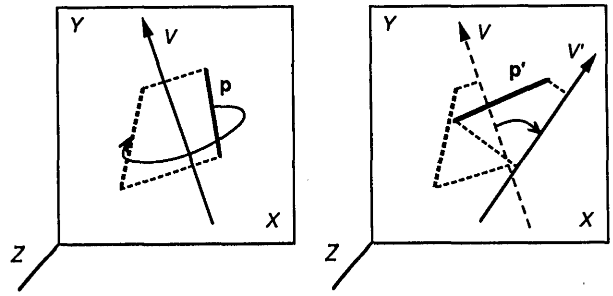


Fig. 2. After the complete rotation, the image of axis V, denoted V', remains in the image plane. The projection of any edge on V (V') before and after rotation is invariant.

Then the relation

$$\mathbf{p}_{i,j} \cdot \mathbf{v} = \mathbf{p}'_{i,j} \cdot \mathbf{v}' \tag{3.2}$$

holds for any i, j, where $\cdot$ denotes the scalar product of vectors in the image plane.

The last equality can be used to find the scale factor, s, and the directions of axes V and V' when the correspondence between points on the images is known. This relation is a homogeneous linear equation for four unknown coordinates in vectors $\mathbf{v}$ and $\mathbf{v}'$. Each corresponding pair of vectors in images P and P′ produces a new equation. As long as $\mathbf{v}$ and $\mathbf{v}'$ are nonzero vectors, the system of equations must have a rank less than 4. It is easy to show[25] that in the general case the rank of the system is equal to the dimension of the linear span of all $\mathbf{r}_{i,j}$. If the object is planar, the system has rank 2, and axes V and V' cannot be determined unambiguously, as was established by Lee and Huang.[24] Thus, at least three pairs of nonplanar edges are necessary for solving the equations. Since the equations are homogeneous, $\mathbf{v}$ and $\mathbf{v}'$ can be found up to a nonzero multiplier; for example, if we choose a solution with $|\mathbf{v}| = 1$, then the scale factor, s, can be obtained from the relation $s = 1/|\mathbf{v}'|$.

Thus, if an object contains four or more noncoplanar points, it is possible to recover from its two images the scale factor and the angle of rotation around an axis normal to the image plane.

Relation (3.2) also provides a consistency criterion for a given correspondence between points on two images, which may be implemented in the following way: scan all pairs of corresponding points in some arbitrary order, obtaining on each image a chain of vectors that passes over each point of the object once. Vectors $\mathbf{v}$ and $\mathbf{v}'$ are calculated by using the first three pairs from the chains. Then one by one, check condition (3.2) for the other points. If a pair does not satisfy this condition, no 3D interpretation exists. If a pair does satisfy this condition, use it to get a more precise estimate for $\mathbf{v}$ and $\mathbf{v}'$ and continue.

An important merit of our scheme is that computational complexity grows linearly with the number of points. Nevertheless, the scheme is an exhaustive one, because when condition (3.2) is valid for edges $\mathbf{r}_{i,i+1}$, it will be valid automatically for other edges (we shall not prove this simple conclusion).

So far, our discussion has depended on the assumption that the correspondence between images is known. However, the consistency criterion actually allows us to calculate this correspondence.

4. SEARCH FOR CORRECT CORRESPONDENCE

In the current literature, there are two hypotheses regarding the relationship between the structure-from-motion problem and the correspondence problem for apparent motion. According to the minimal-mapping theory[14] of apparent motion, correspondence can be established between elements on the basis of their proximity on consecutive projections. Only then, with the correspondence known, can a spatial interpretation be obtained. The second, structural theory,[19,26] states that the processes responsible for correspondence and 3D interpretation operate simultaneously with mutual support. It is possible that the visual system uses both methods to search for correspondence when interpreting motion. (However, when the structural theory is considered in the more general context of visual perception, it provides a description of the search for correspondence between two images of the same volumetric object. This process is an especially important element of recognition.)

The consistency criterion implements the structural hypothesis and solves the correspondence problem. The least squares in relation (3.2) in the course of correspondence verification are a measure of quality of correspondence. In the case of bijectory matching, all possible correspondences can be estimated, and the best correspondence can be found. The direct implementation of this idea is restricted, because the number of cases grows exponentially with the number of points. There are good heuristic methods that reduce the number of scanned correspondences to an appropriate level. But this combinatorial problem is beyond scope of this paper.

Real scenes with numerous objects provide a much more complex task for partial correspondence, because the consistency criterion can be used only for comparisons between correspondences of the same number of points. To circumvent this difficulty, we deal here only with the case of transparent objects. In this case, all feature points are represented on each projection, and the correspondence task can be formally resolved.

5. ONE-DIMENSIONAL FAMILY OF THREE-DIMENSIONAL INTERPRETATIONS

Suppose that the correct correspondence between two images, P and P′, has been established. The world coordinate system, XYZ, as before, is camera centered. This means that the camera is situated in a world origin and is stable while other objects move. Assume without loss of generality that the solution of the system of linear equations (3.2) satisfies the condition $|\mathbf{v}| = 1$. We can then introduce for them new coordinate systems, UVW and $U'V'W'$:

$$\mathbf{e}_u = (-\mathbf{v}_y, \mathbf{v}_x, 0)^T,$$
$$\mathbf{e}_v = (\mathbf{v}_x, \mathbf{v}_y, 0)^T,$$
$$\mathbf{e}_w = (0, 0, 1)^T,$$

and

$$\mathbf{e}_{u'} = s^2(-\mathbf{v}'_y, \mathbf{v}'_x, 0)^T,$$
$$\mathbf{e}_{v'} = s^2(\mathbf{v}'_x, \mathbf{v}'_y, 0)^T,$$
$$\mathbf{e}_{w'} = (0, 0, s)^T.$$

$\{\mathbf{e}_u, \mathbf{e}_v\}$ and $\{\mathbf{e}_{u'}, \mathbf{e}_{v'}\}$ constitute bases for the image plane, and $\mathbf{e}_w$ and $\mathbf{e}_{w'}$ are orthogonal to the image plane (Fig. 3).

Let us transform image P′ so that basis $\{\mathbf{e}_{u'}, \mathbf{e}_{v'}, \mathbf{e}_{w'}\}$ coincides with basis $\{\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}_w\}$. Consequently, by using this basis transformation, we exclude scaling and rotation around the axis orthogonal to the image plane and reduce the situation to the standard one of binocular stereo. We now need only consider the rotation of the object around the V axis, which in the case of binocular stereo is vertical.

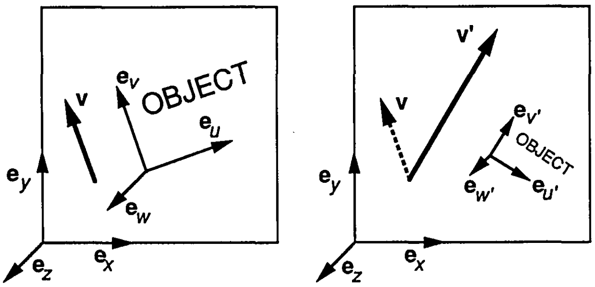


Fig. 3. Effect of scaling. An object is rotated and contracted by a factor of 2. The corresponding vectors **v** and **v'** and bases are shown for the initial (panel at left) and final (panel at right) positions of the object. Note that for scaling with a factor of 0.5, **v'** is twice as long as **v**, but the corresponding basis vectors are twice as short.

V coordinates for $\mathbf{p}_{i,j}$ and $\mathbf{p}'_{i,j}$ are the following:

$$(\mathbf{p}_{i,j})_v = \mathbf{v}\cdot\mathbf{p}_{i,j}, \qquad (\mathbf{p}'_{i,j})_v = \mathbf{v}'\cdot\mathbf{p}'_{i,j}.$$

They are equal to each other, as can be seen from formula (3.2). The U coordinates for $\mathbf{p}_{i,j}$ and $\mathbf{p}'_{i,j}$ can be obtained from the following formula:

$$(\mathbf{p}_{i,j})_u = \mathbf{u}\cdot\mathbf{p}_{i,j}, \qquad (\mathbf{p}'_{i,j})_u = \mathbf{u}'\cdot\mathbf{p}'_{i,j},$$

where $\mathbf{u} = \mathbf{e}_u$, $\mathbf{u}' = s^{-2}\mathbf{e}_{u'}$.

We have now reduced the motion task to a standard stereo task[27] where the axis of rotation corresponding to two projections is known and belongs to the frontal plane. We can now show that, for each pair of projected images and each angle of rotation α about the V axis, there is a unique 3D interpretation. Under the rotation, the vector $\mathbf{r}_{i,j}$ transforms to $\mathbf{r}'_{i,j}$. The V coordinates of these vectors are the same, so we can project edge $\mathbf{r}_{i,j}$ onto plane UW and consider its rotation in this plane. For this rotation we have

$$(\cos\alpha, -\sin\alpha)\begin{bmatrix}(\mathbf{r}_{i,j})_u\\(\mathbf{r}_{i,j})_w\end{bmatrix} = (\mathbf{r}'_{i,j})_u. \tag{5.1}$$

Consequently,

$$(\mathbf{r}_{i,j})_w = (\mathbf{r}_{i,j})_u \cot\alpha - (\mathbf{r}'_{i,j})_u/\sin\alpha. \tag{5.2}$$

Note that $(\mathbf{r}_{i,j})_u = (\mathbf{p}_{i,j})_u$ and $(\mathbf{r}'_{i,j})_u = (\mathbf{p}'_{i,j})_u$ are known, and $(\mathbf{r}_{i,j})_w = (\mathbf{r}_{i,j})_z$.

Thus for each angle, α, there exists a unique 3D interpretation, and the depth coordinates for edges can be obtained from formula (5.2).

Formula (5.2) can be rewritten in the more useful form

$$(\mathbf{r}_{i,j})_z = \lambda\left[\frac{(\mathbf{p}'_{i,j})_u + (\mathbf{p}_{i,j})_u}{2}\right] + \lambda^{-1}\left[\frac{(\mathbf{p}'_{i,j})_u - (\mathbf{p}_{i,j})_u}{2}\right], \tag{5.3}$$

where $\lambda = (\cot\alpha) - 1/(\sin\alpha)$. In this case λ, which can be any nonzero real number, parameterizes the space of all possible interpretations for two given images.

6. INFERENCE OF THREE-DIMENSIONAL STRUCTURE FROM PAIRWISE COMPARISONS

A common approach to depth reconstruction from orthographic projections is based on simultaneous comparison of three images.[14] As we have shown, pairwise comparison is sufficient to establish correspondence between points on images and recover all motion parameters except rotation around the V axis. It would be reasonable to use the information obtained at this stage for depth reconstruction. This can be done in a direct manner. Any successful pairwise comparison provides a one-parameter family of possible 3D interpretations. The families from pairwise comparisons of images of the same object necessarily have a common point: the object. Thus, the intersection of several families must be the unique 3D interpretation that is valid for all given images.

We consider here the case of three projections. (It can be easily generalized to an arbitrary number of projections.) We denote these projections P, P'_λ, and P'_μ and let the one-dimensional families of 3D interpretations for pairs $\{P, P'_\lambda\}$ and $\{P, P'_\mu\}$ be defined by real parameters λ and μ, respectively. We mark by the comparison's real parameter all coordinates obtained in the course of pairwise comparison. As long as the depth coordinate, $(\mathbf{r}_{i,j})_z$, for projection P is the same for both pairwise comparisons, the following equality is valid:

$$\lambda\left[\frac{(\mathbf{p}'_{\lambda,i,j})_u + (\mathbf{p}_{\lambda,i,j})_u}{2}\right] + \lambda^{-1}\left[\frac{(\mathbf{p}'_{\lambda,i,j})_u - (\mathbf{p}_{\lambda,i,j})_u}{2}\right] = \mu\left[\frac{(\mathbf{p}'_{\mu,i,j})_u + (\mathbf{p}_{\mu,i,j})_u}{2}\right] + \mu^{-1}\left[\frac{(\mathbf{p}'_{\mu,i,j})_u - (\mathbf{p}_{\mu,i,j})_u}{2}\right]. \tag{6.1}$$

Thus, if P', P'_λ and P'_μ are projections of one object with constant shape, there exist values for λ and μ such that Eq. (6.1) is valid for any pair i and j. If, on the other hand, appropriate λ and μ do not exist, there does not exist an intersection of interpretation families for pairs $\{P, P'_\lambda\}$ and $\{P, P'_\mu\}$.

To obtain values of unknowns λ and μ, it is sufficient that one have two equations, but it is easier to consider equation (6.1) as a system of linear homogeneous equations for unknowns λ, λ^{-1}, μ, and μ^{-1}. If the system consists of three or more equations, then, in general, it has rank 3, and the solution can be obtained up to a multiplier. Either of the conditions

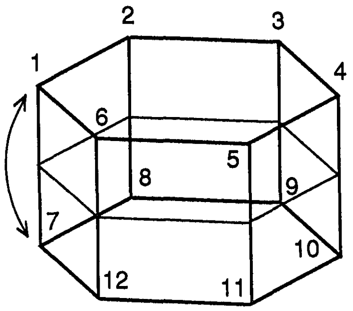
$$\lambda\lambda^{-1} = 1 \quad\text{or}\quad \mu\mu^{-1} = 1 \tag{6.2}$$

will provide a constraint for this multiplier.

Thus the method described needs at least three projections and at least four points in order for three independent equations of system (6.1) to be obtained. This condition is minimal, as was shown by Kontsevich *et al.*[17]

It is important to note that, for the motion case that does not have a rotation component around the Z axis, the system has rank 2. In this case $(\mathbf{p}_{\lambda,i,j})_u$ and $(\mathbf{p}_{\mu,i,j})_u$ are the same, and the system has the additional solution $(1, -1, 1, -1)$. This case also can be resolved (see Ref. 28).

Note that the system of equations (6.1) with constraint (6.2) permits the derivation of values for parameters up to sign. This result means that for each interpretation there exists a dual one that is symmetrical to the first interpretation through the image plane.



1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12
7, 8, 9,10,11,12, 1, 2, 3, 4, 5, 6

Fig. 4. Mirror symmetry defines the nontrivial correct correspondence. In this example the symmetry plane (depicted by a thin line) sets the correspondence (shown in the two lines of numbers at the bottom). The numbers beneath the drawing show the correspondence for vertices.

7. APPLICABILITY OF THE METHOD

Pairwise comparisons in the form described are not specific to the structure-from-motion task; they can be applied to any pair of projections from an object. Proximity in space and time, which is characteristic for motion, can only diminish the number of admissible correspondences in the matching process.

We noted in Section 3 that the existence of correct correspondence between two images cannot be an accidental fact. As noted elsewhere,[29,30] the visual system has a strong tendency to avoid coincidence. For pairwise comparisons, this tendency can be stated in the following assumption:

> If, for some pair of projections, correct correspondence exists, the projections are views of the same object.

With such an assumption, pairwise comparisons can be applied not only to sequential images of a moving object but also to binocular stereo images and to the comparison of sensory-input images with images stored in memory. Note that the latter process not only permits the use of memory for depth perception but might also be fundamental to the process of recognition.

It is an important fact that a unique 3D interpretation can be obtained from pairwise comparisons of projections that have very different origins. For a unique interpretation, the visual system must find at least two distinct correct correspondences. In the case in which only one correct correspondence is found, the visual system cannot find a unique depth and must choose an interpretation from possible solutions. In order to make this choice, the visual system must use some default value of parameter λ and recover depth for this value. In other words, the perceived viewing distance should tend toward a default value. This phenomenon is known as the specific-distance tendency.[31,32]

An advantage of the proposed method is that any additional information about motion or distances can be incorporated easily into the computational scheme. For example, in the case of a moving viewer in a static scene, the parameter sign in a one-parameter family is known, and it can be used to resolve ambiguity of the 3D interpretation (symmetry through the image plane). The precise value of the parameter can be obtained as a solution of system (6.1).

Another example is the case in which the distance between two positions of camera(s) or eye(s) corresponding to two projections of a stable object is known. Here it is possible to determine the volumetric structure of the object in the world-coordinate system. For instance, in the case of the binocular system this distance should be known. These two examples show that the proposed method can be applied to tasks for which additional information about motion is needed for scene interpretation, for example, in navigation.

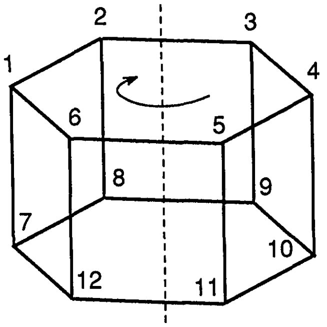
8. DETECTION AND USE OF SPATIAL SYMMETRIES

In some cases two projections or even one projection of an object is sufficient for shape inference without any additional information. This occurs when the object possesses either rotational or mirror symmetry.

When an object possesses mirror symmetry, the reflection of this object around some plane coincides with the object in its initial position. Correspondence of points on projection that is determined by reflection symmetry can be defined also by some rotation of the object (Fig. 4). This rotation is a superposition of reflection symmetry in the object's symmetry plane and reflection symmetry in the projection plane. The first symmetry transforms the object to itself with change of orientation; this symmetry defines point correspondence. The second symmetry recovers orientation without any change of projection.

If an object possesses rotational symmetry, then a nonidentical rotation exists after which the object coincides with itself. This rotation determines nonidentical-points correspondence within a single image, yielding a 3D interpretation (Fig. 5).

In summary, to detect object symmetry from a single projection, one has to compare the projection with itself and find all the correct correspondences. If some nonidentical correct correspondence exists, then a 3D interpretation of the projection possesses nontrivial symmetry. As in the case of comparison of different projections, the existence of a nonidentical correct correspondence in a single projection cannot be accidental. It is not unreason-



1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12
2, 3, 4, 5, 6, 1, 8, 9,10,11,12, 7

Fig. 5. Rotational symmetry correspondence also defines nontrivial correct correspondence.

able to assume that the visual system, while inferring 3D structure, attempts to make the maximally symmetrical interpretation. (There are experimental results supporting this assumption.[33])

Correct nonidentical correspondence inside one image virtually provides a new view of an object, providing a pairwise comparison that can be added to pairwise comparisons with other images. The additional view can be important in the derivation of a unique 3D interpretation.

In some cases, when the object possesses complex symmetry, and two or more nonidentical correspondences within a single image are available, 3D interpretation can be inferred from a single projection. This holds, for example, for images of regular polyhedra.

When only two projections of an object are known, symmetry allows the inference of a unique interpretation, as in the case of binocular stereo. When the object is static, the binocular system possesses only two projections of this object; the angle between the ocular axes is unknown, and the 3D structure cannot be obtained unambiguously. Presence of symmetry resolves this ambiguity.

9. CONCLUDING SUMMARY

The proposed method for 3D shape inference shows that seemingly different mechanisms of depth perception can actually be described as a single mechanism with a relatively simple construction. The method is based on decomposition of the depth perception into two stages: (1) pairwise comparisons of a given image with all images available to the vision system (the given image itself and images from another eye, from short-term memory, and from long-term memory), search for correct correspondences, and definition of each correct correspondence of a one-parameter family of 3D interpretations and (2) search for a unique (up to reflection in frontal plane) interpretation consistent with all correct correspondences.

We have not discussed additional low-level components of the pairwise comparison for different 3D mechanisms. For example, the binocular system uses mainly the element of spatial proximity between the two images to be fused, and the structure-from-motion mechanism is highly sensitive to the temporal component of the stimulation. These heuristics diminish the number of possible correspondences to be tested and help to establish the correct correspondences in real time. All other elements of shape inference might remain similar for all mechanisms. In this case different 3D mechanisms can really work together.

ACKNOWLEDGMENTS

I am indebted to V. V. Turovsky, A. P. Petrov, N. M. Grzywacz, Christopher W. Tyler, L. Barghout, L. K. Hardage, and R. Miller for helpful comments. This study was partly supported by National Institutes of Health grant EY 7890.

REFERENCES AND NOTES

1. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (Freeman, San Francisco, Calif., 1982).
2. W. F. Clocksin, "Perception of surface slant and edge labels from optical flow: a computational approach," Perception **9,** 253–269 (1980).
3. J. A. Webb and J. K. Aggarwal, "Visually interpreting the motions of objects in depth," Comput. J. **14,** 40–49 (1981).
4. N. Sugie and H. Inagaki, "A computational aspect of kinetic depth effect," Biol. Cybern. **50,** 431–436 (1984).
5. B. M. Bennett and D. D. Hoffman, "The computation of structure from fixed axis motion: nonrigid structures," Biol. Cybern. **51,** 293–300 (1985).
6. D. D. Hoffman and B. M. Bennett, "The computation of structure from fixed axis motion: rigid structures," Biol. Cybern. **54,** 71–83 (1986).
7. H. C. Longuet-Higgins, "The role of the vertical dimension in stereoscopic vision," Perception **11,** 377–386 (1983).
8. J. J. Koenderink and A. J. van Doorn, "Local structure of movement parallax of the plane," J. Opt. Soc. Am. **66,** 717–723 (1976).
9. H. C. Longuet-Higgins, "Visual ambiguity of a moving plane," Proc. R. Soc. London B **223,** 165–175 (1984).
10. A. M. Waxman and K. Wohn, "Contour evaluation, neighborhood deformation and global image flow: planar surfaces in motion," Int. J. Robot. Res. **4,** 95–108 (1985).
11. J.-Q. Fang and T. S. Huang, "Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames," IEEE Trans. Patt. Anal. Mach. Intell. **PAMI-6,** 545–554 (1984).
12. T. Broida and R. Chellappa, "Estimating the kinematics and structure of a rigid object from a sequence of monocular images," IEEE Trans. Patt. Anal. Mach. Intell. **13,** 497–513 (1991).
13. J. Roach and J. Aggarwal, "Determining the movement of objects from a a sequence of images," IEEE Trans. Patt. Anal. Mach. Intell. **PAMI-2,** 554–562 (1980).
14. S. Ullman, *The Interpretation of Visual Motion* (MIT Press, Cambridge, Mass., 1979).
15. 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," Opt. Acta **22,** 773–791 (1975).
16. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature **293,** 133–135 (1981).
17. L. L. Kontsevich, M. L. Kontsevich, and A. H. Shen, "Two algorithms of shape recovery." Autometria **No. 5,** 72–77 (1987). (In Russian.)
18. C. J. Poelman and T. Kanade, "A paraperspective factorization nethod for shape and motion recovery." Publication CMU-CS-92-208 (Carnegie-Mellon University, Pittsburgh, Pa., 1992).
19. S. Ullman, "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motion," Perception **13,** 255–274 (1984).
20. M. Braunstein, D. Hoffman, L. Shapiro, G. Andersen, and B. Bennett, "Minimum points and views for the recovery of three-dimensional structure," J. Exp. Psychol. **13,** 335–343 (1987).
21. J. T. Todd, "The perception of three-dimensional structure from rigid and nonrigid motion," Percept. Psychophys. **36,** 97–103 (1984).
22. R. M. Siegel and R. A. Andersen, "Perception of three-dimensional structure from two-dimensional visual motion in monkey and man," Nature **331,** 259–261 (1988).
23. B. M. Bennett, D. D. Hoffman, J. E. Nicola, and C. Prakash, "Structure from two orthographic views of rigid motion," J. Opt. Soc. Am. A **6,** 1052–1069 (1989).
24. C. H. Lee and T. Huang, "Finding point correspondences and determining motion of a rigid object from two weak perspective views," Comput. Vision, Graphics, Image Process. **52,** 309–327 (1990).
25. Let p and p' be linear projective mappings from three-dimensional physical space onto projection planes P and P', respectively. Then $p:\mathbf{r}_{i,j} \rightarrow \mathbf{p}_{i,j}$ and $p':\mathbf{r}_{i,j} \rightarrow \mathbf{p}'_{i,j}$. The vector of the coefficients in relation (3.2) is $(\mathbf{p}_{i,j}^T, -\mathbf{p}_{i,j}'^T)^T$. It is also an image of the linear mapping $\mathrm{p} \oplus (-p'):\mathbf{r}_{i,j} \rightarrow (\mathbf{p}_{i,j}^T, -\mathbf{p}_{i,j}'^T)^T$. In the general case, rank $[p \oplus (-p')] = 4$, and therefore the rank of the mapping image (the rank of the system of equations) is equal to the rank of the linear span

of all $\mathbf{r}_{i,j}$. In some particular cases, for example in the case of scaling only, the rank of the system of equations can be less than the rank of the linear span of all $\mathbf{r}_{i,j}$.

26. N. Grzywacz and A. Yuille, "Massively parallel implementations of theories for apparent motion," Spatial Vis. **3,** 15–44 (1988).
27. L. Kaufman, *Sight and Mind: An Introduction to Visual Perception* (Oxford U. Press, New York, 1974).
28. The nonzero solution of a linear system of homogeneous equations (6.1) that minimizes normalized square error is an eigenvector corresponding to the minimal eigenvalue for the product of the coefficient matrix and its transposed copy. In the case in which the system rank is two, two eigenvalues are equal to zero (or are significantly less than other eigenvalues when noise is present), and the corresponding eigenvectors form a basis for two-dimensional subspace of possible solutions. To find a solution that satisfies condition (6.2) one should select a vector different from $(1, -1, 1, -1)$ in the subspace and find its linear combination with $(1, -1, 1, -1)$ that fits the condition. This is a simple linear task.
29. W. H. Ittelson, *The Ames Demonstrations in Perception* (Hafner, New York, 1952).
30. I. Biederman, "Aspects and extensions of a theory of human image understanding," in *Computational Processes in Human Vision: An Interdisciplinary Perspective,* Z. Pylyshyn, ed. (Ablex, Norwood, N.J., 1988), pp. 370–428.
31. W. C. Gogel, "An indirect measure of perceived distance from oculomotor cues," Percept. Psychophys. **21,** 3–11 (1977).
32. E. B. Johnston, "Systematic distortions of shape from stereopsis," Vision Res. **31,** 1351–1360 (1991).
33. M. King, J. Tangney, G. E. Meyer, and I. Biederman, "Shape constancy and a perceptual bias towards symmetry," Percept. Psychophys. **19,** 129–136 (1976).