



Trifocal tensors for weak perspective and paraperspective projections

Alfred M. Bruckstein^{a,1}, Robert J. Holt^{a,*}, Thomas S. Huang^b, Arun N. Netravali^a

^a*Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA*

^b*Beckman Institute, University of Illinois, Urbana, IL 61801, USA*

Received 7 January 1999; received in revised form 11 November 1999; accepted 11 November 1999

Abstract

Trilinear relationships among the image point coordinates obtained by perspective projection of several feature points over three views have been investigated, and it has been shown that seven point correspondences are sufficient to determine the coefficients involved. We show that analogous trilinear relationships exist for the weak perspective and paraperspective projections, and that only four point correspondences are necessary for the determination of the coefficients. When the trilinear relationships are known, the position of the image of a point in one camera system can be determined from its image in the two other camera systems through linear equations. © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Pose determination; Linear algorithm; Multiple images

1. Introduction

Recently, the trilinear relationships that exist among three perspective views of an object have been discussed by several authors [1–4]. In particular there exists a trio of 3×3 matrices, called the “trifocal tensor” [4], whose elements are the coefficients in a set of linear equations involving the three sets of image coordinates of a feature point. This set of matrices has also appeared [5,6] in the context of determining motion from structure using line correspondences. Shashua [3] and Zhang [4] report that a minimum of seven point correspondences across the three views are necessary in order to solve for all the coefficients.

Shimshoni et al. [7] give a geometric interpretation of the problem of motion recovery from three weak perspective views. The approach provided combines the

results of analyzing pairs of images, a direct and natural approach to such problems. In this paper we show that there exist analogous trifocal tensors for the weak perspective and paraperspective projections. However, only four point correspondences over three views are necessary in order to solve for all the components of the three matrices involved, as several of these components are always zero. As in the perspective case, knowledge of the trifocal tensor and the images of a point in two of the views allows the immediate computation of its position in the third view, as long as the point is in general position.

2. Notation

Let \mathbf{P} , \mathbf{P}' , and \mathbf{P}'' be the coordinates of a point in three-space in the three camera coordinate systems, and let \mathbf{p} , \mathbf{p}' , and \mathbf{p}'' be the corresponding image points, under either weak perspective or paraperspective projection. Let the relations between the systems be given by

$$\mathbf{P}' = \mathbf{R}'\mathbf{P} + \mathbf{T}' \quad \mathbf{P}'' = \mathbf{R}''\mathbf{P} + \mathbf{T}'', \quad (1)$$

* Corresponding author. Tel.: +1-908-582-3177; fax: +1-908-582-7308.

E-mail address: rjh@research.bell-labs.com (R.J. Holt).

¹Permanent address: Department of Computer Science, Technion – IIT 32000, Haifa, Israel.

where

$$\mathbf{R}' = \begin{bmatrix} r'_1 & r'_2 & r'_3 \\ r'_4 & r'_5 & r'_6 \\ r'_7 & r'_8 & r'_9 \end{bmatrix}, \quad \mathbf{T}' = \begin{bmatrix} t'_1 \\ t'_2 \\ t'_3 \end{bmatrix},$$

$$\mathbf{R}'' = \begin{bmatrix} r''_1 & r''_2 & r''_3 \\ r''_4 & r''_5 & r''_6 \\ r''_7 & r''_8 & r''_9 \end{bmatrix}, \quad \mathbf{T}'' = \begin{bmatrix} t''_1 \\ t''_2 \\ t''_3 \end{bmatrix}$$

with \mathbf{R}' and \mathbf{R}'' being rotation matrices.

For the weak perspective and paraperspective projections, we take each camera coordinate system to have its origin \mathbf{O} at the focal point and the Z -axis as its optic axis. For the paraperspective camera we take $Z = F$ as its image plane, where F is the focal length. The focal lengths of the different cameras are assumed to be known but need not all be equal, so we take F' and F'' as the focal lengths of the second and third cameras. The weak perspective projection is an orthographic projection combined with a scaling, and as such does not have a well-defined focal length, so we just set $F = F' = F'' = 1$. If $\mathbf{P} = (X, Y, Z)$ is a point in space, then its weak perspective projection is $\mathbf{p} = (x, y, 1)$, where

$$x = wX, \quad y = wY, \quad (2)$$

where w is a positive constant that equals 1 for ordinary orthographic projection and otherwise is unknown for an uncalibrated camera. These constants may differ among the cameras, and the scaling factors will be denoted by w , w' , and w'' . Knowledge of the values of these is not necessary for the determination of the position of the image of a point in one camera when given its image in the other two, so we will assume w , w' , and w'' are unknown.

The paraperspective projection is a better approximation to true perspective projection than the weak perspective, while still retaining linear relationships between the feature point and image point coordinates. A point $\mathbf{C} = (C_1, C_2, C_3)$, intended to approximate the centroid of the feature points, is selected. A feature point \mathbf{P} is first projected onto the plane $Z = C_3$ in the direction of $\overline{\mathbf{OC}}$, and is then perspective projected onto the image plane. Thus \mathbf{P} is first projected to the point $\mathbf{P} - [(Z - C_3)/C_3]\mathbf{C}$, and then to the point $\mathbf{p} = (F/C_3)\{\mathbf{P} - [(Z - C_3)/C_3]\mathbf{C}\}$. The image point coordinates are then given by

$$x = \frac{F(C_3 X - C_1 Z + C_1 C_3)}{C_3^2},$$

$$y = \frac{F(C_3 Y - C_2 Z + C_2 C_3)}{C_3^2}. \quad (3)$$

Note that the weak perspective projection is the special case of the paraperspective projection when \mathbf{C} is taken to be the point $(0, 0, F/w)$. The chosen centroids may be different in the three views, so we will denote them by \mathbf{C} , \mathbf{C}' , and \mathbf{C}'' (\mathbf{C}' and \mathbf{C}'' are expressed in terms of the second and third camera coordinate systems, respectively). In an ideal situation they will be approximated well enough so that we can assert $\mathbf{C}' = \mathbf{R}'\mathbf{C} + \mathbf{T}'$ and $\mathbf{C}'' = \mathbf{R}''\mathbf{C} + \mathbf{T}''$, but we will not make that assumption as it is not necessary to prove our results.

We also denote by $[\mathbf{p}]_{\times}$ the skew-symmetric matrix such that $[\mathbf{p}]_{\times} \mathbf{v} = \mathbf{p} \times \mathbf{v}$ for all 3×3 vectors \mathbf{v} , so that

$$[\mathbf{p}]_{\times} = \begin{bmatrix} 0 & -F & y \\ F & 0 & -x \\ -y & x & 0 \end{bmatrix}.$$

3. The trifocal tensor

A compact representation of the trilinear constraints is given by Zhang [4]. The trifocal tensor consists of three 3×3 matrices \mathbf{K} , \mathbf{L} , and \mathbf{M} , and they satisfy a relation of the form

$$[\mathbf{p}']_{\times} (x\mathbf{K} + y\mathbf{L} + F\mathbf{M}) [\mathbf{p}'']_{\times} = \mathbf{0}, \quad (4)$$

where $\mathbf{0}$ is the 3×3 zero matrix. Thus each point correspondence gives rise to nine linear equations, although just four of them are independent [3,4]. These equations are all homogeneous in the components of \mathbf{K} , \mathbf{L} , and \mathbf{M} , so at least 26 independent equations, and hence seven points, are necessary to solve for all the components (up to a common scale factor).

4. Weak perspective projection

We seek an identity of the form (4) that holds in the case of weak perspective projection. Such a relation can be found after some algebra applied to Eqs. (1) and (2). One method is to substitute $(x, y, x', y', x'', y'') = (wX, wY, w'X', w'Y', w''X'', w''Y'')$ in Eq. (4), then express X' , Y' , X'' , and Y'' in terms of X , Y , and Z through the use of Eq. (1), and set all the coefficients of terms of the form $X^i Y^j Z^k$ equal to zero. The result, which can be verified by direct substitution, is

$$\mathbf{K} = \begin{bmatrix} (r'_3 r''_1 - r'_1 r''_3)/w & (r'_3 r''_4 - r'_1 r''_6)/w & 0 \\ (r'_6 r''_1 - r'_4 r''_3)/w & (r'_6 r''_4 - r'_4 r''_6)/w & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{L} = \begin{bmatrix} (r'_3 r''_2 - r'_2 r''_3)/w & (r'_3 r''_5 - r'_6 r''_2)/w & 0 \\ (r'_6 r''_2 - r'_5 r''_3)/w & (r'_6 r''_5 - r'_5 r''_6)/w & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M} = \begin{bmatrix} r'_3 t''_1 - r''_3 t'_1 & r'_3 t''_2 - r''_6 t'_1 & r'_3/w'' \\ r'_6 t''_1 - r''_3 t'_2 & r'_6 t''_2 - r''_6 t'_2 & r'_6/w'' \\ -r''_3/w' & -r''_6/w' & 0 \end{bmatrix}. \quad (5)$$

An alternative way of expressing these is

$$\mathbf{K} = (\tilde{\mathbf{R}}'_{(3)} \tilde{\mathbf{R}}''_{(1)}{}^T - \tilde{\mathbf{R}}'_{(1)} \tilde{\mathbf{R}}''_{(3)}{}^T)/w$$

$$\mathbf{L} = (\tilde{\mathbf{R}}'_{(3)} \tilde{\mathbf{R}}''_{(2)}{}^T - \tilde{\mathbf{R}}'_{(2)} \tilde{\mathbf{R}}''_{(3)}{}^T)/w$$

$$\mathbf{M} = \tilde{\mathbf{R}}'_{(3)} \tilde{\mathbf{T}}''^T - \tilde{\mathbf{T}}' \tilde{\mathbf{R}}''_{(3)}{}^T,$$

where $\tilde{\mathbf{R}}'_i$ ($\tilde{\mathbf{R}}''_i$) is the i th column of \mathbf{R}' (\mathbf{R}'') with the third component replaced by 0, $\tilde{\mathbf{T}}' = [t'_1 \ t'_2 \ 1/w']^T$, and $\tilde{\mathbf{T}}'' = [t''_1 \ t''_2 \ 1/w'']^T$.

Since there are 16 nonzero elements of the trifocal tensor, just 15 independent equations are needed to solve for all the components. As is the case with the perspective trifocal tensor [3,4], there are four independent linear equations among the nine present in Eq. (4); for example the equations corresponding to the upper left 2×2 submatrix are linearly independent. Consequently, four points in general position are sufficient to determine all the components of \mathbf{K} , \mathbf{L} , and \mathbf{M} .

Knowing the location of an image point tells us a line in space on which the corresponding feature point lies. A point whose images are (x, y) , (x', y') , and (x'', y'') in the three cameras lies on the lines $(x/w, y/w, 0) + (0, 0, 1)t$, $\mathbf{R}'^{-1}[(x'/w', y'/w', 0) - (t'_1, t'_2, t'_3) + (0, 0, 1)u]$, and $\mathbf{R}''^{-1}[(x''/w'', y''/w'', 0) - (t''_1, t''_2, t''_3) + (0, 0, 1)v]$, where t, u , and v are parameters that may take on any real value. From this, it can be seen that the trifocal tensor approach fails when the optic axes of at least two of the cameras are parallel. This condition may be characterized by $r'_9 = \pm 1$ (which implies $r'_3 = r'_6 = r'_7 = r'_8 = 0$), $r''_9 = \pm 1$ (which implies $r''_3 = r''_6 = r''_7 = r''_8 = 0$), or $r'_3 r''_3 + r'_6 r''_6 + r'_9 r''_9 = \pm 1$ (which implies $(r''_3, r''_6, r''_9) = \pm (r'_3, r'_6, r'_9)$).

Now if the trifocal tensor is known and the image of a point in two views is given, then its position in the third image can be determined as the solution of a set of nine linear equations (four of them independent) in two unknowns. For example, given x, y, x' , and y' , we can determine x'' and y'' as the least-squares solution of these four independent equations:

$$k_5 x + l_5 y - m_8 y' + m_5 = m_6 y'',$$

$$k_4 x + l_4 y - m_7 y' + m_4 = m_6 x'',$$

$$k_2 x + l_2 y - m_8 x' + m_2 = m_3 y'',$$

$$k_1 x + l_1 y - m_7 x' + m_1 = m_3 x'',$$

which is

$$x'' = \frac{m_3(k_1 x + l_1 y - m_7 x' + m_1) + m_6(k_4 x + l_4 y - m_7 y' + m_4)}{m_3^2 + m_6^2}, y'' = \frac{m_3(k_2 x + l_2 y - m_8 x' + m_2) + m_6(k_5 x + l_5 y - m_8 y' + m_5)}{m_3^2 + m_6^2}. \quad (6)$$

Both m_3 and m_6 cannot be zero when no two of the three optic axes are parallel, so this recovery is always possible in this case. Note that since the scale factors w, w' , and w'' do not appear in Eq. (4), their values are not necessary for the determination of the location of the image point in one camera given the corresponding positions in the other two.

5. Paraperspective projection

We proceed just as in the previous section, but here the expressions involved are much more complicated. A relation of the form (4) can be found after a great deal of algebra applied to Eqs. (1) and (3). One direct method is to make the substitutions for x and y , and the analogous ones for $x', y', x'',$ and y'' , given by Eq. (3), into Eq. (4). Then, again express $X', Y', X'',$ and Y'' in terms of $X, Y,$ and Z through the use of Eq. (1), and set all the coefficients of terms of the form $X^i Y^j Z^k$ equal to zero. The result, which can be verified by direct substitution, can be expressed as

$$\begin{aligned} \mathbf{K} &= C_3 [\tilde{\mathbf{C}}'(r''_7 \mathbf{R}''^{(3)T} - r'_7 \mathbf{R}''^{(3)T}) \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C''_3 \tilde{\mathbf{C}}' \mathbf{C}^T (r'_7 \tilde{\mathbf{R}}''^T - \mathbf{R}''^{(3)} \tilde{\mathbf{R}}''_{(1)}{}^T) \\ &\quad + C'_3 (\tilde{\mathbf{R}}'_{(1)} \mathbf{R}''^{(3)T} - r''_7 \tilde{\mathbf{R}}') \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C'_3 C''_3 (\tilde{\mathbf{R}}' \mathbf{C} \tilde{\mathbf{R}}''_{(1)}{}^T - \tilde{\mathbf{R}}'_{(1)} \mathbf{C}^T \tilde{\mathbf{R}}''^T)], \\ \mathbf{L} &= C_3 [\tilde{\mathbf{C}}'(r''_8 \mathbf{R}''^{(3)T} - r'_8 \mathbf{R}''^{(3)T}) \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C''_3 \tilde{\mathbf{C}}' \mathbf{C}^T (r'_8 \tilde{\mathbf{R}}''^T - \mathbf{R}''^{(3)} \tilde{\mathbf{R}}''_{(2)}{}^T) \\ &\quad + C'_3 (\tilde{\mathbf{R}}'_{(2)} \mathbf{R}''^{(3)T} - r''_8 \tilde{\mathbf{R}}') \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C'_3 C''_3 (\tilde{\mathbf{R}}' \mathbf{C} \tilde{\mathbf{R}}''_{(2)}{}^T - \tilde{\mathbf{R}}'_{(2)} \mathbf{C}^T \tilde{\mathbf{R}}''^T)], \\ \mathbf{M} &= \tilde{\mathbf{C}}' [(C_3 r''_9 + t''_3 - C_3) \mathbf{R}''^{(3)T} \\ &\quad - (C_3 r'_9 + t'_3 - C'_3) \mathbf{R}''^{(3)T}] \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C''_3 \tilde{\mathbf{C}}' \mathbf{C}^T [(C_3 r'_9 + t'_3) \tilde{\mathbf{R}}''^T - \mathbf{R}''^{(3)} (C_3 \tilde{\mathbf{R}}'_{(3)}{}^T + \tilde{\mathbf{T}}''^T)] \\ &\quad + C'_3 [(C_3 \tilde{\mathbf{R}}'_{(3)} + \tilde{\mathbf{T}}') \mathbf{R}''^{(3)T} - (C_3 r''_9 + t''_3) \tilde{\mathbf{R}}'] \mathbf{C} \tilde{\mathbf{C}}''^T \\ &\quad + C'_3 C''_3 [\tilde{\mathbf{R}}' \mathbf{C} (C_3 \tilde{\mathbf{R}}'_{(3)}{}^T + \tilde{\mathbf{C}}''^T + \tilde{\mathbf{T}}''^T) \\ &\quad - (C_3 \tilde{\mathbf{R}}'_{(3)} + \tilde{\mathbf{C}}' + \tilde{\mathbf{T}}') \mathbf{C}^T \tilde{\mathbf{R}}''^T], \end{aligned} \quad (7)$$

where $\tilde{\mathbf{R}}'$ is \mathbf{R}' with the third row replaced by zeros, $\mathbf{R}''_{(i)}$ denotes the i th column of \mathbf{R}'' , $\mathbf{R}''^{(i)}$ denotes the transpose of the i th row of \mathbf{R}'' , $\tilde{\mathbf{T}}' = [t'_1 \ t'_2 \ C'_3]^T$, $\tilde{\mathbf{T}}'' = [t''_1 \ t''_2 \ C''_3]^T$, $\tilde{\mathbf{C}}' = [C'_1 \ C'_2 \ 0]^T$, and $\tilde{\mathbf{C}}'' = [C''_1 \ C''_2 \ 0]^T$. This solution reduces to Eq. (5) when $\mathbf{C} = [0 \ 0 \ 1/w]^T$, $\mathbf{C}' = [0 \ 0 \ 1/w']^T$, and $\mathbf{C}'' = [0 \ 0 \ 1/w'']^T$,

with all the coefficients of \mathbf{K} , \mathbf{L} , and \mathbf{M} scaled by the factor $1/(ww'w'')$.

As is the case with the weak perspective projection, the trilinear relationships run into difficulty if the lines in space determined by the image point in two of the cameras are parallel. Here a point whose images are (x, y) , (x', y') , and (x'', y'') in the three cameras lies on the lines $(C_3x/F, C_3y/F, C_3) + (C_1, C_2, C_3)t$, $\mathbf{R}'^{-1}[(C_3x'/F', C_3y'/F', C_3) - (t'_1, t'_2, t'_3) + (C'_1, C'_2, C'_3)]u$, and $\mathbf{R}''^{-1}[(C_3x''/F'', C_3y''/F'', C_3) - (t''_1, t''_2, t''_3) + (C''_1, C''_2, C''_3)]v$. From this, it can be determined that the trifocal tensor approach does not work when at least two of \mathbf{C} , $\mathbf{R}'^{-1}\mathbf{C}'$, and $\mathbf{R}''^{-1}\mathbf{C}''$ are parallel.

The situation now is essentially identical to that with the weak perspective projection. Four points in general position are sufficient to determine all the components of the paraperspective trifocal tensor, and once that is

known, the image of a point in the third camera can be determined from its image in the first two.

6. Experimental results

In order to determine the applicability of the weak perspective and paraperspective trifocal tensors to data obtained from true perspective projections, we ran several experiments on synthetic data. Since the true perspective trifocal tensor requires seven points while the others need just four, their greatest use will be in cases where we have exactly four, five, or six point correspondences.

In our first set of experiments, we took four feature points fixed at the corners of a regular tetrahedron inscribed in the unit sphere, so that these points form an

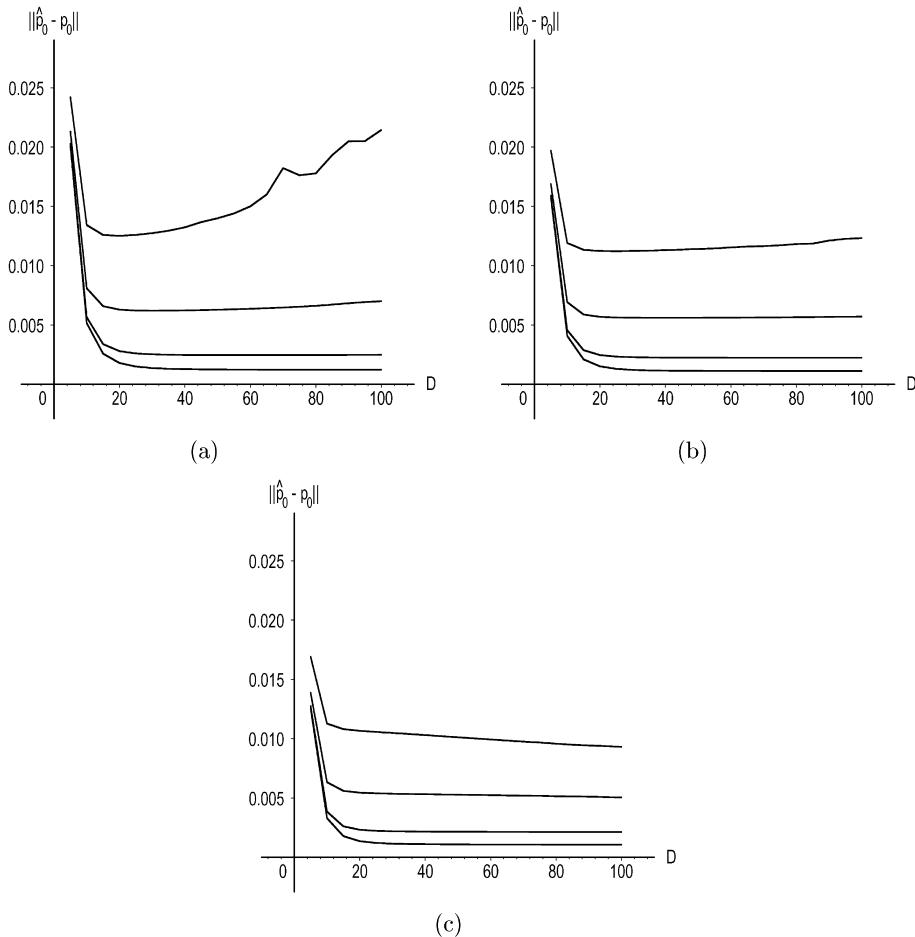


Fig. 1. Average errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) 4, (b) 5, and (c) 6 feature points placed at the vertices of a regular tetrahedron, trigonal bipyramid arrangement, and regular octahedron, respectively. D is the distance from each camera to the center of the sphere from which the feature points are chosen. In each case, from bottom to top the graphs are for 1, 2, 5, and 10 pixel discretization errors.

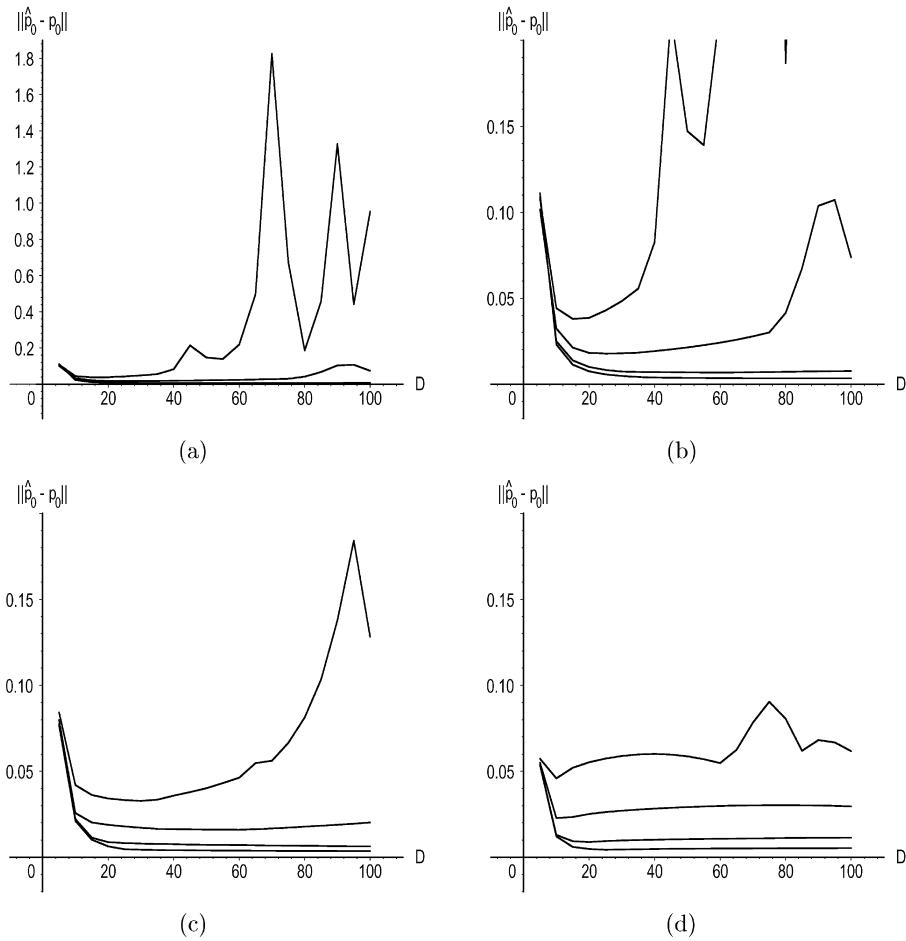


Fig. 2. Maximum errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) and (b) 4, (c) 5, and (d) 6 feature points regularly spaced points. (b) is a close-up of (a) drawn with the same scale as (c) and (d). In each case, from bottom to top the graphs are for 1, 2, 5, and 10 pixel discretization errors.

optimum arrangement for pose determination as in Ref. [8]. This tetrahedron was then given a random 3-D rotation. Then three more random rotations were chosen for the cameras. Trifocal tensors are not useful when the viewing directions of the three cameras are similar, so the viewing directions were chosen to be within 30° of three mutually orthogonal directions. Specifically, one rotation was randomly chosen from a uniform distribution of rotations whose third columns, when expressed as 3×1 vectors, were within 30° of the Z-axis. The other two rotations were chosen in a similar manner, but within 30° of the X- and Y-axis. The translations for the cameras were then taken to be $[0 \ 0 \ D]$, so that the point \mathbf{P} in the world coordinate system is $\mathbf{P}' = \mathbf{R}\mathbf{P} + \mathbf{T}$ in the camera coordinate system. Thus the center of the original tetrahedron is at distance D from each camera. The quantity D took the values $D = 5, 10, \dots, 100$.

A fifth point \mathbf{P}_0 was then chosen randomly from the unit sphere. The collection of five points was then translated by a vector randomly chosen from the ball of radius M centered at the origin, where M was either 0, $D/8$, or $D/4$. The latter corresponds to keeping the object within a 28° field of view of a camera. It was found that the value of M had little impact on the results, so we set $M = D/4$ in all the following experiments.

We assumed a focal length of one for each camera. The images of the first four feature points were computed using true perspective projection, and then we added discretization errors in the x- and y-coordinate of each point. These errors were uniformly selected from the interval $[-K\varepsilon, K\varepsilon]$ for $\varepsilon = 0.001$ and $K = 1, 2, 5, 10$, which corresponds to a K -pixel error in an image with resolution 1000×1000 pixels. With this data, we calculated the trifocal tensor of the form (5) or (7). That is,

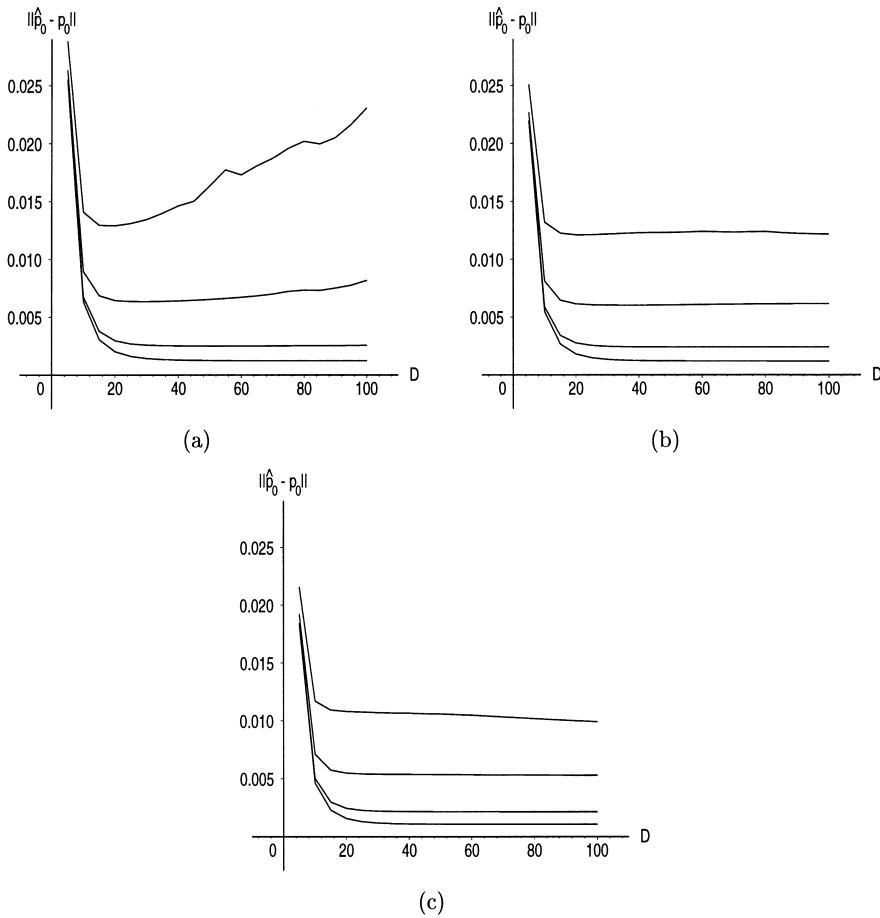


Fig. 3. Average errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) 4, (b) 5, and (c) 6 feature points randomly chosen from balls of radius $1/2$ surrounding the vertices of a regular tetrahedron, trigonal bipyramid arrangement, and regular octahedron, respectively, inscribed in the unit sphere. D is the distance from each camera to the center of the sphere from which the feature points are chosen. In each case, from bottom to top the graphs are for 1, 2, 5, and 10 pixel discretization errors.

we assumed that the 11 elements that are zero in Eqs. (5) and (7) do indeed vanish, and found optimal least-squares fits on the remaining 16 elements. We then took the fifth point and added the same discretization error in two images. Then we computed its image for the third camera by using Eq. (6). The distance from this computed image point to the actual location of the image point assuming no error was then recorded. The results are shown in Figs. 1 and 2. The error in the position of the image of \mathbf{P}_0 in the third camera is denoted by $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ in the graphs. The programs and figures were all produced with the symbolic manipulation program Maple [9,10].

With four original feature points, the results are best for $D > 15$. For smaller values of D the weak perspective projection is not a good approximation for the true perspective. For a one or two pixel error in the image

points, the value of $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ remains small for D up to 100, but with larger errors, particularly when $DK\varepsilon > 0.4$, we occasionally come up with a bad example. With the larger values of D , the discretization error becomes significant in comparison with (at least 25% of) the distance between the image points. These occur when the image points cluster in at least one of the views. When the discretization error is no more than five pixels, $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ is nearly constant for $20 \leq D \leq 70$, and for a one or two pixel error it is nearly constant for $20 \leq D \leq 100$, with a value of approximately $1.3K\varepsilon$.

We repeated these experiments for five and six original feature points. In order to have results as stable as possible, the distances between the points were maximized while keeping them within the unit sphere. The arrangement for five points can be described as placing two of them at the north and south poles, and the other three on

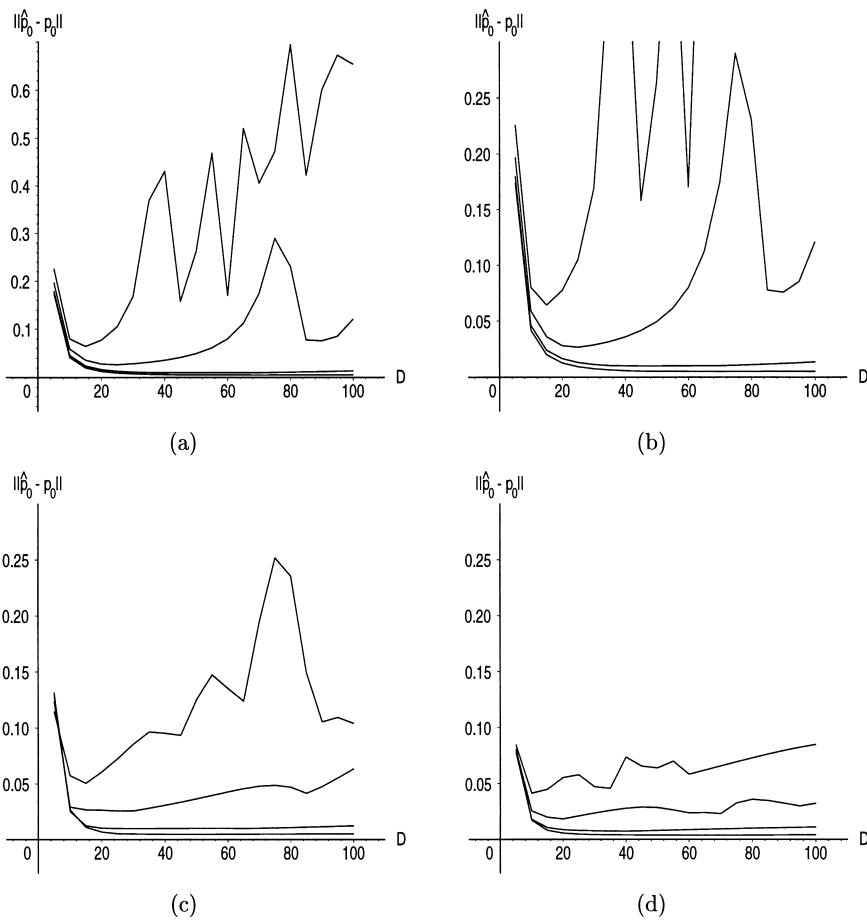


Fig. 4. Maximum errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) and (b) 4, (c) 5, and (d) 6 feature points randomly chosen from balls of radius $1/2$ surrounding the vertices of a regular tetrahedron, trigonal bipyramid arrangement, and regular octahedron, respectively, inscribed in the unit sphere. (b) is a close-up of (a) drawn with the same scale as (c) and (d). In each case, from bottom to top the graphs are for 1, 2, 5, and 10 pixel discretization errors.

the equator 120° apart (trigonal bipyramid). Six points were placed at the vertices of a regular octahedron. The maximum values of $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ decrease by about a factor of two with each additional feature point, while the average errors only decrease slightly. The latter are approximately $1.2K\varepsilon$ and $1.0K\varepsilon$ for five and six original feature points, respectively.

We note that for the purposes of determining the location of an image point in the third camera image, the weak perspective and paraperspective trifocal tensors have the same form insofar as the same 11 components are always zero for each. Thus, it does not matter which of these we say we are using for this image point determination. The difference between the two matters for pose determination, where the rotation and translation between the camera coordinate systems must be computed. Here the paraperspective assumption will yield

better estimates by exploiting some additional information.

In the second set of experiments, rather than being fixed at the vertices of a regular tetrahedron, trigonal bipyramid, or octahedron, the feature points are randomly chosen from spheres of radius 0.5 about the vertices of these objects. The rest of the experiments were the same as for the first set, with 1000 trials run for each value of D . This increases the frequency of bad cases due to the feature points not being so well spaced, but again these are confined to $D < 10$ and $DK\varepsilon > 0.4$. The results are shown in Figs. 3 and 4. For five or six points, the average errors are almost the same as with the well spaced points of the first set of experiments, while with four points, the values of $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ are about 12% greater than the corresponding values in the first set of experiments.

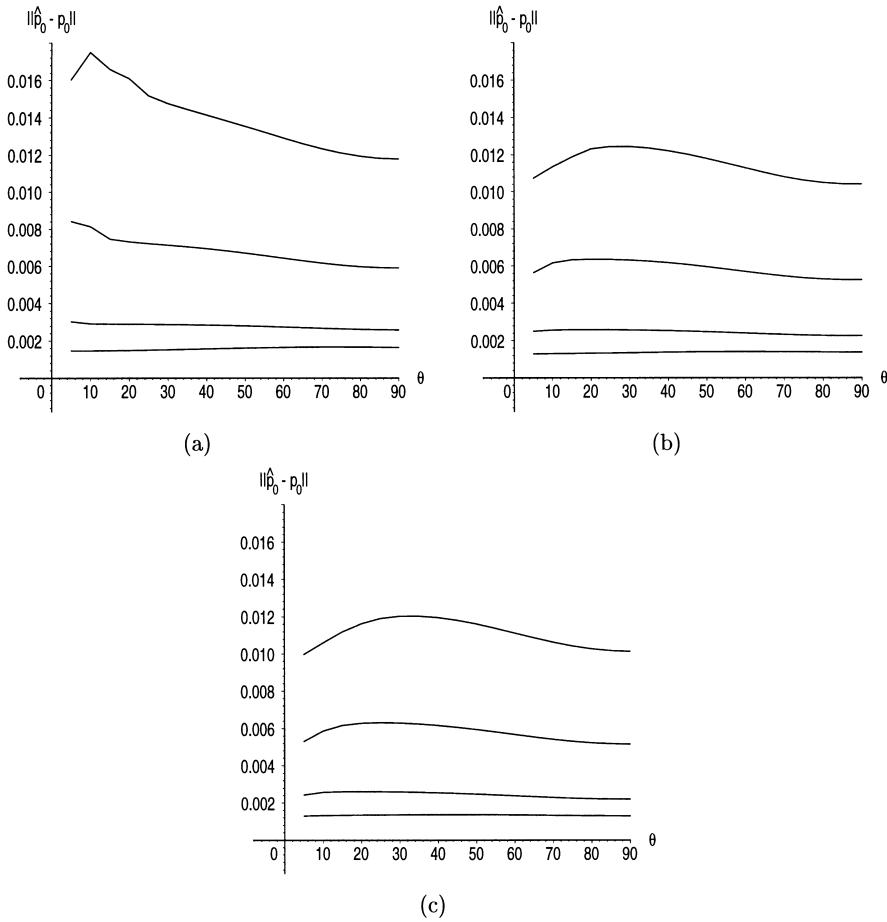


Fig. 5. Average errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) 4, (b) 5, and (c) 6 feature points. θ is the angle formed by the optic axes of each pair of the three cameras. The distance D is fixed at 20. In each diagram, from bottom to top, the graphs are those for which the discretization errors are 1, 2, 5, and 10 pixels.

In the third set of experiments, we investigated how the errors change as a function of the angles between the optic axes of the cameras. We fixed $D = 20$, and let the three optic axes form angles of θ with each other for $\theta = 5^\circ, 10^\circ, \dots, 90^\circ$. The cameras were then given random spins about their optic axes. The rest of the setup was the same as in the first set of experiments, with the feature points forming the vertices of solids inscribed in the unit sphere. Upon adding a test point from the unit sphere, the whole collection of points was translated by a random vector selected from the ball of radius $M = D/4 = 5$. Again, 1000 trials for each value of θ were conducted. The results for four, five, and six feature points are shown in Figs. 5 and 6. The average errors change very little as θ varies, with $\|\hat{\mathbf{p}}_0 - \mathbf{p}_0\|$ being roughly $0.0013K$, $0.0012K$, and $0.0011K$ for four, five, and six feature points, respectively. The maximum errors are close to constant for

$K \leq 2$, while they are erratic and fairly large when $\theta < 30^\circ$ and $K > 2$ for four points. For five and six points, the maximum errors are bounded by $0.0064K$ and $0.0049K$, respectively.

7. Conclusions

We have shown that trifocal tensors that are analogous to that for perspective projection also exist for the weak perspective and paraperspective projections. Since there are only 16 nonzero matrix components to solve for compared to the 27 essential parameters in the perspective trifocal tensor, fewer point correspondences, in fact, four point correspondences across three views instead of seven, are needed in order to determine the trifocal tensor in the weak perspective and paraperspective cases. Thus

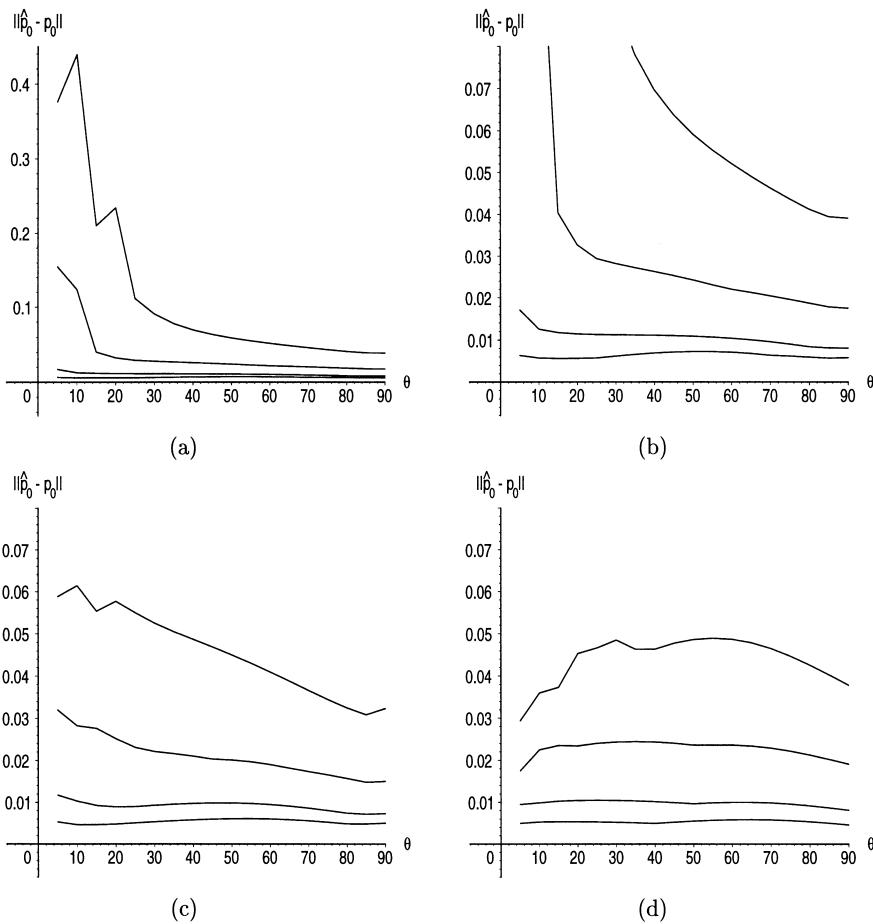


Fig. 6. Maximum errors in the position of an image point \mathbf{p}_0 in the third camera image, as computed by the weak perspective trifocal tensor based on (a) and (b) 4, (c) 5, and (d) 6 feature points. θ is the angle formed by the optic axes of each pair of the three cameras. (b) is a close-up of (a) drawn with the same scale as (c) and (d). The distance D is fixed at 20. In each diagram, from bottom to top, the graphs are those for which the discretization errors are 1, 2, 5, and 10 pixels.

when these approximations to true perspective projection are suitable, these linear algorithms are more likely to be stable than those for true perspective, and to provide better starting points for more accurate nonlinear algorithms for image point determination.

Acknowledgements

We thank Ilan Shimshoni for several useful comments and suggestions.

References

[1] M.E. Spetsakis, A linear algorithm for point and line-based structure from motion, *CVGIP: Image Understanding* 56 (1992) 230–241.

[2] M.E. Spetsakis, J. Aloimonos, A unified theory of structure from motion, *Proceedings of DARPA Image Understanding Workshop*, Pittsburgh, July 1990, pp. 271–283.

[3] A. Shashua, Algebraic functions for recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 17 (1995) 779–789.

[4] Z. Zhang, A tighter lower bound on the Spetsakis-Aloimonos trilinear constraints, *Comput. Vision Image Understanding* 67 (1997) 202–204.

[5] Y. Liu, T.S. Huang, A linear algorithm for motion estimation using straight line correspondences, *Comput. Vision Graphics Image Process.* 44 (1988) 35–57.

[6] M.E. Spetsakis, J. Aloimonos, Closed form solution to the structure from motion problem from line correspondences, *Proceedings AAAI Sixth National Conference on Artificial Intelligence*, Seattle, July 1987, pp. 738–743.

[7] I. Shimshoni, R. Basri, E. Rivlin, A geometric interpretation of weak-perspective motion, *IEEE Trans. Pattern Anal. Mach. Intell.* 21 (1999) 252–257.

- [8] A.M. Bruckstein, R.J. Holt, T.S. Huang, A.N. Netravali, Optimum fiducials under weak perspective projection, *Int. J. Comput. Vision* 35 (1999) 223–244.
- [9] K.M. Heal, M.L. Hansen, K.M. Rickard, *Maple V Learning Guide*, Springer, New York, 1998.
- [10] M.B. Monagan, K.O. Geddes, K.M. Heal, G. Labahn, S.M. Vorkoetter, *Maple V Programming Guide*, Springer, New York, 1998.

About the Author—ALFRED M. BRUCKSTEIN received the B.Sc and M.Sc. Degrees in Electrical Engineering from the Technion, Israel Institute of Technology, in 1977 and 1980, respectively, and the Ph.D. Degree in Electrical Engineering from Stanford University, Stanford, California, in 1984.

In October 1984 he joined the Technion faculty and presently is Professor in the Department of Computer Science, holding Technion's Ollendorff Chair in Sciences.

Professor Bruckstein is a frequent visitor at Bell Laboratories, Lucent Technologies, in Murray Hill, New Jersey. His research interests include image analysis, image processing pattern recognition, ants, robotics and computer graphics. He also has done work in estimation theory, signal processing, algorithmic aspects of inverse scattering, point processes and mathematical models in neurophysiology.

Professor Bruckstein is a member of SIAM, MAA and AMS.

About the Author—ROBERT J. HOLT received the B.S. degree in mathematics from Stanford University, Stanford, California, in 1982 and the Ph.D. Degree in mathematics from the Massachusetts Institute of Technology, Cambridge, Massachusetts, in 1986.

He was Assistant Professor of Mathematics at Northwestern University, Evanston, Illinois, from 1986 to 1988. Since then he worked at Bell Laboratories in Murray Hill, New Jersey, first part of AT&T until 1996, and then part of Lucent Technologies thereafter. He is currently in the Video Communications Research Department of the Multimedia Communications Research Laboratory. His research interests include computer vision, pattern recognition, and image processing.

About the Author—THOMAS S. HUANG received his B.S. Degree in Electrical Engineering from National Taiwan University, Taipei, Taiwan, China; and his M.S. and Sc.D. Degrees in Electrical Engineering from the Massachusetts Institute of Technology, Cambridge, Massachusetts. He was on the Faculty of the Department of Electrical Engineering at MIT from 1963 to 1973; and on the Faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, and Research Professor at the Coordinated Science Laboratory, and Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology.

Dr. Huang's professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 11 books, and over 300 papers in Network Theory, Digital Filtering, Image Processing, and Computer Vision. He is a Fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of American; and has received a Guggenheim Fellowship, an A.V. Humboldt Foundation Senior U.S. Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Acoustics, Speech, and Signal Processing Society's Technical Achievement Award in 1987, and the Society Award in 1991. He is a Founding Editor of the journal *Computer Vision, Graphics, and Image Processing*; and Editor of the Springer Series in Information Sciences, published by Springer-Verlag.

About the Author—ARUN N. NETRAVALI received the B. Tech (Honors) degree from the Indian Institute of Technology, Bombay, India, in 1967 and the M.S. and Ph.D. degrees from Rice University, Houston, Texas, in 1969 and 1970, respectively, all in Electrical Engineering. In 1994, he received an honorary doctorate from the Ecole Polytechnique Federale in Lausanne, Switzerland.

Dr. Netravali is President of Bell Laboratories of Lucent Technologies. He has authored more than 135 technical papers and co-authored three books, and he holds more than 60 patents in the areas of Computer Networks, Human Interfaces to Machines, Picture Processing, and Digital Television.

Dr. Netravali is a Fellow of the American Academy of Arts and Sciences and a member of the United States National Academy of Engineering. Among many awards, he recently received the Japanese C & C Computers and Communications Prize in 1997.