Geotensity Constraint for 3D Surface Reconstruction under Multiple Light Sources

Atsuto Maki and Charles Wiles *

Research and Development Center, TOSHIBA Corporation
1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582 Japan

Abstract. *We tackle the problem of 3D surface reconstruction by a single static camera, extracting the maximum amount of information from gray level changes caused by object motion under illumination by a fixed set of light sources. We basically search for the depth at each point on the surface of the object while exploiting the recently proposed Geotensity constraint [11] that accurately governs the relationship between four or more images of a moving object in spite of the illumination variance due to object motion. The thrust of this paper is then to extend the availability of the Geotensity constraint to the case of multiple point light sources instead of a single light source. We first show that it is mathematically possible to identify multiple illumination subspaces for an arbitrary unknown number of light sources. We then propose a new technique to effectively carry out the separation of the subspaces by introducing the surface interaction matrix. Finally, we construct a framework for surface recovery, taking the multiple illumination subspaces into account. The theoretical propositions are investigated through experiments and shown to be practically useful.*

1 Introduction

3D surface reconstruction of an object has been among the subjects of major interests in computer vision. Given a set of images, in each of which the object is viewed from a different direction, the fundamental issue in extracting 3D information out of 2D images is to match corresponding points in those images so that these points are the projections of an identical point on the surface of the object. For the point correspondence, typically exploited is the constraint that the corresponding parts of the images have equivalent image intensities, regarding the variation in illumination as noise. It has been successfully applied to stereo (see for example [4,5]) where two images are taken simultaneously as the lighting of the object is identical in each image. However, when, as a natural progression, we consider replacing the stereo camera with a single camera observing an object in motion, unfortunately the constraint is nearly always invalid

* Work performed while at Research and Development Center, TOSHIBA Co. The author is presently with Anthropics Technology, Ealing Studios, Ealing Green, Ealing, London W5 5EP, UK.

D. Vernon (Ed.): ECCV 2000, LNCS 1842, pp. 725–741, 2000.

as non-uniform lighting causes the intensity at a specific location on the surface of an object to change as the object moves. Among the few efforts for this issue, whereas *photometric motion* [13] treated the illumination variance due to object motion in terms of optical flow, recently *Geotensity* constraint[1] [11] has been derived to overcome the problem with respect to camera geometry, and to replace the constant intensity constraint. Based on the notion of linear intensity subspaces [14], the Geotensity constraint governs the relationship between four or more images of a moving object, and it can be computed and applied automatically to the task of 3D surface reconstruction. The algorithm for surface reconstruction using Geotensity constraint proceeds basically in two stages. The first stage is to derive the parameters of the Geotensity constraint by analyzing coordinates and image intensities of some sample points on the object in motion. That is, computing structure from motion obtains the geometric parameters of the situation, whereas computing the linear image subspace obtains the lighting parameters of the situation. By combining both sets of parameters we arrive at the Geotensity constraint. Using the same set of images, the second stage is to take each pixel in an arbitrary reference image in turn and search for the depth along the ray from the optical center of the camera passing through the pixel. The depth is evaluated by measuring the agreement of the entire set of projected intensities of a point on the object surface with the Geotensity constraint.

Although the availability of the constraint was limited in principle to the case of a single point light source, the thrust of this paper is to propose a new framework that enables the constraint to be applied to a more general case of multiple point light sources. When multiple light sources exist, computing the lighting parameters in the first stage is not a simple task as was the case with a single light source since most points on the surface are illuminated only by a subset of the light sources and this subset is different for different points. The question is whether we can identify the lighting parameters for different illumination subspace arising from different subset of light sources. Knowing the lighting parameters, we further need to choose appropriate lighting parameters to search for the depth at each point on the object surface.

In this paper, we propose a new method to solve for different illumination subspaces and then to recover the surface of the object. In order to solve for the lighting parameters, in the first stage, we develop a technique to automatically sort the sample points into different clusters according to the illuminating set of light sources. This has been made possible by introducing a matrix representation of property of object surface, which we call the *surface interaction matrix*. The entries of this matrix are computable only from the intensities of the sample points, and transforming it into the canonical form results in segmenting sample points. Once the segmentation is carried out, we simply solve for the lighting parameters individually for each cluster. In the second stage, our mechanism to search for the depth at each point on the surface of the object is to exploit the Geotensity constraint by taking all the possible lighting parameters into account. In principle the proposed constraint is that the correct depth with correct lighting

[1] *Geotensity* stands for "geometrically corresponding pixel intensity."

parameters should best satisfy the Geotensity constraint whereas the agreement with the constraint is measured in the values of pixel intensities in the images themselves. This is important since we employ the Geotensity constraint so that the estimation for surface depth will be optimal in the sense of finding the least square error with respect to image noise (which is generally well understood and well behaved).

Although we will show that our theory is quite general, we will make a number of assumptions in order to present some early results. In particular we will assume that the object has Lambertian surface properties and convex shape (therefore no self-shadowing). Despite such assumptions the key advance is that our formulation allows our results to be obtained fully automatically and for a wide range of well researched stereo algorithms to be applied directly.

2 The Geotensity Constraint

The Geotensity constraint for 3D surface reconstruction was first proposed in [11], applying the notion of linear image basis [14] to object in motion under a single point light source. Here we give a brief summary of the constraint while introducing an additional scheme [10] to explicitly solve for the illuminant direction.

2.1 Issues Respecting Geometry and Intensity

We first consider some issues respecting geometry and intensity that form the basis of the Geotensity constraint. What we initially need is to find some number of corresponding sample points by an independent mechanism. One way to robustly sample proper points between frames is to employ a scheme to extract corners and correlate them automatically while eliminating outliers as seen for example in [1]. For consecutive images sampled at a sufficiently high time frequency we can also use the tracking method of Wiles et al. [17]. Given point correspondence for some sample points, we can derive a constraint on geometry by the coordinates, and also a photometric constraint by observing the intensities on these points.

Solving for Geometry
In this paper, for simplicity, we will concern ourselves with the *affine* and *scaled-orthographic* camera models [12] for projection. Consider the i^{th} world point $\mathbf{X}_i = (X_i, Y_i, Z_i)^\top$ on the surface of an object projected to image point $\mathbf{x}_i(j) = (x_i(j), y_i(j))^\top$ in the j^{th} frame. The affine camera model defines this projection to be

$$\mathbf{x}_i(j) = \mathbf{M}(j)\mathbf{X}_i + \mathbf{t}(j) \ ,$$

where $\mathbf{M}(j)$, an arbitrary 2×3 matrix, and $\mathbf{t}(j)$, an arbitrary 2 vector, encode the motion parameters of the object. The solution to the structure from motion problem using singular value decomposition is well known for this case; given at least four point trajectories $\mathbf{x}_i(j)$ observed through at least two frames

the set $\mathbf{M}(j)$, $\mathbf{X}_i$ and $\mathbf{t}(j)$ can be uniquely recovered up to an arbitrary affine ambiguity [15]. The result is affine structure.

Given the solution to structure from motion using the affine camera model, the Euclidean structure and motion parameters fitting the weak perspective camera model can be recovered [16]. A result of choosing the first frame to be canonical is that the structure vectors have the form, $\mathbf{X}_i = (\mathbf{x}_i^\top(1), Z)^\top$, and we can derive the relationship,

$$\mathbf{x}_i(j) = \mathbf{M}(j) \begin{pmatrix} \mathbf{x}_i(1) \\ Z \end{pmatrix} + \mathbf{t}(j) \ . \tag{1}$$

This relationship effectively describes the *epipolar constraint* between two images.

Solving for Image Intensity

Assuming a static camera and light source, we consider the intensity $I_i(j)$ of the i^{th} point on the surface of a moving object projected into the j^{th} image. For Lambertian surface we can then express $I_i(j)$ in terms of the image formation equation process so that

$$I_i(j) = \max(\mathbf{b}_i{}^\top \mathbf{s}(j), 0) \tag{2}$$

where the maximum operator zeroes negative component [7]. The 3 vector $\mathbf{b}_i$ is the product of the albedo with the inward facing unit normal for the point and the 3 vector $\mathbf{s}(j)$ is the product of the strength of the light source with the unit vector for its direction. The negative components correspond to the shadowed surface points and are sometimes called *attached shadows* [14]. To have no attached shadows, images must be taken with the light source in the *bright cell* (the cell of light source directions that illuminate all points on the object [2]). Note that

$$\mathbf{s}(j) = \mathbf{R}(j)^\top \mathbf{s}(1) \tag{3}$$

where the 3×3 matrix, $\mathbf{R}(j)$, is the rotation of the object from the first canonical frame to the j^{th} frame. Multiplication of $\mathbf{R}(j)^\top$ represents virtually inverse rotation of the light source. The rotation matrix is directly computed from the 2×3 matrix $\mathbf{M}(j)$ that is given above by solving for the structure from motion problem.

Given the correspondence for a small number of n_i pixels through all of n_j images, we record the corresponding pixel intensities, $I_i(j)$, in $\mathbf{I}$, which we call illumination matrix. Then, we can form the matrix equation

$$\mathbf{I} = \mathbf{BS} \tag{4}$$

where $\mathbf{I}$ is a $n_i \times n_j$ matrix containing the elements $I_i(j)$, $\mathbf{B}$ is a $n_i \times 3$ matrix containing the rows $\mathbf{b}_i^\top$, and $\mathbf{S}$ is a $3 \times n_j$ matrix containing the columns $\mathbf{s}(j)$. Equation 4 is then in the familiar form for solution by singular value decomposition to obtain a rank 3 approximation to the matrix $\mathbf{I}$ such that

$$\mathbf{I} = \breve{\mathbf{B}}\breve{\mathbf{S}} \ . \tag{5}$$

In practice we employ RANSAC, a robust random sampling and consensus technique, to ensure that artifacts that are caused by an object not fulfilling the assumed conditions (e.g. specularities and self-shadowing) do not distort the correct solution.

As is well known, the solution is unique up to an arbitrary invertible 3×3 transformation $\mathbf{A}$ and equation 5 is equivalent to

$$\mathbf{I} = \breve{\mathbf{B}}\breve{\mathbf{S}} = (\breve{\mathbf{B}}\mathbf{A}^{-1})(\mathbf{A}\breve{\mathbf{S}}) \ . \tag{6}$$

Therefore, $\breve{\mathbf{S}}$ by arbitrary decomposition contains the information of the light source up to an arbitrary transformation $\mathbf{A}$. Although the matrix $\breve{\mathbf{S}}$ will suffice for surface reconstruction with a single light sources, the parameterization using the lighting vectors is useful in an environment with multiple light sources.

Estimating the Light Source Direction

The problem of estimating the light source direction is attributed to that of solving for matrix $\mathbf{A}$ which transforms $\breve{\mathbf{S}}$ into $\mathbf{S}$ by

$$\mathbf{S} = \mathbf{A}\breve{\mathbf{S}} \ , \tag{7}$$

or for each column $\breve{\mathbf{s}}(j)$ into $\mathbf{s}(j)$ by

$$\mathbf{s}(j) = \mathbf{A}\breve{\mathbf{s}}(j) \tag{8}$$

where $\breve{\mathbf{s}}(j)$ denotes columns of $\breve{\mathbf{S}}$. Substituting equation 8 into equation 3, we have the relation,

$$\mathbf{A}\breve{\mathbf{s}}(j) = \mathbf{R}(j)^{\top}\mathbf{A}\breve{\mathbf{s}}(1) \ , \tag{9}$$

which provides a homogeneous system; three equations for each reference image. In order to solve for the nine elements of $\mathbf{A}$ as a well-posed problem, it is necessary to have reference information so that the resulting light source direction fits to the first canonical input frame. Observing $\mathbf{b}_i$ at an arbitrary sample point i for which the light source is in the bright cell throughout the input images, and substituting equation 8 to equation 2 for this point, we have

$$I_i(j) = \mathbf{b}_i^{\top}\mathbf{A}\breve{\mathbf{s}}(j) \ . \tag{10}$$

Since $\breve{\mathbf{s}}(j)$ is obtained by the singular value decomposition of matrix $\mathbf{I}$ and the set of intensities, $I_i(j)$, is also known, equation 10 provides an additional constraint to matrix $\mathbf{A}$ for each reference image. With the constraints of equation 9 and equation 10 matrix $\mathbf{A}$ can be solved using a minimum of three input images. Once matrix $\mathbf{A}$ is solved, by multiplying $\mathbf{A}$ with $\breve{\mathbf{S}}$ we obtain the explicit light source matrix $\mathbf{S}$ containing the columns $\mathbf{s}(j)$.

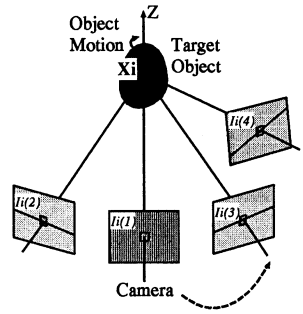


Fig. 1. Geotensity constraint. The intensity of world point $\mathbf{X}_i$ projected into the first image, $I_i(1)$, is represented by a unique linear combination of the intensities of the same point projected in the other three images, $I_i(2) \cdots I_i(4)$ for all points i.

2.2 Depth by the Geotensity Constraint

The term Geotensity constraint accounts for a constraint between four or more images of an object from different views under static lighting conditions. This concept is schematically depicted in Figure 1 by replacing object motion with a coherent motion of the camera and the light source.

The conditions for applying the Geotensity constraint to depth reconstruction are as follows: (i) The scene consists of a single moving object. (ii) The object has Lambertian surface properties and is convex (therefore no self-shadowing) while the surface may or may not be textured. (iii) There is a single distant light source. However, the condition (iii) will be relaxed in Section 3 and 4.

Evaluating the Set of Intensities
At each pixel, $\mathbf{x}$, in the first image, to search for the depth Z we can recall equation 1 for the geometric constraint imposed on a sequence of images so that

$$I(j;\mathbf{x},Z) = I[\mathbf{x}(j)] = I[\mathbf{M}(j)\begin{pmatrix}\mathbf{x}(1)\\ Z\end{pmatrix} + \mathbf{t}(j)] \ . \tag{11}$$

$I(j;\mathbf{x},Z)$ indicates the set of image intensities in the j^{th} frame at the coordinates determined by $\mathbf{x}$ in the first image, guess of depth Z, and the motion parameters $\mathbf{M}(j)$ and $\mathbf{t}(j)$. The task is now to evaluate the set of intensities $I(j;\mathbf{x},Z)$. When full Euclidean lighting conditions have been recovered in advance so that $\mathbf{s}(j)$ is known, with four images we define $\hat{\mathbf{b}}^\top$ as

$$\hat{\mathbf{b}}^\top = [I(1)\ I(2)\ I(3)\ I(4)]\,\mathbf{S}^\top(\mathbf{S}\mathbf{S}^\top)^{-1} \tag{12}$$

where $\mathbf{S}$ is a 3×4 matrix containing the columns $\mathbf{s}(j)(j = 1, 2, 3, 4)$. For a single light source with all images taken with the light source in the bright cell, the estimated values of the intensities are then

$$\hat{I}(j;\mathbf{x},Z) = \hat{\mathbf{b}}^\top\mathbf{s}(j) \ . \tag{13}$$

It should be noted that exactly the same estimation of $\hat{I}(j;\mathbf{x},Z)$ is available also in the case that the light source direction is determined only up to the ambiguity. This is easily confirmed by substituting equation 7 to equation 12 and then equation 8 to 13, where matrix $\mathbf{A}$ turns out to be canceled. Estimating $\hat{I}(j;\mathbf{x},Z)$ by equation 13, we can define the error function to evaluate the set of intensities $I(j;\mathbf{x},Z)$ as

$$E(\mathbf{x},Z)=\sum_{j}(I(j;\mathbf{x},Z)-\hat{I}(j;\mathbf{x},Z))^2 \ . \tag{14}$$

Computing the Depth
At each pixel, $\mathbf{x}$, in the first image we measure the error, E, in the Geotensity constraint at regular small intervals of depth, Z. When the depth is correct we expect the error to approach zero and when it is incorrect we expect the error to be large. The Geotensity constraint can be stated simply as

$$E(\mathbf{x},Z)=0 \ . \tag{15}$$

It is clear that as the depth parameter is varied the location of the corresponding points in each image will trace out the corresponding epipolar line in each image. We then choose such depth Z that minimizes the error $E(\mathbf{x},Z)$ as the depth estimate.

3 Estimation of Multiple Light Sources

Our discussion in the previous section has been limited largely to the case of a single light source. In this section, we extend this by describing a scheme to compute surface depth using the Geotensity constraint for the case of multiple light sources. If all the light sources illuminated all the points on the surface of the object then we could treat the combined light sources as an equivalent single light source and carry out computation of this vector from measurements as before. Unfortunately, most (if not all) points on the surface will be illuminated by a subset of the light sources and this subset will be different for different points. Therefore, the illumination matrix $\mathbf{I}$ will contain intensities derived from different subsets of light sources.

The basic idea we propose for the case with multiple light sources is to first sort the rows of the illumination matrix $\mathbf{I}$ into submatrices, each of which contains intensities derived from an identical subset of light sources. We call each such submatrix an illumination submatrix. Once the segmentation is carried out, we estimate the direction of combined light sources for each subset, applying the technique described in Section 2.1 individually to each illumination submatrix.

3.1 Illumination Submatrix

Although we describe the algorithm mainly for the case of two point light sources for simplicity, the proposed algorithm will turn out to be applicable to the general case of an arbitrary unknown number of light sources. For illustration

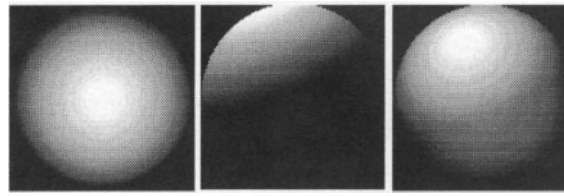


Fig. 2. Synthetic sphere (128 × 128). Left: The first point light source is at infinity in the same direction to the viewing point. Middle: The second in the top-left direction (0.37, - 0.92, -0.18). Right: The sphere is illuminated by both of the point light sources.

we use the synthetic sphere shown in Figure 2. The first point light source illuminating the entire semi-sphere is placed at infinity in the same direction to the viewing point, whereas the second is also at infinity but in the top-left direction causing an attached shadow. For convenience we call the lighting vectors $\mathbf{l}_1$ and $\mathbf{l}_2$, respectively, and the combined lighting vector $\mathbf{l}_0$. I.e, $\mathbf{l}_0 = \mathbf{l}_1 + \mathbf{l}_2$. Thus $\mathbf{l}_0$ illuminates the top-left area while the rest of the area is illuminated by $\mathbf{l}_1$ only. In this example there is no area that is illuminated only by $\mathbf{l}_2$ and we ignore it without loss of generality.

Suppose that n_0 points out of n_i samples stay illuminated by $\mathbf{l}_0$ throughout the image sequence and n_1 points only by $\mathbf{l}_1$. And given that somehow we know the classification of sample points (the method for the classification is described in Section 3.2), we could permute the rows of illumination matrix $\mathbf{I}$ in such a way that each illumination submatrix $\mathbf{I}_l(l = 0, 1)$ respectively contains $n_l(l = 0, 1)$ sample points due to each lighting vector. Analogous to equation 4 in the case of a single point light source, each illumination submatrix could then be rewritten as

$$\mathbf{I}_l \simeq \mathbf{B}_l \mathbf{S}_l \quad (l = 0, 1) \tag{16}$$

and the illumination matrix can be represented in its canonical form $\bar{\mathbf{I}}$:

$$\bar{\mathbf{I}} = \begin{pmatrix} \mathbf{I}_0 \\ \mathbf{I}_1 \end{pmatrix} \simeq \begin{pmatrix} \mathbf{B}_0 & \\ & \mathbf{B}_1 \end{pmatrix} \begin{pmatrix} \mathbf{S}_0 \\ \mathbf{S}_1 \end{pmatrix} = \bar{\mathbf{B}}\bar{\mathbf{S}}, \tag{17}$$

where $n_l \times 3$ matrices $\mathbf{B}_l$ are submatrices of $\mathbf{B}$ containing the rows corresponding to sample points illuminated by $\mathbf{l}_l$. For convenience we use the notations of

$$\bar{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_0 & \\ & \mathbf{B}_1 \end{pmatrix}, \ \bar{\mathbf{S}} = \begin{pmatrix} \mathbf{S}_0 \\ \mathbf{S}_1 \end{pmatrix}. \tag{18}$$

In the following we discuss the acquisition of the above representation $\bar{\mathbf{I}}$, which essentially requires the knowledge about the classification of the sample points so that each row of $\mathbf{I}_l(l = 0, 1)$ contains image intensities generated by a common subset of light sources. Now equations 16 and 17 are not strictly defined as equations because each sample point does not necessarily stay illuminated by a unique subset of light sources throughout the input frames. However, assuming that a majority of the sample points will do, we first include those points illuminated by different sets of light sources in different frames in either of the

illumination submatrices, and then exclude them as outliers by RANSAC prior to the process of factorization in equation 16.

3.2 Segmentation of Illumination Subspace

As is obvious in equation 17, $\mathbf{I}_0$ and $\mathbf{I}_1$ are at most rank 3 and thus full rank of $\bar{\mathbf{I}}$ or unsorted $\mathbf{I}$ is in total 6. In order to consider the non-degenerate case, we assume $n_j \geq 6$, i.e. we utilize a minimum of 6 frames. When we consider an arbitrary row of $\mathbf{I}$, $\mathbf{p}_\alpha^\top$, corresponding to an arbitrary sample point, α, and regard it as representing an n_j dimensional vector whose elements are the pixel intensities of point α observed throughout n_j frames, vector $\mathbf{p}_\alpha^\top$ represents a point in n_j dimensional illumination space R^{n_j}. Therefore, if $\mathbf{p}_\alpha$ belongs to $\mathbf{I}_0$, we have

$$\mathbf{p}_\alpha^\top = \mathbf{b}_\alpha^\top \mathbf{S}_0 \tag{19}$$

where $\mathbf{p}_\alpha^\top$ represents a point in 3-D illumination subspace $\mathcal{L}_0$ spanned by the three row vectors of $\mathbf{S}_0$ with the surface vector $\mathbf{b}_\alpha$ as coefficients. Analogously, each row of $\mathbf{I}_0$ represents a point in a 3-D subspace $\mathcal{L}_0$, and each row of $\mathbf{I}_1$ a point in another 3-D subspace $\mathcal{L}_1$. Therefore, the classification problem of n_i sample points is attributed to segmenting a group of n_i points in R^{n_j} into two different 3-D subspaces. In the following, we first show that the segmentation is mathematically possible on the basis of a theorem, and then propose a technique to practically carry out the segmentation by introducing the *surface interaction matrix*.

Mathematical Background
In the illumination space R^{n_j} under two point light sources we define an $n_i \times n_i$ metric matrix $\mathbf{G} = (G_{\alpha\beta})$ as[2]

$$\mathbf{G} = \mathbf{I}\,\mathbf{I}^\top, \quad G_{\alpha\beta} = (\mathbf{p}_\alpha, \mathbf{p}_\beta) \tag{20}$$

where α and β are indices of two arbitrary sample points, and $\mathbf{p}_\alpha$ and $\mathbf{p}_\beta$ are corresponding rows of $\mathbf{I}$. By definition this is a rank 6 positive semi-definite symmetric matrix [8]. We denote the eigenvalues of $\mathbf{G}$ as $\lambda_1 \geq, ..., \geq \lambda_{n_i}$ (only the first six values are non-zero) and an orthonormal system[3] of corresponding eigenvectors as $\{\mathbf{v}_1, ..., \mathbf{v}_{n_i}\}$. Let us define an $n_i \times n_i$ function matrix $\mathbf{H} = (H_{\alpha\beta})$ with the first six eigenvectors as

$$\mathbf{H} = \sum_{i=1}^{6} \mathbf{v}_i \mathbf{v}_i^\top \ . \tag{21}$$

Then, according to the general theorem for subspace segmentation (see Appendix) we have the following theorem as a special case where the number of subspaces is two ($m = 2$);

[2] We write the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$ as ($\mathbf{a}$,$\mathbf{b}$).
[3] By orthonormal we mean that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ ($i \neq j$) and $\mathbf{v}_i \cdot \mathbf{v}_j = 1$ ($i = j$) for arbitrary set of i and j.

Theorem 1 If $\mathbf{p}_\alpha \in \mathbf{I}_0$ and $\mathbf{p}_\beta \in \mathbf{I}_1$, then $H_{\alpha\beta} = 0$.

Theorem 1 implies the possibility of the classification, i.e. for each pair of sample points, α and β, we can judge if they are illuminated by an identical subset of light sources by computing $H_{\alpha\beta}$. If the value is non-zero, they belong to the same subset of light sources, and if they do not belong to the same subset of light sources, the value is zero.

Surface Interaction Matrix

Based on Theorem 1 we propose to carry out the task of classification systematically for the entire set of sample points. Applying singular value decomposition directly to the unsorted illumination matrix, $\mathbf{I}$, we obtain the familiar form of approximation as $\mathbf{I} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. Matrix $\mathbf{\Sigma}$ is a diagonal matrix consisting of the six greatest singular values whereas $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular matrices, respectively, such that $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{E}_{6\times 6}$ (the identity matrix). Since $\mathbf{U}$ contains eigenvectors of $\mathbf{I}\,\mathbf{I}^\top$, we can compute the $n_i \times n_i$ function matrix $\mathbf{H}$ using its first six columns,

$$\mathbf{H} = \mathbf{U}\mathbf{U}^\top, \tag{22}$$

and we call it the "surface interaction matrix"[4] as it preserves the interactive property of object surface with the light source direction. This matrix $\mathbf{H}$ is by definition computable uniquely from the illumination matrix $\mathbf{I}$. Also, permuting rows of $\mathbf{I}$ does not change the set of values $H_{\alpha\beta}$ that appear in $\mathbf{H}$ though their arrangement in $\mathbf{H}$ does; swapping rows α and β of $\mathbf{I}$ results in swapping corresponding rows α and β of $\mathbf{U}$. Therefore, it results in simultaneously swapping rows α and β and columns α and β in $\mathbf{H}$, but not their entry value.

Since the set of values does not change, to reveal the relevance of Theorem 1, we investigate the character of $\bar{\mathbf{H}}$, the surface interaction matrix for the canonical illumination matrix $\bar{\mathbf{I}}$. Factorizing each illumination submatrix $\mathbf{I}_l (l = 0, 1)$ of $\bar{\mathbf{I}}$ in equation 17 by singular value decomposition in a similar way to above, we have

$$\mathbf{I}_l = \mathbf{U}_l\mathbf{\Sigma}_l\mathbf{V}_l^\top = (\mathbf{U}_l\mathbf{\Sigma}_l^{1/2}\mathbf{A}_l)(\mathbf{A}_l^{-1}\mathbf{\Sigma}_l^{1/2}\mathbf{V}_l^\top), \quad (l = 0, 1) \tag{23}$$

where $\mathbf{A}_l$ represents an arbitrary invertible 3×3 matrix. Denoting also

$$\bar{\mathbf{U}} = \begin{pmatrix} \mathbf{U}_0 & \\ & \mathbf{U}_1 \end{pmatrix}, \quad \bar{\mathbf{\Sigma}} = \begin{pmatrix} \mathbf{\Sigma}_0 & \\ & \mathbf{\Sigma}_1 \end{pmatrix}, \quad \bar{\mathbf{V}} = \begin{pmatrix} \mathbf{V}_0^\top \\ \mathbf{V}_1^\top \end{pmatrix}, \quad \bar{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_0 & \\ & \mathbf{A}_1 \end{pmatrix},$$

we have another factorized form of $\bar{\mathbf{I}}$ as

$$\bar{\mathbf{I}} = (\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}^{1/2}\bar{\mathbf{A}})(\bar{\mathbf{A}}^{-1}\bar{\mathbf{\Sigma}}^{1/2}\bar{\mathbf{V}}^\top) \ . \tag{24}$$

Comparing the first term of equation 24 to that of equation 17 we obtain

$$\bar{\mathbf{B}} = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}^{1/2}\bar{\mathbf{A}} \ . \tag{25}$$

[4] The definition has been inspired by the work of Costeira and Kanade [3] that has proposed the "shape interaction matrix" and applied it successfully to the multibody structure-from-motion problem.

Substituting the relation in equation 25 to 22,

$$\begin{aligned}\bar{\mathbf{H}} &= \bar{\mathbf{U}}\bar{\mathbf{U}}^{\top} \\ &= \bar{\mathbf{B}}\bar{\mathbf{A}}^{-1}\bar{\mathbf{\Sigma}}^{-1}\bar{\mathbf{A}}^{-\top}\bar{\mathbf{B}}^{\top} \\ &= \bar{\mathbf{B}}(\bar{\mathbf{A}}^{\top}\bar{\mathbf{\Sigma}}\bar{\mathbf{A}})^{-1}\bar{\mathbf{B}}^{\top} \\ &= \bar{\mathbf{B}}(\bar{\mathbf{B}}^{\top}\bar{\mathbf{B}})^{-1}\bar{\mathbf{B}}^{\top}. \end{aligned} \tag{26}$$

Further, substituting equation 18 to 26,

$$\begin{aligned}\bar{\mathbf{H}} &= \begin{pmatrix}\mathbf{B}_0 & \\ & \mathbf{B}_1\end{pmatrix}\begin{pmatrix}(\mathbf{B}_0^{\top}\mathbf{B}_0)^{-1} & \\ & (\mathbf{B}_1^{\top}\mathbf{B}_1)^{-1}\end{pmatrix}\begin{pmatrix}\mathbf{B}_0^{\top} & \\ & \mathbf{B}_1^{\top}\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{B}_0(\mathbf{B}_0^{\top}\mathbf{B}_0)^{-1}\mathbf{B}_0^{\top} & \\ & \mathbf{B}_1(\mathbf{B}_1^{\top}\mathbf{B}_1)^{-1}\mathbf{B}_1^{\top}\end{pmatrix}. \end{aligned} \tag{27}$$

This means that the canonical $\bar{\mathbf{H}}$ matrix to the sorted $\bar{\mathbf{I}}$ has a very defined block-diagonal structure as can be expected by Theorem 1. Thus, each block of $\bar{\mathbf{H}}$ provides important information; sample points illuminated by a common set of light sources belong to an identical block in $\bar{\mathbf{H}}$.

Segmentation in Practice

The problem of segmenting the sample points under different subsets of light sources now has been reduced to sorting the entries of matrix $\mathbf{H}$ into $\bar{\mathbf{H}}$ by swapping pairs of rows and columns until it becomes block diagonal. Once the sorting is achieved, the corresponding permutation of rows of $\mathbf{I}$ will transform it to its canonical form, $\bar{\mathbf{I}}$, where sample points under common subset of light sources are grouped into contiguous rows. We can then derive the light source matrix $\mathbf{S}_0$ and $\mathbf{S}_1$ from each submatrix of $\bar{\mathbf{I}}$ independently by the same technique used in the case of a single light source.

In practical segmentation, with presence of points illuminated by different sets of light sources in different frames, perfect diagonalization may not be possible. As stated earlier, however, it is not crucial either since those points will anyway be excluded by RANSAC in computation of lighting parameters. It is important that the borders between the blocks are roughly found so that each block contains a sufficient number of sample points to compute the lighting parameters. With noisy measurements, a pair of sample points under different subsets of light sources may exhibit a small non-zero entry in $\mathbf{H}$. We can regard $H_{\alpha\beta}^2$ as representing the energy of the surface interaction, and the block diagonalization of $\mathbf{H}$ can be achieved by minimizing the total energy of all possible off-diagonal blocks over all sets of permutations of rows and columns of $\mathbf{H}$. A simple iterative minimization procedure suffices for our purpose, but more efficient approaches such as using the hill-climbing method can be employed as proposed in [3] for sorting the shape interaction matrix. When the search over the set of permutations is explosive, an approach based on genetic algorithm can be efficiently adapted.

3.3 Generalized Segmentation of Illumination Subspace

The discussion for the two points light sources applies to the case with an arbitrary number of light sources (see Appendix). In the general case where m subsets of the light sources exist, basically a minimum of $3m$ frames will be necessary. Although the number of the blocks, namely the number of illumination subspaces, m, that reside on the surface of the object will be required prior to the segmentation, it can also be identified by computing the rank r of the illumination matrix $\mathbf{I}$ since $r = 3m$.

Regarding the condition to have non-degenerate $\mathbf{I}$, the sample points in each lighting classification should have surface normals that span 3D. Also, the rotation axis with respect to the object motion should not coincide with the light source direction, nor be coplanar throughout the input frames.

4 Surface Reconstruction

In this section we describe our algorithm for dense surface recovery under multiple light sources. With the light source matrix $\mathbf{S}_0$ and $\mathbf{S}_1$ acquired using the technique discussed in the previous section, we can exploit equations 12, 13 and 14. Although $\mathbf{S}_0$ and $\mathbf{S}_1$ are valid as they are insofar as the surface stays illuminated by an identical set of light sources throughout the image sequence, it is not the case in general. The fundamental difficulty is in dealing with the subset of light sources that is not constant throughout the frames, different for each surface point, and unavailable in advance. For each surface point, i, nevertheless, a specific light source matrix must exist and be formed by an appropriate combination of $\mathbf{S}_0$ and $\mathbf{S}_1$. We denote such a matrix $\mathbf{S}^i$. Each column of $\mathbf{S}^i$ is either from $\mathbf{S}_0$ or $\mathbf{S}_1$ depending on which subset of light sources illuminates point i in each frame. Hence, the number of possible candidates for $\mathbf{S}^i$ is m^{n_j} where m is the number of the possible subsets of the light sources. In order to search for the depth Z, correct error in the Geotensity constraint defined by equation 14 should be measured with the light source matrix $\mathbf{S}^i$. Since multiple candidates exist for $\mathbf{S}^i$, we measure the error using all the possible candidates at each image point, $\mathbf{x}$. As in the case of a single light source, we expect the error to approach zero when the depth is correct and the error to be large when the depth is incorrect. We also expect the error to be large when an incorrect candidate of $\mathbf{S}^i$ is used. Because we compute the error for all the candidates of $\mathbf{S}^i$ at regular small intervals of depth, we regard the smallest error as $E(\mathbf{x}, Z)$ and choose as the depth estimate such depth Z that minimizes the error $E(\mathbf{x}, Z)$. Including the steps to compute the light source matrices, the algorithm for estimating the depth can be summarized as:

1° Decompose the illumination matrix $\mathbf{I}$ using SVD and yield $\mathbf{I} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ and compute the rank $r = 3m$.
2° Using the first r columns of $\mathbf{U}$, compute the surface interaction matrix $\mathbf{H} = \mathbf{U}\mathbf{U}^\top$ and block-diagonalize it.
3° For each of m blocks in $\mathbf{H}$, permute matrix $\mathbf{I}$ into submatrices, $\mathbf{I}_l (l = 1, ..., m)$, and compute $\mathbf{S}_l$ accordingly.

4° At point $\mathbf{x}$, measure $I(j; \mathbf{x}, Z)$ using equation 11 for a particular guess of depth Z.
5° Estimate $\hat{\mathbf{b}}^\top$ with $I(j; \mathbf{x}, Z)$ by equation 12 considering all the combinations of $\mathbf{S}_l$ for $\mathbf{S}$, and then $\hat{I}(j; \mathbf{x}, Z)$ by equation 13.
6° Computing the error $E(\mathbf{x}, Z)$ by equation 14, choose such depth Z that minimizes $E(\mathbf{x}, Z)$ as the depth estimate.

5 Experiments

In order to investigate the performance of the proposed algorithms, we first utilize the synthetic sphere shown in Figure 2 as an example of an input object that satisfies the required assumptions. The synthetic sphere is illuminated by two point light sources at infinity. The upper-left part is illuminated by both of the light sources whereas the rest of the surface is illuminated only by one of them. Since the appearance of the sphere would not change as it rotates, we acquire several identical input images. As we need some point correspondence to solve for the parameters of imaging geometry as well as the lighting, in the simulation we give some preliminary sample points arbitrarily on the sphere as shown in Figure 3 and map the coordinates from one frame to the other while giving some rotation to the sphere.

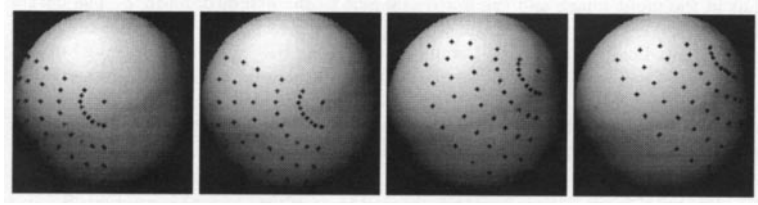


Fig. 3. Synthetic sphere (128 × 128) illuminated by two point light sources at infinity. Corresponding coordinates are marked by the cross. 4 images are shown out of 8 used.

Figure 4 shows the surface interaction matrix $\mathbf{H}$ computed for the 46 sample points on the sphere surface and its diagonalised form $\tilde{\mathbf{H}}$. It is observed in the diagonalized form that the sample points can be divided roughly into two clusters. It is also observed that the result involves non-zero off-diagonal elements due to some sample points which do not exactly belong to either of the clusters while illuminated by different sets of light sources in different frames. However, those points are excluded as outliers by RANSAC. In this example, a total of 17 points are excluded whereas the entire set of points are divided into 21 and 25 points in each cluster. From each cluster we then compute the light source direction illuminating the corresponding area of the sphere surface.

The surface structure of the sphere was computed by measuring the depth at every pixel in the first image. For illustration a vertical scan line in the middle of the sphere ($x = 64$) is first examined in detail. We compute the error $E(Z)$ by equation 14 at each assumed depth Z for each pixel on the scan line to generate

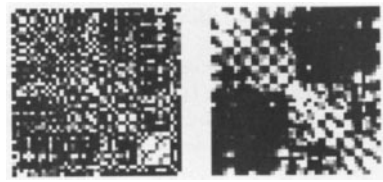
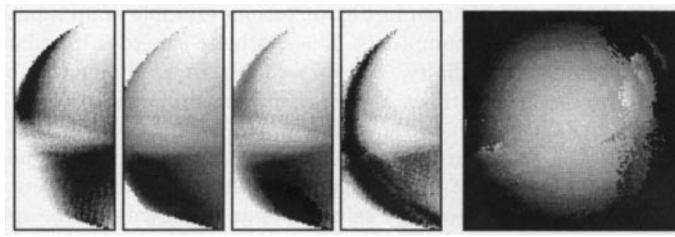


Fig. 4. Left: The surface interaction matrix $\mathbf{H}$ before sorting. Right: Diagonalized surface interaction matrix $\bar{\mathbf{H}}$. The lighter the gray scale, the greater is the value.



(a) (b) (c) (d) (e)

Fig. 5. The error map by the Geotensity constraint. The vertical axis corresponds to that in the input image and the horizontal axis to the assumed depth Z. The darker the gray scale is, the smaller is the error $E(Z)$. The gray scale is histogram-equalized. (a) Computed only using $\mathbf{s}_0$. (b) Computed only using $\mathbf{S}_1$. (c) Computed using both $\mathbf{S}_0$ and $\mathbf{S}_1$. (d) Computed taking all the combination of columns in $\mathbf{S}_0$ and $\mathbf{S}_1$ into account. (e) The surface depth map. The lighter, the closer. The estimate in the extreme-right part should be ignored since due to the rotation that part of the sphere is invisible in most of the input image sequence.

an error map for the line. Figure 5 shows the resulting error map where $\mathbf{S}_0$ and $\mathbf{S}_1$ are utilized interchangeably in order to clarify their relevance. The darker the image intensity, the smaller is the error $E(Z)$. The gray scale is histogram-equalized so that regions with smaller error can be studied more easily. It should be noted that $E(Z)$ has been computed without using a template and only by referring to a pixel intensity in each frame. By searching for the depth Z along the horizontal axis to find the point where the error $E(Z)$ approaches zero and then tracing this Z through each vertical coordinate, a curve that reflects a continuous estimation of depth should be generated.

Figure 5 (a) and (b) shows maps of $E(Z)$ computed using only $\mathbf{S}_0$ or $\mathbf{S}_1$, respectively. Correct shape of the surface is estimated only for the upper part of the scan line in Figure 5 (a), and in (b) only for the lower part. This is expected since $\mathbf{S}_0$ and $\mathbf{S}_1$ are valid as they are only where the surface stays illuminated by the corresponding set of light sources. Figure 5 (c) shows another map obtained using both $\mathbf{S}_0$ and $\mathbf{S}_1$ and choosing the smaller $E(Z)$. A reasonable result is obtained for the upper and lower part of the scan line where either $\mathbf{S}_0$ or $\mathbf{S}_1$ is valid. However, the result for the part in between those two parts is not as desired since the correct set of light sources for this part should be some combination of $\mathbf{S}_0$ and $\mathbf{S}_1$. Finally, Figure 5 (d) shows the map obtained by considering all

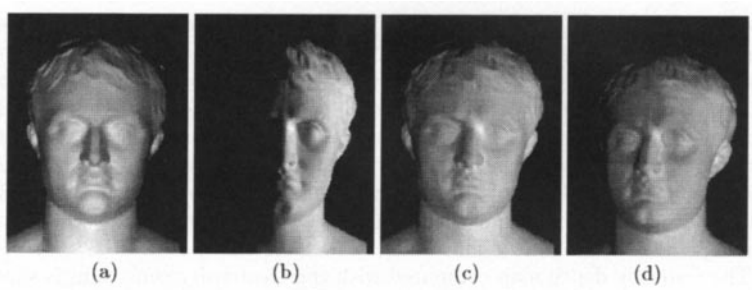
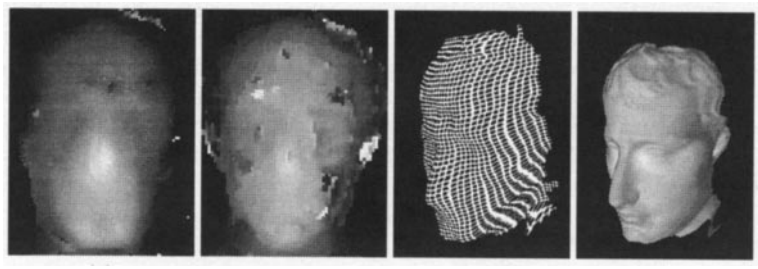


Fig. 6. Statue of Julius Caesar. (a) Illuminated by a point light source in the viewing direction. (b) Illuminated by another point light source placed at the right-hand side. (c) Illuminated by both of the point light sources. (d) Illuminated by both of them but in a different pose.



(a) (b) (c) (d)

Fig. 7. 3-D reconstruction of Julius Caesar. (a) The depth map by the Geotensity constraint. The lighter, the closer. (b) The depth map by correlation. (c) The recovered surface by the Geotensity constraint shown with mesh. (d) The recovered surface as in (c) but the surface is texture-mapped.

the combinations of $\mathbf{S}_0$ and $\mathbf{S}_1$ as the candidates of the light source matrix $\mathbf{S}$. Although the error map appears to be more noisy since different combinations of $\mathbf{S}_0$ and $\mathbf{S}_1$ are taken into account, the error is minimized at correct depth for the entire vertical coordinates. Figure 5 (e) shows the result obtained by performing the procedure for the entire surface. In the extreme-right part of the sphere the estimation is not properly obtained since that part of the sphere does not stay visible throughout the input images due to its rotation. Also, some noise appears where an incorrect combination of $\mathbf{S}_0$ and $\mathbf{S}_1$ happens to minimize the error $E(Z)$ at incorrect depth. As a whole, however, it is observed that the sphere surface is effectively recovered by investigating the Geotensity constraint under multiple light sources.

We have applied the scheme also to a statue of Julius Caesar shown in Figure 6. The images of Julius Caesar were taken under two point light sources placed in different directions, one in the viewing direction and the other at the right-hand side, both about two meters away. Figure 6 (a) and (b) illustrate the image radiance due to each point light source whereas Figure 6 (c) shows that produced by both of the light sources. Seven other images of the statue, each

obtained in a different pose, were used, and thus, a total of eight images were used in the experiment. Figure 6 (d) shows one of them. As can be observed, one of the light sources only illuminates part of the statue and the part varies due to the motion of Julius Caesar. For instance the central part of the forehead is illuminated by both of the light sources in Figure 6 (c) but by only one of them in Figure 6 (d). In this way the appearance of the statue changes dramatically, making the correspondence problem very difficult for conventional methods such as using a constant intensity constraint.

The resulting depth map computed with the Geotensity constraint is shown in Figure 7 (a). For the purpose of comparison, a depth map computed in the same framework but using cross-correlation is shown in Figure 7 (b). In both cases we used a 15×15 template for the search to suppress the error arising from the image noise. Obviously, less error is involved in the estimation by the Geotensity constraint, especially in the area where the surface has little texture such as around the forehead and cheek. It implies the advantage of the Geotensity constraint that basically works regardless of the surface albedo. Figures 7 (c) shows recovered surface with mesh to assess the depth accuracy. The recovered surface is smoothed using Gaussian operator (with standard deviation $\sigma = 1.0$). Figures 7 (d) shows the surface with texture. The result demonstrates the validity of the proposed scheme for a real object.

6 Summary and Discussion

For the problem of 3D object surface reconstruction by a single static camera, we have considered to extend the Geotensity constraint to the case of multiple light sources. We have first shown that it is mathematically possible to sort the sample points into different clusters according to the subset of relevant light sources. Introducing the surface interaction matrix as a technique to carry out the task of practical segmentation, we have proposed that the object surface be computed by solving for the combined light source direction from each cluster and taking the combinations of different sets of light sources into account. Although the demonstration has been limited to the case of two point light sources, the algorithm is in principle applicable to the general case of an arbitrary unknown number of light sources. When a large number of subsets of light sources is involved in the presence of noise, a practical mechanism would be required in order to estimate the clustering of sample points and also reduce the search for the combination of light sources. Future work will be directed at developing such a mechanism, e.g. by employing a statistical approach for the grouping of points as proposed in [6], and also at extending the algorithm to treat more general lighting conditions, including shadowing or inter-reflection.

A Appendix: Subspace Segmentation

Consider N points $\mathbf{p}_\alpha$ in n-dimensional space R^n ($\alpha = 1, ..., N$) and decompose the group of indices $\mathcal{I} = \{1, ..., N\}$ into m subgroups as:

$$\mathcal{I}_1 \cup \cdots \cup \mathcal{I}_m = \mathcal{I}, \;\; \mathcal{I}_1 \cap \cdots \cap \mathcal{I}_m = \emptyset.$$

Define a rank r_k subspace $\mathcal{L}_k$ spanned by the k-th group $\mathbf{p}_\alpha(\alpha \in \mathcal{I}_k)$. When m subspaces $\mathcal{L}_k(k = 1, ..., m)$ are linearly independent, $\mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_m$ represents an r-dimensional subspace of R^n so that $r = \sum_{k=1}^{m} r_k$. Assuming $N \geq r$, let us define an $N \times N$ metric matrix $\mathbf{G} = (G_{\alpha\beta})$ as

$$G_{\alpha\beta} = (\mathbf{p}_\alpha, \mathbf{p}_\beta) \ , \tag{28}$$

where $(\beta = 1, ..., N)$ is an index to another sample point $\mathbf{p}_\beta$ in R^n. By definition this is a rank r positive semidefinite symmetric matrix [8]. We denote the eigenvalues as $\lambda_1 \geq, ..., \geq \lambda_N$ (only the first r values are non-zero) and an orthonormal system of corresponding eigenvectors as $\{\mathbf{v}_1, ..., \mathbf{v}_N\}$. Let us define an $N \times N$ function matrix $\mathbf{Q} = (Q_{\alpha\beta})$ as

$$\mathbf{Q} = \sum_{i=1}^{r} \mathbf{v}_i \mathbf{v}_i^\top .$$

Theorem 2 If $\alpha \in \mathcal{I}_k$ and $\beta \notin \mathcal{I}_k$, then $Q_{\alpha\beta} = 0$. (See Kanatani [9] for the proof.)

Acknowledgments

The authors wish to thank Hiroshi Hattori, Kazuhiro Fukui and Mutsumi Watanabe for their valuable discussions.

References

1. P.A. Beardsley, P. Torr, and A.P. Zisserman. 3D model acquisition from extended image sequences. In *4rd ECCV*, pages 683–695, Cambridge, UK, 1996.
2. P.N. Belhumeur and D.J. Kriegman. What is the set of images of an object under all possible illumination conditions? *IJCV*, 28:3:245–260, 1998.
3. J. Costeira and T. Kanade. A multi-body factorization method for independently moving objects. *IJCV*, 29:3:159–179, 1998.
4. F. Devernay and O. Faugeras. Computing differential properties of 3D shapes from stereoscopic images without 3D models. In *CVPR*, pages 208–213, 1994.
5. P. Fua. Object-centered surface reconstruction: combining multi-image stereo and shading. *IJCV*, 16:35–56, 1995.
6. G. W. Gear. Multibody grouping from motion images. *IJCV*, 29:2:133–150, 1998.
7. B.K.P. Horn. *Robot Vision*. The MIT Press, 1992.
8. K. Kanatani. *Geometric computation for machine vision*. Oxford University Press, Oxford, 1992.
9. K. Kanatani. Factorization without factorization: Multibody segmentation. Vol.98 no.395, IEICE, PRMU98-117, 1998.
10. A. Maki. Estimation of illuminant direction and surface reconstruction by Geotensity constraint. In *11th SCIA*, pages 71–78, 1999.
11. A. Maki, M. Watanabe, and C.S. Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. In *6th ICCV*, pages 1053–1060, 1998.
12. J.L. Mundy and A. Zisserman, editors. *Geometric invariance in computer vision*. The MIT Press, 1992.
13. A.P. Pentland. Photometric motion. *IEEE-PAMI*, 13:9:879–890, 1991.
14. A. Shashua. *Geometry and photometry in 3D visual recognition*. PhD thesis, Dept. Brain and Cognitive Science, MIT, 1992.
15. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9:2:137–154, 1992.
16. D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *4th ICCV*, pages 675–682, 1993.
17. C.S. Wiles, A. Maki, N. Matsuda, and M. Watanabe. Hyper-patches for 3d model acquisition and tracking. In *CVPR*, pages 1074–1080, 1997.