Constructing Illumination Image Basis from Object Motion

Akiko Nakashima, Atsuto Maki, and Kazuhiro Fukui

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

Abstract. We propose to construct a 3D linear image basis which spans an image space of arbitrary illumination conditions, from images of a moving object observed under a static lighting condition. The key advance is to utilize the object motion which causes illumination variance on the object surface, rather than varying the lighting, and thereby simplifies the environment for acquiring the input images. Since we then need to re-align the pixels of the images so that the same view of the object can be seen, the correspondence between input images must be solved despite the illumination variance. In order to overcome the problem, we adapt the recently introduced geotensity constraint that accurately governs the relationship between four or more images of a moving object. Through experiments we demonstrate that equivalent 3D image basis is indeed computable and available for recognition or image rendering.

1 Introduction

In appearance-based object recognition, lighting variation is one of the most significant issues. As Moses et al.[10] pointed out in the context of face recognition, "the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity." In order to deal with this problem several approaches using image bases have been proposed, and successful results were reported. See, for example, [2,5]. The advantage of these methods is due to the fact that an arbitrary image of an object under a distant light source has low-dimensional representation of the object. The observation concerning the low-dimensional representation originated in Shashua's works[14,15]. He showed that images of an object under arbitrary lighting conditions are represented by linear combinations of three base images, given that the surface of the object follows the Lambertian model without shadows. Yuille and Snow [21] further considered ambient background illumination, Belhumeur and Kriegman [3] extended the representation to the case with attached shadows. They proved that a set of images with attached shadows under arbitrary lighting conditions forms a convex polyhedral cone which is again based on three base images. As is obvious from these observations, the notion of base images provides a foundation for representing images under arbitrary illumination. In this paper, we call the set of the three base images *illumination image basis*.

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2352, pp. 195–209, 2002.

In previous approaches the illumination image basis has been obtained from images of an object in a fixed pose under various lighting conditions. To acquire such images, therefore, the object and the camera must be kept stationary, and the direction of illumination be varied. As opposed to this situation, if we could equally construct the illumination image basis from images of an object that is in motion under a stationary lighting condition, the process of capturing input images would be substantially simplified. This is especially true for an object such as a human head, which we mainly deal with in this article, since the only required action is to move the object in 3D space. However, to our knowledge there has been no attempt to consider the possibility, and the purpose of this paper is to propose a technique for constructing the illumination image basis from object motion.

Given images of an object captured in different poses, the set of pixels at the same coordinate in the images do not correspond to an identical point in the 3D surface. For the construction of illumination image basis, thus, we basically need to re-align the pixels in the images as if the object were seen from an identical direction, and the problem therefore becomes that of finding the corresponding image points.

For the point correspondence, typically exploited is the constraint that the corresponding parts of the images have equivalent intensity, regarding the variation in illumination as noise. Unfortunately, however, the constraint is nearly always invalid for an object in motion, since relatively non-uniform lighting causes the intensity at a specific location on the surface of the object to change as the object moves [13]. Among the few efforts to address this issue, recently the *geotensity* constraint[1] [8] 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.

The algorithm for point correspondence using the geotensity constraint proceeds basically in two stages. The first stage is to derive the parameters of the Geotensity constraint by analyzing coordinates and image intensity of some sample points on the object in motion. That is, computing structure from motion gives the geometric parameters of the situation, whereas computing the linear image subspace gives the lighting parameters of the situation. By combining both sets of parameters we arrive at the Geotensity constraint. Using the same set of input 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 correspondence 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 geotensity constraint was proposed originally for the task of 3D surface reconstruction, we will show that the constraint can be applied directly for our porpose, that is, to obtain the illumination image basis from images of an object in motion under stationary illumination. Further, once we obtain the illumination image basis, images of the object under arbitrary lighting con-

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

ditions can be generated simply by linear combinations of the basis as stated earlier. Since the Geotensity constraint can be applied to any choice of reference frame, different choices allow us to construct illumination bases corresponding to different viewing directions and thereby we can synthesize variations of object images both in terms of lighting conditions and object poses. While such variations are obviously useful for recognition problems in general, it should be noted that they become available from a minimum of four input images. Although the resulting variations may look similar to what is obtained by the method in [5], we should point out that the entire scheme is different in that we explicitly solve for the point correspondence instead of estimating the surface geometry in the framework of photometric stereo [20].

This paper is organized as follows: First, we introduce the low-dimensional representation for images captured under arbitrary illumination in Section 2 and the geotensity constraint in Section 3. Then, in Section 4 we describe an algorithm to construct the illumination image basis by applying the geotensity constraint. Experiments are shown in Section 5. In Section 6, the final section, discussions and conclusions are presented.

2 Illumination Image Basis

In this section we briefly describe the low-dimensional representations for variations of illumination which have been successfully utilized in appearance-based object recognition.

Consider the case of a convex object with a surface conforming to a Lambertian reflectance model with varying albedo, the ratio of outgoing to incoming light intensity, under a single point light source at infinity. Let us assume further that both the camera and object are stationary, but that the light source can be moved. The key advantage of this assumption is that the correspondence problem between frames is solved in advance since the same view of the object is always seen.

If we have three pictures of the object $\mathbf{I}(j)(j = 1, 2, 3)$ from light source directions $\mathbf{s}(j)(j = 1, 2, 3)$, respectively, then the intensities in any fourth frame, taken from a novel setting of the light source, must simply be a linear combination of the intensities in the first three frames. Mathematically,

$$\mathbf{I}(4) = \sum_{j=1}^{3} a(j)\mathbf{I}(j)$$

for some coefficient $a(j)$ [14].

Epstein et al. [4] showed empirically that this relationship is generally quite good, but it was left to Belhumeur and Kriegman [3] to prove that the correct relationship assuming no self-shadowing is in fact

$$\mathbf{I}(4) = \max(\sum_{j=1}^{3} a(j)\mathbf{I}(j), 0) \ , \tag{1}$$

Fig. 1. Images of an object under variable lighting and an illumination basis.

Fig. 2. Images of an object in motion and an illumination basis.

where images $\mathbf{I}(j)(j = 1, 2, 3)$ must be taken with the light source in the *bright cell*, which is the cell of light source directions that illuminate all points on the object, and the maximum operator is effected independently on each pixel and establishes whether or not that pixel was illuminated at all by the light source.

These first three images form a 3D image basis from which all other images can be derived and hence can be replaced by any equivalent basis of images. Hence Equation 1 can be replaced by,

$$\mathbf{I}(j) = \max([\breve{\mathbf{I}}(1) \;\; \breve{\mathbf{I}}(2) \;\; \breve{\mathbf{I}}(3)]\mathbf{a}(j), 0) \;\; , \tag{2}$$

where a set of $\breve{\mathbf{I}}(j)(j = 1, 2, 3)$ is an image basis, which we call *illumination image basis*, and the 3D vector $\mathbf{a}(j)$ contains linear coefficients that map the basis to the j^{th} image.

An example is shown in Figure 1. The four images on the left are original images of a statue of Caesar in frontal view under variable lighting. Applying the principle component analysis (PCA) to these images, we obtain an illumination image basis. It is shown in the three images on the right in Figure 1 arranged in the order of the eigenvalues. In theory, the illumination image basis can be obtained directly as three images captured in a fixed pose under three different lighting conditions. In this case, however, the illumination image basis remains to be contaminated if the captured images contain any noise. In order to obtain an illumination image basis in which the noise is attenuated, PCA is often applied to a larger number of images which are captured as inputs. Although we illustrate it in Figure 1 using only four images as the minimum number for the inputs, it is to show the relevancy of the method, and also for the consistency with the remaining examples. The concept of illumination image basis is quite general, and has been employed to develop further sophisticated algorithms for object recognition as well as image rendering as seen for example in [16].

In this paper, we deal with images of an object in motion captured by a fixed camera under stationary illumination. In such a situation, the object is observed differently according to the motion. Figure 2 shows four images of the statue of Caesar in various poses illuminated from its frontal direction. From these images we wish to construct an illumination image basis as shown in the right-hand part of Figure 2 that is a counterpart to the illumination image basis obtained in the conventional way in Figure 1. Methods for object recognition which give successful result for arbitrary lighting conditions should then become analogously available without collecting images under various illumination. Unlike the conventional situation, however, it is not so straightforward to construct an illumination image basis from images of an object in motion. While this is due to the correspondence problem between frames which is not solved in advance, we do overcome the problem by adapting the geotensity constraint which was originally proposed for 3D surface reconstruction in [8]. We introduce the geotensity constraint in the context of constructing an illumination image basis in the next section, and describe the actual algorithm constructing the illumination image basis in Section 4.

3 Geotensity Constraint

The geotensity constraint is a constraint on geometrically corresponding pixel intensity between four or more images of an object seen from different views under static lighting conditions. In this section, we introduce the geotensity constraint for a single light source from the viewpoint of constructing an illumination image basis; a detailed explanation of the geotensity constraint can be found in [8]. In general, the geotensity constraint consists of two parts, the geometry constraint and the intensity constraint, from which the geotensity constraint is derived.

3.1 Intensity Constraint

As shown in Equation 2, images of an object in a fixed pose captured under arbitrary illumination can be represented by three base images, and the representation is also valid for an intensity at each pixel. That is, the key advantage of Equation 2 is that the same view of the object is always seen and the correspondence problem between frames is solved in advance.

Now let us assume a static camera and light source and a moving object. Although we no longer have correspondence between images, we consider the intensity $I_i(j)$ of the i^{th} point on the surface of the object projected into the j^{th} image. $I_i(j)$ must then satisfy

$$I_i(j) = \max([\breve{I}_i(1) \quad \breve{I}_i(2) \quad \breve{I}_i(3)]\mathbf{a}(j), 0) \ , \tag{3}$$

where a set of $\breve{\mathbf{I}}(j)(j = 1, 2, 3)$ represents the intensity at corresponding points in the three base images. Equation 3 is the intensity constraint between images.

3.2 Geometry Constraint

Regarding the first image as the reference frame, for simplicity we discuss the *affine* and *scaled-orthographic* camera models [11] 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 as

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

where $\mathbf{M}(j)$, an arbitrary 2×3 matrix, and $\mathbf{t}(j)$, an arbitrary 2-dimensional vector, encode the motion parameters of the object.

The scaled orthographic camera model replaces $\mathbf{M}(j)$ with the first two rows of a scaled 3×3 rotation matrix $\mathbf{R}(j)$ so that

$$\mathbf{x}_i(j) = \lambda \begin{pmatrix} \mathbf{R}_1^\top(j) \\ \mathbf{R}_2^\top(j) \end{pmatrix} \mathbf{X}_i + \mathbf{t}(j) \ .$$

Note that $\mathbf{R}(j)$ represents the rotation of the object from the first frame to the j^{th} frame. The Euclidean structure and motion parameters fitting the weak perspective camera model can be recovered by the technique of structure from motion [18]. There is an arbitrary choice of affine or Euclidean frame which can be partially fixed by choosing the first frame to be *canonical* [7] such that

$$\mathbf{x}_i(1) = \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \end{pmatrix} \mathbf{X}_i + \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{4}$$

A result of choosing the canonical frame is that the structure vectors have the form, $\mathbf{X}_i = (\mathbf{x}_i^\top(1), Z_i)^\top$ and we can derive the relationship,

$$\mathbf{x}_i(j) = \mathbf{M}(j) \begin{pmatrix} \mathbf{x}_i(1) \\ Z_i \end{pmatrix} + \mathbf{t}(j) \ . \tag{5}$$

This relationship effectively describes the *epipolar constraint* between two images. This constraint says that given a point, $\mathbf{x}_i(1)$, in one image then the corresponding point in another image, $\mathbf{x}_i(j)$, will lie on a line defined by the object motion parameters, $\mathbf{M}(j)$ and $\mathbf{t}(j)$, between the two frames and that the exact position along that line depends on the correct depth, Z_i, of the point. The corresponding point on an epipolar line is shown in Figure 3.

Indeed, given a point in one image and the motion parameters for a number of other frames, then the corresponding points in all the other images are uniquely determined by a particular choice of Z_i. We will use Equation 5 as a practical affine form of the geometric constraint later. A similar form of this geometric (epipolar) constraint can be easily derived for the more complex camera models.

3.3 Geotensity Constraint

Consider a set of n_j images, $\mathbf{I}(j)(j = 1, \ldots, n_j)$, and a point $(\mathbf{x}_i^T, Z_i)^T$ on the surface of an object with depth Z_i that projects to point, $\mathbf{x}_i(1)$, in the first image.

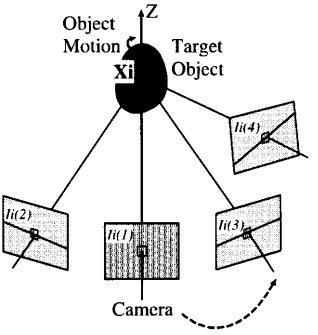

Fig. 3. Given a point in one image, according to the choice of depth Z, corresponding points in the other images are searched on the epipolar lines while investigating the intensity constraint.

Under object motion and affine imaging conditions we can recall Equation 5 for the geometric constraint imposed on a sequence of images and the corresponding pixel in the j^{th} image, $\mathbf{x}_i(j)$, is given by Equation 5. Hence, the corresponding pixel intensity, $I_i(j)$, is given by

$$I_i(j) = \mathbf{I}(j)[\mathbf{x}_i(j)] = \mathbf{I}(j)[\mathbf{M}(j) \begin{pmatrix} \mathbf{x}_i \\ Z_i \end{pmatrix} + \mathbf{t}(j)] \ . \tag{6}$$

If the motion parameters, $\mathbf{M}(j)$ and $\mathbf{t}(j)$, and the depth of the point on the object surface, Z_i, are given, $I_i(j)$ is measured from the intensity at the corresponding pixels in the j^{th} image.

On the other hand, when the corresponding pixel intensity $I_i(j)$ is non-zero, it has another representation by the intensity constraint defined by Equation 3 which we can rewrite as

$$I_i(j) = \left[\breve{I}_i(1) \quad \breve{I}_i(2) \quad \breve{I}_i(3) \right] \mathbf{a}(j) \ . \tag{7}$$

For n'_j measured intensities $I_i(j)$ which are non-zero, it follows

$$\mathbf{I}^{\top}_{n'_j} = \left[\breve{I}_i(1) \quad \breve{I}_i(2) \quad \breve{I}_i(3) \right] \mathbf{a}_{n'_j}(j) \ , \tag{8}$$

where $\mathbf{I}_{n'_j}$ and $\mathbf{a}_{n'_j}$ are n'_j-dimensional vector consisting of the non-zero measured intensity and a $3 \times n'_j$ matrix containing the columns $\mathbf{a}(j)$, respectively. When the linear coefficients $\mathbf{a}(j)$ are known in advance, using the non-zero measured intensity $\mathbf{I}_{n_k}$, $\breve{I}_i(j)(j = 1, 2, 3)$ can be computed by the solution of Equation 8 in the form

$$\left[\breve{I}_i(1) \quad \breve{I}_i(2) \quad \breve{I}_i(3) \right] = \mathbf{I}^{\top}_{n_k} \mathbf{a}^{\top}_{n_k} (\mathbf{a}_{n_k} \mathbf{a}^{\top}_{n_k})^{-1}. \tag{9}$$

Then, the other representation of the corresponding pixel intensity, $\hat{I}_i(j)$, is given by

$$\hat{I}_i(j) = \max(\left[\breve{I}_i(1) \quad \breve{I}_i(2) \quad \breve{I}_i(3) \right] \mathbf{a}(j) \ , 0). \tag{10}$$

Now let us define an error between the measured and estimated intensities as

$$E_i = \sum_j (I_i(j) - \hat{I}_i(j))^2 \ . \tag{11}$$

If the estimated values $\hat{I}_i(j)$ are correct, they agree with the measured intensity $I_i(j)$. Hence, the constraint on geometrically corresponding intensity between captured images, namely the geotensity constraint, can be stated simply as $E_i = 0$. Since the measured intensity $I_i(j)(j = 1, \cdots, n_j)$ includes noise in practice, the corresponding intensity between captured images, $\hat{I}_i(j)(j = 1, \cdots, n_j)$, is estimated so as to minimize the error in Equation 11.

4 Construction of Illumination Image Basis

In this section we propose a method to construct an illumination image basis from an image sequence by applying the geotensity constraint. The basic concept is to search for the depth, Z_i, of the surface of the object at each pixel, $\mathbf{x}_i$, in a reference image while computing the error E_i in the geotensity constraint. Then, the corresponding pixels are searched along the epipolar lines in the other images as schematically depicted in Figure 3. The algorithm for constructing an illumination image basis is detailed, preceded by descriptions of practical schemes for solving for the geometry and the intensity constraints.

4.1 Solving for Geometry

Computing geometric correspondence requires that $\mathbf{M}(j)$ and $\mathbf{t}(j)$ in Equation 5 be known. To recover these components we solve the well-known affine structure from motion problem using SVD [17].

An initial measurement matrix containing the coordinates of a few corresponding sample points is thus needed and these must be provided by an independent mechanism. One way to sample proper points through a sequence is to employ a scheme to extract corners and find their correspondence automatically, as seen for example in [1,19]. For the obtained corresponding sample points, we solve not only for motion parameters but also compute the correspondences between pixel intensity that are required for the next stage of solving for intensity.

4.2 Solving for Intensity

Using the sample corresponding points obtained in the process of solving for motion parameters, we must now acquire the linear coefficients $\mathbf{a}(j)$ acting on the illumination image basis in Equations 9 and 10. We assume Lambertian surface properties, a convex object, and a single point light source. Observing n_i sample corresponding pixels in each image through n_j frames, we can form the matrix equation

$$\mathbf{I} = \breve{\mathbf{I}}\mathbf{a} \tag{12}$$

Fig. 4. Re-assigned images in frontal view.

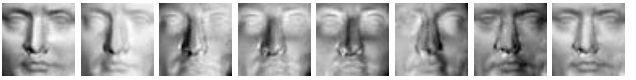

Fig. 5. Synthesized images under various illumination.

where $\mathbf{I}$ is an $n_i \times n_j$ matrix containing the corresponding sample pixel intensity $\mathbf{I}'_i(j)$, $\breve{\mathbf{I}}$ is an $n_i \times 3$ matrix containing a 3D basis for the sparse elements in $\breve{\mathbf{I}}'(j)(j = 1, 2, 3)$ such that

$$\breve{\mathbf{I}} = [\breve{\mathbf{I}}'(1) \;\; \breve{\mathbf{I}}'(2) \;\; \breve{\mathbf{I}}'(3)]$$

and $\mathbf{a}$ is a $3 \times n_j$ matrix whose columns are the linear coefficients $\mathbf{a}(j)$ as

$$\mathbf{a} = [\mathbf{a}(1) \;\; \mathbf{a}(2) \;\; \ldots \;\; \mathbf{a}(n_j)].$$

We record the corresponding sample pixel intensity, $I'_i(j)$, in $\mathbf{I}$, which we call illumination matrix. Equation 12 is then in the familiar form for solution by singular value decomposition to obtain a rank 3 approximation to the matrix $\mathbf{I}$. Hence, the linear coeffecints are acquired from the intensity of the small number of sample pixels in advance.

In order to solve for singular value decomposition of $\mathbf{I}$, in this case, at least four sample points are required through four images in the sequence. The astute reader would have noticed that Equation 12 can only be used if all the pixels used in forming $\mathbf{I}$ were illuminated in all j images. This is easy to ensure since under a single light source these pixels will all have non-zero intensity values and thus it is not necessary for the images to have been taken with the light in the bright cell. In practice we employ a robust random sampling and consensus technique (RANSAC) to ensure that artifacts that are caused by an object not fulfilling the assumed conditions (eg. 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 $\mathbf{A}$ since

$$\breve{\mathbf{I}}\mathbf{a} = (\breve{\mathbf{I}}\mathbf{A})(\mathbf{A}^{-1}\mathbf{a}). \tag{13}$$

However, we may leave $\mathbf{A}$ undetermined, since $\mathbf{A}$ is canceled when computing the estimated intensity using Equations 9 and 10.

4.3 Constructing Illumination Image Basis

We describe an algorithm for constructing illumination image basis by way of dense search for the corresponding pixel intensity applying the geotensity constraint. Note that we have acquired $\mathbf{M}(j)$, $\mathbf{t}(j)$ and $\mathbf{a}(j)$ through the techniques

discussed in the previous two sections. At each pixel in the reference image, $\mathbf{x}_i(1)$, we measure the corresponding pixel intensity, $I_i(j)$, estimate $\hat{I}_i(j)$, and compute the error, E_i, in the geotensity constraint at depth, Z_i, at regular small intervals. The corresponding pixel intensities are computed from Equations 6, 9, and 10 armed with $\mathbf{M}(j)$, $\mathbf{t}(j)$ and $\mathbf{a}(j)$. As is apparent in Equation 6, the pixel coordinate where we acquire the intensity depends on the depth map. When the depth is correct we expect the error to approach zero. The effective search range for the depth is determined according to the pointalistic model of the object computed in Section 4.1 so that it is limited to the range where the object may exist in space.

By the minimization of the error at each pixel in the reference frame, $\mathbf{x}_i(1)$, we acquire the intensity at the corresponding pixels in the other frames, $I_i(j)$. With the acquired intensity in the pixels $\mathbf{x}_i(j)$, we can construct the images of the object seen from the same view as the reference frame. Figure 4 shows the images constructed from four images on the left shown in Figure 2 in various views. The viewing direction in the constructed images is aligned to that in the reference frame and the images appear to be illuminated by variable lighting source. We can either employ a set of three images from the constructed images directly as an illumination image basis, or apply the PCA to the constructed images and form an illumination basis by the three eigenvectors corresponding to the highest eigenvalues. The three images on the right in Figure 2 compose an illumination image basis obtained by applying the PCA to the constructed images shown in Figure 4. Indeed, images under arbitrary illumination can be synthesized by the resulting illumination basis as shown in Figure 5.

Given n_j images of an object in motion, the framework of the algorithm is summarized as follows:

1° We set a reference image as the first one, $\mathbf{I}(1)$.
2° For a pointalistic model, we compute the motion parameters $\mathbf{M}(j)(j = 1, \cdots, n_j)$ and $\mathbf{t}(j)(j = 1, \cdots, n_j)$, and linear coefficients, $\mathbf{a}(j)(j = 1, \cdots, n_j)$.
3° For a particular guess of depth, Z_i, at point $\mathbf{x}_i$, we measure $I_i(j)(j = 1, \cdots, n_j)$ using Equation 6 with $\mathbf{M}(j)$, $\mathbf{t}(j)$, and $\mathbf{a}(j)$.
4° These measurements are fed into Equation 9 to calculate the values $\breve{I}_i(j)(j = 1, 2, 3)$.
5° The estimated values, $\hat{I}_i(j)(j = 1, \cdots, n_j)$, are computed using Equation 10.
6° The error E_i can be obtained using Equation 11.
7° Do 3° – 6° while searching for a depth that minimizes the error E_i.
8° Do 3° – 7° for all pixels in the reference frame, $\mathbf{x}_i(1)$.
9° The corresponding pixel intensity $I_i(j)(j = 2, \cdots, n_j)$ is aligned to the pixel in the reference image $\mathbf{x}_i(1)$.
10° We apply the PCA to the aligned images and the reference frame. Then, an illumination basis is acquired as three eigenvectors corresponding to the highest eigenvalues.

In the process 10°, alternatively, we may choose three images directly from the reference and aligned images as an illumination basis.

Fig. 6. Images of a face under variable lighting and an illumination basis.

Fig. 7. Input images of a face in motion and an illumination basis in frontal view.

Fig. 8. Images of another face under variable lighting and an illumination basis.

Fig. 9. Input images of a face in motion and an illumination basis in frontal view.

5 Experiments

In this section, we show several experiments to confirm the validity of the proposed algorithm. We apply the algorithm to a statue of Caesar and human faces. First, we compare the results obtained by the proposed algorithm with those obtained in the conventional way. Then, we explore the possibility of utilizing the proposed method for face recognition by comparing the difference between the obtained illumination image bases for faces of different subjects. It is also shown that the method makes it possible to construct illumination image bases in various views according to the choice of the reference frame.

5.1 Illumination Image Bases of a Statue of Caesar and Human Faces

Let us begin by confirming the validity of the proposed method. Using a statue of Caesar and human faces as objects, we compare illumination bases constructed from images of the objects under motion with those constructed from images under variable illumination. Although the results for Caesar have already been referred to in the previous sections for explanation, we discuss more of the details here.

Table 1. (a) Similarities between the subspaces of the same subjects spanned by illumination image bases constructed in the conventional and proposed methods. (b) Similarities between person A and B.

subjects	Caesar	person A	person B
similarities	0.99	0.99	0.99

(a)

person A \ B	illumination	object motion
illumination	0.94	0.95
object motion	0.95	0.94

(b)

In Figures 1, 6, and 8, four images on the left are captured under variable lighting. We extract the central area including eyes, nose, and mouth from the images. By applying PCA to the extracted areas, each illumination image basis is obtained by three eigenvectors. It is shown in the right-hand parts of Figures 1, 6, and 8 and arranged according to eigenvalues. As observed in [6], the first illumination base image is illuminated from the front, the second from the right or left, and the third from the top or bottom. In Figures 2, 7, and 9, on the other hand, four images on the left are input images of the objects in motion. The left image in frontal view is set to be a reference frame. By applying the geotensity constraint to the input images, we search correspondences between the images, and then images in frontal view under variable illumination are obtained by re-aligning the intensity in the process 9° in Section 4.3. The re-alinged images for Caesar are shown in Figure 4. It is observed that searching of correspondences is successful. These images are the analogues to the four left-hand images of the object under variable lighting shown in Figure 1. As in Figures 1, 6, and 8 PCA is applied to the re-aligned images to obtain an illumination image basis. The results are shown in the right-hand parts of Figures 2, 7, and 9. The first base image is illuminated from the front, the second from the right or left, and the third from the top or bottom. These aspects are similar to those in Figures 1, 6, and 8 except the reverse of white and black which means that only the signs of elements of the base image are opposite. Hence, each of the base images in Figures 2, 7, and 9 appears to correspond with each of those in Figures 1, 6, and 8.

Mathematically, however, since there exist infinite illumination image bases which span the illumination subspace unique to one object in one pose, it is not prerequisite for an illumination image basis constructed from images under object motion to exactly correspond with the counterpart from images under varying illumination. In order to evaluate the illumination image basis acquired by the proposed method, thus, we investigate its closeness to the counterpart due to varying illumination by comparison in the level of the subspaces. For computing the similarities between the subspaces spanned by illumination image bases constructed in the conventional and proposed methods in quantity, we introduce a canonical angle θ. It represents an angle between two spaces, S_1 and S_2, and defined by

$$cos^2\theta = \sup_{\substack{s_1 \in S_1, s_2 \in S_2 \\ ||s_1|| \neq 0, ||s_2|| \neq 0}} \frac{(s_1, s_2)}{||s_1||^2 ||s_2||^2}. \tag{14}$$

If S_1 exactly corresponds to S_2, then $cos^2\theta$ is equal to 1. The results of computing Equation 14 between the illumination subspaces represented by illumination image bases shown in Figures 1 and 2, 6 and 7, and 8 and 9 are 0.99 as listed in Table 1 (a). From the results, it can be inferred that an illumination image basis is constructed from object motion by the proposed algorithm as well as from variable illumination. As for the human faces, the second and third base images in Figures 7 and 9 have some destroyed parts in their appearance, which are seemingly caused by the non-rigidity of the subjects especially around the eyes (the statue of Caesar is rigid on the other hand). Nevertheless, the resulting figures of similarity in terms of the illumination subspaces are not deteriorated and thereby show the validity of the proposed algorithm.

We also explore the possibility of utilizing the illumination image bases obtained by the proposed method for face recognition, by computing the differences between the illumination subspaces for two different human subjects. The resulting angles computed by Equation 14 for all possible combinations of the illumination subspaces derived from the two different human subjects are listed in Table 1 (b). The computed similarities for the two subjects, shown in Figures 6 through 9, are 0.94 or 0.95 for different cases of utilizing varying illumination and object motion interchangeably. The values are smaller than those for the similarity between the illumination image bases of the same subject, 0.99. In the same way, illumination image bases are constructed for several different subjects, and similar results are obtained. That is, the similarities for the same subjects are 0.99 and for different subjects are around 0.94. Hence, the difference between the two subjects can be regarded as reasonably well conserved. Although the difference between 0.94 and 0.99 may seem trivial, it is in practice a significant difference when considered in algorithms for face recognition. In Fisher face [2], for example, the ratio of dispersion of images classified in a correct category to dispersion of images classified in a false category is minimized. Since the minimization has the effect of emphasizing differences between categories, the observed difference will be feasible recognized in the algorithm.

5.2 Illumination Image Bases in Various Views

In the proposed method, illumination image bases in various views can be constructed according to the selected reference frames. Here, we use the input images in Figure 7 again. The four input images are set to be a reference frame in turn, and illumination image bases are constructed. Then, we synthesize images under various illumination represented by the linear combination of the obtained illumination base images with various linear coefficients. The synthesized images in views corresponding to the reference frames are shown in Figure 10. It can be seen that the proposed method makes construction of illumination image bases in various views possible only by replacing the reference image with other images

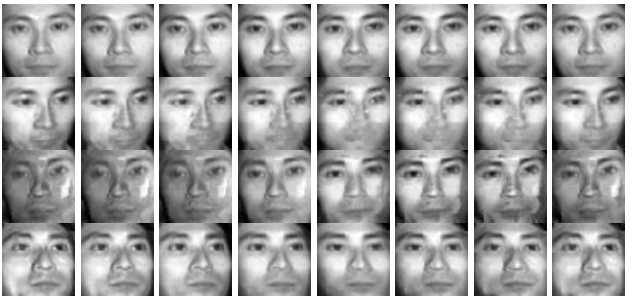

Fig. 10. Illumination bases in various views. The reference images are selected from the input images in Figure 7.

in different views. Therefore, constructing illumination image bases from object motion is useful to deal with variations of not only illumination but also views in the appearance-based object recognition.

When replacing the reference frame, we do not need to search the small number of sample points and to recompute motion parameters and linear coefficients. Once the motion parameters and linear coefficients are computed, the new motion parameters for a different reference image are easily derived from the previous motion parameters which represent the relative motion between the frames and the linear coefficients can be used directly as they are.

6 Discussions and Conclusions

We have proposed a method to construct an illumination image basis from images of an object in motion under fixed illumination. The experiments showed the validity of the proposed method for a statue of Caesar and human faces. The advantages of the method are:

1. An illumination image basis is constructed from images of an object in motion without changing illumination or camera view.
2. Images under arbitrary illumination are synthesized by linear combination of the obtained illumination base images equivalently to the conventional method.
3. Illumination image bases in various poses are constructed from an image sequence only by replacing the reference frame.
4. By utilizing the illumination image bases in various poses, we can enjoy the advantages of an object recognition algorithm that considers the varieties of illumination and poses.

Although we dealt with the geotensity constraint in the case of a single lighting source for simplicity in this paper, extension to the case of multiple light sources is possible by applying the geotensity constraint under multiple light sources [9]. In general, illumination image bases in various views constructed by the proposed method can be utilized for image rendering [16], and for object recognition [12], although occlusion problem is left as the future work.

Acknowledgments. The authors wish to thank Hiroshi Hattori for providing the RANSAC program to find corner correspondence.

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, J.P. Hespanha, and Kriegman D.J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE-PAMI*, 19(7):711–720, 1997.
3. 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.
4. R. Epstein, P.W. Hallinan, and A. Yuille. 5 ± 2 Eigenimages suffice: an empirical investigation of low-dimensional lighting models. In *Proc. IEEE Workshop on Physics-based Modeling in CV*, 1995.
5. A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *CVPR*, pages 52–58, 1998.
6. P.W. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *CVPR*, pages 995–999, 1994.
7. R.I. Hartley. Euclidean reconstruction from uncalibrated views. In J.L. Mundy, A. Zisserman, and D. Forsyth, editors, *Proc. Applications of Invariance in computer vision*, pages 238–256. Springer-Verlag, 1994. Lecture Notes in Computer Science 825.
8. A. Maki, M. Watanabe, and C.S. Wiles. Geotensity: Combining motion and lighting for 3d surface reconstruction. In *6th ICCV*, pages 1053–1060, 1998.
9. A. Maki and C. Wiles. Geotensity constraint for 3d surface reconstruction under multiple light sources. In *6th ECCV*, pages 725–741, 2000.
10. Y. Moses, Y. Adini, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. In J.O. Eckland, editor, *3rd ECCV*, pages 286–296, Stockholm, Sweden, 1994. Springer-Verlag.
11. J.L. Mundy and A. Zisserman, editors. *Geometric invariance in computer vision.* The MIT Press, 1992.
12. H. Murase and S.K. Nayer. Visual learning and recognition of 3-d objects from appearance. *IJCV*, 14(1):5–24, 1995.
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. A. Shashua. On photometric issues in 3d visual recognition from a single 2d image. *IJCV*, 21:99–122, 1997.
16. A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE-PAMI*, 23(2):129–139, 2001.
17. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *IJCV*, 9:2:137–154, 1992.
18. D. Weinshall and C. Tomasi. Linear and incremental acquisition of invariant shape models from image sequences. In *4th ICCV*, pages 675–682, 1993.
19. C.S. Wiles, A. Maki, and N. Matsuda. Hyper-patches for 3d model acquisition and tracking. *IEEE-PAMI*, 23(12):1391–1403, 2001.
20. R.J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.
21. A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *CVPR*, pages 158–164, 1997.