Augmented Reality with Automatic Illumination Control Incorporating Ellipsoidal Models

Jürgen Stauder

***Abstract*— In applications of augmented reality like virtual studio TV production, multisite video conference applications using a virtual meeting room and synthetic/natural hybrid coding according to the new ISO/MPEG-4 standard, a synthetic scene is mixed into a natural scene to generate a synthetic/natural hybrid image sequence. For realism, the illumination in both scenes should be identical. In this paper, the illumination of the natural scene is estimated automatically and applied to the synthetic scene. The natural scenes are restricted to scenes with nonoccluding, simple, moving, mainly rigid objects. For illumination estimation, these natural objects are automatically segmented in the natural image sequence and three-dimensionally (3-D) modeled using ellipsoid-like models. The 3-D shape, 3-D motion, and the displaced frame difference between two succeeding images are evaluated to estimate three illumination parameters. The parameters describe a distant point light source and ambient light. Using the estimated illumination parameters, the synthetic scene is rendered and mixed to the natural image sequence. Experimental results with a moving virtual object mixed into real video telephone sequences show that the virtual object appears naturally having the same shading and shadows as the real objects. Further, shading and shadow allows the viewer to understand the motion trajectory of the objects much better. Demos are available at http://www.irisa.fr/prive/Jurgen.Stauder or http://www.tnt-uni-hannover.de/~stauder.**

***Index Terms*— Augmented reality, lighting estimation, maximum likelihood estimation**

I. Introduction

FOR virtual studio TV production as in RACE MONA LISA project [1], for multisite video conference applications using a virtual meeting room [12], and for synthetic/natural hybrid coding according to the new ISO/MPEG-4 standard [7], [2], synthetic and natural scenes are mixed to generate a synthetic/natural hybrid image sequence. This may be a natural object like a speaker in front of a computer-generated synthetic background.

If the illumination of the natural scene is nondiffuse, e.g., by a spot light or sun light, shading effects [6] will be visible in the natural image sequence. Shading is the spatial change of appearing brightness of a natural object. For example, a speaker's face will appear brighter at the side facing a spot light. Shading may change temporally. For example, the speakers face will get darker if he rotates his head away from the spot light. In addition to shading, cast shadows will be visible [20], [23]. For example, a speaker causes a cast shadow on the scene background.

A synthetic/natural hybrid image sequence appears realistic only if the viewing geometry as well as the illumination of the synthetic and the natural scenes are identical. To align the viewing geometry, the camera parameters of the real camera have to be estimated [15]. This paper is concerned with the illumination aspect. The case of one synthetic and one natural scene is discussed. Further, the natural scenes will be restricted to scenes with nonoccluding, simple, moving, mainly rigid objects, e.g., a speaker in front of a background. To let the natural and synthetic objects appear as being illuminated by the same illumination, the illumination of the natural scene has to be estimated from the natural image sequence. Then, the estimated illumination can be applied to the synthetic scene. This paper addresses the illumination estimation problem. The application of the estimated illumination to the synthetic scene and the final image synthesis will be solved by standard computer graphics tools.

The first contribution of this paper is the combination of several algorithms from the literature to an automatic system for illumination-controlled augmented reality (AR). In known AR systems, the illumination is either set manually by a user [18] or determined from photographs with a lot of user interaction [27]. By a restriction to simple scenes as a speaker's moving head or a single moving rigid object, this paper proposes for the first time a system with automatic illumination control for AR. Therefore, the illumination has to be estimated automatically from the natural image sequence. First approaches to automatic illumination estimation evaluate only a single video image and assume unicolored shaded objects in the natural image sequence, see, e.g., [16] or [28] for one or [26] for several light sources. More recent approaches evaluate two video images and allow for arbitrarily colored objects, see [11] for a single light source and [21] for a single light source with ambient light. In this paper, the approach of [21] will be chosen because it considers ambient light, too. In [21], the scene illumination is described by three illumination parameters, two for the position and one for the intensity of a single light source.

The proposed system contains further algorithms from the literature for automatic segmentation of natural objects, automatic three-dimensional (3-D) shape estimation, and automatic 3-D motion estimation. Shape and motion are needed in

Manuscript received August 26, 1998; revised March 5, 1999. This work was carried out at University of Hannover, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung (http://www.tnt.uni-hannover.de). The associate editor coordinating the review of this paper and approving it for publication was Dr. Minerva Young.

The author is with IRISA/INRIA, Project TEMICS, Campus Universitaire de Beaulieu, F-35042 Rennes Cedex, France (e-mail: jurgen.stauder@irisa.fr).

Publisher Item Identifier S 1520-9210(99)04098-5.

addition to the natural images themselves by the mentioned recent illumination estimation methods. To facilitate the choice of algorithms, the natural scene will be restricted to nonoccluding, simple, moving, mainly rigid objects in front of a rigid background.

A second contribution of this paper is the enhancement of the chosen illumination method from [21]. First, it is shown that the method is a maximum-likelihood estimator. Second, the minimization of the cost function is enhanced. Whereas in [21] the three illumination parameters are estimated one after the other, this paper proposes a single joint estimation step.

The paper is organized as follows. In Section II, the natural scene and its illumination is described by a 3-D scene model and an illumination model. Based on these models, Section III reviews the chosen algorithms for automatic segmentation, shape, and motion estimation for natural objects. In Section IV, the new illumination estimator will be derived that uses the segmentation, shape, and motion results. In Section V, the final synthesis of the synthetic/natural image sequence using standard computer graphics methods will be described. Hereby, the illumination is automatically controlled using the illumination estimation results from Section IV. Sample results are shown in Section VI followed by a conclusion in Section VII.

II. Three-Dimensional Scene Model

In order to estimate the illumination, the 3-D natural scene that is visible in the natural sequence and its illumination have to be described by a physical model. In this section, an illumination model (Section II-A) and a scene model (Section II-B) are briefly introduced. A more detailed description can be found in [21].

A. Illumination Model

In this paper, the scene illumination is described by the photometric quantity of the irradiance. The irradiance is the incident light power per illuminated object surface measured in Lumen. The irradiance is described by a *geometric* parametric illumination model. A *spectral* illumination model [8] is not discussed in this paper. Instead, any light is assumed to have the same color. The geometric illumination model of distant point light sources with ambient diffuse light according to Phong [17] is used. It is shown in Fig. 1 for the case of a single point light source. It describes the irradiance

$$E(\mathbf{N}) = E_D + E_{P0}\max\{0, \cos\angle(\mathbf{L}_0, \mathbf{N})\} \tag{1}$$

as a function of the normal $\mathbf{N}$ of the illuminated surface. The irradiance consists of two terms.

The first term is the constant irradiance E_D describing time invariant, diffuse, local, ambient light. Ambient light may be caused indoors by reflection from walls and outdoors by reflections from the cloudy sky.

The second term describes a single, time invariant, local, distant point light source. Indoors, this may be a lamp and outdoors, the sun. The point light source is distant such that its light incidents in the whole scene from the direction $\mathbf{L}_0$.

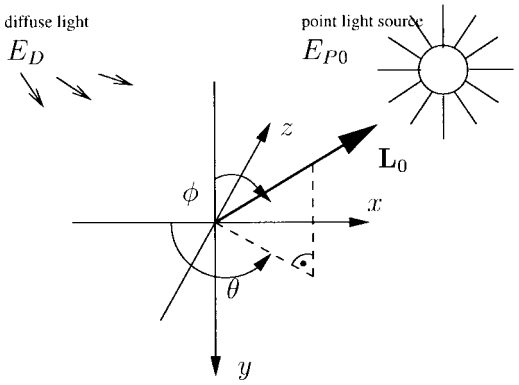


Fig. 1. Illumination model: The illumination parameters θ, ϕ define the direction $\mathbf{L}_0$ of a distant point light source, and parameter $e = E_{P0}/E_D$ defines the ratio of irradiances E_{P0} of point light source and E_D of the ambient light.

The irradiances E_D and E_{P0} can usually be derived from an image signal only relatively, because the amplification of the camera is unknown. Thus, the illumination model (5) is reorganized according to

$$E(\mathbf{N}) = E_D\left(1 + \frac{E_{P0}}{E_D}\max\{0, \cos\angle(\mathbf{L}_0, \mathbf{N})\}\right) \tag{2}$$

where the factor E_D cannot be derived absolutely from the image signal.

The illumination model is defined by three parameters. The ratio of point light source and ambient light irradiance

$$e = \frac{E_{P0}}{E_D} \tag{3}$$

is the first illumination parameter. For example, a value of $e = 3/4 = 0.75$ indicates at the brightest image point an irradiance contribution of 75% from the point light source and of 25% from the ambient light. The second and third illumination parameters are the polar angle θ and the azimuth angle ϕ of the illumination direction

$$\mathbf{L}_0^T = (\cos\phi\sin\theta, \cos\theta, \sin\phi\sin\theta) \tag{4}$$

according to Fig. 1. Equation (2) can be simplified to

$$E(\mathbf{N}) = E_D\left(1 + \max\left\{0, \mathbf{L}^T\mathbf{N}\right\}\right) \tag{5}$$

with

$$\mathbf{L} = e\mathbf{L}_0 \tag{6}$$

the illumination direction $\mathbf{L}_0$ weighted by e and $e\cos\angle(\mathbf{L}_0, \mathbf{N}) = \mathbf{L}^T\mathbf{N}$ a vector point product.

B. Scene Model

Because the illumination estimation method proposed in this paper will use as input data both a natural image sequence as well as information on shape and motion of natural objects, in

(a) (b)

Fig. 2. Shape model for natural objects visualized as (a) wireframe or (b) artificially shaded.

this subsection a shape model, a motion model, and a camera model will be introduced.

1) Shape Model: The shape of a natural object is approximated by a closed rigid 3-D wireframe consisting of triangles [13]. Fig. 2 shows an example of a wireframe.

The wireframe is defined by a number of 3-D control points and their relation to the triangles. The mean of all control points is the center of gravity $\mathbf{C}$ of the wireframe. From the control points, surface normals for each triangle can be calculated. The surface normals are important because the irradiance in (5) depends on it directly.

2) Motion Model: The motion of a natural object between two time instants $k-1$ and k is described as a 3-D translation and a 3-D rotation of the wireframe.

A sample point on the surface of a wireframe with the 3-D position $\mathbf{P}_{k-1}$ at time instant $k-1$ is moved to the position $\mathbf{P}_k$ at time instant k according to

$$\mathbf{P}_k = \underline{\mathbf{R}}(\mathbf{P}_{k-1} - \mathbf{C}_{k-1}) + \mathbf{C}_{k-1} + \mathbf{T}. \tag{7}$$

Here, $\mathbf{T}$ is the 3-D translation vector and $\underline{\mathbf{R}}$ the 3-D rotation matrix [10].

3) Camera Model: The camera model consists of a geometrical and a photometric camera model. The geometric camera model describes the two-dimensional (2-D) image position where a 3-D point is projected. The photometric camera model describes the value of the emerging image signal. In this paper, the geometric camera model is perspective projection, also called pin-whole model. The optical axis is the z-axis from the (x, y, z) world coordinate system. A point at the 3-D position $\mathbf{P} = (P_x, P_y, P_z)^T$ in the world coordinate system is projected onto the 2-D position $\mathbf{p} = (p_X, p_Y)^T$ in the image plane coordinate system (X, Y)—that is parallel to the x, y-plane—according to

$$\mathbf{p} = \begin{pmatrix} p_X \\ p_Y \end{pmatrix} = \frac{f}{P_z} \begin{pmatrix} P_x \\ P_y \end{pmatrix}. \tag{8}$$

Here, $z = f$ is the image plane and f is the focal length of the camera.

The photometric camera model assumes diffuse reflecting object surfaces and neglects lens distortions. It describes the image luminance $s(\mathbf{p})$ at an image position $\mathbf{p}$ according to

$$s(\mathbf{p}) = \eta(\mathbf{P})E(\mathbf{N}) \tag{9}$$

as a product of surface albedo η and irradiance E [24], [21], where $\mathbf{N}$ is the surface normal at position $\mathbf{P}$ on the object surface. The albedo represents the object reflectance in the luminance channel of the camera.

III. Automatic Segmentation and Three-Dimensional Modeling of Moving Natural Objects

The method for illumination estimation proposed in this paper requires as input both a natural image sequence as well as the 3-D object shape and 3-D object motion. Therefore, this section presents an automatic system for segmentation of natural objects (Section III-A), shape estimation for natural objects (Section III-B) and motion estimation for natural objects (Section III-C). The algorithms are combined for the first time to a complete, automatic system.

A first criterion for the choice of specific algorithms is the consideration of directed light in the natural scene. This is important especially for object segmentation (shadows) and motion estimation (temporal luminance changes). A second criterion is a low complexity that give already results sufficiently good to ensure a realistic augmented reality result. The chosen algorithms are closely related to approaches from the literature and will be therefore only shortly reviewed.

A. Segmentation of Moving Natural Objects

For segmentation of natural objects in the natural image sequence, the objects are assumed to move with respect to the background, to be opaque and to not cover each other. The background is assumed to be dominant. A nondiffuse scene illumination is considered, such that moving cast shadows may occur on the background. The segmentation algorithm is a slight modification of [23], as will be explained.

In a first step of segmentation, the motion of the dominant background is estimated and compensated between a current image and its previous image.

In a second step, a change detection mask [5] is estimated from the current and the previous natural image by a relaxation scheme using a locally adapted threshold.

In a third step, image regions changed by moving cast shadows on the background are eliminated from the change detection mask. Those image regions are identified by three criteria. First, moving shadows are assumed only on the dominant background that is detected by a dominant motion approach [14], [23]. Further, changes caused by shadows are detected by a smooth displaced frame ratio (DFR). The DFR is the pel-by-pel ratio between the previous and current image after dominant motion compensation [14], [23]. Finally, shadows are detected by smooth edges at their contour caused by the penumbra [22], [23].

In a fourth step, the background that has been uncovered by object motion is deleted from the change detection mask using a displacement vector field (DVF). Instead of a block matcher as used in [23], in this paper the DVF is calculated according to [13]. Using the 3-D shape estimate for the previous image, if available (see Section III-B), the 3-D motion of the natural object from the previous to the current image is estimated (see Section III-C). From object shape and motion, the DVF

is calculated. The advantage compared to blockmatching is the usage of the rigid motion constraint.

In a last step, the resulting object mask is adapted to luminance edges in the current image.

B. Estimation of 3-D Object Shape

An estimate for the 3-D shape of a natural object will be used in this paper for

- motion compensation between succeeding video images (necessary for illumination estimation (Section IV) and object segmentation [Section III-A)] and as
- 3-D shape information for illumination estimation (Section IV).

Especially, the 3-D shape estimate will not be used to generate images from different viewpoints. It has been found that for these purposes a rough shape approximation is sufficient. This can be further justified by the capabilities of the segmentation algorithm (Section III-A) that assumes objects without occlusion.

The chosen shape estimation algorithm from [13] calculates automatically a 3-D wireframe from the object mask. Hereby, image plane symmetry and an ellipsoid-like shape are assumed. For wireframe calculation, the 3-D shape is assumed to follow a distance transformation. This transformation defines the depth of an object for each point in the image plane inside the object mask evaluating the horizontal distance to the border of the object mask. The hereby defined shape is then approximated by a closed 3-D wireframe. A sample result can be seen in Fig. 2. During the image sequence, the wireframe is segmented into several rigid parts regarding local 3-D motion.

C. Estimation of 3-D Object Motion

The chosen motion estimation method is taken from [21]. The method is based on the classical optical flow equation using additional a rigid motion constraint as described in [13]. The temporal change of luminance caused by nondiffuse scene illumination has been considered in the optical flow equation by using the illumination estimation result from the previous image as described in [21]. The consideration of scene illumination increases 3-D motion accuracy [3].

IV. Illumination Estimation from Two Images

In this section, a method for automatic estimation of three illumination parameters e, ϕ, θ [see (3) and (4)] is presented. The unknown illumination parameters are accessible only by so-called observations. In this paper, an illumination estimation method allowing for arbitrarily colored natural objects is desired. Such methods from the literature [11], [21] observe the displaced frame difference (DFD), the frame-to-frame difference after motion compensation. The observations may contain stochastic errors such as camera noise. The illumination estimator computes estimates $\hat{e}$, $\hat{\phi}$, $\hat{\theta}$ for the illumination parameters from one realization (one measurement) of the observations.

In this paper, illumination estimation is based on the approach in [21]. Whereas in [11] only a single point light source is considered, the approach in [21] considers additionally ambient light that is often present in natural scenes. In the following, the illumination estimation method from [21] will be reformulated as maximum-likelihood (ML) estimator. Therefore, an observation model will be developed in Section IV-A. The observation model describes the relation between the DFD observations and the unknown illumination parameters. From the observation model, a ML estimator will be derived in Section IV-B. The estimator is based on an iterative optimization scheme. Whereas in [21] the three illumination parameters are optimized one after the other, in this paper a joint optimization is proposed.

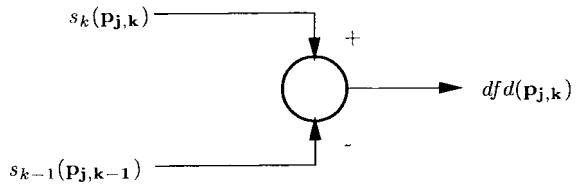


Fig. 3. Calculation of the DFD from two images.

A. Observation Model

The observation model describes the relation between the DFD observations on the one hand and other input data—shape and motion parameters—as well as the unknown illumination parameters e, ϕ, θ on the other hand.

The DFD is the difference of luminances $s_{k-1}(\mathbf{p}_{j,k-1})$ and $s_k(\mathbf{p}_{j,k})$ of two succeeding images at time instants $k-1$ and k according to

$$dfd(\mathbf{p}_{j,k}) = s_k(\mathbf{p}_{j,k}) - s_{k-1}(\mathbf{p}_{j,k-1}) \tag{10}$$

at the two 2-D image positions $\mathbf{p}_{j,k-1}$ and $\mathbf{p}_{j,k}$. These 2-D image positions correspond to the 3-D positions $\mathbf{P}_{j,k-1}$ and $\mathbf{P}_{j,k}$ in 3-D space of a j-th so-called observation point on the surface of a natural object before and after object motion. The DFD indicates the luminance difference after motion compensation. The practical computation of the DFD is shown in Fig. 3.

According to (9), the luminances

$$\begin{aligned} s_k(\mathbf{p}_{j,k}) &= \eta(\mathbf{P}_{j,k})E(\mathbf{N}_{j,k}) \\ s_{k-1}(\mathbf{p}_{j,k-1}) &= \eta(\mathbf{P}_{j,k-1})E(\mathbf{N}_{j,k-1}) \end{aligned} \tag{11}$$

depend on the albedo η of the observation point and the incident irradiance E. $\mathbf{N}_{j,k-1}$ and $\mathbf{N}_{j,k}$ are the surface normals of the observation point before and after motion. The albedo does not change over time, thus

$$\eta(\mathbf{P}_{j,k}) = \eta(\mathbf{P}_{j,k-1}) \tag{12}$$

holds. Inserting (11) and (12) into (10) gives

$$dfd(\mathbf{p}_{j,k}) = s_{k-1}(\mathbf{p}_{j,k-1})\frac{E(\mathbf{N}_{j,k})}{E(\mathbf{N}_{j,k-1})} - s_{k-1}(\mathbf{p}_{j,k-1}). \tag{13}$$

Equation (13) shows that the DFD is nonzero only, if the irradiance incident at the observation point changes over time.

To further develop (13), the surface normals can be related by

$$\mathbf{N}_{j,k} = \underline{\mathbf{R}}\,\mathbf{N}_{j,k-1} \tag{14}$$

with $\underline{\mathbf{R}}$ the object rotation matrix from (7). Inserting (14) and the irradiance from (5) in (13) leads to

$$dfd(\mathbf{p}_{j,k}) = s_{k-1}(\mathbf{p}_{j,k-1})\frac{1+\max\{0,\mathbf{L}^T\underline{\mathbf{R}}\mathbf{N}_{j,k-1}\}}{1+\max\{0,\mathbf{L}^T\mathbf{N}_{j,k-1}\}} - s_{k-1}(\mathbf{p}_{j,k-1}). \tag{15}$$

To simplify (15), the binary value

$$\mu_{j,k} = \begin{cases} 1, & \text{if } \mathbf{L}^T\mathbf{N}_{j,k} \geq 0 \\ 0, & \text{else} \end{cases} \tag{16}$$

is introduced. It is equal to one only if the j-th observation point is illuminated by the point light source at time instant k. With (16), the DFD is

$$dfd(\mathbf{p}_{j,k}) = s_{k-1}(\mathbf{p}_{j,k-1})\frac{1+\mu_{j,k}\mathbf{L}^T\underline{\mathbf{R}}\mathbf{N}_{j,k-1}}{1+\mu_{j,k-1}\mathbf{L}^T\mathbf{N}_{j,k-1}} - s_{k-1}(\mathbf{p}_{j,k-1}). \tag{17}$$

Assuming few object rotation, the illumination geometry changes few and $\mu_{j,k-1} \approx \mu_{j,k}$ holds. Then, the DFD can by finally written by

$$dfd(\mathbf{p}_{j,k}) = s_{k-1}(\mathbf{p}_{j,k-1})\mu_{j,k-1}\frac{\mathbf{L}^T(\underline{\mathbf{R}}-\underline{\mathbf{I}})\mathbf{N}_{j,k-1}}{1+\mathbf{L}^T\mathbf{N}_{j,k-1}} \tag{18}$$

with $\underline{\mathbf{I}}$ the 3×3 unit matrix.

Equation (18) relates the observed DFD with the illumination parameters to be estimated that are contained in $\mathbf{L}$ [see (6)]. To smooth the discontinuities of shading as visible in Fig. 2, Phong interpolation [17], [25] is applied to the surface normals in (18). The observation model is based on the following input data to estimation.

- The object shape, i.e. the surface normals $\mathbf{N}_j$
- The object motion, i.e. the rotation matrix $\underline{\mathbf{R}}$
- The image luminance of two image s_{k-1} and s_k

B. Maximum-Likelihood Illumination Estimator

In this section, a maximum-likelihood (ML) estimator is developed from the observation model derived in the preceding section. First, a cost function will be developed in Section IV-B1. Actually, the cost function of the ML estimator is shown to be a theoretically proved reformulation of the cost function used in [21]. Then, an optimization method will be proposed in Section IV-B2 that minimizes the cost function. Whereas in [21] the illumination parameters are optimized one after the other, this paper proposes a joint optimization.

1) Cost Function: For illumination estimation according to the ML principle [9], those illumination parameter estimates should be taken that cause most probably the DFD observations. To see which estimates these are, the DFD has to be described by a stochastic process. The process describes the vector $\mathcal{DFD}_k$ of DFD observations $(\mathcal{DFD}_{0,k},\ldots,\mathcal{DFD}_{J-1,k})^T$ actually measured at time instant k at J different image positions. Assuming that the image luminance is superimposed by a stochastic, Gaussian, zero mean, stationary camera noise of the variance σ^2, the vector $\mathcal{DFD}_k$ is also jointly Gaussian distributed and its probability density is

$$p(\mathcal{DFD}_k) = \left(\prod_{j=0}^{J-1}\frac{1}{2\sqrt{\pi}\sigma}\right)e^{-\sum_{j=0}^{J-1}\frac{(\mathcal{DFD}_{j,k}-dfd(\mathbf{p}_{j,k}))^2}{4\sigma^2}}. \tag{19}$$

The ML estimation rule is then

$$(\hat{e},\hat{\phi},\hat{\theta}) = \arg\max_{(e,\phi,\theta)} p(\mathcal{DFD}_k). \tag{20}$$

By inserting the probability density of (19) into (20), application of the natural logarithm and suppression of constant terms, the estimator can be formulated as

$$(\hat{e},\hat{\phi},\hat{\theta}) = \arg\min_{(e,\phi,\theta)} \epsilon^T\epsilon \tag{21}$$

i.e., a minimization of a sum of squared residuals

$$\epsilon = \begin{pmatrix} \mathcal{DFD}_{0,k} \\ \cdots \\ \mathcal{DFD}_{J-1,k} \end{pmatrix} - \begin{pmatrix} dfd(\mathbf{p}_{0,k}) \\ \cdots \\ dfd(\mathbf{p}_{J-1,k}) \end{pmatrix} \tag{22}$$

where each residual is the difference between an actual observation and the observation model. The J image positions $\mathbf{p}_{j,k}$ are chosen inside the object mask in regions of low spatial luminance gradient [21].

2) Optimization: Here, a strategy for the minimization of the squared residuals in (21) with respect to the illumination parameters e,ϕ,θ is described. In opposite to [21], the minimization is done jointly for all three illumination parameters. Therefore, the Gauss-Newton optimization method has been chosen, that is suitable for non linear least squares problems [19]. It searches for the minimum by the necessary criterion

$$\nabla\epsilon^T\epsilon = \mathbf{0} \tag{23}$$

with $\nabla = (\partial/\partial e, \partial/\partial\phi, \partial/\partial\theta)^T$ the gradient operator. To find this zero crossing, the method of Newton-Kantorowitsch is applied. Starting with initial values $\hat{e}^{(0)}$, $\hat{\phi}^{(0)}$, $\hat{\theta}^{(0)}$ of the estimates, the well-known iteration rule is

$$\begin{pmatrix} \hat{e}^{(m+1)} \\ \hat{\phi}^{(m+1)} \\ \hat{\theta}^{(m+1)} \end{pmatrix} = \begin{pmatrix} \hat{e}^{(m)} \\ \hat{\phi}^{(m)} \\ \hat{\theta}^{(m)} \end{pmatrix} + (\underline{\mathbf{H}}^T\underline{\mathbf{H}})^{-1}\underline{\mathbf{H}}^T\epsilon \tag{24}$$

with the matrix

$$\underline{\mathbf{H}}^T = (\nabla dfd(\mathbf{p}_{0,k}),\ldots,\nabla dfd(\mathbf{p}_{J-1,k})) \tag{25}$$

and ϵ the vector of residuals from (22). The matrix $\underline{\mathbf{H}}$ and the vector ϵ are computed using the estimates of the m-th iteration. The iteration is stopped as soon as the cost function decreases only marginally.

The derivations contained in the gradient of (25) are

$$dfd(\mathbf{p}_{j,k})' = \frac{\mu_{j,k-1}s_{k-1}(\mathbf{p}_{j,k-1})}{(1+\mathbf{L}^T\mathbf{N}_{j,k-1})^2}[(1+\mathbf{L}^T\mathbf{N}_{j,k-1})\mathbf{L}'^T - (\mathbf{L}'^T\mathbf{N}_{j,k-1})\mathbf{L}^T](\underline{\mathbf{R}}-\underline{\mathbf{I}})\mathbf{N}_{j,k-1} \tag{26}$$

(a) (b) (c)

Fig. 4. Sample images of natural image sequences: (a) "Matthias" (image 2), (b) "Tai" (image 30) and (c) "Ball" (image 3). The sequences are in common intermediate format (CIF) 352 × 288 pel, 10 Hz frame rate.

(a) (b)

Fig. 5. Sample results for (a) automatic object segmentation considering cast shadows and (b) automatic 3-D shape estimation corresponding to the sample image in Fig. 4(a).

with $'$ indicating a derivation with respect to e, ϕ or θ and

$$\begin{aligned}\frac{\partial}{\partial e}\mathbf{L} &= (-\cos\phi\sin\theta, -\cos\theta, -\sin\phi\sin\theta)^T \\ \frac{\partial}{\partial \phi}\mathbf{L} &= e(\sin\phi\sin\theta, 0, -\cos\phi\sin\theta)^T \\ \frac{\partial}{\partial \theta}\mathbf{L} &= e(-\cos\phi\cos\theta, \sin\theta, -\sin\phi\cos\theta)^T\end{aligned} \tag{27}$$

the derivations of the weighted illumination direction $\mathbf{L}$.

For optimization, initial estimates $\hat{e}^{(0)}$, $\hat{\phi}^{(0)}$, $\hat{\theta}^{(0)}$ for the illumination parameters have to be known. Therefore, initial values are systematically searched at about 25 positions in the solution space $\{e, \phi, \theta / 0.1 < e < 9 \wedge 0 < \phi < 180° \wedge 0 < \theta < 180°\}$ [21].

V. Image Synthesis with Automatic Illumination Control

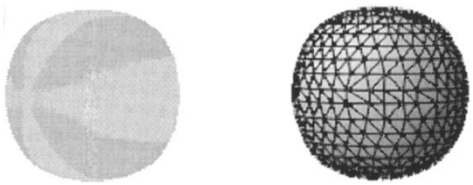
For synthetic/natural hybrid image sequence synthesis, standard computer graphic tools are used, the synthesis is not in the scope of this paper.

Using the tools, the natural objects and the synthetic objects are placed in a synthetic/natural scene. The synthetic objects are animated artificially, the natural objects are motion compensated according to their estimated motion. For the experiments in this paper, the static background is modeled as a plane parallel to the image plane. Without any further user interactivity, the synthetic/natural scene is rendered applying the automatically estimated illumination parameters. Using a z-buffer algorithm [25], the renderer considers cast shadows generated by synthetic objects as well as cast shadows that affect synthetic objects [4].

VI. Experimental Results

In this section will be shown that the proposed illumination estimation can be successfully applied to automatic control of illumination in augmented reality. Therefore, a synthetic ball is mixed into three natural image sequences. The sequences are shown in Fig. 4. The shown natural scenes are illuminated by a single spot light. Further, walls and ceiling are bright and reflect some light back into the scene. In Fig. 4(a), a cast shadow moving together with the person can be seen on the background. Because this paper is not concerned with the geometric aspect of synthetic/natural image sequence synthesis [15], the camera in all sequences is static. For all shown results, the same algorithms with fixed parameters have been used.

In Fig. 5, sample results of object segmentation and shape estimation can be seen. The segmentation mask in Fig. 5(a) has been derived automatically from the first and second image of the image sequence "Matthias" in Fig. 4(a) according to Section III-A. The two visible wholes disappear along the sequence by temporal integration. Image regions changed by the moving cast shadow are not included in the object mask up to the artifact at the persons neck. Fig. 5(b) shows the wireframe automatically calculated from the segmentation mask according to Section III-B. The results shown in Fig. 5 for image 2 of "Matthias" are the worse results compared to results for later images. Nevertheless, they are sufficient for the purpose of this paper as will be seen in the following results.



(a) (b)

Fig. 6. (a) Virtual textured ball and (b) its 3-D wireframe, shaded artificially for presentation purposes.

As synthetic scene, the synthetic ball in Fig. 6(a) is used. Its 3-D wireframe is shown in Fig. 6(b). The ball follows an animated motion: Coming from the one side, it bumps against the background and passes to the other side.

At the left side of Fig. 7, the synthetic/natural hybrid image sequences *without* consideration of illumination is shown. The image sequence lacks in realism because the virtual object is rendered assuming diffuse illumination. At the right side of Fig. 7, the same images of the synthetic/natural hybrid image sequences *with* applied illumination parameters are shown. The synthetic ball and the natural person are shaded and generate a cast shadow on the background as being illuminated by the same illumination.

VII. Conclusion

This paper proposes a method for automatic illumination control in augmented reality. Natural objects from a natural scene and synthetic objects from a synthetic scene are mixed into a hybrid image sequence such that the synthetic objects are illuminated by the same illumination as the natural objects. Therefore, the illumination of the natural scene is estimated automatically from the natural image sequence and applied to the synthetic scene. The illumination is described by three illumination parameters. Two spheric angles define the direction of a distant point light source. A third parameter is the ratio of irradiances of the point light source and ambient, diffuse light.

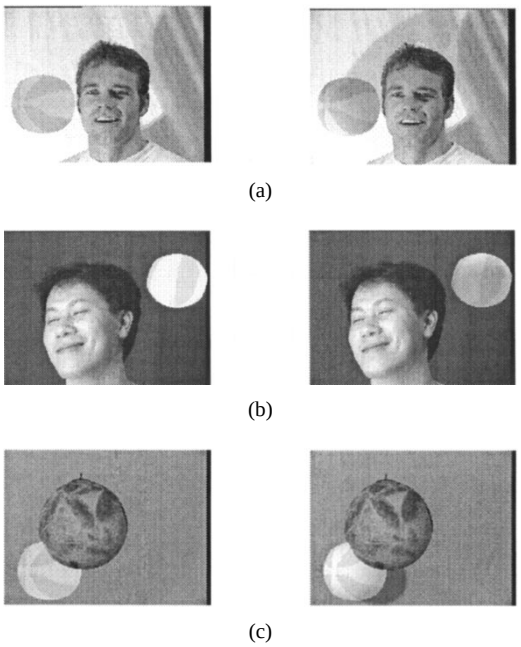


Fig. 7. Sample results of augmented reality left *without* and right *with* automatic illumination control using the three test sequences (a) "Matthias, (b) "Tai" and (c) "Ball" from Fig. 4. At the right side, the synthetic ball looks more realistic. Furthermore, the cast shadow allows a much better identification of the balls motion trajectory. These results can be regarded at http://www.irisa.fr/prive/Jurgen.Stauder or at http://www.tnt.uni.hannover.de/~stauder.

For illumination estimation, a system has been built by combination and refinement of known algorithms. The system performs automatically the following tasks: Natural, nonoccluded objects in a natural image sequence are detected. Hereby, a rigid dominating background is assumed. Furthermore, cast shadows on the background are explicitly considered and not detected as objects. From each object segmentation mask, a 3-D ellipsoid-like wireframe is generated. Using the wireframes, the 3-D motion of the natural objects is estimated. For motion estimation, spatial as well as temporal image signal gradients—caused by the scene illumination—are considered. Finally, the illumination parameters are estimated from the image sequence, from the object shape and from the object motion. Assuming Gaussian camera noise, a maximum-likelihood least square algorithm is developed. The method allows for mainly-rigid, simple, natural objects of arbitrary, unknown texture. It assumes that the objects undergo a rotation from image to image.

Experiments with a synthetic object mixed into natural video telephone sequences show that shading and cast shadow of the synthetic ball matches shading and cast shadow of the natural object resulting in a realistic impression of the synthetic/natural hybrid image sequence. Furthermore, the presence of shading and cast shadow allows a viewer to understand much better the motion trajectories of the objects in the sequence.

This paper focuses on the case where *synthetic* objects are mixed into a natural scene. Thus, illumination effects are mainly applied to *synthetic* objects. It has not been solved yet to mix *natural* objects from different natural image sequences into a natural scene. Here, shading and shadows have to be applied to *natural* objects. Because the objects may be already shaded, future work have to focus on the compensation of illumination effects on the surface of natural objects. Further, the presented method is restricted to simple, mainly rigid objects that are modeled by ellipsoid-like models. To allow for a larger variety of natural, nonrigid objects, the chosen 3-D modeling algorithm may be replaced by a more sophisticated tool in future work.

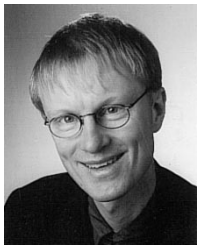
ACKNOWLEDGMENT

The author thanks J. Ostermann for the contribution of a software tool for 3-D wireframe generation from segmentation masks.

REFERENCES

[1] L. Blondé, M. Buck, R. Galli, W. Niem, Y. Paker, W. Schmidt, and G. Thomas, "A virtual studio for live broadcasting, The MONA LISA project," *IEEE Multimedia*, vol. 3, no. 2, Summer 1996.
[2] L. Chiariglione, "MPEG and multimedia communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 5–18, Feb. 1997.
[3] P. Eisert and B. Girod, "Model-based 3D-motion estimation with illumination compensation," in *6th Int. Conf. Image Processing and Its Applications*, vol. 1, pp. 194–198.
[4] M. Fouad, "Object based image synthesis based on the model of global illumination," Tech. Rep., Univ. Hannover, Germany, 1993.
[5] M. Hötter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation," *Signal Process.*, vol. 15, no. 3, pp. 315–334, Oct. 1988.
[6] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1987.
[7] R. Koenen, Ed., "Overview of the MPEG-4 standard," MPEG document ISO/IEC JTC1/SC29/WG11/N2459, Atlantic City, NJ, Oct. 1998.
[8] G. J. Klinker, S. A. Shafer, and T. Kanade, "A physical approach to color image understanding," *Int. J. Comput. Vis.*, vol. 4, no. 7, pp. 7–38, July 1990.
[9] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Englewood Cliffs, NJ, Prentice-Hall, 1995.
[10] K. Meyberg and P. Vachenhauer, *Höhere Mathematik*. Berlin, Germany: Springer-Verlag, vol. 1, 1993.
[11] N. Mukawa, "Estimation of light source information from image sequence," *Syst. Comput. Jpn.*, vol. 23, no. 10, pp. 92–99, Oct. 1992.
[12] H. Nicolas and J. Motsch, "Very low bitrate coding using hybrid synthetic/real images for multi-sites videoconference applications," in *Proc. Visual Communication and Image Processing Conf.*, San Jose, CA, 12.-14.2.97, vol. 3024, pp. 1330–1341.
[13] J. Ostermann, "Object-based analysis-synthesis coding (OBASC) based on the source model of moving rigid 3D objects," *Signal Process.: Image Commun.*, vol. 6, no. 2, pp. 143–161, May 1994.
[14] ———, "Segmentation of image areas changed due to object motion considering shadows," in *Multimedia Communications and Video Coding*, Y. Wang u.a., Ed. New York: Plenum, 1996.
[15] J.-I. Park and C. W. Lee, "Robust estimation of camera parameters from image sequence for video composition," *Signal Process.: Image Commun.*, vol. 9, no. 1, pp. 43–52, Nov. 1996.
[16] A. P. Pentland, "Finding the illumination direction," *J. Opt. Soc. Amer.*, vol. 72, no. 4, pp. 448–455, Apr. 1982.
[17] B.-T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, June 1975.
[18] P. Poulin, K. Ratib, and M. Jacques, "Sketching shadows and highlights to position lights," in *Proc. Comput. Graphics Int.*, June 1977.
[19] L. E. Scales, *Introduction to Non-Linear Optimization*. London, U.K.: Macmillan, 1985.
[20] S. A. Shafer, *Shadows and Silhouettes in Computer Vision*. Dordrecht, The Netherlands: Kluwer, 1985.
[21] J. Stauder, "Estimation of point light source parameters for object-based coding," *Signal Process.: Image Commun.*, vol. 7, pp. 355–379, Nov. 4–6, 1995.
[22] ———, "Segmenation of moving objects in presence of moving shadows," in *Int. Workshop Coding Techniques Very Low Bit-Rate Video*, Linköping, Sweden, July 28–30, 1997, pp. 41–44.

[23] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Trans. Multimedia*, vol. 1, pp. 65–76, Mar. 1999.
[24] J. Stauder, "Schätzung der Szenenbeleuchtung aus Bewegtbildfolgen," Ph.D. dissertation, Univ. Hannover, Germany, 1999.
[25] A. Watt, *3D Computer Graphics*, 2nd ed. Reading, MA: Addison-Wesley, 1993.
[26] Y. Yang and A. Yuille, "Sources from shading," in *Int. Conf. Computer Vision*, 1990, pp. 534–539.
[27] Y. Yu and J. Malik, "Recovering photometric properties of architectural scenes from photographs," in *Proc. SIGGRAPH'98 Conf.*, pp. 207–218.
[28] Q. Zheng and R. Chelappa, "Estimation of illuminant direction, albedo, and shape from shading," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 680–702, July 1991.



Jürgen Stauder received the Dipl.-Ing. degree in electrical engineering from the University of Darmstadt, Germany, in 1990.

From 1990 to 1998, he was with the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung at the University of Hannover, Germany, as a Research Assistant, where he finished his Ph.D. thesis. Since 1998, he has been with the IRISA Laboratory of INRIA in Rennes, France, as a Visiting Researcher. He contributes to the European research project COST 211 and participates in the European TMR program. His research and education interests are in the field of image sequence processing, where he works on illumination estimation, object segmentation, shadow detection, augmented reality, and 3-D object-based video coding.