An Illumination Estimation Method for 3D–Object–Based Analysis–Synthesis Coding

J. Stauder
Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung
Universität Hannover, Appelstraße 9A, D–30167 Hannover 1, FRG
Phone: +49–511–762–5311
Email: stauder@tnt.uni–hannover.de

Abstract

In video telephone scenes to be transmitted at low data rates, non–diffuse illumination has to be expected. Therefore, in this work the scene illumination is automatically estimated from a monocular image sequence in addition to shape and motion of moving objects. Image analysis uses the source model "moving rigid 3D objects with nonuniform local illumination". The irradiance is modelled by a discrete 3D irradiance map. Experiments report an interframe prediction error gain over a former method ignoring illumination. This gain depends on the amount of object motion and the nonuniform portion of scene illumination. A first experiment analysing a real image sequence states an interframe prediction gain of about 1dB.

1 Introduction

In Object–Based Analysis–Synthesis Coding (OBASC) [2], three dimensional information has to be automatically extracted from two dimensional projections of a real 3D scene containing moving 3D objects. Working on a monocular image sequence, a 3D model based image analysis algorithm was presented [3][4] based on the source model "moving rigid 3D objects with diffuse illumination" (3Dd). The images are segmented into static background and moving objects. For each object a silhouette is evaluated. Each moving real object is modelled by a 3D model object. A model object is defined by its shape, position and colour. The shape is rigid and represented by control points connected to a wireframe of triangles, the position is changed by motion that is represented by 3D motion parameters for rotation and translation, the colour is represented by texture mapped onto each triangle. For every real image of the image sequence a model image is synthesized by projecting the model objects into a model image plane. Based on the source model, the 3D image analysis estimates and improves the 3D model object parameters by analyzing the differences between real and model image. In [4], illumination was assumed to be diffuse. If this assumption does not hold in the real 3D scene, as expected for video telephony, model failures occur, i.e. regions of a real image differ significantly from the model image. Thus, nonuniform illumination has to be estimated.

For image analysis a lot of work has been done in the field of illumination estimation based on the model of uniform coloured and still 3D objects [7]. Going away from uniform coloured and still objects, a few approaches were made using 2D displacement and 2D illumination and reflection models [5] or 2D displacement and 3D illumination models [6][8]. Lee and Pearson [1] proposed a 3D irradiance map as illumination model. The author presented a first estimation algorithm using a 3D irradiance map as 3D illumination model [9]. The latter approach assumes known shape and motion of the objects. In this work the latter approach is refined and embedded in the image analysis proposed by Ostermann [4]. The shape and motion of the objects and the scene illumination are automatically estimated. A new source model "moving rigid 3D objects with nonuniform local illumination" (3Dn) is introduced.

2 Source Model

The model objects belong to the model scene viewed by a model camera and illuminated by arbitrarily shaped distant light sources, thus parallel light incidence can be assumed. Here the photometric aspects of the source model are described. The geometric aspects did not change and are described in [4].

Assuming local, white and time–invariant illumination, diffuse reflection, and time–invariant viewing direction, the radiance $L_{r,\lambda}$ reflected from an object surface depends on the incident irradiance E_i and the object surface reflectance function R according to

$$L_{r,\lambda}(N,\lambda) = R(\lambda)\ E_i(N) \tag{2.1}$$

N is the surface normal vector and λ the wave length. The luminance s_Y of the output signal of camera is

$$s_Y = t_Y \, E_i(N) \qquad with \quad t_Y = \tilde{c} \int_{\lambda=-\infty}^{\infty} \alpha(\lambda) \; R(\lambda) \; d\lambda \qquad \tilde{c} = const. \tag{2.2}$$

Here $\alpha(\lambda)$ is the camera luminance responsivity and t_Y the effective luminance reflectance.

The luminance colour parameter of an object surface patch is taken at time instant 0 from the first image $s_{Y,0}$ of an image sequence and stored together with the surface patch normal vector N_0. To generate a model image at time instant $k+1$, the colour parameters are projected onto the model image plane. They are irradiance compensated as shown for the luminance parameter:

$$s'_{Y,k+1} = s_{Y,0} \;\; \gamma(N_{k+1}, N_0) = t_Y \;\; E_i(N_{k+1}) \;\; with \;\; \gamma(N_{k+1}, N_0) = \frac{E_i(N_{k+1})}{E_i(N_0)} \tag{2.3}$$

Here, $s'_{Y,k+1}$ is the model camera output luminance, N_{k+1} the normal vector of the surface patch at time instant $k+1$ and $\gamma(N_{k+1}, N_0)$ the irradiance compensation due to surface patch rotation from N_0 to N_{k+1}.

The irradiance $E_i(N)$ is modelled using a discrete irradiance map. The irradiance map is a reflectance map for a matt surface. The irradiance map M(i,j) of I $\times$ I fields is defined as an array of camera output signal luminances depending on the discretisized surface normal vector of a viewed fictive matt sphere providing uniform reflectance $R(\lambda) = R_0(\lambda)$ all over its surface:

$$M(i,j) := c \;\; E_i(\sigma^{(i)}, \tau^{(j)}) \qquad c = \tilde{c} \int_{-\infty}^{\infty} \alpha(\lambda) R_0(\lambda) d\lambda \qquad i,j \in \; [0;I-1] \tag{2.4}$$

$\sigma^{(i)}$ and $\tau^{(j)}$ are the in an π–interval uniformly quantized polar and azimuth angles σ and τ of the surface normal vector N of a half Gaussian sphere, i and j are the corresponding indices of the irradiance map. The described illumination model does not include pre–knowledge about the geometry of the scene illumination.

3 Image Analysis

The shape parameters of the model objects are evaluated from their silhouettes by a distance transform [4]. The colour parameters of the model objects are taken from the first image of the sequence. The 3D motion is estimated for each real image $s_{Y,k+1}$, minimizing the MSE between real image and model image $s'_{Y,k+1}$. A gradient method is applied [4]. The remaining differences are assumed to be due to nonuniform illumination only. The illumination is estimated as follows. For a moving model–object–surface–patch with the surface normal vector (σ_0, τ_0) at colour parameter storage time and $(\sigma_{k+1}, \tau_{k+1})$ after motion, the average camera–output–signal luminances $\overline{s_{Y,k+1}}$ and $\overline{s'_{Y,k+1}}$ of real and model image, respectively, are evaluated. For this purpose the model image $s'_{Y,k+1}$ is synthesized without irradiance compensation, i.e. setting $\gamma(N_{k+1}, N_0)$ in Eq. (2.3) to 1. Only image points belonging to the patch are taken into account according to

$$\overline{s_{Y,k+1}} = \frac{1}{A_{k+1}} \sum_{(X,Y) \in A_{k+1}} s_{Y,k+1}(X,Y) \tag{3.1}$$

X and Y are image plane coordinates and A_{k+1} is the image plane area covered by the surface patch at time instant $k+1$. To make use of illumination information from former analysis steps, Eq. (3.1) is extended accortding to

$$\overline{s_{Y,k+1}} = \frac{\sum_{(X,Y)\in A_{k+1}} s_{Y,k+1}(X,Y) + \sum_{n=0}^{M-1} A_{l(n)} \quad \overline{s_{Y,k}}}{\sum_{n=0}^{M} A_{l(n)}} \tag{3.2}$$

Here *l(n)* is one of the say *M* former time instants where the discretisized surface normal vector of the patch was the same as at the current time instant *l(M)=k+1*. $\overline{s'_{Y,k+1}}$ is evaluated equivalently. The ratio of the averaged luminances $\overline{s_{Y,k+1}}$ and $\overline{s'_{Y,k+1}}$ is a hint to the change of incident irradiance of that surface patch due to its motion. Attaching the ratio to the corresponding fields of the irradiance map leads to

$$\frac{E_i(N_{k+1})}{E_i(N_0)} \approx \frac{M(i_{k+1},j_{k+1})}{M(i_0,j_0)} \stackrel{!}{=} \frac{\overline{s_{Y,k+1}}}{\overline{s'_{Y,k+1}}} \tag{3.3}$$

The indices $i_0,\ j_0,\ i_{k+1},\ j_{k+1}$ correspond to $\sigma_0,\ \tau_0,\ \sigma_{k+1},\ \tau_{k+1}$, respectively, and describe the surface patch normal vector changing from N_0 to N_{k+1}. Collecting dependencies according to Eq. (3.3) for all patches of the object surface, a linear homogeneous equation system for the fields of the irradiance map arises:

$$M(i,j) - a_{ijvw}\, M(v,w) = 0 \qquad a_{ijij} = 1 \qquad i,j,v,w \in [0, I\text{–}1] \tag{3.4}$$

Not all equations (3.4) cover all fields of the irradiance map. This may happen due to two reasons. First if there are edges in the shape of the model object, where the surface normal changes more than π/I. Secondly if a rotation angle is bigger than π/I . Eqs. (3.4) are solved for least error grouped to clusters of equations joining the same unknown irradiance map fields. For each cluster one field is set to 1 as a side condition. The clusters are adjusted to each other by applying scaling factors. Undefined fields are interpolated. Adjustment and interpolation are necessary due to the applied Phong shading technique for model image synthesis. Such techniques require a group of neighboured irradiance map fields that may belong to diffrent clusters or include undefined fields. Fig. 4.4 shows the clusters of an estimated irradiance map.

To lower influence of slight motion and shape estimation errors, heuristic spatial and temporal weighting factors, respectively, are introduced into Eq. (3.2) according to

$$\overline{s_{Y,k+1}} = \frac{m \sum_{(X,Y)\in A_{k+1}} w_{k+1}(X,Y)\ s_{Y,k+1}(X,Y) + (1-m) \sum_{n=0}^{M-1} m(1\text{–}m)^{M-1-n} \sum_{(X,Y)\in A_{l(n)}} w_{l(n)}(X,Y) \quad \overline{s_{Y,k}}}{\sum_{n=0}^{M} m\ (1\text{–}m)^{M-n} \sum_{(X,Y)\in A_{l(n)}} w_{l(n)}(X,Y)} \tag{3.5}$$

$m \in [0,1]$ is a temporal weighting factor and $w_l(X,Y), l \in [0, k+1]$, are spatial weighting factors. Eq. (3.5) becomes Eq. (3.2) setting *m=0.5* and $w_l(X,Y)=1$. Influence of slight shape estimation errors can be reduced by temporal weighting. Note that a wrong model object shape appears to the image analysis as unstationary illumination, if the corrosponding object rotates. Applying an adequate temporal weighting factor *m*, the speed of adaption to temporally changing illumination can be set. The influence of slight motion estimation errors can be reduced by spatially weighting. In case of motion estimation errors the model image does not match the real image properly. The spatial weighting factors $w_l(X,Y)$ are derived from the reciprocal of the absolute amplitude of the image gradient at position (X,Y). This lowers the influence of points near texture edges that give wrong information in presence of slight motion estimation errors.

4 Experimental results

Interframe predicion experiments were done using real and animated image seqeuences shown in Fig. 4.1. The real image sequence "Ball" of 5 images shows a moving textured ball

illuminated by a spot light and ambient light. The artificially animated image sequences contain each 30 images. These sequences are designed to be close to real video telephone sequences. They show a person like rigid 3D model object rotating around several axes up to a maximum angle α_{max} away from the initial position. This artificial scene is illuminated by a fixed point light with a relative intensity of c compared to ambient light with the relative intensity $1-c$. The surface colour of the model object is taken from a real video telephone sequence. Sequences for several values of α_{max} and c were generated.

(a) (b)

Figure 4.1: First image of the real image sequence "Ball" (a) and first image of one of the artificially animated image sequences (b)

The image sequences were automatically analyzed based on the source models 3Dd or 3Dn for comparison, respectively. For the real image sequence the object shape was automatically estimated [4], for the animated sequences the object shape was known. The object colour was taken from the first image of a seqeuence. In case of source model 3Dd motion was estimated and compensated for each image. In case of source model 3Dn motion and illumination were estimated and compensated for each image. Using the compensated model objects, object–based interframe prediction was carried out for each image. The mean square prediction error was avaraged over each image sequence. Analyzing the real image sequence m was set to 0.9 else to 0.5. The size of the irradiance map is $I=9$.

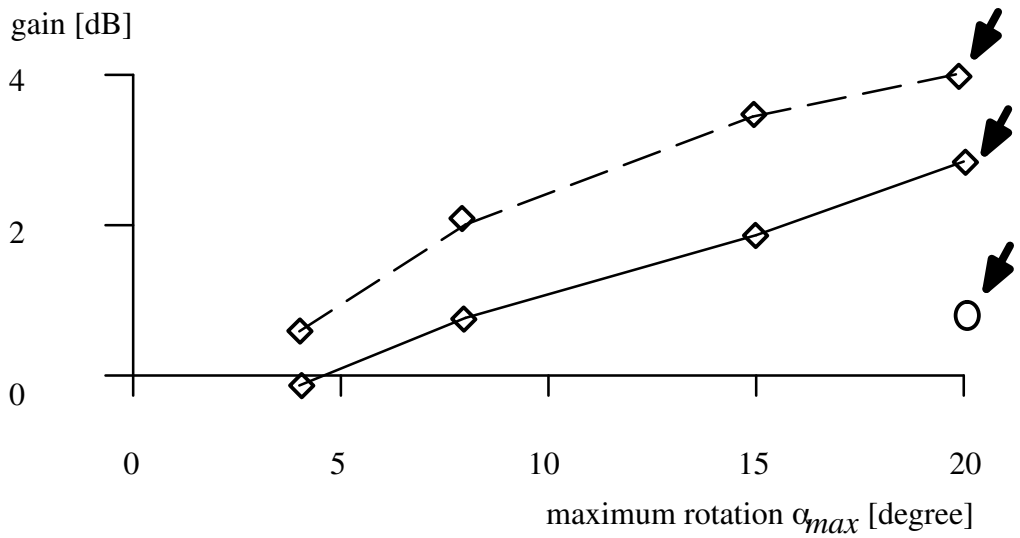


Figure 4.2: Achieved (——) and maximum possible (– –) 3D–object–based interframe prediction gain of source model 3Dn over source model 3Dd for $c=0.3$ analyzing artificially animated (◇) and real (○) image sequences. The text discusses the measurements marked by ➔.

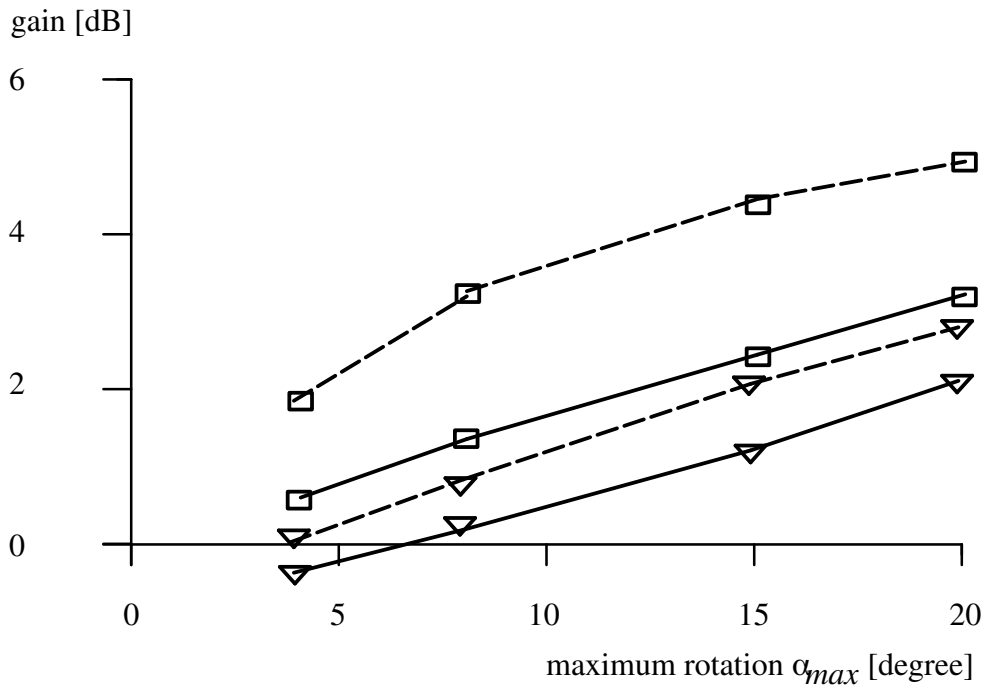
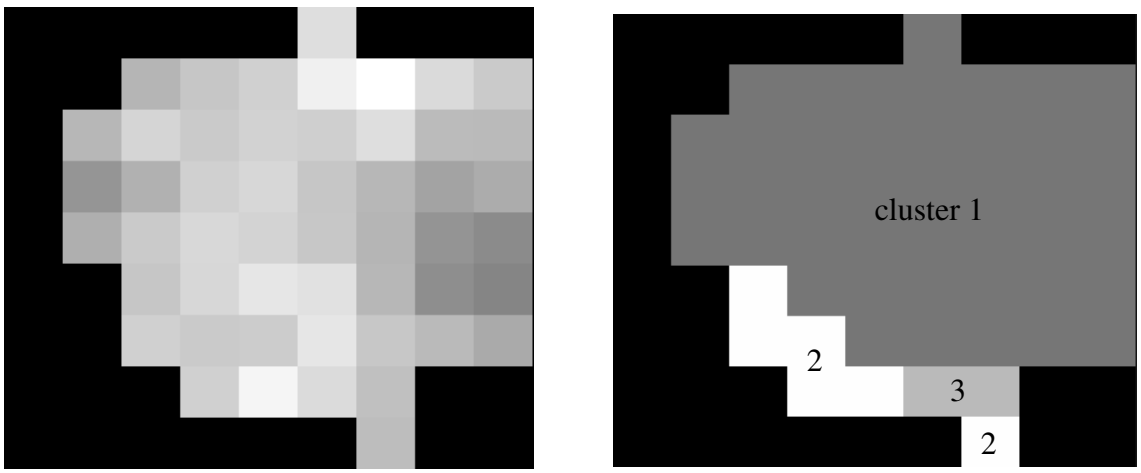


Figure 4.3: Achieved (——) and maximum possible (– –) 3D–object–based interframe prediction gain of source model 3Dn over source model 3Dd for *c=0.2* (▽) and *c=0.4* (□) analyzing artificially animated image sequences



(a) (b)

Figure 4.4: An irradiance map (a) automatically estimated from 5 images of the real image sequence "Ball" consisting of three clusters (b) marked by grey levels. Black areas have to be interpolated.

In Figs. 4.2 and 4.3 achieved and maximum possible object–based interframe prediction gain of source model 3Dn over 3Dd is shown for the animated image sequences. The reached gain is the comparison between the prediction errors applying the estimated motion for 3Dd and the estimated motion and illumination for 3Dn. For the maximum possible gain the true illumination is applied for 3Dn instead of the estimated.

For the experiment analyzing the real image sequence "Ball" c and $\alpha_{\max}$ were approximately evaluated from the estimated motion parameters and the estimated irradiance map shown in Fig. 4.4.

For the real image sequence "Ball" using source model 3Dn a gain in interframe prediction error of 0.9 dB compared to source model 3Dd is stated (See arrows in Fig. 4.2). If the object shape is known and distortions by the camera, interreflections and other neglected effects are zero, 2 dB more can be gained, prooved by animated image sequences. Another 1 dB can be gained by an improved estimation technique. Generally a positive gain can be expected, if the nonuniform portion of scene illumination exceeds 20% and the objects rotate more than 10^{o} over the image sequence.

5 Conclusion

A new method for the automatic estimation of nonuniform scene illumination is proposed. The illumination is modelled by a discrete 3D irradiance map, including no pre–knowledge about shape and number of light sources. The irradiance map is estimated evaluating remaining errors after motion compensated prediction. The estimation technique is embedded in the image analysis of OBASC extending the source model from "moving rigid 3D objects with diffuse illumination" (3Dd) to "moving rigid 3D objects with nonuniform local illumination" (3Dn). Interframe object–based prediction experiments state a gain in prediction error, if the source model 3Dn is applied instead of 3Dd. The gain depends on the amount of object motion and the portion of nonuniform scene illumination. For a real image sequence showing a moving rigid object a gain of 1dB is stated.

6 References

[1] Lee, K.Y., Pearson, D.E.: "Luminance compensation in model–based–prediction", PCS 1991, pp. 275–276.

[2] Musmann, H. G.; Hötter, M.; Ostermann, J.: "Object–oriented analysis–synthesis coding of moving images", Signal Processing: *Image Communication*, Vol. 1, No. 2, November 1989 pp 117–137.

[3] Ostermann, J.: "Modelling of 3D moving objects for an analysis–synthesis coder", Proc. SPIE, Vol. 1260, 1990, pp. 240–249

[4] Ostermann, J.: "Object–Based Analysis–Synthesis Coding (OBASC) based on the source model of moving rigid 3D objects", submitted to Signal Processing Image Communication

[5] Moloney, C. R., Dubois, E.: "Estimation of motion fields from image sequences with illumination variation", Proc. ICASSP, Vol. 4, 1991, pp. 2425–2428

[6] Mukawa, N.: "Estimation of light source information from image sequence", Systems and Computers in Japan, Vol. 23, No. 10, 1992, pp. 92–99

[7] Pentland, A. P.: "Linear shape from shading", Int'l J. of Comp. Vis., No. 4, 1990, pp. 153–162

[8] Pentland, A.: "Photometric motion", IEEE Trans. on Patt. Anal. and Mach. Intell., Vol. 13, No. 9, September 1991, pp. 879–890

[9] Stauder, J.: "Estimation of nonuniform irradiance for object–oriented coding based on 3D objects", Picture Coding Symposium, Lausanne, Switzerland, March 1993, p. 18.2