Joint Estimation of Object Motion and Point Light Illumination for 3D–Object–Based Analysis–Synthesis Coding

Jürgen 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 scene illumination has to be expected. In this contribution, non–diffuse scene illumination and 3D object motion is estimated concurrently from a monocular image sequence. Image analysis uses the source model "3D rigid moving objects with non–diffuse illumination". The illumination is modeled by a distant point light source and ambient light. Experimental results of estimating and compensating non–diffuse scene illumination and 3D object motion using real and artificially animated image sequences are given. For a real image sequence, a prediction gain of about 0.6 dB can be stated compared to the source model "3D rigid moving objects with diffuse illumination".

1 Introduction

In Object–Based Analysis–Synthesis Coding (OBASC) [3] three dimensional information has to be 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 [4] based on the source model "3D rigid moving objects with diffuse illumination" (3Dd). The images are segmented into static background and moving objects. 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 by a mesh of triangles, the position is changed by motion that is represented by six 3D motion parameters for rotation and translation according to x–, y– and z–axis, the colour is represented by luminance and chrominance parameters and is mapped on each triangle. For every real image of the image sequence a model image is synthesized by projection of 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 in video telephony, model failures occur, i.e. regions of a real image differ significantly from the model image. Thus non–diffuse scene 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 [5]. Going away from uniform coloured and still objects, a few approaches exist using 2D displacement and 2D illumination and reflection models [1] or 2D displacement and 3D illumination models [2][6]. The author proposed an estimation algorithm using 3D motion and an irradiance map as 3D illumination model [9]. In this work 3D motion and a distant point light source as 3D illumination model are applied. Only the shape of the objects is assumed to be known, whereas their motion and the scene illumination is automatically estimated. A shape estimation technique for OBASC was proposed by Ostermann [4].

2 Source model

The source model "3D rigid moving objects with non–diffuse illumination" (3Dn) is applied. The model objects build a model scene viewed by a model camera and illuminated by a model illumination. Here the photometric part of the source model is described. 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

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

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

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

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

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

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

Here, $s'_{Y,k+1}$ is the model camera output signal, 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 as a distant point light source combined with diffuse ambient light according to

$$E_i(N) = c_{AMBIENT} + c_{LAMBERT} \;\; \max\{\; 0, LN \;\} \tag{2.4}$$

Here L is the unit point light direction vector, N the unit surface normal vector, $c_{AMBIENT}$ is the ambient irradiance and $c_{LAMBERT}$ the maximum point light source irradiance in case of perpendicular incidence. Eq. (2.4) is part of the Phong reflection model used in early computer graphics. Eq. (2.4) is a local illumination model, because neither the distance between light source and surface patch nor shadowing or mutual illumination are included.

3 Image Analysis

In this section only the estimation of motion and illumination parameters is derived. Shape and colour parameters of the model object are assumed to be already estimated [4] and correct. The four illumination parameters L, $c_{AMBIENT}$ and $c_{LAMBERT}$ (L is a unit vector defined by two spherical angles) are estimated together with the motion parameters. Therefore *J+I* observation points $O^{(j)} = (P_k^{(j)}, N_k^{(j)}, N_0^{(j)}, s_{Y,0}^{(j)}, g_0^{(j)})$, *j∈[0,J+I–1]* are selected on the model object surface, where $P_k^{(j)}$ is the current 3D position, $N_k^{(j)}$ the current surface normal vector, $s_{Y,0}^{(j)}$ the object luminance parameter, $g_0^{(j)}$ the gradient measured at the corresponding 2D position $p_k^{(j)}$ in the image that was used for colour parameter storage, and N_0 the surface normal vector at colour parameter storage time. N_k and N_0 are interpolated over a triangle using Phong Shading [7]. J observation points with high and I observation points with low image gradients are selected giving cues for motion and illumination, respectively.

For each observation point an equation is set up applying a gradient method, used earlier for 3D motion estimation [4]. The equation describes the relation between the current 2D position $p_k^{(j)}$ of an observation point $O^{(j)}$ and the unknown new position $p_{k+1}^{(j)}$. The luminance $s_{Y,k+1}(p_k^{(j)})$ is known, whereas the luminance $s_{Y,k+1}(p_{k+1}^{(j)})$ is unknown but can be stated to be

$$s_{Y,k+1}(p_{k+1}^{(j)}) = s_{Y,0}^{(j)} \;\; \gamma \;\; (N_0^{(j)}, N_{k+1}^{(j)}) \tag{3.1}$$

with N_{k+1} the unknown rotated normal vector N_k. Eq. (3.1) expresses an irradiance compensation for the observation point luminance due to an unknown rotation. The difference between these luminances can be explained by applying a linear signal model to $s_{Y,k+1}$ using the unknown displacement $p_{k+1}^{(j)} - p_k^{(j)}$ and the gradient $g^{(j)} = constant$ of the image $s_{Y,k+1}$:

$$s_{Y,0}^{(j)} \;\; \gamma \;\; (N_0^{(j)}, N_{k+1}^{(j)}) \;\; - \;\; s_{Y,k+1}(p_k^{(j)}) \;\; = \;\; g^{(j)} \;\; \left(p_{k+1}^{(j)} - p_k^{(j)}\right) \quad j \in [0, J + I - 1] \tag{3.2}$$

Eq. (3.2) has to be solved for the motion parameters hidden in $N_{k+1}^{(j)}$ and $p_{k+1}^{(j)}$ and for the illumination parameters hidden in γ. Because E_i is used in γ only relatively, one illumination parameter can be eliminated by estimating only the ratio $c = \frac{c_{LAMBERT}}{c_{AMBIENT}}$ and L, resulting in three free illumination parameters.

Further on it is advisable to reduce the number of illumination parameters to two. Schurbohm [8] showed that estimating three parameters leads to instable results. For explanation, in Fig. 3.1 a rotating object under two different illuminations is shown. If the illumination direction is parallel

to the rotation axis, the irradiance on any object surface patch will not change due to rotation. If the illumination direction is perpendicular to the rotation axis, the irradiance on an object surface patch will change heavily due to motion. Generally the component of the illumination direction parallel to the rotation axis has less effect on the change of irradiance than the other components. Therefore L is represented by spheric coordinates in an orthogonal rotation axis oriented coordinate system $U = (u_1, u_2, u_3)$, with the base vector u_3 parallel to the rotation axis of the unknown rotation, according to $L = \cos\phi \sin\theta u_1 + \sin\phi \sin\theta u_2 + \cos\theta u_3$. Focussing on the irradiance compensation γ in Eq. (3.2), it can be shown, that c and θ are not independent, whereas θ represents the component parallel to the rotation axis and has less effect than c. Thus θ is held constant during estimation, only c and ϕ are estimated.

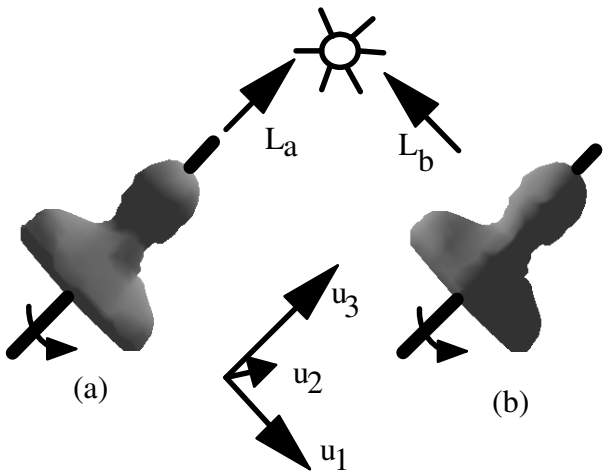

Figure 3.1: An illuminated rotating object under two illumination situations is shown: The illumination direction is (a) parallel to the rotation axis or (b) perpendicular. A rotation axis oriented coordinate system is shown.

Motion and illumination are estimated alternating iteratively. For each motion iteration, J equations from Eqs. (3.2) are approximated and linearized assuming little motion. They are solved for 6 motion parameters for translation and rotation using the method of least squares. After each motion iteration the estimated motion is applied to the model objects. For each illumination iteration, c, ϕ are substituted by $c^0 + c', \phi^0 + \phi'$ in I equations from Eqs. (3.2) with c^0, ϕ^0 the current illumination parameters and c', ϕ' the changes to be estimated. At the beginning of the image sequence the illumination parameters are set to $c^0 = 0.1$ and L^0 = viewing direction. The I equations are linearized with respect to c', ϕ' and solved using the method of least squares. The iteration is necessary due to the mentioned approximation and linearizations. Especially Eq. (2.4) is heavily nonlinear.

4 Results

Interframe prediction experiments were carried out using real and artificially animated image sequences. The real image sequence "Ball" containing 5 images shows a moving textured ball illuminated by a spot light and ambient light. The artificially animated sequences contain each 30 images and are designed to be close to real video telephone sequences. They show a person like 3D rigid model object rotating around several axes up to a maximum angle $\alpha_{\max}$ 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 colour of the model object is taken from a real video telephone sequence.

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 sequence. 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, model–based interframe prediction was carried out for each image. The mean square prediction error was avaraged over each image sequence. The prediction experiments differ according to the maximum angle of rotation $\alpha_{\max}$ and the relative intensity c of the point light source.

In Fig. 4.1 achieved and maximum possible model–based interframe prediction gain of source model 3Dn over 3Dd is shown. The achieved 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 assumed to be known in advance for 3Dn. Depending on the maximum rotation angle $\alpha_{\max}$ the prediction gain is about 2.8dB for a relative intensity

c=*0.6* of the point light source. The maximum possible gain is up to 7.5dB. For the real image sequence "Ball" using source model 3Dn a gain in interframe prediction error of 0.6 dB compared to source model 3Dd was measured. For the real image sequence "Ball" $c \approx 0.6$ and $\alpha_{max} \approx 20$ were approximately evaluated from the estimated motion and illumination parameters. For each estimation cycle about 30 iterations are necessary. Convergence problems cause the gap between achieved and maximum possible gain in Fig. 4.1. Errors in the estimated shape is the main reason for the gap between the achieved gain for the real and the artificially animated image sequences.

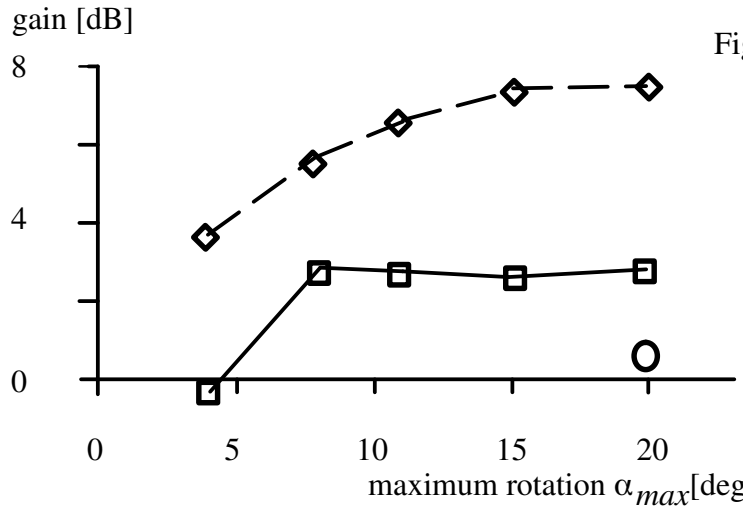

Figure 4.1: Achieved (□) and maximum possible (◇) 3D–object–based interframe prediction gain of source model 3Dn over source model 3Dd for c=*0.6* analyzing artificially animated image sequences; Achieved gain (○) for the real image sequence "Ball" for $c \approx 0.6$

5 Conclusion

A new method for the automatic estimation of 3D object motion and non–diffuse scene illumination is proposed. The illumination is modelled by a distant point light source and ambient light. Motion and illumination are estimated concurrently by an iterative gradient method. 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 non–diffuse 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 non–diffuse scene illumination. For a real image sequence showing a moving rigid object a gain of 0.6dB is stated.

6 References

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

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

[3] 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.

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

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

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

[7] Phong., B.: "Illumination for computer generated pictures", Commun. ACM, vol. 18, 1975, pp. 311–317.

[8] Schurbohm, E.: "Schätzung von 3D–Bewegung und Beleuchtung unter Verwendung des Modells Lambertscher Reflexion und lokaler Beleuchtung", diploma thesis, Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, Germany, 1993.

[9] Stauder, J.: "An illumination estimation method for 3D–Object–Based Analysis–Synthesis Coding", COST 211 ter European Workshop on New Techniques for Coding of Video Signals at Very Low Bitrates, Hannover, Germany, December 1993.