OBJECT ROTATION AXIS FROM SHADING

Jürgen Stauder

IRISA/INRIA
Campus Universitaire de Beaulieu, F-35042 Rennes Cedex, France
jurgen.stauder@irisa.fr, http://www.irisa.fr/temics

(Presented at ICASSP'99, 15.-19.3.1999, Phoenix, Arizona, USA, Vol. 6, pp. 3285-3288)

ABSTRACT

In this paper, a non-complex estimator is developed for the tilt angle of the rotation axis of an object that is illuminated by a point light source. The tilt angle defines the orientation of the 2D projection of the 3D object rotation axis in the image plane and is a strong clue for image understanding. The estimator evaluates two images of a video image sequence showing a moving object. Additionally, only a displacement vector field and the 2D object silhouette are required. No 3D information is required. Therefore, the object is assumed to be rigid, to be matte, to have equally distributed surface normals and to be illuminated by a distant point light source and ambient light. For estimation, the displaced frame ratio (DFR), i.e. the frame ratio after motion compensation, is evaluated statistically. The DFR depends only on the photometric effect of temporally changing object shading. Experimental results with real images show the proper performance of the derived estimator for real objects. A demo can be found at http://www.irisa.fr/prive/Jurgen.Stauder

1. INTRODUCTION

The estimation of rigid object motion relatively to a capturing video camera is a classical problem. The rigid motion is therefore described by 3D rotation and 3D translation, i.e. by six degrees of freedom.

The 3D object motion and 3D object shape can be estimated from the 2D positions of corresponding image points in succeeding images of a video image sequence in case of known and unknown intrinsic camera parameters, see [5] and [18] for reviews, respectively. In these approaches, pure geometric information is exploited. The object motion estimation problem can be simplified, if the 3D object shape is described by a parametric shape model [6] or if the 3D object shape is even known in advance [7][12]. In these approaches, additionally to pure geometric information, the image signal itself is exploited directly by applying a gradient method [12].

In the approaches mentioned until now, photometric effects in the video images like shading, cast shadows and specular reflections have been either avoided (by using robust point correspondences) or neglected (during exploitation of the image signal). It is assumed that a moving object preserves its brightness during motion, see [11] for a discussion.

In presence of a strong, point-like light source, such photometric effects are strong and contain information on motion and shape of the objects, shown e.g. in [14].

To utilize photometric effects, the 3D object shape may be estimated from shading [9][20], from specular reflections [19] or from cast shadows [15][8][4]. In these approaches, the objects are either static [9][20][8][4] or the motion is known in advance from a turntable [19]. Further, the objects are often assumed to be unicolored, i.e. having a spatially uniform reflectance [9][20][8].

To utilize photometric effects for 3D object motion estimation, the usage of temporally changing object shading has been proposed. For example, the shading of a persons face changes, if the person turns his face away from the point light source. Pentland [14] presented in 1991 an approach for shape estimation from nothing than temporally changing shading. For 3D motion estimation, in [16][3] temporally changing shading has been taken into account during evaluation of the image signal by a gradient method. Thus, photometric *and* geometric effects are used. In these approaches, the 3D object shape is assumed to be known in advance. Because a gradient method tracks the motion of a local image region independently from other image regions, the objects are allowed to be multi-colored, i.e. they may have a spatially varying reflectance.

In this paper, the 3D motion of an object is estimated from the photometric effect of temporally changing object shading, only. More specifically, one of the six degrees of freedom of object motion is estimated. The estimated entity is the tilt angle of the object rotation axis that defines the orientation of the projection of the object rotation axis in the image plane. The projected object rotation axis is a powerful key to image understanding.

The approach and its relation to the literature are as follows. Whereas in [5][18][16][3] mainly geometric effects are evaluated for motion estimation, in this approach nothing than the photometric effect of temporally changing object shading is evaluated. Whereas in [5][18] 3D motion and 3D shape are estimated jointly, in this paper only motion is estimated by a less complex method. In opposite to [7][12][16][3], the 3D object shape is not assumed to be known in advance. Instead, the assumption of equally distributed surface normals is applied, known from illumination estimation approaches [13][9][20]. To overcome the spatially varying, but unknown object reflectance, two images will be evaluated. Inspired by Pentland [14], the images will be motion compensated (using a given displacement vector field) and combined by a non-linear processing to eliminate the unknown reflectance. In this paper, as non-linear processing the *displaced frame ratio* (DFR) is introduced. It will be shown theoretically, that there is a strong link between the mean spatial gradient of the DFR inside the 2D object silhouette and the object rotation axis tilt.

The paper is organized as follows. In Section 2, the displaced frame ratio (DFR) will be introduced. In Section 3, a method for rotation axis tilt estimation from DFR observations will be developed. In Section 4, experimental results for real images will be presented.

2. DISPLACED FRAME RATIO

In this section, the displaced frame ratio (DFR) will be introduced as observation entity of the estimator. To measure the DFR, a previous image at time instant $k-1$ with the luminance

$$s_{k-1}(\mathbf{p}) \;=\; \eta(\mathbf{P})\, E(\mathbf{N}) \qquad (1)$$

and a current image at time instant k with the luminance $s_k(\mathbf{p})$ are needed. In Eq. 1, $\mathbf{p}^T = (x, y)$ is a 2D image position, $\mathbf{N}$ the normal of the object surface patch at the 3D position $\mathbf{P}$ that is visible at $\mathbf{p}$, $\eta(\mathbf{P})$ the reflectance of the surface patch and $E(\mathbf{N})$ the irradiance, i.e. the perceived light power per object surface. The irradiance term $E(\mathbf{N})$ describes the shading on the object surface depending on the normal $\mathbf{N}$ [17].

To divide out the unknown but spatially varying object reflectance from the image luminances, the current image is motion compensated by a given displacement vector field $\mathbf{d}_k(\mathbf{p})$ and related to the previous image resulting to the DFR

$$dfr(\mathbf{p}) \;=\; \frac{s_k(\mathbf{p}-\mathbf{d}_k(\mathbf{p}))}{s_{k-1}(\mathbf{p})} \;=\; \frac{E(\underline{\mathbf{R}}\mathbf{N})}{E(\mathbf{N})}\,, \qquad (2)$$

with $\underline{\mathbf{R}}$ the rotation matrix describing the rotational part of the object motion from time instant $k-1$ to k. The rotation matrix can be described by a unit rotation axis vector $\mathbf{R}$ and a rotation angle β [10]. The vector $\mathbf{R}$, further denoted as rotation axis, can be defined by a tilt angle τ_R and a slant angle η_R as shown in Fig. 1. The DFR describes the photometric effect of temporally changing object shading.

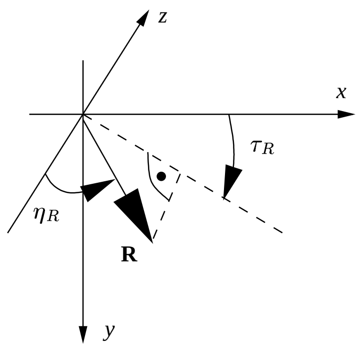

Figure 1: Definition of the rotation axis vector $\mathbf{R}$ in the word coordinate system (x, y, z): The tilt angle τ_R defines the orientation of the projection of $\mathbf{R}$ into the xy-plane (i.e. image plane), whereas the slant η_R is the angle between the z-axis (i.e. viewing direction) and $\mathbf{R}$.

For a scene illumination by a distant point light source and ambient diffuse light, the irradiance

$$E(\mathbf{N}) \;=\; c\; (1 + \mu\, \mathbf{L}\mathbf{N}) \qquad (3)$$

at an object surface patch is defined - up to a camera dependent constant c - by the surface normal $\mathbf{N}$ and the weighted illumination direction $\mathbf{L}$. The direction of $\mathbf{L}$ is defined by the angles τ_L and η_L as shown for $\mathbf{R}$ in Fig. 1. $|\mathbf{L}| = e$ is the ratio between point light source and ambient light intensities [17]. In Eq. 3, the factor μ equals one, if the point light source illuminates the object surface. Else, at surface regions of self shadow, μ equals zero.

3. OBJECT ROTATION AXIS TILT ESTIMATION

In this section, an estimator of the rotation axis tilt angle will be derived analytically. Therefore, the de-

$$\frac{\overline{g_y}}{\overline{g_x}} = \frac{\begin{array}{c} 2e\cos\eta_L\sin\eta_L\cos\tau_L\cos\eta_R + 2e\sin\eta_R\cos\tau_R\cos^2\eta_L + 2e\sin^2\eta_L\sin^2\tau_L\sin\eta_R\cos\tau_R \\ +3\cos\eta_L\sin\eta_R\cos\tau_R + 3\sin\eta_L\cos\tau_L\cos\eta_R - 2e\sin^2\eta_L\sin\tau_L\cos\tau_L\sin\eta_R\sin\tau_R \end{array}}{\begin{array}{c} -2e\cos^2\eta_L\sin\eta_R\sin\tau_R - 2\sin\eta_L e\cos\eta_L\sin\tau_L\cos\eta_R + 2e\sin^2\eta_L\cos\tau_L\sin\tau_L\sin\eta_R\cos\tau_R \\ -2e\sin^2\eta_L\cos^2\tau_L\sin\eta_R\sin\tau_R - 3\cos\eta_L\sin\eta_R\sin\tau_R - 3\sin\eta_L\sin\tau_L\cos\eta_R \end{array}} \tag{4}$$

scription of the DFR according to Eq. 2 is simplified using the assumption that the

- *object rotation is small,*

such that Eq. 2 can be linearized with respect to the rotation angle β. Using the further assumption that the

- *object surface can be locally approximated by a spheric patch,*

the spatial gradient $\nabla dfr(\mathbf{p}) = (g_x(\mathbf{p}), g_y(\mathbf{p}))^T$ of the DFR can be derived similarly as shown in [9] for the image luminance. $\nabla^T = (\partial/\partial x\,, \partial/\partial y)$ is the gradient operator.

Following an idea of Pentland [13], who calculated the spatial means of luminance gradients to recover the point light source direction tilt, the DFR gradients are averaged all over the 2D object silhouette that has to be known in advance. Using Pentland's assumption that the

- *object surface normals are equally distributed*

(see [9] for formulation), assuming further that the

- *point light source is weak,*

such that $1 + \mathbf{LN} \approx 1$ holds, and assuming finally that

- *self shadowing is negligible,*

such that $\mu = 1$ in Eq. 3 (assumption used implicitly also in [13][9]), the DFR gradient can be analytically integrated over the 2D object silhouette. If the y component $\overline{g_y}$ of the mean gradient is divided by its x component $\overline{g_x}$, the analytic calculation gives the ratio shown in Eq. 4 at top of this page. Assuming that

- *illumination is in viewing direction,*

i.e. $\eta_L = 0$, the ratio in Eq. 4 simplifies to $\overline{g_y}/\overline{g_x} = -\cos\tau_R/\sin\tau_R$, such that the final rotation axis tilt estimator is

$$\hat{\tau}_R = \frac{\pi}{2} + \begin{cases} atan(\hat{\overline{g_y}}/\hat{\overline{g_x}}) & \\ atan(\hat{\overline{g_y}}/\hat{\overline{g_x}}) + \pi & \\ \pi/2 & \\ 3\pi/2 & \end{cases} \quad if \quad \begin{array}{l} \hat{\overline{g_x}} > 0 \\ \hat{\overline{g_x}} < 0 \\ \hat{\overline{g_x}} = 0 \text{ or } \hat{\overline{g_y}} > 0 \\ \hat{\overline{g_x}} = 0 \text{ or } \hat{\overline{g_y}} < 0 \end{array} \tag{5}$$

with

$$\begin{pmatrix} \hat{\overline{g_X}} \\ \hat{\overline{g_Y}} \end{pmatrix} = \sum_{\text{object mask}} w(\mathbf{p})\; \nabla\; dfr(\mathbf{p}) \tag{6}$$

the estimate of the mean DFR gradient. The weight $w(\mathbf{p}) = 1/\sqrt{|\nabla dfr(\mathbf{p})|}$ reduces influence of motion compensation errors and neglected self shadow contours.

4. RESULTS

In Fig. 2, sample results with a real image sequence are shown. They are obtained by a) automatic object segmentation using the COST 211^{quat} image analysis model [2], b) by automatic estimation of a displacement vector field using hierarchical block matching with half pel resolution [1] and c) application of the developed rotation axis tilt estimator.

The results show a good performance even, if the object shading is weak and the illumination is absolutely not in viewing direction as assumed in Section 3. As long as the 2D object silhouettes were roughly found (80% of the images), the accuracy for the shown experiments is about ± 10 degrees by visual inspection. An analytic error analysis and a larger variety of experiments will follow in future work.

5. REFERENCES

[1] M. Bierling: "Hierarchische Displacementschätzung zur Bewegungskompensation in digitalen Fernsehbildsequenzen", Dissertation, Universität Hannover, Fortschritt-Berichte VDI, Reihe 10, No. 179, Hannover, 1991, ISBN 3-18-147910-1.

[2] COST 211^{quat} Analysis Model, http://www.teltec.dcu.ie/cost211

[3] P. Eisert, B. Girod: "Illumination compensated motion estimation for analysis synthesis coding", in B. Girod, H. Niemann, H.-P. Seidel: "3D image analysis and synthesis '96", Proceedings, 18.-19. November 1996, Erlangen, pp. 61-66.

[4] A.M. Bruckstein, R.J. Holt, Y.D. Jean, A.N. Netravali: "On the use of shadows in stance recovery", Bell Labs, Lucent Technologies, Techn. Mem., BL0113380-971222-04TM, December 1997.

Figure 2: Two sequences of three succeeding images of the test sequence "Tai", format CIF, frame rate 12.5Hz: (a) images 24,25,26 and (b) images 41,42,43. For each sequence, the second and third images have been evaluated for estimation. For each sequence, the estimated 2D projection of the 3D rotation axis is superimposed on the third image. It can be seen that the rotation axis follows the rotation of the head.

[5] T.S. Huang, A.N. Netravali: "Motion and structure from feature correspondences: a review", Proceedings of the IEEE, Vol. 82, No. 2, February 1994, pp. 252-268.

[6] Y.-P. Hung, D.B. Cooper, B. Cernuschi-Frias: "Asymptotic Bayesian surface estimation using an image sequence", Int'l Journal of Computer Vision, Vol. 6, No. 2, 1991, pp. 105-132.

[7] F. Kappei: "Modellierung und Rekonstruktion bewegter dreidimensionaler Objekte in einer Fernsehbildfolge", Dissertation, Universität Hannover. Germany, 1989.

[8] M.S. Langer, S.W.Zucker: "Shape from shading on a cloudy day", Journal of the Optical Soc. of America, Vol. 11, No.2, February 1994, pp. 467-78.

[9] C.H. Lee, A. Rosenfeld: "Improved methods of estimating shape from shading using the light coordinate system" in B.K.P. Horn, M.J. Brooks (Ed.): "Shape from shading", MIT Press, Cambridge 1989, pp. 323-347.

[10] K. Meyberg, P. Vachenhauer: "Höhere Mathematik", Volume 1, Springer-Verlag, Berlin 1993.

[11] H. Nicolas, C. Labit: "Motion and illumination variation estimation using a hierarchy of models: Application to image sequence coding", Journal of Visual Communication and Image Representation, Vol. 6, No. 4, December 1995, pp. 303-316.

[12] J. Ostermann: "Object-based analysis-synthesis coding (OBASC) based on the source model of moving rigid 3D objects", Signal Processing: Image Communication, Vol. 6, No. 2, May 1994, pp. 143-161.

[13] A.P. Pentland: "Finding the illumination direction", Journal of the Optical Society of America, Vol.. 72, No. 4, April 1982, pp. 448-455.

[14] A.P. Pentland: "Photometric Motion", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9, September 1991, pp. 879-890.

[15] S. A. Shafer: "Shadows and silhouettes in computer vision", Kluwer Academic Publishers, Dordrecht, Niederlande, 1985.

[16] J. Stauder, "Joint estimation of object motion and point light illumination for 3D-object-based analysis-synthesis coding", Int'l Workshop on Coding Techniques for Very Low Bit-rate Video, Colchester, England, 7.-8. April 1994, p. 4.1.

[17] J. Stauder: "Estimation of point light source parameters for object-based coding", Signal Processing: Image Communication, Vol. 7, No. November 4-6, 1995, pp. 355-379.

[18] Z. Zhang: "Determining the epipolar geometry and its uncertainty", Int'l Journal of Computer Vision, Vol. 27, No. 2, 1998, pp. 161-195.

[19] J.Y. Zheng, Y. Fukagawa, N. Abe: "3D surface estimation and model construction from specular motion in image sequences", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 5, May 1997, pp. 513-520.

[20] Q. Zheng, R. Chelappa: "Estimation of illuminant direction, albedo, and shape from shading", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 7, July 1991, pp. 680-702.