



Object Rotation Axis from Shading

Jürgen Stauder, Henri Nicolas

► **To cite this version:**

Jürgen Stauder, Henri Nicolas. Object Rotation Axis from Shading. [Research Report] RR-3865, INRIA. 2000. <inria-00072789>

HAL Id: inria-00072789

<https://hal.inria.fr/inria-00072789>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Object Rotation Axis from Shading

Jürgen Stauder, Henri Nicolas

No 3865

January 2000

———— THÈME 3 ————



*Rapport
de recherche*



Object Rotation Axis from Shading

Jürgen Stauder, Henri Nicolas

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Temics

Rapport de recherche n° 3865 — January 2000 — 17 pages

Abstract: In this report, an simple and new estimator is developed for the tilt angle of the rotation axis of a rigid object that is illuminated by a point light source. The tilt angle defines the orientation of the 2D projection of the 3D object unit rotation axis in the image plane. The angle is one of the 6 degrees of freedom of 3D rigid motion. It is analytically shown that the tilt angle can be estimated just by exploiting the error of motion compensation between two images due to object shading. Required input data are a dense displacement vector field for motion compensation and a 2D binary object mask. The estimator assumes rigid, Lambertian objects with equally distributed surface normals as well as a scene illumination by a distant point light source and ambient light. Experiments with head-and-shoulder video sequences yield qualitatively good results that are derived fully automatically. A performance analysis shows that the estimator works best if light comes from viewing direction. Finally, an application to video indexing shows that head-and-shoulder sequences can be fully automatically segmented into temporal segments of homogeneous object motion. A demo can be found at <http://www.irisa.fr/temics/Demos>.

Key-words: Motion estimation, shading, illumination

(Résumé : [tsvp](#))

Supported by a European Marie Curie Research Training Grant

Axe de rotation d'objet à partir de l'ombrage

Résumé : Ce rapport présente une nouvelle méthode simple permettant l'estimation de la projection dans le plan image de l'axe de rotation 3D d'un objet rigide illuminé par une source ponctuelle. Cette projection, définie par un angle tilt, correspond à un des six degrés de liberté du mouvement 3D rigide. Il est montré ici que cet angle peut être estimé en exploitant l'erreur de compensation de mouvement issue des variations d'ombrage de l'objet considéré. Cette approche utilise un champ dense de vecteurs de vitesse et un masque 2D d'objets préalablement calculés. De plus, la méthode proposée suppose que les objets sont rigides et Lambertiens avec des normales de surface distribuées également. L'illumination est supposée se composer d'une lumière ambiante et d'une source lumineuse ponctuelle à l'infini. Cette approche a été validée sur des séquences de type visioconférence pour lesquelles des résultats satisfaisants ont été obtenus. Une analyse d'efficacité montre que l'estimateur marche mieux si la direction de la source lumineuse se situe proche de l'axe de visé de la caméra. Cette méthode a été appliquée au problème de l'indexation vidéo, ce qui a permis de montrer que des séquences de type visioconférence peuvent être segmentées entièrement automatiquement en segments temporels de mouvement homogène au sens de l'angle de tilt. Différents résultats expérimentaux sont disponibles à l'adresse: <http://www.irisa.fr/temics/Demos>.

Mots-clé : Estimation de mouvement, ombrage, illumination

1 Introduction

The estimation of rigid object motion relatively to a capturing video camera is a classical problem. The 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 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 [24] 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][13]. In these approaches, additionally to pure geometric information, the image signal itself is exploited directly by applying a gradient method [13].

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 [12] for a discussion.

In presence of a dominating, point-like light source, such photometric effects are shown in [15] to be strong and to contain information on motion and shape of the objects.

To utilize photometric effects, the 3D object shape may be estimated from shading [9][26], from specular reflections [25] or from cast shadows [16][8][4]. In these approaches, the objects are either static [9][26][8][4] or the motion is known in advance from a turntable [25]. Further, the objects are often assumed to be unicolored, i.e. having a spatially uniform reflectance [9][26][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 light source. Pentland [15] presented in 1991 an approach for shape estimation from nothing than temporally changing shading. For 3D motion estimation, in [17][3] temporally changing shading has been taken into account during evaluation of the image signal by a gradient method. In these approaches, photometric *and* geometric effects are used. Further, 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 report [22][20], the 3D motion of an object is estimated from nothing than the photometric effect of temporally changing object shading. 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 2D projection of the 3D 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][24][17][3] mainly geometric effects are evaluated for motion estimation, this approach evaluates nothing than the photometric effect of temporally changing object shading. Whereas in [5][24] 3D

$$\mathbf{R} = \begin{pmatrix} \cos \beta + (1 - \cos \beta)R_x^2 & (1 - \cos \beta)R_x R_y - R_z \sin \beta & (1 - \cos \beta)R_x R_z + R_y \sin \beta \\ (1 - \cos \beta)R_x R_y + R_z \sin \beta & \cos \beta + (1 - \cos \beta)R_y^2 & (1 - \cos \beta)R_y R_z - R_x \sin \beta \\ (1 - \cos \beta)R_x R_z - R_y \sin \beta & (1 - \cos \beta)R_y R_z + R_x \sin \beta & \cos \beta + (1 - \cos \beta)R_z^2 \end{pmatrix} \quad (1)$$

motion and 3D shape are estimated jointly, in this report only motion is estimated by a less complex method. In opposite to [7][13][17][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 [14][9][26]. To overcome the spatially varying and unknown object reflectance, two images will be evaluated. Inspired by Pentland [15], 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 report, as non-linear processing the *displaced frame ratio* (DFR) is introduced. It will be shown theoretically, that there is a strong link between the DFR inside the 2D object silhouette and the object rotation axis tilt. To show that the estimated tilt angle is a useful information, the estimated tilt will be applied to video indexing [21].

The report 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 and its performance will be discussed. In Section 4, experimental results and an application to video indexing will be presented. Section 5 gives a conclusion.

2 Displaced frame ratio

In this section, the displaced frame ratio (DFR) is introduced. The DFR will be the observation for the estimator in the next section. To observe the DFR, a previous image at time instant $k - 1$ with the luminance

$$s_{k-1}(\mathbf{p}) = \eta(\mathbf{P}) E(\mathbf{N}) \quad (2)$$

and a current image at time instant k with the luminance $s_k(\mathbf{p})$ are needed. In Eq. 2, $\mathbf{p} = (x, y)^T$ 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} [18][19].

To divide out the unknown and 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. This results in the DFR

$$dfr(\mathbf{p}) = \frac{s_k(\mathbf{p} - \mathbf{d}_k(\mathbf{p}))}{s_{k-1}(\mathbf{p})} = \frac{E(\mathbf{RN})}{E(\mathbf{N})}, \quad (3)$$

with \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 $\mathbf{R} = (R_x, R_y, R_z)^T$ and a rotation angle β [11] according to Eq. 1 at the top of the previous page. The vector \mathbf{R} , further denoted as rotation axis, can be defined by a tilt angle τ_R and a slant angle η_R according to

$$\mathbf{R} = \begin{pmatrix} R_x \\ R_y \\ R_z \end{pmatrix} = \begin{pmatrix} \sin \eta_R \cos \tau_R \\ \sin \eta_R \sin \tau_R \\ -\cos \tau_R \end{pmatrix} \quad (4)$$

(see Fig). 1. The DFR describes the photometric effect of temporally changing object shading.

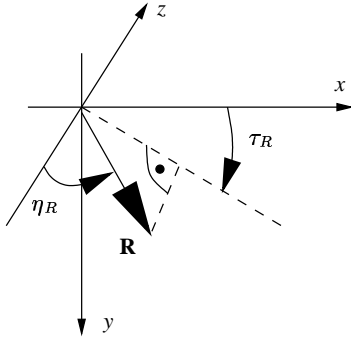


FIG. 1 – Definition of the rotation axis vector \mathbf{R} in the word coordinate system (x, y, z) : The tilt angle τ_R defines the orientation of the projection of \mathbf{R} into the xy -plane (i.e. image plane), whereas the slant η_R is the angle between the z -axis (i.e. viewing direction) and \mathbf{R}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

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

$$E(\mathbf{N}) = c (1 + \mu \mathbf{L}\mathbf{N}) \quad (5)$$

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 [18][19]. In Eq. 5, $\mu = 1$, if the point light source illuminates the object surface, else $\mu = 0$.

3 Object Rotation Axis Tilt Estimation

In this section, an estimator of the rotation axis tilt angle will be derived analytically. Therefore, the description of the DFR according to Eq. 3 is simplified using the assumption that

$$\begin{aligned}
g_x &= \frac{\beta}{N_z r} \frac{-L_x R_y N_x + L_y R_z N_z + L_y R_x N_x - L_z R_y N_z}{1 + \mathbf{L N}} \\
&+ \frac{\beta}{N_z r} \frac{(L_x N_z - L_z N_x)(-L_z R_x N_y - L_y R_z N_x + L_x R_z N_y + L_z R_y N_x - L_x R_y N_z + L_y R_x N_z)}{(1 + \mathbf{L N})^2} \\
g_y &= \frac{\beta}{N_z r} \frac{-L_x R_z N_z + L_y R_x N_y - L_x R_y N_y + L_z R_x N_z}{1 + \mathbf{L N}} \\
&+ \frac{\beta}{N_z r} \frac{(L_y N_z - L_z N_y)(-L_z R_x N_y - L_y R_z N_x + L_x R_z N_y + L_z R_y N_x - L_x R_y N_z + L_y R_x N_z)}{(1 + \mathbf{L N})^2} \tag{6}
\end{aligned}$$

A1: the object rotation is small

such that the rotation matrix from Eq. 1 can be linearized with respect to the rotation angle β yielding

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & -R_z \beta & R_y \beta \\ R_z \beta & 1 & -R_x \beta \\ -R_y \beta & R_x \beta & 1 \end{pmatrix}. \tag{7}$$

By insertion of $\tilde{\mathbf{R}}$ in Eq. 3, also the DFR

$$d\tilde{f}r(\mathbf{p}) = \frac{E(\tilde{\mathbf{R}}\mathbf{N})}{E(\mathbf{N})}, \tag{8}$$

is linear with respect to the rotation angle β . As a next step, the spatial gradients of the DFR will be derived. Using the assumption that

A2: the object surface can be locally approximated by a spherical patch,

the surface normal $\mathbf{N}(x + \Delta x, y + \Delta y)$ in a local neighborhood of a considered image position $\mathbf{p} = (x, y)^T$ can be written as

$$\mathbf{N} = \frac{1}{r} \begin{pmatrix} \Delta x \\ \Delta y \\ -\sqrt{r^2 - \Delta x^2 - \Delta y^2} \end{pmatrix}, \tag{9}$$

with r the radius of the spherical patch. This approximation has been used by Rosenfeld in a similar manner for the image luminance [9]. Now, the surface normal in Eq. 8 is replaced by the spherical approximation 9. Then Eq. 8 is derived with respect to Δx and Δy , respectively. Finally, the spherical approximation 9 is eliminated in the derivative result giving the spatial gradient

$$\nabla d\tilde{f}r(\mathbf{p}) = (g_x(\mathbf{p}), g_y(\mathbf{p}))^T \tag{10}$$

with $\nabla^T = (\partial/\partial x, \partial/\partial y)$ the gradient operator and $g_x(\mathbf{p}), g_y(\mathbf{p})$ as shown in Eq. 6.

$$\frac{\overline{g_y}}{\overline{g_x}} = \frac{2e \cos \eta_L \sin \eta_L \cos \tau_L \cos \eta_R + 2e \sin \eta_R \cos \tau_R \cos^2 \eta_L + 2e \sin^2 \eta_L \sin^2 \tau_L \sin \eta_R \cos \tau_R + 3 \cos \eta_L \sin \eta_R \cos \tau_R + 3 \sin \eta_L \cos \tau_L \cos \eta_R - 2e \sin^2 \eta_L \sin \tau_L \cos \tau_L \sin \eta_R \sin \tau_R}{-2e \cos^2 \eta_L \sin \eta_R \sin \tau_R - 2 \sin \eta_L e \cos \eta_L \sin \tau_L \cos \eta_R + 2e \sin^2 \eta_L \cos \tau_L \sin \tau_L \sin \eta_R \cos \tau_R - 2e \sin^2 \eta_L \cos^2 \tau_L \sin \eta_R \sin \tau_R - 3 \cos \eta_L \sin \eta_R \sin \tau_R - 3 \sin \eta_L \sin \tau_L \cos \eta_R} \quad (11)$$

To further develop the calculation, an idea of Pentland [14] is used. He proposed to calculate the spatial means of luminance gradients to recover the point light source direction tilt. We apply this idea to the DFR gradients to recover the object rotation axis tilt. We average the DFR gradients in the image region covered by the moving object. To do so, the object shape has to be known, because the surface normal \mathbf{N} appears in the DFR gradients in Eq. 6. To avoid this, we apply Pentland assumption that

A3: the object surface normals are equally distributed.

This restricts the type of video scenes to scenes with ball-like objects. As the results in section 4 will show, this assumption allows even for human heads. By this assumption, the surface normal

$$\mathbf{N} = (\sin \eta_N \cos \tau_N, \sin \eta_N \sin \tau_N, -\cos \tau_N)^T, \quad (12)$$

defined by the slant and tilt angles η_N, τ_N corresponding to \mathbf{R} in Fig. 1, has the probability density function

$$p_{\mathbf{N}}(\eta_N, \tau_N) = \frac{\sin \eta_N \cos \eta_N}{\pi}, \quad (13)$$

see [9]. The linear mean of the DFR gradients is

$$\overline{g_x} = \int_0^{2\pi} \int_0^{\pi/2} g_x p_{\mathbf{N}}(\eta_N, \tau_N) d\eta_N d\tau_N \quad (14)$$

and correspondingly for $\overline{g_y}$. The integral in Eq. 14 covers the whole visible surface of the object, whereas Eq. 5 contains the switch μ to decide whether a surface is illuminated or not by the point lights source. We use thus the assumption that

A4: self shadowing is negligible

and set $\mu = 1$. This assumption has been used implicitly also in [14][9]. Assuming further that

A5: the point light source is weak,

enables us to solve the integral in Eq. 14. This means that $1 + \mathbf{LN} \approx 1$ holds and the denominators in Eq. 6 can be drastically simplified.

The idea is that the mean gradient of the DFR is related to the the orientation of the projected rotation axis. To motivate this idea, Fig. 2 shows examples for the DFR. It can be seen that the 2D projection of the true 3D object rotation axis seems to be orthogonal to the mean gradient of the DFR. To recover the orientation of the mean DFR gradient, the y

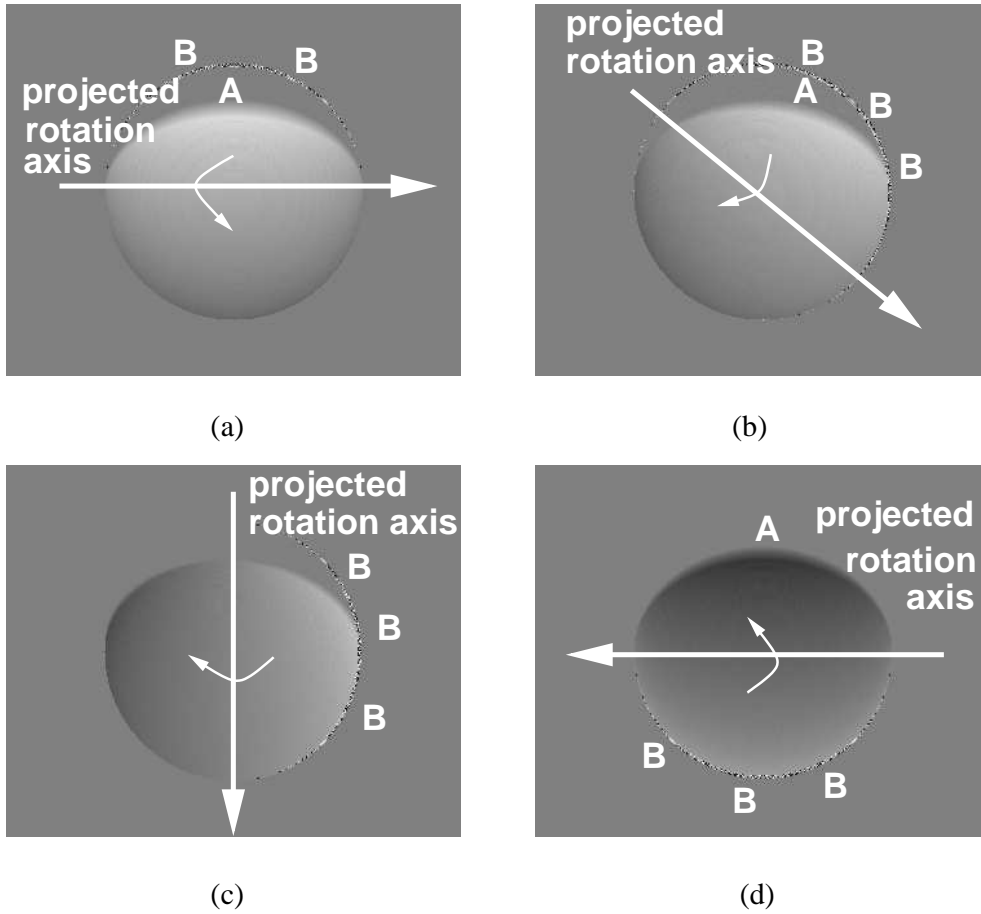


FIG. 2 – *Displaced frame ratio (DFR) of two images of a unicolored sphere rotating around an axis in the image plane ($\eta_R = \pi/2$): The point light source is located "at the feet of the viewer", i.e. $\eta_L = \pi/2$ and $\tau = \pi/2$. The rotation axis tilt is (a) $\tau_R = 0$, i.e. the x-axis, (b) $\tau_R = \pi/4$ (c) $\tau_R = \pi/2$ or (d) $\tau_R = \pi$. The motion compensation is perfect. The region on top of the object marked by "A" is not illuminated by the point light source (self shadow). The region at the occluding object border marked with "B" is arising new image content that can not be predicted by motion compensation. The background grey level corresponds to a frame ratio of one.*

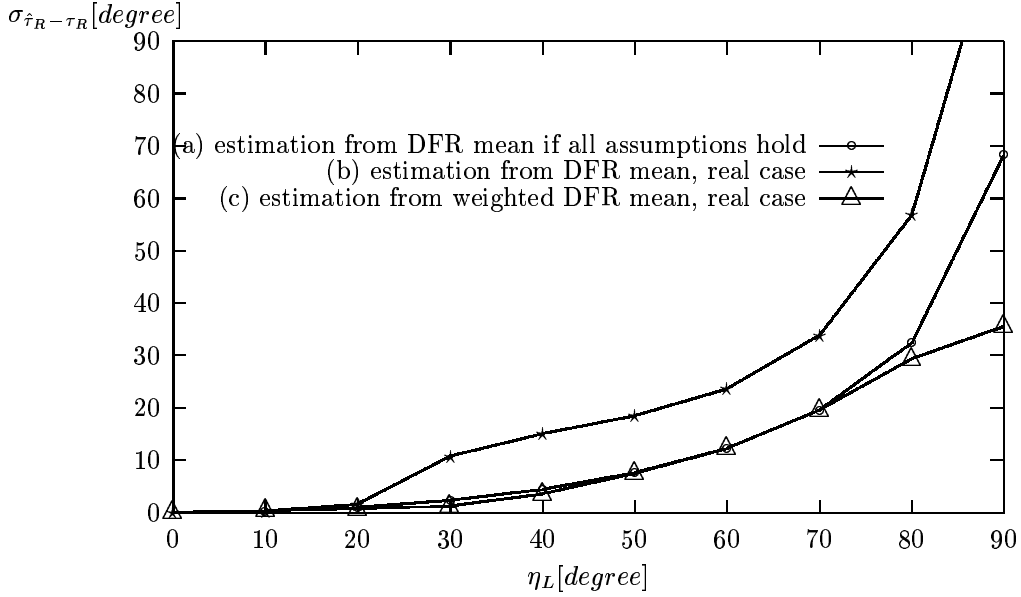


FIG. 3 – Results of rotation axis tilt estimation: Standard deviation $\sigma_{\hat{\tau}_R - \tau_R}$ of estimation error of rotation axis tilt versus illumination slant η_L , measured at $\eta_R = \pi/4, \tau_R = \pi/4$ and averaged over $\tau_L = 0.. \pi/2$; The curves show (a) the estimation error in the ideal case for synthesized images that behave really according to all used assumptions (see text), (b) the estimation error for realistically synthetic images and (c) the estimation error for realistic synthetic images using a weighted mean (see text).

component $\overline{g_y}$ of the mean gradient is divided by its x component $\overline{g_x}$ giving the ratio shown in Eq. 11 at top of the next page. To find the close relationship to the rotation axis, an

A6: illumination in viewing direction

is assumed, i.e. $\eta_L = 0$. Then, the ratio in Eq. 11 simplifies to $\overline{g_y}/\overline{g_x} = -\cos \tau_R / \sin \tau_R$, such that the rotation axis tilt can be estimated by

$$\hat{\tau}_R = \frac{\pi}{2} + \begin{cases} \text{atan}(\hat{g}_y/\hat{g}_x) & \hat{g}_x > 0 \\ \text{atan}(\hat{g}_y/\hat{g}_x) + \pi & \hat{g}_x < 0 \\ \pi/2 & \hat{g}_x = 0 \text{ or } \hat{g}_y > 0 \\ 3\pi/2 & \hat{g}_x = 0 \text{ or } \hat{g}_y < 0. \end{cases} \quad (15)$$

The estimation of the rotation axis tilt can be carried out by Eq. 15 using the means $\overline{\hat{g}_X}, \overline{\hat{g}_Y}$ of the DFR frame ratio inside the 2D binary object mask, the image region that corresponds to the object.

To have a closer look at the performance of the estimator, the estimation has been carried out for the simple scene described in Fig. 2. The conditions of the experiments are as follows:

- Object shape: fixed, object is a sphere
- Scene illumination
 - Intensity $|\mathbf{L}|$: fixed to $|\mathbf{L}| = 1.0$ (maximum, most contradictory to assumption A5)
 - Slant angle η_L : fixed per estimation, several estimations with varying slants $0 \leq \eta_L \leq \pi/2$, see horizontal η_L -axis in Fig. 3
 - Tilt angle τ_L : Fixed per estimation, several estimations with varying tilt, estimation error averaged over $0 \leq \tau_L \leq \pi/2$, standard deviation of estimation errors is vertical axis in Fig. 3
- Object motion
 - Tilt angle τ_R : Fixed to $\tau_R = \pi/4$ (without restriction of any generality, because results are averaged over τ_L)
 - Slant angle η_R : Fixed to $\eta_R = \pi/4$ (restriction to a representative case)

All experiments have been carried out using Eq.15. Note that experiments for all possible illumination slants η_L have been derived, even if the estimator assumes $\eta_L = 0$ (assumption A6). Note also, that the point light source is not weak, but at maximum amplitude, even if the estimator assumes a weak point light source (assumption A5).

Fig.3(a) shows the result, if all used assumptions hold. This theoretical case means that the mean DFR gradient really behaves as the simplified Eq. 11 says. The error standard deviation is zero for an illumination in viewing direction ($\eta_L = 0$, assumption A6) and increases the more the more the illumination direction of the point light source inclines to the image plane (contradictory to assumption A6). The maximum standard deviation in case of an illumination direction parallel to the image plane is $\sigma_{\hat{\tau}_R - \tau_R} = 70 \text{ degrees}$ which is quite a lot.

Fig. 3(b) shows the result for a more realistic experiment: an estimation from actual images. The images has been synthesized as described from a moving unicolored sphere. From the images, the DFR gradients measured and then used for estimation according to Eq.15. The standard deviation is higher than before, because in this more realistic experiment a part of the object is self shadowed (contradictory to assumption A4) and the point light source is not weak (contradictory to assumption A5). The maximum standard deviation in case of an illumination direction parallel to the image plane passes in this case even $\sigma_{\hat{\tau}_R - \tau_R} > 90 \text{ degrees}$.

To refine the estimator, a weighted mean of the DFR gradients is proposed according to

$$\begin{pmatrix} \hat{g}_X \\ \hat{g}_Y \end{pmatrix} = \sum_{\text{shrunked object mask}} w(\mathbf{p}) \nabla dfr(\mathbf{p}) \quad (16)$$

with the weight

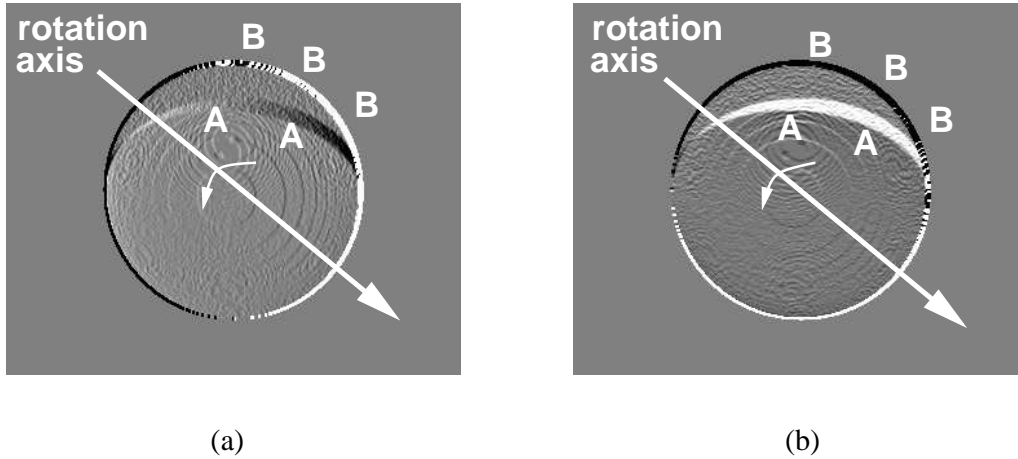


FIG. 4 – Gradients (a) in x and (b) in y direction of the displaced frame ratio (DFR) between two images of a unicolored sphere rotating around a rotation axis in the image plane ($\eta_R = \pi/2$) with an angle of $\tau_R = \pi/4$ to the x -axis; Strong gradients can be seen in area marked with "A" at the self shadow border and in the area marked with "B" at that part of the occluding object border where new image contents arises.

$$w(\mathbf{p}) = \frac{1}{\sqrt{|\nabla dfr(\mathbf{p})|}}. \quad (17)$$

The weight can be motivated by regarding the DFR gradients in Fig. 4. As can be seen, the border of the self shadowed region (the region which is not illuminated by the point light source) is surrounded by strong gradients that are not considered by the estimator (assumption A4). Further, there are strong gradients at the border of the object due to two reasons. First, areas of new scene contents can not be predicted. Second, displacement estimates at the object border are erroneous [23]. For these reasons, the gradients in Eq. 16 are weighted by the inverse of their amplitude and they are averaged only inside a shrunk object mask.

In Fig. 3(c) can be seen that this modification reduces the influence of motion compensation errors and neglected self shadow contours. The standard deviation is now comparable to the ideal case in Fig. 3(a). For the worst case of an illumination direction parallel to the image plane ($\eta_L \rightarrow \pi/2$) the result is even better. The maximum error standard deviation is now $\sigma_{\hat{\tau}_R - \tau_R} = 35$ degrees.

4 Results and Applications

In this section, results for the estimation of the rotation axis tilt from real video sequences are presented. Further, it is shown that the developed estimator derives results with an accuracy high enough to enable video indexing based on the tilt estimates.

4.1 Results of rotation axis tilt estimation

In Figs. 5 and 6, sample results with real image sequences are shown. The image sequences show approximately rigid objects with a sphere-like shape. The sequences are in CIF format (352x288 pels) and 12.5 Hz frame rate. The estimated object rotation axis tilt is visualized by a white arrow. The results 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 from Section 3.

The results show a good performance even, if the object shading is weak (contradictory to assumption A5 in Section 3) and the illumination is not in viewing direction (contradictory to assumption A6 in Section 3). As long as the 2D object silhouettes were roughly found (from about the 10th image on), the accuracy for the shown experiments is about ± 10 degrees by visual inspection. This size of errors corresponds well to the performance analysis in Section 3.



FIG. 5 – Two temporal segments of homogeneous object motion automatically found for the test sequence “Tai”, format CIF, frame rate 12.5Hz: (a) images 16 ... 31 and (b) images 36 ... 50. Superimposed is the estimated 2D projection of the 3D object rotation axis.

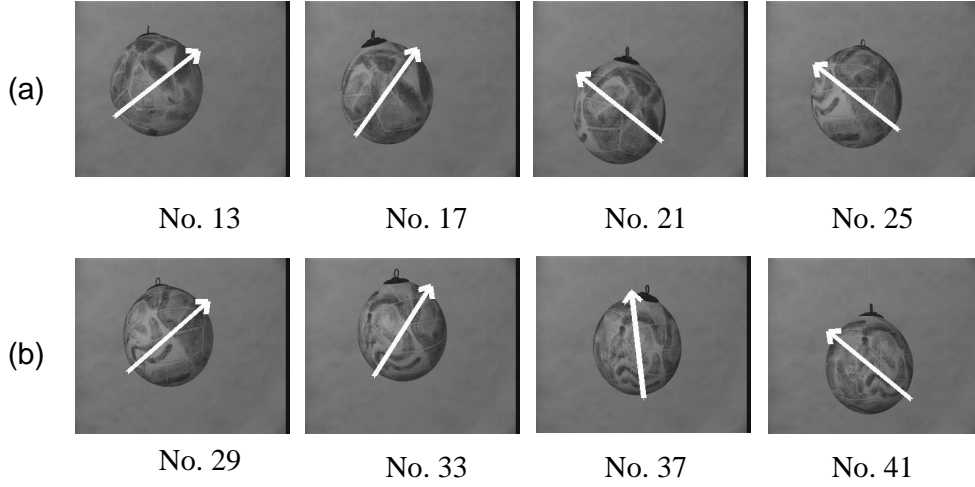


FIG. 6 – Two temporal segments of homogeneous object motion automatically found for the test sequence “Ball”, format CIF, frame rate 12.5Hz: (a) images 13 ... 26 and (b) images 29 ... 41. Superimposed is the estimated 2D projection of the 3D object rotation axis.

4.2 Application to video indexing

This section presents a new method for the temporal segmentation of a video sequence based on estimates of the tilt angle of the rotation axis of a rigid object. The temporal segments are identified regarding the temporal evolution of the tilt angle. Therefore, a Kalman filter with a physical motion model is applied to the tilt angle.

To describe the object motion just by the rotation axis tilt angle, for a time instant k a motion state vector

$$\mathbf{x}(k) = (\tau_R(k), \dot{\tau}_R(k))^T \quad (18)$$

is introduced, where $\dot{\tau}_R(k)$ is the temporal change (velocity) of the tilt angle. For the Kalman filter, this report follows the notation of [10]. The object motion is described by the motion model

$$\mathbf{x}(k) = \underline{\Phi} \mathbf{x}(k-1) + \mathbf{w}(k), \quad (19)$$

where

$$\underline{\Phi} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad (20)$$

is the system matrix for homogeneous motion and $\mathbf{w}(k)$ is a 2D motion state noise vector. From a rotation axis tilt $\hat{\tau}_R(k)$ that has been estimated at time instant k according to Section 3, the following motion measurement vector is derived

$$\mathbf{z}(k) = \begin{pmatrix} \hat{\tau}_R(k) \\ \hat{\tau}_R(k) - \hat{\tau}_R(k-1) \end{pmatrix} \quad (21)$$

that is assumed to follow the measurement model

$$\mathbf{z}(k) = \mathbf{x}(k) + \mathbf{v}(k) \quad (22)$$

with $\mathbf{v}(k)$ a 2D measurement noise vector.

For each time instant k , the Kalman filter needs as input a measurement vector according to Eq. 21. Further, it needs the former motion state vector estimate and its covariance matrix. To initialize the filter at the time instant $k = -1$, a zero motion state vector with a diagonal motion state covariance matrix with elements much larger than one are used. Finally, the Kalman filter needs to know the covariance matrices of measurement noise $\mathbf{v}(k)$ and system noise $\mathbf{w}(k)$. Ignoring Eq. 20, they are set to unit matrices. Thus, the noise for the elements of the motion state vector and the measurement vector are assumed to be equal in power and decorrelated.

The Kalman filter derives for each time instant k four outputs. A prediction as well as an estimate for both, the motion state vector and its covariance matrix.

In this report, the Kalman filter output is used for two purposes. First, the Kalman estimate for the object rotation tilt angle describes the object motion more reliable than the direct result $\hat{\tau}_R(k)$ from the method in Section 3. The Kalman tilt estimate is the first element $[\hat{\mathbf{x}}(k|k)]_0$ of the motion state estimate $\hat{\mathbf{x}}(k|k)$. The notation $k|k$ indicates that the estimate corresponds to the time instant k using all measurements including the measurement at time instant k [10].

Second, temporal segments of homogeneous object motion are identified by checking the validity of the assumed motion model defined in Eq. 19. At a time instant k , the object is said to undergo a homogeneous rotation if the object rotation tilt angle behaves as the Kalman filter predicts. As criterion, the inequality

$$|\hat{\tau}_R(k) - [\hat{\mathbf{x}}(k|k-1)]_0| \leq \alpha \sqrt{[\mathbf{P}(k|k)]_{00}} \quad (23)$$

has to be fulfilled. Here, $\hat{\mathbf{x}}(k|k-1)$ is the Kalman prediction for the motion state vector for time instant k using the measurements from all time instants up to $k-1$. Further, $\mathbf{P}(k|k)$ is the covariance matrix of the motion state vector that has been estimated by the Kalman filter and $[\cdot]_{00}$ indicates the first diagonal element. The factor α is a constant for all video sequences and is set to 40 by manual optimization.

In Figs. 5 and 6, automatically found temporal segments of two test sequences are shown. The identified segments fulfill the constraint Eq. 23 for all time instants. They correspond well to the main temporal segments of object motion observable in the video sequences. The results have been obtained fully automatically. The results show that

1. Indexing of video sequences showing simple rigid objects is possible using only one motion parameter, the object rotation axis tilt.
2. The object rotation axis tilts estimated by the method proposed in this report are accurate enough to ensure a correct indexing result.

5 Conclusion

In many existing approaches to motion estimation, illumination effects in the exploited video sequences are neglected. It is assumed that a moving object preserves its brightness during motion. For video scenes with a strong, point-like light source, illumination effects are strong and may disturb those algorithms.

To the contrary, this report shows that illumination effects contain useful information on object motion. An estimator is developed for the tilt angle of the rotation axis of a rigid 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.

It is analytically shown that the projected rotation axis can be estimated by exploiting only temporal changes of object shading. 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. The developed estimator is based on a link between the projected rotation axis and the mean spatial gradient of the displaced frame ratio (DFR). The DFR is the pel-by-pel frame ratio after motion compensation.

The estimator is of low complexity. He is based on a pel-by-pel division of two images followed by a calculation and averaging of spatial image gradients over an image area. Required input data are a dense displacement vector field (DVF) for motion compensation and a 2D binary object mask to define the image region of the object.

Experiments using automatically derived DVFs and object masks show that the estimator yields qualitatively good results for video scenes with simple objects like head-and-shoulder scenes. A performance analysis shows that the estimator works best if the main illumination is in viewing direction. In the worst case of illumination parallel to the image plane, theoretically a maximum standard error deviation of *35 degrees* can be attained. In the experiments, a visual inspection of the tilt estimation results revealed an uncertainty of about ± 10 *degrees*.

An application of the estimator to video indexing shows that the estimator is accurate enough to obtain correct indexing results. For indexing, a temporal Kalman filter is applied to the temporal evolution of the object rotation axis tilt angle. As long as the tilt angle follows the temporal motion model of the Kalman filter, the images are said to form a temporal segment. As soon as the tilt angle does not behave as predicted, the current temporal segment is said to be completed and the indexing algorithm waits for the start of the next segment. Applied to real video sequences showing simple objects like head-and-shoulder scenes, the algorithm finds correctly the temporal segments of the image sequences within the object moves homogeneously.

Whereas the estimation algorithm works well for video scenes with diffuse ambient light and a dominating point light source, the algorithm will fail if there is no predominating light source in the scene. This is because the algorithm is based only on the exploitation of photometric effects, here temporal changes of object shading. In future, it should be therefore combined with an algorithm exploiting geometric effects as motion of feature points. This report may serve as motivation to make use of the DFR and its useful information on 3D object shape and on 3D object motion.

Références

- [1] M. Bierling, Displacement estimation by hierarchical blockmatching, 3rd SPIE *Symposium on Visual Communications and Image Processing*, Cambridge, USA, November 1988, pp. 942-951.
- [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.
- [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 Fernsehbilddfolge", 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] J.M. Mendel: "Lessons in estimation theory for signal processing, communications and control", Prentice Hall, Englewood Cliffs, New Jersey, 1995.
- [11] K. Meyberg, P. Vachenhauer: "Höhere Mathematik", Volume 1, Springer-Verlag, Berlin 1993.
- [12] 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.
- [13] 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.
- [14] A.P. Pentland: "Finding the illumination direction", Journal of the Optical Society of America, Vol.. 72, No. 4, April 1982, pp. 448-455.
- [15] A.P. Pentland: "Photometric Motion", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9, September 1991, pp. 879-890.
- [16] S. A. Shafer: "Shadows and silhouettes in computer vision", Kluwer Academic Publishers, Dordrecht, Netherlands, 1985.

-
- [17] 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.
 - [18] 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.
 - [19] J. Stauder: "Augmented Reality with Automatic Illumination Control Incorporating Ellipsoidal Models", IEEE Trans. on Multi Media, Vol. 1, No. 2, June 1999, pp. 136-143.
 - [20] J. Stauder: "Object rotation axis from shading", ICASSP'99, Phoenix, Arizona, USA, 15.-19.3.1999, Vol. 6, pp. 3285-3288.
 - [21] J. Stauder, H. Nicolas: "Motion-based video indexing evaluating object shading", ICIP'99, Kobe, Japan, 24.-28.10.1999.
 - [22] J. Stauder: "Object Rotation Axis from Shading", submitted to the IEEE Transactions on Pattern Analysis and Machine Intelligence, October 1999.
 - [23] M. Wollborn: "Object-based displacement estimation and motion compensated prediction for hybrid video coding", VLBV'99, Kyoto, Japan, 29.-30.10.1999.
 - [24] Z. Zhang: "Determining the epipolar geometry and its uncertainty", Int'l Journal of Computer Vision, Vol. 27, No. 2, 1998, pp. 161-195.
 - [25] 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.
 - [26] 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.



Unit ´e de recherche INRIA Lorraine, Technople de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit ´e de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit ´e de recherche INRIA Rhne-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit ´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit ´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

diteur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399