Int J Comput Vis (2010) 90: 304–312
DOI 10.1007/s11263-010-0353-7

Illuminance Flow Estimation by Regression

Stefan M. Karlsson · Sylvia C. Pont · Jan J. Koenderink · Andrew Zisserman

Received: 17 August 2009 / Accepted: 7 May 2010 / Published online: 9 June 2010

Abstract We investigate the estimation of illuminance flow using Histograms of Oriented Gradient features (HOGs). In a regression setting, we found for both ridge regression and support vector machines, that the optimal solution shows close resemblance to the gradient based structure tensor (also known as the second moment matrix).

Theoretical results are presented showing in detail how the structure tensor and the HOGs are connected. This relation will benefit computer vision tasks such as affine invariant texture/object matching using HOGs.

Several properties of HOGs are presented, among others, how many bins are required for a directionality measure, and how to estimate HOGs through spatial averaging that requires no binning.

Keywords Illuminance flow · Surface 3D texture · Histogram of oriented gradients · Illuminant estimation

S.M. Karlsson (✉)
IDE, Halmstad University, 30118 Halmstad, Sweden
e-mail: Stefan.Karlsson@hh.se

S.C. Pont
π-lab, Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands

J.J. Koenderink
π-lab, Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2028 CD Delft, The Netherlands

A. Zisserman
Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

1 Introduction

In this paper, meso-scale stochastic variation of an object's surface is not considered part of the shape, but is treated as 3D texture. This texture makes it possible to estimate (image) *illuminance flow*, an axial (bi-directional) flow field in the image that results from projecting the light vector first into the objects tangential plane, and then into the image plane.

A shape, as the one illustrated in Fig. 1, is modeled as a smooth differentiable manifold (here a sphere). The appearance of the texture is dependent on the direction of light relative to the tangential plane of the manifold. This modeling is reminiscent of bump-mapping in computer graphics (Blinn 1978). Illuminance flow can in general only be observed if 3D texture is present.

Vectors $(a, b)^T$ and $(-a, -b)^T$ describe the same flow at a given position, so the flow field in the image is described

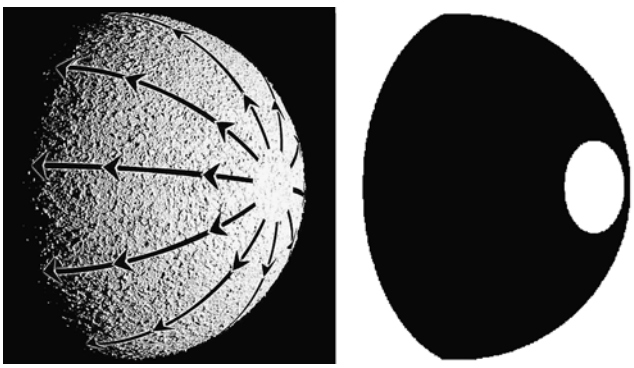

Fig. 1 *Left*: Example photograph from the database used with flow direction superimposed. A textured sphere is illuminated from the right (tilt = 0°, slant = 50°). *Black arrows* illustrate illuminance flow (note that both directions along the *black lines* are valid flow lines). *Right*: Binary image indicating where the flow is well defined

by local angles $\phi \in [0, 180)$. The main application for illuminance flow is as a shape cue of the surface, but this is *not* the focus of this paper. Shape from illuminance flow can be seen as an approach to make use of 3D texture for shape inference, but, again, this is not a topic that will be further investigated in this paper. In this paper we focus on the estimation of the field from images.

This work is in line with the study by Pont and Koenderink (2005) where a theory for analyzing the illumination orientation from 3D texture was presented. Generalizations to oblique viewing (Pont and Koenderink 2005; Karlsson et al. 2009), anisotropic surfaces (Karlsson et al. 2008, 2009) and non-uniform albedo (Varma and Zisserman 2004) have been made.

In this paper, we investigate real-world rough objects viewed from an arbitrary direction, and using standard regression methods, estimate the illuminance flow over their surfaces (see Fig. 1). We focus on a contemporary and in fashion low-level feature: the Histograms of Oriented Gradients (HOGs) (Dalal and Triggs 2005; Lowe 1999). Histogram-like statistics with directionality have a long tradition (good examples are Picard and Minka 1995; Flickner et al. 1995 and Michel et al. 1996). The HOGs are the low-level features of the key points in the Scale Invariant Feature Transform (SIFT) (Lowe 1999). They perform well for human pose recognition from video (Dalal and Triggs 2005) without the scale optimization and key point detection of the SIFTs. We will use the HOGs in a low-level and local fashion as a way of measuring directionality at a position in an image.

Directionality can be defined in several ways, one way (several ones will be discussed) is by the structure tensor (2nd moment matrix), which can be seen as three coarse descriptors of the distribution of the gradient.

Experimental results on real world surfaces show that the structure tensor yields promising results (Koenderink and Pont 2003) for estimating the illuminance flow, with estimates within a few degrees of the veridical orientation (in normal viewing). In previous works (Pont and Koenderink 2005; Karlsson et al. 2008, 2009) we have made theoretical predictions based on modeling the imaging process of 3D texture. This approach has shown the structure tensor to be appropriate for estimating illuminance flow. However, the issue of learning an optimal estimator based on observations with ground-truth from arbitrary viewpoints has never been posed, which is the focus of the current paper.

Several local and unsupervised illuminant estimators have been suggested (Pentland 1982; Knill 1990; Zheng and Chellappa 1992; Chantler and Delguste 1997; Llado 2003). Iterative non-local algorithms (such as Brooks and Horn 1985), are not considered, neither are supervised algorithms for a finite set of textures (such as Chantler et al. 2005). These algorithms are either identical to or correlate strongly with the gradient based structure tensor. The premise is a surface texture normally viewed; the goal is a local estimate of the illuminant tilt. In this setting, the illuminant tilt is the same as the illuminance flow, but this is not true for arbitrary viewing of the texture, as illustrated in Fig. 1. The tilt is relative to the camera frame, while the illuminance flow angle is relative to the tangential frame of the object, and can change locally within the image, even for collimated beams (point source at infinity).

In this paper, we focus on estimating the illuminance flow of the image of convex objects with rough surface texture. We use the HOG features in a regression setting, where we try both linear ridge regression and the support vector machine (SVM). To connect the results to previous work we show how the HOGs can be used to achieve a similar measure to that of the tensor (but also how this measure differs). This will enable us to show further properties of the HOGs, including how they are connected to the tensor, how many bins are required to encode a directionality measure, a different algorithm without binning to calculate the HOGs, and show how affine invariant approaches affect them.

2 Theoretical Background

The position of the light source relative to the camera is given by the tilt and the slant angles, as specified in Fig. 2.

The theory of illuminance flow estimation has been based on the structure tensor (second moment matrix), which in normal viewing for isotropic, uniformly Lambertian, low relief surface textures will give the flow (Koenderink and Pont 2003). The structure tensor (2nd moment matrix), is defined as

$$G = E[\nabla I \nabla^T I] = \begin{pmatrix} E[I_x I_x] & E[I_x I_y] \\ E[I_x I_y] & E[I_y I_y] \end{pmatrix}$$

where I_x is the partial derivative of the image and $E[\cdot]$ indicates expected value. The highest eigenvector will yield the directionality of the image. G contains the second moment description of the stochastic 2D variable ∇I as well as the second moments of the image power-spectrum (Bigun and Granlund 1987).

Measuring directionality can also be done by building a histogram of oriented gradients (HOG). For each gradient in

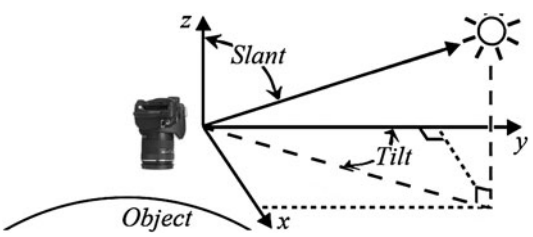

Fig. 2 The imaging geometry

an image a bin is increased in value. The angle of the gradient determines which bin, and the magnitude how much is added to it. A generalization is the Parzen window method (Parzen 1962) where many positions (bins) in an angular vicinity are updated. Assuming that such estimation is performed, there is no complication from grouping involved, i.e. one can freely choose a large number of bins based on a small number of data. Of course, the density estimation will be less reliable as data amount decreases.

The HOG can be made invariant to the sign of the gradient. The bins will then only need to cover the orientational (axial) interval $[0°, 180°)$. We will refer to the invariant version as the orientational HOG and to the regular HOG as directional. Discrete periodic sequences of the HOGs are denoted $\bar{f}_d(\frac{n2\pi}{N})$ and $\bar{f}_o(\frac{n\pi}{N})$ (for directional and orientational HOGs respectively with N bin values). These are samples of slightly different density functions, that are both related to the bivariate probability density function (pdf) $f(\boldsymbol{x})$ for the gradient ∇I.

We will use a probabilistic approach to analyze both the HOGs and G, and note on similarities with power-spectrum moments. The main drawback with frequency based analysis is the enforced toroidal topology and dependency on smooth convex window functions. Instead, we will assume that the pdf for the gradient, $f(\boldsymbol{x})$, always exists. We mention that what follows could be achieved in an elegant way using Lebesgue theory. This would also generalize our expressions to probability distributions (here we assume density functions), but while adding elegance, it would do little for insights for our specific problem. We will use only basic probability theory and will still come to the same conclusions.

There are three coarse descriptors of $f(\boldsymbol{x})$ found in G. The HOGs are similarly coarse descriptors for $f(\boldsymbol{x})$ because each bin is the average of $|\nabla I|$, *given* some direction $\operatorname{atan}(\nabla I) = \theta$. In other words, HOGs are estimates of conditional expectations. We consider the change of variables implied by polar coordinates:

$$R = |\nabla I|,$$
$$T = \operatorname{atan}(\nabla I).$$

It will have a bivariate pdf $f_{pol}(r, \theta) = r f(r\cos\theta, r\sin\theta)$. The directional HOG is an estimation of $f_d(\theta) = E[R|T = \theta] = \int_0^\infty r f_{pol}(r, \theta)\, dr$. We have:

$$f_d(\theta) = \int_0^\infty r^2 f(r\cos\theta, r\sin\theta)\, dr,$$
$$f_o(\theta) = f_d(\theta) + f_d(\theta + \pi)$$

where f_d has period 2π and f_o has period π, and are the population versions of $\bar{f}_d$ and $\bar{f}_o$.

3 Directionality by Complex Change of Variables

With the foundation laid in the previous section, we will be able to show concretely how the HOGs and the structure tensor differ in the information they are representing. In the end of this section we will arrive at 5 properties for HOGs. To reach this point, we need to consider the following complex expected values $\rho_\gamma(k)$, with corresponding estimations $\bar{\rho}_\gamma(k)$:

$$\rho_\gamma(k) = E[|\nabla I|^\gamma \exp(-\mathrm{i}k \operatorname{atan}(\nabla I))],$$
$$\bar{\rho}_\gamma(k) = \frac{1}{P}\sum_{p=1}^{P} \frac{(I_x(\boldsymbol{r}_p) - \mathrm{i}I_y(\boldsymbol{r}_p))^k}{(I_x^2(\boldsymbol{r}_p) + I_y^2(\boldsymbol{r}_p))^{\frac{k-\gamma}{2}}} \tag{1}$$

for $\gamma \in \mathbb{R}^+$ and $k \in \mathbb{Z}$, where $\mathrm{i} = \sqrt{-1}$ and the $\boldsymbol{r}_p$s are pixel positions. We can normalize it by $\hat{\rho}_\gamma(k) = \rho_\gamma(k)/\rho_\gamma(0)$ so that $|\hat{\rho}_\gamma(k)| \in [0, 1]$. The $\rho_\gamma(2)$ for different γ are different measures of directionality. $|\hat{\rho}_\gamma(2)| = 1$ always occurs for images consisting entirely of isolines in the $\frac{\angle\hat{\rho}_\gamma(2)}{2}$ orientation. When estimating $\hat{\rho}_\gamma$ by $\bar{\rho}_\gamma(k)/\bar{\rho}_\gamma(0)$, we can say that we are performing kth order voting with a γ-correction term. This is similar to the theory of Bigun and Granlund (1987), where the differential operator $(D_x + \mathrm{i}D_y)$ and its powers are analyzed. Powers of $(D_x + \mathrm{i}D_y)$ include higher order derivatives of the image, which in turn are used to obtain higher orders of complex moments of the power-spectrum. We use normalized powers of $(I_x + \mathrm{i}I_y)$ which use only first derivatives of the image. The sequence in Eq. 1 are *not* power-spectrum moments.

A special case which connects to the Bigun-Granlund theory is $\rho_2(2) = E[(I_x - \mathrm{i}I_y)^2]$ and $\rho_2(0) = E[I_x^2 + I_y^2]$. They encode G completely:

$$\rho_2(2) = (\lambda_{max} - \lambda_{min}) \exp(-\mathrm{i}2 \operatorname{atan}(\boldsymbol{v}_{max})),$$
$$\rho_2(0) = \lambda_{max} + \lambda_{min}$$

where λ_{max} and $\boldsymbol{v}_{max}$ are the highest eigenvalue and corresponding eigenvector of G.

Another special case is $\rho_0(k) = E[\exp(-\mathrm{i}k \operatorname{atan}(\nabla I))]$. This corresponds to the characteristic function (Mardia and Jupp 2000) of the circular variable: $T = \operatorname{atan}(\nabla I)$. The characteristic function is equivalent to a Fourier transform of the pdf of T. Thus, $|\rho_0(2)|$ is a fit of the second harmonic to the pdf of T, and $\frac{\angle\rho_0(2)}{2}$ is the orientation (the phase on the unit circle) of the second harmonic. For $\gamma = 0$, the magnitude of the gradient is ignored, which is one extreme way of measuring directionality.

A third special case is that of $\gamma = 1$, which is strongly connected to the HOGs, as we shall see. In general, for all γ, the change of variable formula (Mardia and Jupp 2000)

gives the relation:

$$\rho_\gamma(k) = \iint_{-\infty}^{\infty} |\boldsymbol{x}|^\gamma \exp(-\mathrm{i}k \operatorname{atan}(\boldsymbol{x}))\, f(\boldsymbol{x})\, d\boldsymbol{x}$$
$$= \int_{-\pi}^{\pi} \exp(-\mathrm{i}k\theta) \int_0^{\infty} r^{\gamma+1} f(r\cos\theta, r\sin\theta)\, dr\, d\theta,$$
$$\rho_1(k) = \int_{-\pi}^{\pi} \exp(-\mathrm{i}k\theta)\, f_d(\theta)\, d\theta, \tag{2}$$
$$\rho_1(2k) = \int_0^{\pi} \exp(-\mathrm{i}2k\theta)\, f_o(\theta)\, d\theta. \tag{3}$$

Equation 3 is found by evaluating Eq. 2 for $k \to 2k$ as the sum of two integrals, one over $[-\pi, 0]$, the other over $[0, \pi]$, and then using $\exp(-\mathrm{i}2k\pi) = 1$ and $f_o(\theta) = f_d(\theta) + f_d(\theta \pm \pi)$. Equations 2 and 3 yield Fourier series coefficients for f_d and f_o:

$$f_d(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho_1(k) \exp(\mathrm{i}k\theta),$$
$$f_o(\theta) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \rho_1(2k) \exp(\mathrm{i}2k\theta).$$

If the population versions f_d, f_o, ρ_1 are replaced with the sample versions $\bar{f}_d$, $\bar{f}_o$ and $\bar{\rho}_1$, then Eqs. 2 and 3 will turn into discrete Fourier transforms. For the orientational HOG we have:

$$\bar{\rho}_1(2k) = \sum_{n=0}^{N-1} \exp\left(-\mathrm{i}2k\frac{n\pi}{N}\right) \bar{f}_o\left(\frac{n\pi}{N}\right). \tag{4}$$

Some properties of the HOGs that emerge from these observations are:

(1) If a directionality measure needs to be explicitly calculated using HOGs, then a best matching sinusoidal yields it. The second harmonic approximation of the directional HOGs is equivalent to the first harmonic approximation of the orientational HOGs and constitutes the measure that is closest possible to G (assuming no other information of $f(\boldsymbol{x})$ is available).

(2) The minimum number of bins required to yield such a directionality measure is given by the Nyquist-Shannon sampling theorem (the sampling frequency is $\frac{N}{2\pi}$). For the orientational HOGs, we require $N = 3$, and for the directional HOGs $N = 5$. Note that we are considering the number of bins to be just the number of height samples of the estimated pdf. In traditional histograms, there is an additional problem of grouping involved. More bins will mean also a reduction in the quality of the estimation. This problem disappears if one assumes a Parzen window method (Parzen 1962).

(3) The directionality inherent in the HOGs is strongly correlated with that of G. They differ in γ-correction only. The structure tensor has $\gamma = 2$, while the HOGs have $\gamma = 1$. Algorithmically speaking, in G higher magnitude gradients are weighted more than in the HOGs. If the magnitudes of the gradients would be fixed to one ($f(\boldsymbol{x})$ is nonzero only on a circle), then the directionality of the HOGs and G would be identical.

(4) An alternative to calculating the HOGs is to calculate $\bar{\rho}_1(k)$, and then to Fourier transform it. This approach avoids the grouping procedure (the 'binning') inherent in the conventional histogram approach. K elements of the sequence ($k \in [0, K-1]$) yields $N = 2K - 1$ bin values (samples in $\bar{f}_d$). For estimating $\bar{f}_o(\theta)$ the sequence $\bar{\rho}_1(2k)$ is used in the same way. This is equivalent to using a wrapped sinc function as a Parzen window (Parzen 1962). The equivalent to a Gaussian Parzen window can be achieved by multiplying $\bar{\rho}_1(k)$ with a Gaussian (because multiplication in Fourier gives convolution and because a Gaussian function transforms back to a Gaussian).

(5) If images are affine normalized using G, as is proposed in several works (Mikolajczyk et al. 2005), then there is little or no discriminant information available in $\bar{\rho}_1(0)$ and $\bar{\rho}_1(2)$. There are a total of three degrees of freedom in $\rho_1(0)$ and $\rho_1(2)$ (real and complex valued resp.) that correspond closely to $\rho_2(0)$ and $\rho_2(2)$ (that encode G). If one uses the HOGs as low level features, it might be prudent to use $\bar{\rho}_1(0)$ and $\bar{\rho}_1(2)$ for affine normalization, instead of G. However, if HOGs are estimated on smaller regions within a larger affine normalized region, then $\bar{\rho}_1(0)$ and $\bar{\rho}_1(2)$ can still hold valuable information. Also note that HOGs are usually normalized to unit mean which corresponds to enforcing $\bar{\rho}_1(0) = 1$ regardless of affine normalization.

We emphasize that we are not suggesting a new set of low-level features here, but rather we suggest an analysis of the existing ones (HOG) that makes the connection to the structure tensor readily available, and sheds some light onto what the HOGs actually do in terms of non-parametric density estimation.

4 Axial Regression

We now turn our attention to the specific topic of illuminance flow estimation using the HOGs as low-level features. We want to once more remind the reader that we are here not interested in any other task then illuminance flow estimation. The topic is covered in previous works (Pont and Koenderink 2005; Karlsson et al. 2009). In the present paper, we are not going to attempt shape inference with our regression (e.g. regressing for the normals of the surface).

We tried two standard approaches, first a linear model with ridge regression, then a support vector machine where several kernels were considered. Because illuminance flow is an axial (orientational) property, we used the orientational HOG.

4.1 Linear Model

We first phrased the problem in a linear setting as $y = \boldsymbol{f}^T \boldsymbol{w}$, where y is the illuminance flow at a point, $\boldsymbol{f} = \{\bar{f}_o(\frac{0\pi}{N}), \bar{f}_o(\frac{1\pi}{N}), \ldots, \bar{f}_o(\frac{(N-1)\pi}{N})\}^T$ is the feature vector, and $\boldsymbol{w} = \{w_0, w_1, \ldots, w_{N-1}\}^T$ is the weight vector. Collecting all feature vectors in matrix F, and all ground-truths in vector $\boldsymbol{y}$, ridge regression is a regularized version of LSE minimization, resulting in the pseudo-inverse: $\boldsymbol{w} = (F^T F + c_{lin}\mathcal{I})^{-1} F^T \boldsymbol{y}$, where $\mathcal{I}$ denotes the identity matrix, and c_{lin} is the ridge parameter for the regression. This corresponds to minimizing the objective function:

$$E_{lin} = c_{lin}\|\boldsymbol{w}\|^2 + \frac{1}{2}\|F\boldsymbol{w} - \boldsymbol{y}\|^2$$

Using the illuminance flow angle ϕ directly to represent the flow is not suitable for regression because of the angular discontinuity (for axial data, 0° is equivalent with 180°). Instead, one can choose to regress towards $\cos(2\phi)$ and $\sin(2\phi)$ separately, arriving at 2 weight vectors independently (when using the models for predictions, one needs to divide the output angle by two). This can be eloquently phrased as one single regression, by using complex numbers where $y = \exp(\mathrm{i}2\phi)$:

$$y = \boldsymbol{f}^T \boldsymbol{w} = \exp(\mathrm{i}2\phi) = \sum_{n=0}^{N-1} \bar{f}_o\left(\frac{n\pi}{N}\right) w_n \tag{5}$$

where $w_n \in \mathbb{C}$. The definition of the pseudo inverse allows for complex numbers (replacing $\boldsymbol{w}^T$ with conjugate transpose $\boldsymbol{w}^*$). We are minimizing one single consistent error E_{lin} (we have $\|\boldsymbol{w}\|^2 = \boldsymbol{w}^*\boldsymbol{w}$). That the regression on $\cos(2\phi)$ and $\sin(2\phi)$ is done separately does not matter for the outcome, nor does the coordinate frame we choose for ϕ.

In this model, if $w_n = \exp(-\mathrm{i}2\pi\frac{n}{N})$ then, following Eq. 4, $y = \bar{\rho}_1(2)$ which correlates with the structure tensor (it differs only in γ-correction, and is the closest possible to the tensor we can get using only the HOG).

4.2 Support Vector Model

The Support Vector Machine (SVM) was originally suggested by Vapnik (1995) for classification and regression (Vapnik et al. 1999). The SVM fits the linear function $y = \boldsymbol{f}^T \boldsymbol{w} + b$ to the data by solving the following convex optimization problem:

$$\begin{aligned} &\text{minimize } E_{svm} = \frac{1}{2}\|\boldsymbol{w}\|^2 + c_{svm}\sum_{\forall k}\zeta_k, \\ &\text{subject to } \begin{cases} |(\boldsymbol{f}_k^T.\boldsymbol{w}) + b - y_k| \le \epsilon + \zeta_k, \\ \zeta_k \ge 0. \end{cases} \end{aligned} \tag{6}$$

The constants c_{svm} and ϵ are for tweaking the regression. The SVM solves this by optimizing a dual formulation, arrived at through forming the Lagrangian of E_{svm}. New variables are introduced, $\boldsymbol{\alpha} = \{\alpha_k\}$, where $\boldsymbol{w} = \sum_{k=1}^K \alpha_k \boldsymbol{f}_k$, and thus $y = b + \sum_{k=1}^K \alpha_k \boldsymbol{f}_k^T \boldsymbol{f}$. With a new objective function and constraints with respect to $\boldsymbol{\alpha}$, the dual problem gives solutions to the original (primal) problem (Vapnik 1995). The set of training feature vectors that contribute to the output is small because $\boldsymbol{\alpha}$ is usually sparse (these are the support vectors of the machine).

Generalizing to non-linear functions is done by providing a substitute scalar product (the kernel), $\kappa(\boldsymbol{f}_1, \boldsymbol{f}_2)$. Each possible kernel corresponds to performing transformations of the data before the regression, such that a linear regression in the transformed space corresponds to non-linear regression in the original one (Aizerman et al. 1964). If the objective function makes use of the kernel instead of scalar products, then the output of the machine will be $y = b + \sum_{k=1}^K \alpha_k \kappa(\boldsymbol{f}_k, \boldsymbol{f})$.

The linear SVM is very similar to ridge regression, but the objective function we minimize is essentially different. E_{lin} and E_{svm} have the same smoothness term $\|\boldsymbol{w}\|^2$, which is independent of the data, they differ in how they penalize data deviation from the model. Any deviations that is within the ϵ bound is not penalized in the SVM regression, and is penalized linearly for the amount above that threshold, with slope equal to the parameter c_{svm}. Ridge regression minimizes a squared error, with no ϵ insensitive region. However, squaring also entails penalizing lower deviations less then higher ones, so the two methods should be expected to yield similar output. This argument holds despite the fact that ridge regression and SVMs require widely different algorithms (matrix inverse vs. iterative search). Both are convex optimization problems, where a unique global optimum is guaranteed. The power of SVMs lies in their ability to introduce nonlinearities. For this reason, we compared SVM regression with several popular kernels with the result we achieved with the ridge regression.

We phrased the illuminance flow problem in a similar fashion as with the ridge regression. We regress towards $\cos(2\phi)$ and $\sin(2\phi)$ separately, and will get two SVMs that are combined to make predictions (output angle is divided by two). In the ridge regression setting, this can be phrased as one single regression using complex numbers, with a consistent error. We are not aware of any method to provide the same property for SVMs. However, because the ridge regression and the linear SVM are similar, we argue that the linear SVM has a roughly consistent error, but the same cannot be said for any kernel.

4.3 Implementation and Data

For data, we used a set of 28 images of the same textured sphere photographed under controlled laboratory con-

ditions. The lighting was approximately collimated beams (one point source, far from the sphere), where the position of the light source was varied in 14 slant directions $(10°, 20°, \ldots, 140°)$. For each slant direction, the sphere was photographed with lighting from the right (tilt direction 0°) and the left (tilt direction 180°). Furthermore, the images were rotated to simulate more tilt directions, in total 16 directions $(0, 22.5, \ldots, 337.5)$ for every slant angle.

The photographs are 8 bit gray depth, 600 by 600 pixels. The sphere was carefully positioned such that the center of the image corresponds to the center of the sphere. The sphere is roughly 0.5 meters in diameter, and was positioned 2 meters away from the camera. With a maximum variation in visible height profile of 0.25 meters, we can assume that the camera has an orthogonal projection as far as the sphere is considered. The height profile is a sphere and with collimated beams of known direction, we calculated the ground-truth illuminance flow angles ϕ for every valid position in the image. Positions outside of the silhouette, in the shadow of the sphere or where the light hits the surface very close to its normal direction were not considered, as these positions do not have illuminance flow well defined (illustrated in Fig. 1). This amounted to over 260.000 data points in 32 dimensional feature space, with associated ground-truth illuminance flow $\phi \in [0°, 180°)$.

Orientational HOGs were calculated on square, 8 by 8 pixel wide cells in the image. We used 8 bins, representing the angular intervals of $\{[0°, 22.5°), \ldots, [157.5°, 180°)\}$. At the vertex of each cell are the positions to be described by the features. For each position, the HOG bin values of all 4 cells sharing that vertex are collected into a feature vector of 32 elements. These are normalized to unit mean. Each feature vector (for each position considered) gets values from 4 HOGs, so the outer scale is 16 by 16 pixel wide blocks. Each HOG contributes to 4 blocks, so there is overlap between the feature vectors. Two such blocks are illustrated in Fig. 3.

The software for calculating HOGs was Bill Triggs' (Dalal and Triggs 2005) implementation, which uses 2-point derivative filters. We also implemented the structure tensor algorithm with the same derivative filters and size of blocks as with the HOGs. Linear interpolation is used in the binning algorithm of the HOGs to improve accuracy, which approximately corresponds to a triangular Parzen window (Parzen 1962). Ridge regression was performed with the statistics toolbox in the Matlab environment. For the SVM, we used the "SVM^{light} v. 6.01" software implementation (Joachims 1999). We tried 4 different kernels: linear, polynomial, sigmoidal (tanh) and radial basis function (rbf).

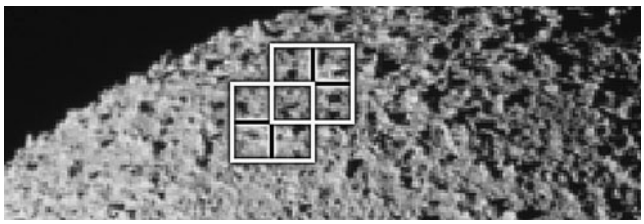

Fig. 3 Illustration of blocks and cells of the HOGs. Two blocks are illustrated as *white squares*. Each Block has 4 cells, but only 7 unique cells are covered due to feature overlap. The figure displays the same size as was used in the experiments

5 Results

Interestingly, all models (ridge and SVM) yielded near identical results to the $\bar{\rho}_1(2)$ measure, which is visually indistinguishable from the structure tensor on these images. Typical output is illustrated in Fig. 4 together with ground-truth. Performance is evaluated through the average angular deviation (E_{ad}) of the estimate to the ground-truth illuminance flow angle (averaged over all images in the database, over all valid positions). All the regressed models had an E_{ad} lower than 16.5°, with the best results for the (second order) polynomial kernel at 16.43° and the worst being the rbf kernel at 16.48° (further tweaking of kernel parameters and selecting different training sets will change this ordering). The left panel of Fig. 5 illustrates the performance across the image (different positions on the sphere). The corresponding illustrations for the other models are not shown, because they are all visually indistinguishable.

We also implemented the regular structure tensor, with equivalent derivative filters and averaging areas as with the HOGs (2-point derivative filters and 16 by 16 square blocks). The structure tensor yields $E_{ad} = 16.7°$. If the weights of the linear model are fixed to those corresponding

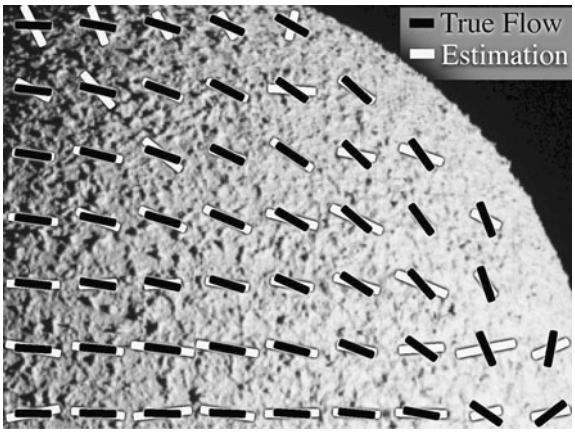

Fig. 4 Typical output of the regressed models (*white line* segments) compared to ground-truth (*black*). Samples are taken of every third valid x and y coordinate (1/9 of the valid positions shown). Estimation is by the linear ridge predictor, but is visually indistinguishable from any other model regressed. Note that the positions in the lower right corner occupy a domain where illuminance flow is not well defined, as illustrated by Fig. 1 (incoming light is near to being parallel to the surface normal)

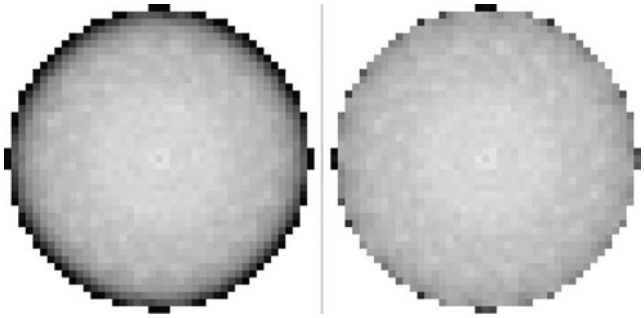

Fig. 5 Average angular deviation of the prediction to ground-truth, averaged over all images. *White*: no deviation, *black*: deviation >45°, gray: linear scale in [0, 45]. *Left panel*: Result achieved with HOG features (approx. same image for all models). Maximum average deviation ≈ 79°. *Right panel*, equivalent image for when the normal is appended to the feature vector, for the rbf kernel SVM

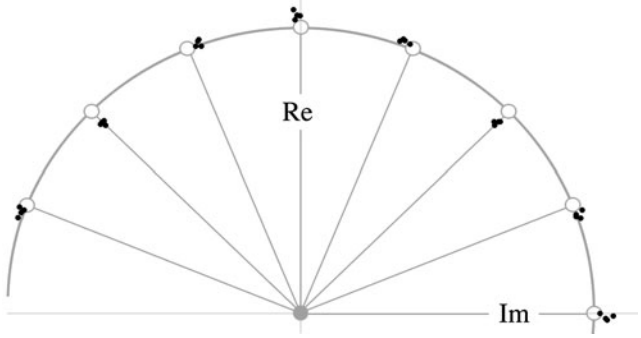

Fig. 6 Weights of the linear model from ridge regression. *Black dots*: weights, *gray circles*: centered around the positions of the $\bar{\rho}_1(2)$ predictor. Weights have their angular part divided by two, so that each gray radial line corresponds to the center of the bin of the HOG. There are four HOGs used for each position in the image thus four weights for each direction

to the structure tensor, then the same error occurs (difference is smaller than 0.05°). The small reduction in performance when going from the regressed models to the pure structure tensor is explained by the directional bias in the estimation (square blocks, and 2-point filters). The regressed models (both SVMs and linear) have compensated for directional bias in the HOG features.

Ridge regression: Inspection of the regressed weights in Fig. 6 verifies that $w_n \approx C \exp(-\mathrm{i}2\pi \frac{n}{N})$, which corresponds to the $\bar{\rho}_1(2)$ measure. Figure 6 depicts the weights with their angle divided by two (thus w_n occupy half a circle) which is the equivalent to dividing the output of the model by two (which is required for an estimate of the illuminance flow angle). Incorporating more features in the training set made the weights converge towards $w_n \approx C \exp(-\mathrm{i}2\pi \frac{n}{N})$ (closer to the circles of Fig. 6). Changing the parameter c_{lin} changes the magnitude of the regressed weights (C above) but not their directional component, except for when c_{lin} is close to zero. For $c_{lin} = 0$ the regression is pure LSE minimization which will be (for this problem) unstable and prone to over-fitting. We note that the results show some directional bias, especially in the directions (±45° and ±135°). These are weighted slightly less than the horizontal and vertical directions. This is because 2-point derivative filters and square regions are used in the HOG calculations, which give rise to directional bias. If derivatives of Gaussians are used as filters and roughly circular cells implemented, then this effect disappears. Smoothing the image before the filtering reduces but does not eliminate the directional effects of 2-point derivative filters.

SVM models: The output is almost identical to the structure tensor, and visually indistinguishable from the ridge regression for all kernels used. We used a training set of 12.000 randomly selected features for training, and the remaining features for verification. Training is on a mere 4% of the data but took nearly a day to complete per kernel. The output of the SVM models is indistinguishable from that of the ridge regression, for both training and verification set. All the regressed predictors had very similar performance on the training and the verification set, with a difference in E_{ad} smaller than 2°. Using different kernels yields different result only for much smaller subsets of the training data, but as the training set becomes bigger they all converge towards structure-tensor-like behavior.

It seems that the HOG features are not indicative of illuminance flow beyond what is in $\rho_1(2)$. Unfortunately, with our SVM modeling, we are not minimizing a consistent error which weakens this conclusion. To illustrate this issue, if we change the frame in which the illuminance flow angle is described (the direction of the zero axis) then the results of the SVM regression change. For small training sets, a noticeable change in output is evident (this is not so with the ridge regression which is totally invariant to the frame used). As the training set becomes bigger the output becomes less dependent on the frame (in the training set used for the final results, no directional bias can be seen).

5.1 Accounting for Oblique Viewing

A major source of error from the estimators comes from oblique viewing of the texture, i.e. where image patches are taken closer to the silhouette of the object. This is clearly evident in the left panel of Fig. 5 where E_{ad} actually goes above 45° when taken close to the silhouette (random guesses results in $E_{ad} = 45°$ for axial data). We appended the local normal of the sphere (relative to the camera frame) to the feature vector, to see if the regression would improve. We note that it is unfeasible to assume that the normal is available in real-world applications. We also encoded the normals differently by e.g. projecting them into the camera plane (yielding 2D vectors with magnitude less than one), and by doubling the angle of the projected vector (which is equivalent to considering it as an axis rather than a vector).

We also used the normal in a preprocessing scheme, where we performed an affine transform on the image region before the HOG features were calculated.

The affine transform aims at minimizing the effects of oblique viewing of the texture. We performed first a contraction of the patch equivalent to the foreshortening but in the orthogonal direction of the foreshortening. After that, a uniform up-sampling of the texture patch was performed. We note that modeling oblique viewing as an affine transform is an approximation to begin with, and that we lose fine-scale information doing the normalization.

All the models gained performance with the affine normalization but they regressed nonetheless to the same structure-tensor-like behavior. The best performance was with the polynomial kernel ($E_{ad} = 11.9°$) and the worst was the sigmoidal kernel ($E_{ad} = 12.1°$), a difference we accredit to variability in training set selection and kernel parameters.

Ridge regression: We achieved no improvement with the appended normals. Inspection of the regressed weights showed that the ones corresponding to the normals equaled zero, independently of how the normals where encoded. The linear model is not powerful enough to make use of this kind of information.

SVM models: In contrast with the ridge regression, the SVM has the capability to make use of the normals. When appending them to the feature vectors, an improvement was noticed that was dependent on which kernel was used. The best improvement was achieved with the rbf kernel ($E_{ad} = 12.8°$) with the projected normal coded with double angle. The second order polynomial kernel has $E_{ad} = 13.4°$ and the sigmoidal has $E_{ad} = 15.7°$. The performance over different positions in the image of the rbf kernel SVM is illustrated in the right panel of Fig. 5. All the other results follow the same pattern: an improvement close to the silhouette, but the closer to the center of the image (normal viewing) the closer the models agree with the structure tensor algorithm. The rbf kernel SVM performs very much like the affine normalization scheme.

6 Discussion

This paper has 2 major contributions; (1) deriving theoretical properties of the Histogram of Oriented Gradients (HOGs), and (2) estimate illuminance flow through regression on the HOG features.

6.1 HOG Properties

Our theory uses spatial averaging over a set of non-linear mappings of the gradient (Eq. 1). We have shown how the resulting sequence is equivalent to a Fourier series expansion of the HOG features, where the second harmonic is strongly correlated with the eigenvector of the structure tensor (2nd moment matrix). The only difference between the second harmonic of the HOG and the structure tensor is a γ-correction of the gradients in the corresponding spatial averaging. In affine invariant texture and object matching the structure tensor is often used in a normalizing procedure, and our theory predicts how this will affect the HOGs. It also shows how many bins are needed of the HOG to calculate a similar measure as the structure tensor, as well as an alternative way of calculating HOGs, without binning.

The structure tensor is not the only way of achieving affine normalization. It entails a γ-correction of two in our spatial averaging. Better results might be achieved if a directionality measure is used that is consistent with the low-level features (HOGs), that involves a γ of one (i.e. enforcing the second harmonic to have zero energy). Further investigation into this will be a subject of future work.

One could naturally ask whether the gradient mappings might yield even more efficient features then the HOGs. This is an interesting topic as well, but beyond the scope of this paper. We have not suggested a new set of features, but rather, an analysis of the existing ones (HOG) that makes the connection to the structure tensor readily available, and explains what the HOGs actually do in terms of non-parametric density estimation.

6.2 Illuminance Flow Regression

We used the HOGs as low-level features in a regression setting and tried different methods to train an illuminance flow estimator. All methods yield approximately the same estimator: the second harmonic of the HOGs. Because this is the closest possible to the structure tensor that can be achieved using only HOGs, as well as visually indistinguishable on the images used, we conclude that the structure tensor is near optimal for the images in this study.

A natural question is whether the results generalize to arbitrary texture, which could be composed of different fine-scale surface geometry and any form of local variation in surface reflectance. We have in earlier work discussed the applicability of the structure tensor for deviations from the uniform plaster type of texture (Karlsson et al. 2008). Essentially, as long as the height profile has sufficiently low average height and is reasonably smooth, then a less Lambertian reflectance will not affect the outcome of the structure tensor. This is especially true if the light source is more elongated than a point source. For flat spatially varying albedo texture (on which the vast majority of computer vision theory is based) Illuminance flow is not observable. In the case where both fine-scale surface texture (say plaster) and flat albedo texture (say a flat painted pattern) are present simultaneously, things become more troublesome. Oblique viewing and anisotropy for flat texture can be modeled as one

single affine transform (a fact that makes popular computer vision algorithms possible).

We tried both linear ridge regression and Support Vector Machines(SVM) with several kernels. We were unable to find any significant improvement in performance using the more powerful SVM. This is an indication that for the estimation of illuminance flow there is no more useful information in HOGs other then what is in the second harmonic. Essentially, we found that the gradient structure tensor is the optimal estimator in our setting (albeit with a lower γ-correction than what is usually suggested).

This conclusion is weakened by the fact that our SVM regression is not minimizing a consistent error. We could not find a way to do the regression on both the x and the y component of the ground-truth simultaneously as was done with the ridge regression. We contended with doing 2 separate SVM regressions, one for each component of the illuminance flow vector.

We formulated the linear ridge regression through complex numbers, which easily shows that the error is independent of the particular frame we use to describe ground-truth angles. If, similarly, the SVM framework could be generalized to deal with complex numbers, then it should be able to use SVMs with a consistent error for axial regression problems of this kind. Alternatively, the two SVM regressions could be performed simultaneously, with an additional constraint that the combined output should be on the unit circle (which should still be a convex optimization problem), thus coupling the models. This will be a worthwhile subject for future work.

Acknowledgements Oscar van Hoof is gratefully acknowledged for providing the images used for the data collection. We thank professors Josef Bigun and Christoph Schnoerr for valuable discussions. This work has been funded by EU project VISIONTRAIN (MRTN-CT-2004-005439). Sylvia C. Pont was supported by the Netherlands Organization for Scientific Research (NWO).

References

Aizerman, M., Braverman, E., & Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, *25*, 821–837.

Bigun, J., & Granlund, G. H. (1987). Optimal orientation detection of linear symmetry. In *Proceedings of ICCV* (pp. 433–438).

Blinn, J. F. (1978). Simulation of wrinkled surfaces. *SIGGRAPH*, *12*, 286–292.

Brooks, M., & Horn, B. (1985). Shape and Source from Shading. In *Proceedings of 9th international joint conference on artificial intelligence* (pp. 932–936).

Chantler, M., & Delguste, G. (1997). Illuminant-tilt estimation from images of isotropic texture. *Proceedings of Vision, Image and Signal Processing*, *144*, 213–219.

Chantler, M., et al. (2005). Classifying surface texture while simultaneously estimating illumination direction. *International Journal of Computer Vision*, *62*, 83–96.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of CVPR* (pp. 886–893).

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: the QBIC system. *Computer*, *28*, 23–32.

Joachims, T. (1999). *Making large-scale SVM learning practical, advances in kernel methods—support vector learning*. Software available online at http://svmlight.joachims.org/, accessed June 2008.

Karlsson, S., Pont, S. C., & Koenderink, J. J. (2008). Illuminance flow over anisotropic surfaces. *Journal of the Optical Society of America A*, *25*, 282–291.

Karlsson, S., Pont, S. C., & Koenderink, J. J. (2009). Illuminance flow over anisotropic surfaces with arbitrary viewpoint. *Journal of the Optical Society of America A*, *26*, 1250–1255.

Knill, D. (1990). Estimating illuminant direction and degree of surface relief. *Journal of the Optical Society of America A*, *7*, 759–775.

Koenderink, J. J., & Pont, S. C. (2003). Irradiation direction from texture. *Journal of the Optical Society of America A*, *20*, 1875–1882.

Llado, X. (2003). Simultaneous surface texture classification and illumination tilt angle prediction. In *British machine vision conference*

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of ICCV* (pp. 1150–1157).

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. *Wiley series*. New York: Wiley.

Michel, S., Karoubi, B., Bigun, J., & Corsini, S. (1996). Orientation radiograms for indexing and identification in image databases. *Eusipco*, *96*, 1693–1696.

Mikolajczyk, K., et al. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, *65*, 43–72.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, *33*, 1065–1076.

Pentland, A. P. (1982). Finding the illuminant direction. *Journal of the Optical Society of America A*, *72*, 448–455.

Picard, R. W., & Minka, T. P. (1995). Vision texture for annotation. *Multimedia Systems*, *3*, 3–14.

Pont, S. C., & Koenderink, J. J. (2005). Irradiation orientation from obliquely viewed texture. In *Proceedings of the DSSCV05* (pp. 205–210).

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Berlin: Springer.

Vapnik, V. N., Golowich, S., & Smola, A. (1999). Support vector method for multivariate density estimation. *Advances in neural information processing systems* (Vol. 12, pp. 659–665).

Varma, M., & Zisserman, A. (2004). Estimating illumination direction from textured images. In *Proceedings of CVPR* (pp. 179–186).

Zheng, Q., & Chellappa, R. (1992). Estimation of illuminant direction, albedo and shape from shading. In *Physics-based vision: shape recovery* (pp. 39–61).