

Histogram of Directions by the Structure Tensor

Josef Bigun
Halmstad University
IDE SE-30118
Halmstad, Sweden
Josef.Bigun@hh.se

Stefan M. Karlsson
Halmstad University
IDE SE-30118
Halmstad, Sweden
Stefan.Karlsson@hh.se

ABSTRACT

Many low-level features, as well as varying methods of extraction and interpretation rely on directionality analysis (for example the Hough transform, Gabor filters, SIFT descriptors and the structure tensor). The theory of the gradient based structure tensor (a.k.a. the second moment matrix) is a very well suited theoretical platform in which to analyze and explain the similarities and connections (indeed often equivalence) of supposedly different methods and features that deal with image directionality. Of special interest to this study is the SIFT descriptors (histogram of oriented gradients, HOGs). Our analysis of interrelationships of prominent directionality analysis tools offers the possibility of computation of HOGs without binning, in an algorithm of comparative time complexity.

Categories and Subject Descriptors

I.4.7 [image processing and computer vision]: Feature Measurement; I.5.0 [Pattern Recognition]: General

General Terms

Algorithms, Performance

Keywords

Histogram of Oriented Gradients, Structure Tensor, Complex Weighting

1. INTRODUCTION

Directionality can be defined in several ways. With this paper, we will make a case for using the structure tensor as the natural analytic tool to bind several definitions and derived algorithms together into one fold. In some cases, different directionality measures are identical, but a tool such as the structure tensor is required to show this. The theoretical exercise of doing this is worthwhile because it reduces redundancy in research (avoids duplicate terms for the same

entity). Also, by using the approach of trying to reduce directionality measures to the structure tensor, insights are to be gained. This is especially true for the study of the histogram of oriented gradient (HOGs) features (the descriptor of the SIFT algorithm[12]). We will present both how these are very similar to the structure tensor, but also detail how they differ, and in the process present a different algorithm for computing them without binning. In this paper, we will limit ourselves to the study of 3 kinds of definitions of directionality, and their associated features: 1) the structure tensor, 2) HOGs, and 3) Gabor filters. The results of relating the Gabor filters to the tensor have been studied earlier [3], [9], and so for brevity, more attention will be given to the HOGs.

There are many other features that can be the object of a comparative study using the structure tensor, but the Gabor filters and HOGs are the most widely used. The generalized Hough transform[1] has been discussed before in this context ([2] Ch. 10.16, 11.6). Likewise, the Harris corner/singularity detector[8] can be fully explained in terms of the structure tensor([2] Ch. 10.9) (rather than a presence-of-corner specified by a corner model).

It is straightforward to apply the principles outlined here on several works, e.g. [10], and find that they are identical to the structure tensor, and varies in output through parameter selection. This does not lessen the value of these works, quite the contrary. They are unique in how they have been derived, from valuable viewpoints. However, they all beg the question of what is the size of their common denominator. Can these descriptors be used to complement each other? If so then how, and what is gained?

1.1 Structure Tensor

The structure tensor (2^{nd} moment matrix), can be seen as three coarse descriptors of the distribution of the gradient:

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

where I_x is the partial derivative of the image and $E[\cdot]$ indicates expected value[3]. The most significant eigenvector will yield the orientation of the directionality of the image. The amount of directionality can be defined in terms of the eigenvalues as $\frac{\lambda^{max} - \lambda^{min}}{\lambda^{max} + \lambda^{min}} = |\hat{\rho}_2(2)|$ (this notation will be useful in later sections). Because the tensor is positive definite we have that $|\hat{\rho}_2(2)| \in [0, 1]$, where maximum occurs for the maximally directional image (such as the left illustration of Fig. 1), and zero for an image of no directionality. $|\hat{\rho}_2(2)| = 0$ occurs when there is no single preferred direc-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISABEL '11, October 26-29, Barcelona, Spain

Copyright 2011 ACM ISBN 978-1-4503-0913-4/11/10 ...\$10.00.

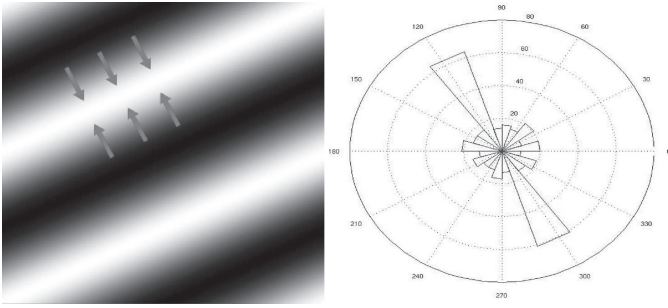


Figure 1: Left, Example of a highly directional image, with the gradients superimposed as arrows. Right, the HOG of the image on the left, with tiny amounts of added white noise.

tion where energy is concentrated in the power spectrum (for example, when all directions have equal energy), or equivalently, the gradient distribution has principal directions of equal variance (for example, no preferred direction in terms of variance).

In the following we will denote the image power-spectrum as $p(\omega)$, where ω are the 2D frequency coordinates. Note that this is with respect to the original image $I(\mathbf{x})$ before differentiation into $\nabla I(\mathbf{x})$. We will denote with $f(\omega)$ the bivariate probability density function of ∇I , where it is understood that ω in this context is a 2D vector whose magnitude and direction-angle (r, θ) represent the magnitude and direction of the gradient. These two functions will be analyzed in very similar ways in terms of their moments, yet we want to be clear that the two entities are very different.

In one interpretation of the structure tensor the image is analyzed on a topological torus and G analyzed as the second moment description of $p(\omega)$ [3]. In a nearly identical interpretation, G contains the second moment description of $f(\omega)$. This implies that two quite different 2D entities, will have equal second order moments. This is in fact not true in general! One can, however, easily show that this is always true if a toroidal topology is assumed. More specifically, we have:

Lemma 1. A sufficient condition that the pdf of ∇I ($f(\omega)$) share second order moments with the power-spectrum ($p(\omega)$) of $I(\mathbf{x})$ is that $E[\nabla I] = \mathbf{0}$.

Proof. Consider the covariance matrix (the dispersion matrix) for $f(\omega)$:

$$E[\nabla I \nabla I^T] - E[\nabla I] E[\nabla I^T] =$$

This is clearly equal to the structure tensor if and only if $E[\nabla I] = \mathbf{0}$. The central moment matrix of the power spectrum, on the other hand, will always equal the structure tensor, because all odd ordered moments vanish due to Hermitian symmetry[3]. ■

We will henceforth assume $E[\nabla I] = \mathbf{0}$ when discussing Gabor filters or the power spectrum interpretation of the structure tensor. In practice, to ensure $E[\nabla I] = \mathbf{0}$, we would have either a torus or a region of interest that has been reduced to go to zero smoothly on its boundaries (using, e.g. a Gaussian window function).

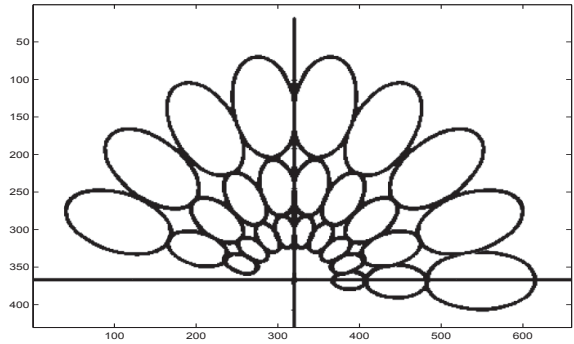


Figure 2: Example of rosette like partitioning of the Log-Gabor filters. They give a sparse estimation of the local spectrum.

1.2 Gabor Filter Magnitudes

Widely used for directional analysis are the 2D Gabor Filters [7],[11],[5]. Gabor filters are Gaussians tessellating the Fourier domain according to some uniform partitioning scheme. We will assume the Gabor filter partitioning scheme originally suggested in[6] which is similar to that of [11] and is a perfectly direction-isotropic scheme, as illustrated in Fig. 2. Convolution with the image and the filter can be thought of as a sequence of scalar products. On a specific position in the image, such a scalar product constitutes a coarse estimation of the local Fourier domain. The full set of filter scalar products at a specific position estimates the local spectrum (coarsely) and is sometimes called the Gabor Jet. The square magnitude, is the coarse estimation of the local power spectrum.

We will use the notation $\bar{p}(r, \frac{n\pi}{N})$ to denote the Gabor square magnitude, of radial frequency (tune-on) position r , and orientation n (assuming a total number of N orientations in the filter bank of R radial frequency bands). In other words r and n encode the coordinates of tune-on frequencies in the Fourier domain. We will collapse the Gabor responses to a function of only n by a weighted averaging in the radial direction, and we will write simply $\bar{p}(\frac{n\pi}{N}) = \sum_r r^2 \bar{p}(r, \frac{n\pi}{N})$ where we make explicit the "abusive" notation of $\bar{p}(\cdot)$ with one argument being the average of $\bar{p}(\cdot, \cdot)$ with two arguments. In the same spirit, this will be an estimation of $p(\theta) = \int r^3 p(r, \theta) dr$. Note that because of Hermitian symmetry, $p(\theta)$ is periodic with period π .

1.3 Histogram of Oriented Gradients

Measuring directionality can also be done by building a histogram of oriented gradients (HOG). For each gradient in 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 (Fig. 1 illustrates this). A histogram is a crude form of non-parametric density estimation. A generalization is the Parzen window method[15] where many positions (bins) in an angular vicinity are updated (this is often called kernel-based estimation of the histogram). Assuming that such estimation is performed, there is no technical 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. For an investigation how HOGs (without SIFT) perform for supervised human detection from video,

see[4].

The HOG can be made invariant to the sign of the gradient. The bins will then only need to cover the orientational (axial) interval $[0^\circ, 180^\circ)$. 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) $f(\boldsymbol{\omega})$ for the gradient ∇I . We remind that $\boldsymbol{\omega}$ here is a 2D vector whose magnitude and direction-angle are r, θ and represent the gradient.

Similar to the structure tensor, the HOGs are coarse descriptors for $f(\boldsymbol{\omega})$, but instead of being moments, the HOGs estimate samples of the densities:

$$\begin{aligned} f_d(\theta) &= \int_0^\infty r^2 f(r \cos \theta, r \sin \theta) dr \\ f_o(\theta) &= f_d(\theta) + f_d(\theta + \pi) \end{aligned}$$

f_d has period 2π and f_o has period π , and are the population versions of \bar{f}_d and \bar{f}_o . To make the analogy even clearer with the Gabor magnitude responses: one could implement a HOG algorithm in two steps. Firstly, one estimates $f(\boldsymbol{\omega})$ (denoted $\bar{f}(\boldsymbol{\omega})$) by e.g. a 2D Parzens window technique or a 2D histogram. Secondly, HOGs are built by collapsing $\bar{f}(\boldsymbol{\omega})$ into a 1D discrete signal, by weighted summing in the radial direction.

2. DIRECTIONALITY BY CHANGE OF REAL VARIABLES TO COMPLEX

Consider the following complex expected values $\rho_\gamma(k)$, with corresponding estimations $\bar{\rho}_\gamma(k)$:

$$\begin{aligned} \rho_\gamma(k) &= E[|\nabla I|^\gamma \exp(-ik \text{atan}(\nabla I))] \\ \bar{\rho}_\gamma(k) &= \frac{1}{N} \sum_{n=1}^N \frac{(I_x(\mathbf{x}_n) - iI_y(\mathbf{x}_n))^k}{(I_x^2(\mathbf{x}_n) + I_y^2(\mathbf{x}_n))^{\frac{k-\gamma}{2}}} \end{aligned} \quad (1)$$

For $\gamma \in \mathbb{R}^+$ and $k \in \mathbb{Z}$, where $i = \sqrt{-1}$. 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 k^{th} order voting with a γ -correction term. For the analogous approach of connecting the Gabor magnitudes with the structure tensor (Fourier expansion of $p(\theta)$), we would use the differential operator $(D_x + iD_y)$ and its higher powers as described in [9]. Powers of $(D_x + iD_y)$ include higher order derivatives, which in turn correspond to higher orders of complex moments of the power-spectrum (not of $f(\boldsymbol{\omega})$). For the HOGs, we use normalized powers of $(I_x + iI_y)$ which use only first derivatives.

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

$$\begin{aligned} \rho_2(2) &= (\lambda_{max} - \lambda_{min}) \exp(-i2 \text{atan}(\mathbf{v}_{max})) \\ \rho_2(0) &= \lambda_{max} + \lambda_{min} \end{aligned}$$

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

Another special case is $\rho_0(k) = E[\exp(-ik \text{atan}(\nabla I))]$. This corresponds to the so-called characteristic function[13] of the circular variable: $\text{atan}(\nabla I)$. The characteristic function is equivalent to a Fourier transform of the pdf of $\text{atan}(\nabla I)$. Thus, $|\rho_0(2)|$ is a fit of the second harmonic to the pdf of $\text{atan}(\nabla I)$, 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[13] gives the relation:

$$\begin{aligned} \rho_\gamma(k) &= \iint_{-\infty}^{\infty} |\mathbf{x}|^\gamma \exp(-ik \text{atan}(\mathbf{x})) f(\boldsymbol{\omega}) d\boldsymbol{\omega} = \\ &= \int_{-\pi}^{\pi} \exp(-ik\theta) \int_0^\infty r^{\gamma+1} f(r \cos \theta, r \sin \theta) dr d\theta \end{aligned}$$

$$\rho_1(k) = \int_{-\pi}^{\pi} \exp(-ik\theta) f_d(\theta) d\theta \quad (2)$$

$$\rho_1(2k) = \int_0^\pi \exp(-i2k\theta) f_o(\theta) d\theta \quad (3)$$

Eq. 3 is found by evaluating Eq. 2 for $k \rightarrow 2k$ as the sum of two integrals, one over the interval $[-\pi, 0]$, the other over $[0, \pi]$, and then using $\exp(-i2k\pi) = 1$ and $f_o(\theta) = f_d(\theta) + f_d(\theta \pm \pi)$. Eq. 2 and 3 yield Fourier series coefficients for f_d and f_o :

$$\begin{aligned} f_d(\theta) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho_1(k) \exp(ik\theta) \\ f_o(\theta) &= \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \rho_1(2k) \exp(i2k\theta) \end{aligned}$$

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 Eq. 2 and 3 will turn into discrete Fourier transforms. For the orientational HOG we have:

$$\bar{\rho}_1(2k) = \sum_{n=0}^{N-1} \exp\left(-i2k \frac{n\pi}{N}\right) \bar{f}_o\left(\frac{n\pi}{N}\right) \quad (4)$$

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 (the second harmonic approximation of the directional HOGs, or equivalently, the first harmonic approximation of the orientational HOGs) yields the desired measure. Other methods, such as using functions not strictly sinusoidal or methods to measure the bi-modality of a circular function, can be devised, but the harmonic will yield the measure that is closest possible to G (assuming no other information of $f(\boldsymbol{\omega})$ 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 $\frac{\#bins}{2\pi}$). For

the orientational HOGs, we require 3 bins, and for the directional HOGs, 5 bins.

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 ($f(\omega)$ is nonzero only on a circle), then the directionality of the HOGs and G would be identical.

4) If images are affine normalized using G (as is proposed in several works[14]), 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.

5) 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 $2K - 1$ 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[15]. The equivalent to a Gaussian Parzen window can be achieved by multiplying $\bar{\rho}_1(k)$ with a Gaussian (because multiplication in Fourier domain gives convolution and because a Gaussian function transforms back to a Gaussian).

Regarding the Gabor filters, many similar conclusions can be drawn. For points 1 to 4 above, the Gabor filter magnitudes have fully equivalent properties to that of the orientational HOGs. Point 5 differs however, because the theory for deriving the Gabor filter properties require different weighting in the radial direction of the power spectrum. This can be partly overcome by considering the different bands separately, as done in multi-scale approaches to image analysis problems, and is indeed the approach taken in [9].

3. DISCUSSION

We have shown how the platform of the structure tensor can be used to describe and relate seemingly different approaches of low level directionality features. 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 (2^{nd} 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.

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 at all, or dispensing with HOGs all together if the goal is to use them as descriptors discriminating regions. This is because mutual distances in a set of vectors are invariant to a discrete Fourier transformation of the vectors themselves.

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). As a corollary question, one could naturally ask whether the non-linear mapping of gradients might yield even more efficient features than the currently available descriptive features. Further investigation into these issues will be a subject of future work.

4. REFERENCES

- [1] B. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–112, 1981.
- [2] J. Bigun. *Vision with Direction*. Springer, Heidelberg, 2006.
- [3] J. Bigun and G. Granlund. Optimal orientation detection of linear symmetry. In *First International Conference on Computer Vision, ICCV, London, June 8–11*, pages 433–438. IEEE Computer Society, 1987.
- [4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages I: 886–893, 2005.
- [5] J. Daugman. Six formal properties of two-dimensional anisotropic visual filters: Structural principles and frequency / orientation selectivity. *IEEE trans. on Systems, Man and Cybernetics*, 13:882–887, 1983.
- [6] D. Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, A4:2379–2394, 1987.
- [7] D. Gabor. Theory of communication. *Journal of the IEE*, 93:429–457, 1946.
- [8] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the fourth Alvey Vision Conference*, pages 147–151, 1988.
- [9] S. Karlsson and J. Bigun. Multiscale complex moments of the local power spectrum. *JOSA A*, 24(3):618–625, 2007.
- [10] D. Knill. Estimating illuminant direction and degree of surface relief. *JOSA A*, 7:759–775, 1990.
- [11] H. Knutsson. *Filtering and reconstruction in image processing*. PhD Thesis no:88, Linköping University, 1982.
- [12] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999.
- [13] K. V. Mardia and P. E. Jupp. *Directional Statistics*. Wiley Series, 2000.
- [14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. G. Matas, F. Schaffalitzky, T. Kadir, and L. J. V. Gool. A comparison of affine region detectors. *IJCV*, 65(1-2):43–72, Nov. 2005.

- [15] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.