Energy filters, motion uncertainty, and motion sensitive cells in the visual cortex: a mathematical analysis

R. S. Jasinschi

Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742, USA

Received March 27, 1991/Accepted in revised form May 21, 1991

Abstract. Energy filters are tuned to space-time frequency orientations. In order to compute velocity it is necessary to use a collection of filters, each tuned to a different space-time frequency. Here we analyze, in a probabilistic framework, the properties of the motion uncertainty. Its lower bound, which can be explicitly computed through the Cramér–Rao inequality, will have different values depending on the filter parameters. We show for the Gabor filter that, in order to minimize the motion uncertainty, the spatial and temporal filter sizes cannot be arbitrarily chosen; they are only allowed to vary over a limited range of values such that the temporal filter bandwidth is larger than the spatial bandwidth. This property is shared by motion sensitive cells in the primary visual cortex of the cat, which are known to be direction selective and are tuned to space-time frequency orientations. We conjecture that these cells have larger temporal bandwidth relative to their spatial bandwidth because they compute velocity with maximum efficiency, that is, with a minimum motion uncertainty.

1 Introduction

In the space-time filtering scheme (Ahumada and Watson 1985; Adelson and Bergen 1985; Fleet and Jepson 1985; Heeger 1988; Sperling and Van Santen 1984), the process of extracting velocity (optical flow) uses the convolution of a sequence of images with a collection of space-time filters each of which is tuned to a given space-time orientation. In particular, in the case of the space-time energy filters (Adelson and Bergen 1985; Heeger 1988), which were shown to be equivalent to the correlation model used to describe the extraction of velocity by insects (Reichardt and Hassenstein 1956), each individual energy filter is oriented along a given space-time frequency band. The individual energy filters are not velocity tuned, and therefore it is necessary to use a collection of them in order to extract velocity. One of the consequences of this is that the extraction of velocity through these filters can only be done with limited precision, and so there always exists a non-zero motion uncertainty.

In this paper we will analyze the properties of space-time energy filters, and in particular we will compute and analyze the motion uncertainty. One of the main motivations in studying the structure of the motion uncertainty for the extraction of velocity through energy filters comes from the fact that, by requiring it to be minimum we will generate constraints on the range of values assumed by the various filter parameters. An elegant way to study the properties of the motion uncertainty is through the use of the Cramér–Rao inequality (VanTrees 1968) which gives us the lower bound on the mean square error for the estimation of the velocity. To deduce this inequality we use estimation theory (VanTrees 1968), and this requires the knowledge of a conditional probability function. By assuming that the images are corrupted by additive, white and Gaussian noise we are able to show that the conditional probability function for the energy noise$^1$, is given by the $\chi^2$ distribution. The resulting motion uncertainty lower bound will depend on the velocity and on the filter parameters. We show that this lower bound is minimum, that is, in a given range of velocities, it has the smallest value such that the filter parameters can only assume a limited number of values. In particular, we compute explicitly the motion uncertainty lower bound for the Gabor energy filter and show that this lower bound is minimum if the temporal bandwidth of the filter is larger than its spatial bandwidth.

The fact that the temporal bandwidth is larger than the spatial bandwidth is in accord with physiological data obtained from the primary visual cortex of cats (Holub and Morton-Gibson 1981). We therefore conjecture that this difference between the spatial and temporal bandwidths has the purpose of making the

---

$^1$ The energy noise is defined as the error in the measurement of the energy associated with the convolution of a sequence of images with a space-time oriented filter
extraction of velocity through cells in the primary visual cortex more efficient, that is, it minimizes the motion uncertainty. We are not interested in describing a detailed model for motion sensitive cells in the primary visual cortex, but instead to analyze, in a qualitative way, some common properties of these cells and energy filters. Here we will focus mainly on the mathematical properties of the energy filters.

In Sect. 2 we deduce, in the framework of estimation theory, the conditional probability function for the energy noise, and we compute in terms of this probability the Cramér-Rao lower bound for the motion uncertainty. Next, in Sect. 3, we apply this formalism to the specific case of the Gabor filter, and deduce an explicit expression for the motion uncertainty lower bound. Also, we show that, in order to minimize this lower bound, the filter parameters are required to vary inside a specific range of values. We show in Sect. 4 that the constraint involving the spatial and temporal filter sizes is such that the temporal bandwidth is larger than the spatial bandwidth. Finally, in Sect. 5 we draw conclusions and discuss some open questions.

2 Energy filters and motion uncertainty

2.1 Probabilistic model for the extraction of velocity

The process of extracting velocity (optical flow) through energy filters can be formulated in the framework of estimation theory (VanTrees 1968). Given a sequence of images \(I(x, y, t)\) we convolve it with a collection of \(N\) space-time filters \(\{U_i(x, y, t)\}_{i=1, \ldots, N}\) which are oriented in \(N\) different space-time directions. The square of the modulus of this convolution, for an arbitrary filter \(U_i(x, y, t)\), is given by

\[
II(x,y,t) = \sum x^2 dxdydt|I(x,y,t) * U_i(x,y,t)|^2,
\]

where \(\ast\) represents the operation of convolution. Since the images are corrupted by noise, which we model as additive, white and Gaussian with zero mean noise \(W(x, y, t)\) we have

\[
I(x, y, t) = I_0(x, y, t) + W(x, y, t).
\]

Using (2.2) in (2.1) and taking the expected value of (2.1) we get

\[
E[I(x, y, t) * U_i(x, y, t)]^2 = E[I_0(x, y, t)
+ W(x, y, t) * U_i(x, y, t)]^2.
\]

and by assuming that \(I_0\) is uncorrelated with \(W\), we obtain

\[
E[I(x, y, t) * U_i(x, y, t)]^2 = E[I_0(x, y, t) * U_i(x, y, t)]^2
+ E[W(x, y, t) * U_i(x, y, t)]^2.
\]

The total energy associated with (2.4) is given by

\[
G_i = F_i + N_i,
\]

where

\[
G_i = E\left[\int \int dx dy dt I(x, y, t) * U_i(x, y, t)\right]^2.
\]

\[
F_i = E \left[\int \int dx dy dt |I_0(x, y, t) * U_i(x, y, t)|^2\right],
\]

and

\[
N_i = E \left[\int \int dx dy dt |W(x, y, t) * U_i(x, y, t)|^2\right].
\]

By using Rayleigh's theorem (Bracewell 1965) in (2.7) we get

\[
F_i = \int \int dk_x dk_y dw |I_0(k_x, k_y, w) \times \hat{U}_i(k_x, k_y, w)|^2,
\]

where \(\hat{U}_i(k_x, k_y, w)\) and \(\hat{I}_0(k_x, k_y, w)\) correspond to the Fourier transforms of \(U_i(x, y, t)\) and \(I_0(x, y, t)\) respectively, and \(k_x, k_y, w\) are the \(X, Y, T\) frequency coordinates.

By assuming that, locally in the images, the velocity vector is purely translational, the contrast function \(I_0(x, y, t) = I_0(r - vt)\) has a planar support in the frequency domain, that is, \(I_0(k_x, k_y, w) = \delta(w - v \cdot k) \times \hat{I}_0(k_x, k_y)\). Inserting this into (2.9) we obtain

\[
F_i = \int \int dk_x dk_y dw \left|\hat{I}_0(k_x, k_y, w) \times \hat{U}_i(k_x, k_y, w)\right|^2.
\]

At this point we assume, similarly to Heeger (Heeger 1988), that \(I_0\) has a flat power spectrum, which is a good approximation for highly textured images. So, by using \(E[|\hat{I}_0(k_x, k_y)|^2] = \kappa^2\) in (2.10), where the image power spectrum \(\kappa^2\) is constant, we have

\[
F_i = \kappa^2 \int \int dk_x dk_y dw \left|\hat{U}_i(k_x, k_y, w)\right|^2 |_{w - z \cdot v - k}.
\]

Since, \(\hat{U}_i\) is a function of the filter parameters, which we represent by \(\{\psi_i\}_{i=1, \ldots, M}\), we have that \(F_i\) is also a function of these parameters, as a consequence of (2.11).

The extraction of velocity through energy filters is a two-stage process. The first stage is given by the convolution of \(N\) filters oriented at different space-time orientations (or space-time frequencies) with a temporal sequence of images. In the second stage we compute the velocity, at each pixel in the image, by using the output of the convolutions performed in the previous stage. This second stage can be described in the framework of estimation theory (VanTrees 1968), as will be shown next.

At each pixel in the image we want to estimate the velocity of translation \(v\) by maximizing the posterior probability density function \(P(v|G)\) where \(G\) is a \(N\)-dimensional vector whose components represent the measured energies given by (2.6). Since the logarithm is a

\[\text{2 The computation of (2.10) for an arbitrary } I_0 \text{ is difficult, unless we assume that the image can be described by some particular model which is more general than the flat power spectrum model, but nevertheless makes it possible to calculate (2.10) explicitly}\]