Computer Vision and Image Understanding 114 (2010) 681–699

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A compact harmonic code for early vision based on anisotropic frequency channels ☆

Silvio P. Sabatini [a,*], Giulia Gastaldi [a], Fabio Solari [a], Karl Pauwels [b], Marc M. Van Hulle [b], Javier Diaz [c], Eduardo Ros [c], Nicolas Pugeault [d,1], Norbert Krüger [e]

[a] Dipartimento di Ingegneria Biofisica ed Elettronica, University of Genoa, Italy
[b] Laboratorium voor Neuro- en Psychofysiologie, K.U.Leuven, Belgium
[c] Departamento de Arquitectura y Tecnologia de Computadores, University of Granada, Spain
[d] School of Informatics, University of Edinburgh, UK
[e] The Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Odense, Denmark

A R T I C L E I N F O

Article history:
Received 1 September 2008
Accepted 24 March 2010
Available online 29 March 2010

Keywords:
Early vision
Phase-based image analysis
Multichannel filtering
Image representations
Stereo
Motion
Bio-inspired vision processing

A B S T R A C T

The problem of representing the visual signal in the harmonic space guaranteeing a complete characterization of its 2D local structure is investigated. Specifically, the efficacy of anisotropic versus isotropic filtering is analyzed with respect to general phase-based metrics for early vision attributes. We verified that the spectral information content gathered through channeled oriented frequency bands is characterized by high compactness and flexibility, since a wide range of visual attributes emerge from different hierarchical combinations of the same channels. We observed that constructing a multichannel, multiorientation representation is preferable than using a more compact one based on an isotropic generalization of the analytic signal. Maintaining a channeled (i.e., distributed) representation of the harmonic content results in a more complete structural analysis of the visual signal, and allows us to enable a set of "constraints" that are often essential to disambiguate the perception of the different features. The complete harmonic content is then combined in the phase-orientation space at the final stage, only, to come up with the ultimate perceptual decisions, thus avoiding an "early condensation" of basic features. The resulting algorithmic solutions reach high performance in real-world situations at an affordable computational cost.

1. Introduction

Although the basic ideas underlying early vision appear deceptively simple and their computational paradigms are known for a long time, early vision problems are difficult to quantify and solve. Such a difficulty is often related on the representation we adopt for the visual signal, which must be capable of capturing, through proper "channels", *what* is *where* in the visual signal, that is the structural ("what") and the positional ("where") information from the images impinging the retinas. Ever since the initial formulation of the channel concept, the problem arises of jointly handling the existence of spatial frequency channels on the one hand, and of orientation channels on the other. At a local operator level, the two-dimensional (2D) Gabor filter (proposed by J. Daugman [1] and S. Marcelja [2], as an extension of its one-dimensional (1D) counterpart [3]) retains the optimal joint information resolution in both the domains and meets thoroughly this demand, by underlining the 2D nature of the frequency representation and thus being isomorphic to the 2D character of the spatial manifold of the visual/retinal image. In this way, 2D Gabor filters reconciled the "atomistic" description of early vision, based on local feature detection in the space domain with the "undulatory" interpretation, based on a Fourier-like decomposition into spatial-frequency components. Yet, the picture remained still incomplete, if one considers the representation problem as a whole. Indeed, a hybrid approach asserted itself, consisting in an energy-based multichannel feature extraction (i.e., a parametric analysis in the image domain at different frequency bandwidths), which is still reminiscent of the more intuitive atomistic description, rather than a coarse, local, frequency analysis embedded within the global space-domain mapping. For years, the phase was the missing concept, and, even when it recovered its computational significance, its role within a unifying perspective of the optimal representation of the visual signal has never been fully explored.

In this paper, we propose a general and fully conciliatory position between the spatial and spectral (i.e., frequency) approaches to early vision. Specifically, if we include the local phase as a key

☆ This work has matured from a preliminary and shorter communication presented at the 2nd International Conference on Computer Vision Theory and Applications (VISAPP'07), 8-11 March, 2007 Barcelona, Spain.

* Corresponding author. Fax: +39 010 3532289.
E-mail address: silvio.sabatini@unige.it (S.P. Sabatini).
1 Present address: University of Surrey, UK.

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.03.008

descriptive element of the visual signal, we gain a complete representation and the orientation becomes an integral part of the harmonic representation. Indeed, the extraction of local phase information in two dimensions is an intrinsically *anisotropic* problem that refers to a *selected* orientation, unless one introduces an isotropic extension of the analytic signal (Hilbert transform) and thereby the concept of *dominant* orientation. We will remark the conceptual complementarity of the two (spatial and spatial-frequency) descriptions, yet pointing out the optimal economy and the major richness of the information conveyed in the harmonic domain by anisotropic filters, which allow us to derive, with relative ease, more complex (higher-order) visual descriptors without resorting to complicated relational and/or symbolic constructors, but still operating at the signal level.

The rest of the paper is organized as follows: in Section 2, we introduce the problem of the representation of the visual signal in the harmonic space, according to different sets of band-pass filters. On that ground, we define/qualify early vision features in terms of specific phase properties and phase relationships. In Section 3, channel interaction is formalized. Section 4 presents a comparative analysis of the experimental results on both synthetic and natural image sequences. Concluding remarks in Section 6 are preceded by a general discussion in Section 5.

2. Visual features as measures in the harmonic space

The goal of early vision is to extract as much information as possible about the structural properties of the visual signal. As pointed out by [4,5], an efficient internal representation is necessary to guarantee all potential visual information can be made available for higher level analysis. The measurement of specific, significant visual "elements" in a local neighborhood of the visual signal has led to the concept of "feature" and of "feature extraction". An image feature can be defined in terms of attributes related to the visual data. Though, in practice, many features are also defined in terms of the particular procedure used to extract information about that feature, and, in more general terms, on the specific scheme adopted for the representation of the visual signal. At an early level, feature detection occurs through initial local *quantitative* measurements of basic image properties (e.g., edge, bar, orientation, movement, binocular disparity, color) referable to spatial differential structure of the image luminance and its temporal evolution (cf. linear visual cortical cell responses, see e.g. [6–8]). Later stages in vision can make use of these initial measurements by combining them in various ways, to come up with categorical *qualitative* descriptors, in which information is used in a non-local way to formulate more global spatial and temporal predictions (e.g., see [9]).

The receptive fields of the cells in the primary visual cortex have been interpreted as fuzzy differential operators (or local *jets* [4]) that provide regularized partial derivatives of the image luminance along different directions and at several levels of resolution, simultaneously. The jets characterize the local geometry in the neighborhood of a given point $\mathbf{x} = (x, y)$. The order of the jet determines the amount of geometry represented. Given the 2D nature of the visual signal, the spatial direction of the derivative (i.e., the orientation of the corresponding local filter) is an important "parameter". Within a local jet, the directionally biased receptive fields are represented by a set of similar filter profiles that merely differ in orientation.

Alternatively, considering the space/spatial-frequency duality [3,1], the local jets can be described through a set of 2D spatial-frequency channels, which, for each spatial orientation, are selectively sensitive to a different limited range of spatial frequencies. These oriented spatial-frequency channels are equally apt as the spatial ones [10]. From this perspective, it is formally possible to derive, on a local basis, a complete harmonic representation (amplitude, phase, and orientation) of any visual stimulus, by defining the associated analytic signal in a combined space-frequency domain through filtering operations with complex-valued 2D band-pass kernels. Since spatial information is being linearly transformed from the space domain at the level of pixels, into a combined space-frequency domain at a cortical-like representation level, no actual analysis is taking place at this level, and the information is merely being put into another (presumably more useful) form [11].

Formally, due to the impossibility of a direct definition of the analytic signal in two dimensions, a full harmonic characterization of the 2D spatial vision channels in the Fourier domain requires explicit (1D) reference axes, and their association with the orientation of the spatial frequency channel must be discussed. Basically, this association can be handled either (1) 'separately', for each orientation, by using Hilbert pairs of band-pass filters that display symmetry and antisymmetry about a steerable axis of orientation or (2) 'as-a-whole', by introducing a 2D isotropic generalization of the analytic signal: the monogenic signal [12], which allows us to build isotropic harmonic representations that are independent of the orientation (i.e., omnidirectional). By definition, the monogenic signal is a 3D phasor in spherical coordinates and provides a framework to obtain the harmonic representation of a signal respect to the dominant orientation of the image that becomes part of the representation itself. In the first case, for each orientation θ, an image $I(\mathbf{x})$ is filtered with a complex-valued filter:

$$f_A^\theta(\mathbf{x}) = f^\theta(\mathbf{x}) - i f_{\mathscr{H}}^\theta(\mathbf{x}) \qquad (1)$$

where $f_{\mathscr{H}}^\theta(\mathbf{x})$ is the Hilbert transform of $f^\theta(\mathbf{x})$ with respect to the axis orthogonal to the filter's orientation:

$$f_{\mathscr{H}}^\theta(\mathbf{x}) = f_{\mathscr{H}}(x_\theta, y_\theta) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{f(\xi, y_\theta)}{\xi - x_\theta} \mathrm{d}\xi$$

with x_θ and y_θ the principal axes of the energy distribution of the filter in the spatial domain.

This results in a complex-valued *analytic image*:

$$Q_A^\theta(\mathbf{x}) = I * f_A^\theta(\mathbf{x}) = C_\theta(\mathbf{x}) + i S_\theta(\mathbf{x}), \qquad (2)$$

where $C_\theta(\mathbf{x})$ and $S_\theta(\mathbf{x})$ denote the responses of the quadrature filter pair. For each spatial location, the amplitude $\rho_\theta = \sqrt{C_\theta^2 + S_\theta^2}$ and the phase $\phi_\theta = \mathrm{atan2}(S_\theta, C_\theta)$ envelopes measure the harmonic information content in a limited range of frequencies and orientations to which the channel is tuned (see Fig. 1a).

In the second case, the image $I(\mathbf{x})$ is filtered with a *spherical quadrature filter* (SQF):

$$f_M(\mathbf{x}) = f(\mathbf{x}) - (i, j) \cdot \mathbf{f}_{\mathscr{R}}(\mathbf{x}) \qquad (3)$$

defined by a radial bandpass filter $f(\mathbf{x})$ (i.e., rotation invariant even filter) and a vector-valued isotropic odd filter $\mathbf{f}_{\mathscr{R}}(\mathbf{x}) = (f_{\mathscr{R},1}(\mathbf{x}), f_{\mathscr{R},2}(\mathbf{x}))^T$, obtained by the Riesz transform of $f(\mathbf{x})$ [12]:

$$\mathbf{f}_{\mathscr{R}}(\mathbf{x}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\xi}{|\xi|^3} f(\mathbf{x} - \xi) \mathrm{d}\xi \qquad (4)$$

This results in a *monogenic image*:

$$Q_M(\mathbf{x}) = I * f_M(\mathbf{x}) = C(\mathbf{x}) + (i, j)\mathbf{S}(\mathbf{x}) = C(\mathbf{x}) + i S_1(\mathbf{x}) + j S_2(\mathbf{x}) \qquad (5)$$

where using the standard spherical coordinates,

$$C(\mathbf{x}) = \rho(\mathbf{x}) \cos \varphi(\mathbf{x})$$
$$S_1(\mathbf{x}) = \rho(\mathbf{x}) \sin \varphi(\mathbf{x}) \cos \vartheta(\mathbf{x})$$
$$S_2(\mathbf{x}) = \rho(\mathbf{x}) \sin \varphi(\mathbf{x}) \sin \vartheta(\mathbf{x}).$$

The amplitude of the monogenic signal is the vector norm of $f_M : \rho = \sqrt{C^2 + S_1^2 + S_2^2}$, as in the case of the analytic signal, and,