Neurocomputing 150 (2015) 227–237

Contents lists available at ScienceDirect

Neurocomputing

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

BIMP: A real-time biological model of multi-scale keypoint detection in V1

Kasim Terzić *, João M.F. Rodrigues, J.M. Hans du Buf

Vision Lab (LARSyS), FCT, University of the Algarve, Gambelas Campus, 8000 Faro, Portugal

ARTICLE INFO

Article history:
Received 18 January 2014
Received in revised form
3 July 2014
Accepted 15 September 2014
Available online 27 October 2014

Keywords:
Computer vision
Gabor filter
V1
Keypoint
Categorization

ABSTRACT

We present an improved, biologically inspired and multiscale keypoint operator. Models of single- and double-stopped hypercomplex cells in area V1 of the mammalian visual cortex are used to detect stable points of high complexity at multiple scales. Keypoints represent line and edge crossings, junctions and terminations at fine scales, and blobs at coarse scales. They are detected by applying first and second derivatives to responses of complex cells in combination with two inhibition schemes to suppress responses along lines and edges. A number of optimisations make our new algorithm much faster than previous biologically inspired models, achieving real-time performance on modern GPUs and competitive speeds on CPUs. In this paper we show that the keypoints exhibit state-of-the-art repeatability in standardised benchmarks, often yielding best-in-class performance. This makes them interesting both in biological models and as a useful detector in practice. We also show that keypoints can be used as a data selection step, significantly reducing the complexity in state-of-the-art object categorisation.

1. Introduction

Accurate detection of stable interest points is a central task in many object detection and recognition approaches, and an important part of early human visual processing. While many computer vision algorithms have been motivated by insights gained from biological vision, including image processing with Gabor wavelets and current work on deep hierarchies, existing biologically plausible keypoint detection algorithms are limited to a single scale [1], or are computationally too complex to run in real time on a CPU [2]. Furthermore, no comparative benchmarking of biological keypoint models is available in the literature. In this paper, we present an optimised keypoint extraction algorithm based on existing models of end-stopped cells in the mammalian striate cortex and evaluate its performance.

Early processing in the area V1 of the mammalian visual cortex has been extensively studied in the literature. The image signal from retina enters V1 via the Lateral Geniculate Nucleus (LGN) and is then processed by layers of the so-called simple cells, complex cells and hypercomplex (or end-stopped) cells. Simple cells, often modelled using oriented Gabor filters, respond to lines and edges. Complex cells provide more position-invariant responses to both. End-stopped cells respond to line terminations (single-stopped cells), as well as to corners and blobs (double-stopped cells). Earlier work has shown that models of these cells can act as a general-purpose keypoint detector, but they require convolutions with large filter kernels, making them prohibitively slow for most applications in computer vision and cognitive robotics.

The main contributions of this paper are (i) a new and optimised algorithm which is fast enough to run on a CPU and which runs in real time on GPUs due to its parallel nature; and (ii) extensive benchmarking of the algorithm, showing state-of-the-art performance compared to best available algorithms, and setting several records in terms of repeatability and precision. To the best of our knowledge, this is the first extensive comparison of a biological model with the state of the art in computer vision. We have released the CPU and GPU implementations of our detector as Free Software, so others can use them for real-world applications.

1.1. Related work

There exist a number of approaches for detecting interest points in images which are stable under a wide range of transformations, including scaling, translation and rotation. Early work on corner detection used structure tensors [3,4], which have recently been extended to provide scale invariance [5]. Other computational approaches include Difference of Gaussians [6] and the Determinant of

* Correspondence to: CINTAL, University of the Algarve, Gambelas Campus, 8000 Faro, Portugal. Tel.: +351 962846514.
E-mail addresses: kterzic@ualg.pt (K. Terzić), jrodrig@ualg.pt (J.M.F. Rodrigues), dubuf@ualg.pt (J.M.H. du Buf).
URLS: http://w3.ualg.pt/~kterzic (K. Terzić), http://w3.ualg.pt/~jrodrig (J.M.F. Rodrigues), http://w3.ualg.pt/~dubuf (J.M.H. du Buf).

http://dx.doi.org/10.1016/j.neucom.2014.09.054
0925-2312/

Hessian [7]. Meaningful blobs have been detected using region-based methods [8,9] and by other affine-invariant region detectors. Additional interest point and region detectors are described in [10].

Biologically inspired approaches to keypoint detection attempt to model early processing stages of the mammalian visual cortex (area V1), consisting of layers of cells. The so-called simple cells are modelled using a bank of bandpass filters, usually Gabor wavelets. Beyond this, responses of complex and end-stopped cells are often represented implicitly, as a spatial combination of simple-cell responses [11,12]. However, there have also been efforts to model complex and end-stopped cells directly, in order to obtain an explicit representation of keypoints corresponding to strong activations of complex cells [1,13], but there is no comparative benchmarking of such models against state-of-the-art keypoint detection in computer vision.

Our model follows the early single-scale model of Heitger et al. [14], which consists of single and double end-stopped cells and two inhibition schemes. Several extensions have been proposed [2,15], capable of detecting keypoints at multiple scales and adapting NCRF inhibition [16] to keypoints. Our new model is inspired on the one by Rodrigues and du Buf [2], which is too slow for practical use, taking hours on large images. In this paper, we expand on our earlier work presented in [17].

We completely reformulate and re-implement the algorithm. Instead of modelling individual cells, as in [2], we model populations of cells as activation maps, obtained by parallel filtering operations which can be efficiently evaluated on modern CPUs and GPUs. We use a Gaussian pyramid combined with sub-pixel localisation and show that this step significantly improves repeatability compared to [2]. We also introduce a scale selection method which reduces the redundancy of detected keypoints. These changes result in a significant improvement in both speed and accuracy, such that biologically inspired keypoints are now suitable for real-time applications. We benchmark our improved approach on standard datasets, showing that it improves on both [2] and the state of the art in computer vision.

2. An optimised computational model of V1

Our approach is based on area V1 of the mammalian visual cortex, with layers of specialised cells responding to increasingly complex patterns. At the highest level, responses of single and double end-stopped cells are used to detect stable events (corners, blobs and terminators) at all scales.

Basically, the keypoint model works as follows. Simple and complex cells respond to lines and edges. Assume that there is a corner formed by a vertical and a horizontal edge, and the goal is to detect only the corner position. Complex cells tuned to the edge orientations will produce a maximum response at the edge positions. Cells tuned to other orientations will also respond at the edges, but less. Now, single and double end-stopped cells are modelled by first and second derivatives of the responses of complex cells, in the same orientations as those of the complex cells; see Fig. 1 (top).

This implies that first derivatives (single-stopped cells) will produce responses astride the edges: on both sides but zero in the middle. Second derivatives (double-stopped cells) will also produce responses, but these are maximum at the edge positions and they decrease on both sides. Hence, when all responses are summed over all orientations, there will be a peak at the corner, where all cells respond, but also significant responses (derivatives) at and astride the two edges. Responses along edges are a common problem in keypoint detection. For example, difference of Gaussian blob detection used by the SIFT algorithm produces strong responses along edges, just like complex cells in our model, which results in poorly localised features. SIFT uses the ratio of eigenvalues of the Hessian matrix to discard keypoints along edges. In our model, we apply two inhibition schemes to the responses of complex cells to suppress such responses when applying end-stopped cells. Tangential inhibition serves to suppress all responses astride the edges. Radial inhibition suppresses responses at the edges but, because of the orthogonal kernels, not at the corner. For a detailed explanation of the inhibition schemes used in our algorithm, we refer to Fig. 10 in [14].

2.1. Multi-scale filter kernels as V1 model

The multi-scale extension of the Heitger et al. model [2] applies the same derivation and inhibition schemes. Obviously, at coarser scales the sizes of all cell models are bigger, and this makes the multi-scale model so expensive in terms of computations.

In our new model, each layer of cells is modelled as a linear filtering operation, where the kernel corresponds to a typical weight profile of a particular type of cell. Unlike the original computational approach [2], this formulation allows for easy parallel implementation on GPUs and consistent use of filtering in the frequency domain. As is common, we define simple cells using complex Gabor filters

$$g_{\lambda,\sigma,\theta}(x,y)=\exp\left(-\frac{\tilde{x}^2+\gamma\tilde{y}^2}{2\sigma^2}\right)\exp\left(i\frac{2\pi\tilde{x}}{\lambda}\right), \tag{1}$$

with $\tilde{x}=x\cos\theta+y\sin\theta$, $\tilde{y}=y\cos\theta-x\sin\theta$ and $\gamma=0.5$. λ is the wavelength (in pixels) and sigma is the receptive field size (in pixels), which are related by $\sigma/\lambda=0.56$. θ determines the filter orientation (typically eight orientations are used). Simple cell responses R are obtained by convolving the image I with the complex Gabor filter, and complex cells C are defined as the moduli of the simple cell responses

$$R_{\lambda,\theta}=I*g_{\lambda,\theta};\quad C_{\lambda,\theta}=|R_{\lambda,\theta}|. \tag{2}$$

Simple cells respond to line and edge stimuli, complex cells respond to both and exhibit more spatial invariance. All other cells are defined by employing combinations of Gaussian filter kernels. Let $G(\hat{\sigma})$ be a 2D Gaussian function with standard deviation $\hat{\sigma}$ centered at the origin, and $G(x,y,\hat{\sigma})$ its equivalent centered at x,y. Let $ds=0.6\lambda\sin\theta$ and $dc=0.6\lambda\cos\theta$ be offsets from the kernel centre. Then kernels representing single- and double-stopped cells are defined by

$$k^S_{\lambda,\theta}=G(ds,-dc,\hat{\sigma})-G(-ds,dc,\hat{\sigma}), \tag{3}$$

$$k^D_{\lambda,\theta}=G(\hat{\sigma})-\tfrac{1}{2}G(-2ds,2dc,\hat{\sigma})-\tfrac{1}{2}G(2ds,-2dc,\hat{\sigma}). \tag{4}$$

The parameters used in this step were carefully selected in order to obtain best results. $\hat{\sigma}$ is used to control the amount of smoothing performed at this step, which is useful for reducing the effect of noise, and is typically set to $\sigma/2$. When $\hat{\sigma}$ approaches zero, the kernels become a combination of Dirac functions. This can be a useful optimisation at the expense of noise sensitivity, so we use this in our CPU-based implementation. End-stopped cell response maps are then computed by convolutions

$$S_{\lambda,\theta}=C_{\lambda,\theta}*k^S_{\lambda,\theta};\quad D_{\lambda,\theta}=C_{\lambda,\theta}*k^D_{\lambda,\theta}. \tag{5}$$

In order to suppress responses along lines and edges, tangential and radial inhibition are used, as in [2]. Each one is modelled as a layer of inhibition cells represented by the two kernels

$$k^{IT}_{\lambda,\theta}=-2G(\hat{\sigma})+G(dc,ds,\hat{\sigma})+G(-dc,-ds,\hat{\sigma}), \tag{6}$$

$$k^{IR}_{\lambda,\theta}=G(dc/2,ds/2,\hat{\sigma})+G(-dc/2,-ds/2,\hat{\sigma}). \tag{7}$$

Inhibition cell response maps are obtained by convolving the responses of complex cells with these kernels

$$I^T_{\lambda,\theta}=C_{\lambda,\theta}*k^{IT}_{\lambda,\theta};\quad I^R_{\lambda,\theta}=C_{\lambda,\theta}*2G(\hat{\sigma})-C_{\lambda,\theta+\pi/2}*Ak^{IR}_{\lambda,\theta} \tag{8}$$

where A determines the inhibition strength, usually set between 4 and 16. Note that radial inhibition uses two response maps of complex