Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Generative part-based Gabor object detector[☆]

Ekaterina Riabchenko[a], Joni-Kristian Kämäräinen[b,*]

[a] *Department of Mathematics and Physics, Lappeenranta University of Technology, Finland*
[b] *Department of Signal Processing, Tampere University of Technology, Finland*

ARTICLE INFO

Article history:
Received 20 March 2015
Available online 20 August 2015

Keywords:
Gabor feature
Gaussian mixture model
Object detection
Visual classification
Generative learning

ABSTRACT

Discriminative part-based models have become the approach for visual object detection. The models learn from a large number of positive and negative examples with annotated class labels and location (bounding box). In contrast, we propose a part-based generative model that learns from a small number of positive examples. This is achieved by utilizing "privileged information", sparse class-specific landmarks with semantic meaning. Our method uses bio-inspired complex-valued Gabor features to describe local parts. Gabor features are transformed to part probabilities by unsupervised Gaussian Mixture Model (GMM). GMM estimation is robustified for a small amount of data by a randomization procedure inspired by random forests. The GMM framework is also used to construct a probabilistic spatial model of part configurations. Our detector is invariant to translation, rotation and scaling. On part level invariance is achieved by pose quantization which is more efficient than previously proposed feature transformations. In the spatial model, invariance is achieved by mapping parts to an "aligned object space". Using a small number of positive examples our generative method performs comparably to the state-of-the-art discriminative method.

1. Introduction

Discriminative part-based models have become the approach for visual object detection and achieve state-of-the-art for various datasets, e.g., Caltech-101 [10], Caltech-256 [16] and Pascal VOC [9]. Part-based models have two detection stages: detection of object parts and verifying detected parts' spatial configuration (constellation). The first methods with explicit spatial models were generative [11,43], but recent methods are based on discriminative learning from a large number of positive and negative examples with manually annotated class labels and location, e.g., the deformable part-based model (DPM) by [13,14].

The recent "big visual data" datasets, such as the ImageNet ILSVRC [8,34], provide sufficient number of data for training deep architectures with millions of parameters to be optimized [22,37] and which are superior to the previous part-based models. However, with limited data and in specific applications part-based models and hybrids of deep architectures and part-based models perform extremely well [25,41,42]. Despite dominance of discriminative learning in visual classification, generative models have desirable properties such as prior probabilities, learning from unlabeled data and visual synthesis, and therefore provide an alternative approach to be investigated.

In this work, we propose a part-based generative model (Fig. 1) that learns from a small number of positive examples. This is achieved by utilizing "privileged information", sparse class-specific landmarks with semantic meaning. Our method uses bio-inspired complex-valued Gabor features to describe local parts. Gabor features are transformed to part probabilities by unsupervised Gaussian Mixture Model (GMM) probability densities. GMM estimation is robustified for a small amount of examples by novel randomized training inspired by random forests. GMMs are also used to represent spatial probabilities of the part configurations. Our detector is invariant to translation, rotation and scaling. On the part level, this is achieved by pose quantization which is more efficient than the previously proposed feature transformations [20]. In the spatial level, invariance is achieved by mapping parts to an "aligned object space". Using a small number of positive examples our generative method performs comparably to the state-of-the-art discriminative method.

2. Related work

Part-based models – The visual Bag-of-Words (BoW) [6,38] methods are omitted here since their spatial models are not explicit (see the recent survey by [19]). The first part-based methods with constellation models were generative [11,12,43], but since then the field has been dominated by discriminative learning and, in particular, the deformable part-based model (DPM) by [13,14]. Recently, [45] introduced a generative FRAME model which also uses Gabor features, but the model is computationally intensive, window-based and

[☆] This paper has been recommended for acceptance by R. Davies.
[*] Corresponding author. Tel.: +358 50 3001851.
E-mail address: joni.kamarainen@tut.fi, First.Second@tut.fi (J.-K. Kämäräinen).

http://dx.doi.org/10.1016/j.patrec.2015.08.004
0167-8655/© 2015 Elsevier B.V. All rights reserved.

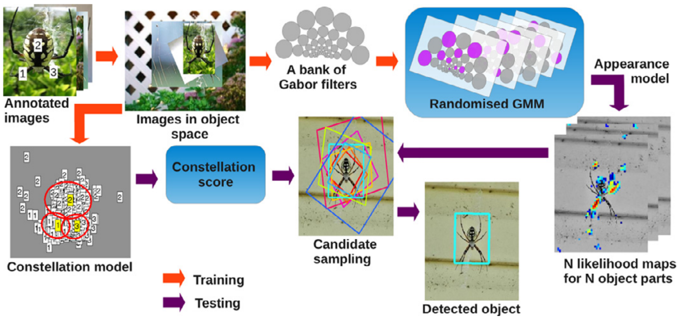

Fig. 1. Workflow of our generative learning and detection.

embeds geometry variation to appearance. [1] proposed a generative poselet model for part-based human pose detection, but its generality to other classes is unclear. Our method differs rather strongly from the above by the facts that it is generative, is generic, and has explicit models for the parts and constellation.

In particular, we extend our previous works of Gabor feature extraction [23] and Gaussian mixture model probabilistic part descriptor [30]. Our quantized object pose space (Section 5.1) avoids the computationally expensive matrix shifts in [23] and the proposed randomized Gaussian mixture model (Section 4.3) can exploit a large Gabor filter bank and still learn a model from a few examples instead of hundreds required in [30]. Preliminary results have been published in two conference papers, part detector in [33] and spatial model in [32], while this work refines the theory to form a single probabilistic framework, simplifies computation and improves performance by the quantized pose space, reports results from extensive experiments along with the full source code available in a public repository[1].

Contributions – We make the following contributions:

- A probabilistic (generative) local part descriptor using complex-valued multi-resolution Gabor features.
  - In contrast to a small size Gabor bank used in the literature, we use a large bank and propose a method to identify a part-specific subset of the filters.
  - We avoid using heuristic prior distributions to learn from a small number of training examples by a novel random forest inspired generative learning procedure: *randomized Gaussian mixture model*.
  - We propose a likelihood-driven part detection procedure with efficient non-maximum suppression.
- A probabilistic part spatial constellation model in “aligned object space”
  - The model combines the probability terms of parts and their constellation.
  - The aligned space is formed by quantizing object appearances over rotation and scales.
- In extensive experiments on Caltech and ImageNet images our method performs favorably to the popular DPM.

3. Local Gabor descriptor

Gabor features have been successful in many vision applications such as iris and face recognition [7,36]. They are considered as texture descriptors [2,18,26], but local part description was one of the first applications [24,44]. We adopt the multi-resolution Gabor feature - “simple Gabor feature space” - by [20,23]:

$$\begin{aligned}\psi(x,y) &= \frac{f^2}{\pi\gamma\eta} e^{-\left(\frac{f^2}{\gamma^2}x'^2+\frac{f^2}{\eta^2}y'^2\right)} e^{j2\pi f x'} \\ x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \quad . \end{aligned} \tag{1}$$

f is the discrete tuning frequency, θ the rotation angle, γ the sharpness (bandwidth) of the major axis, and η of the minor axis. The spatial domain filter in (1) is a complex plane wave (a 2D Fourier basis function) multiplied by a Gaussian, and in the frequency domain it is a single real-valued Gaussian centered at f. The multi-resolution form and parametrisation in (1) enforces self-similarity: filters are scaled and rotated versions of each other, “Gabor wavelets”.

Multi-resolution Gabor features are constructed from responses of filters tuned to the multiple frequencies f_m and orientations θ_n. Scales (f) are drawn from the exponential scale

$$f_m = k^{-m} f_{max}, \quad m = \{0, \ldots, M-1\} \tag{2}$$

where f_m is the mth frequency, $f_0 = f_{max}$ is the highest frequency, and $k > 1$ is the frequency scaling factor. The filter orientations are uniformly sampled:

$$\theta_n = \frac{n2\pi}{N}, \quad n = \{0, \ldots, N-1\} \tag{3}$$

where θ_n is the nth orientation and N is their total number.

The multi-resolution Gabor parameters f_{max}, k, M, N, γ and η are redundant and an intuitive parametrisation is to set the filter cross points to $p = 0.5$ when the filter envelopes cross at the half magnitude providing sufficient “shiftability” [35]. In that case, the adjustable parameters are the highest frequency f_{max}, the number of frequencies m and the number of orientations n. The bandwidths γ and η are automatically set.

Descriptor invariance – The simple Gabor feature space part descriptor at the location (x_0, y_0) forms a Gabor response matrix:

$$\mathbf{G} = \begin{pmatrix} r(x_0,y_0;f_0,\theta_0) & r(x_0,y_0;f_0,\theta_1) & \cdots & r(x_0,y_0;f_0,\theta_{n-1}) \\ r(x_0,y_0;f_1,\theta_0) & r(x_0,y_0;f_1,\theta_1) & \cdots & r(x_0,y_0;f_1,\theta_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_0,y_0;f_{m-1},\theta_0) & r(x_0,y_0;f_{m-1},\theta_1) & \cdots & r(x_0,y_0;f_{m-1},\theta_{n-1}) \end{pmatrix} \tag{4}$$

where rows denote different frequencies and columns orientations. The first row is the highest frequency $f_0 = f_{max}$ and the first column $\theta_0 = 0^\circ$.

Column and row shifts of the response matrix provide invariance to geometric transformations, scaling and rotation [20]. For example, anti-clockwise rotation of an image by $\frac{\pi}{N}$ corresponds to a single shift operation:

$$\begin{pmatrix} r(x_0,y_0;f_0,\theta_{n-1})^* & r(x_0,y_0;f_0,\theta_0) & \Rightarrow & r(x_0,y_0;f_0,\theta_{n-2}) \\ r(x_0,y_0;f_1,\theta_{n-1})^* & r(x_0,y_0;f_1,\theta_0) & \Rightarrow & r(x_0,y_0;f_1,\theta_{n-2}) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_0,y_0;f_{m-1},\theta_{n-1})^* & r(x_0,y_0;f_{m-1},\theta_0) & \Rightarrow & r(x_0,y_0;f_{m-1},\theta_{n-2}) \end{pmatrix} \tag{5}$$

A similar shift operation exists for scaling, but in Section 5.1 we show that object poses are heavily “quantized” in the datasets and we propose invariant matching without the shift operations.

Importance of complex phase – Unlike the most other works which use only the magnitude information, our Gabor feature descriptor is complex-valued which is justified by the three important findings: (1) the phase information plays a dominant role for visual representation (Fig. 2) [29]; (2) complex representation provides superior performance (see the experiments section); and (3) complex covariance matrix is more compact in our Gaussian mixture model probability density (Section 4.2).

4. Learning and detecting object parts

Our generative model builds upon the probabilistic models of object parts F_i, $p(\boldsymbol{G}|F_i)$, where $\boldsymbol{G}$ is the local Gabor descriptor computed at location (x_0, y_0). Our workflow has three processing stages

[1] https://bitbucket.org/EkaterinaRiabchenko/gabor_object_detector_code/.