

# Learning Generative Models of Object Parts from A Few Positive Examples

Ekaterina Riabchenko

Machine Vision and Pattern Recognition Laboratory  
Lappeenranta University of Technology

Joni-Kristian Kämäräinen and Ke Chen

Department of Signal Processing  
Tampere University of Technology

**Abstract**—A number of computer vision problems such as object detection, pose estimation, and face recognition utilise local parts to represent objects, which include the distinguished information of objects. In this work, we introduce a novel probabilistic framework which automatically learns class-specific object parts (landmarks) in generative-learning manner. Encouraged by the success in learning and detecting facial landmarks, we employ bio-inspired multi-resolution Gabor features in the proposed framework. Specifically, complex-valued Gabor filter responses are first transformed to landmark specific likelihoods using Gaussian Mixture Models (GMM), and then efficient response matrix shift operations provide detection over orientations and scales. We avoid the undesirable characteristic of generative learning, a large number of training instances, with the novel concept of randomised Gaussian mixture model. Extensive experiments with public benchmarking Caltech-101 and BioID datasets demonstrate the effectiveness of our proposed method for localising object landmarks.

## I. INTRODUCTION

Providing detection of specific object parts, such as “a tyre of a car” has its significance in a number of computer vision problems such as object detection [1]–[4], pose estimation [5]–[7], and face recognition [8], [9] utilising the part-based representation. The motivation of detecting object parts are under the assumption that a small set of parts contain sufficient discriminative information. However, localising object parts remains challenging due to the variant appearance of object parts as well as the changes of illumination and viewing angles.

Recently, part-based methods [1]–[3] implicitly modelling the spatial relationship between object parts have proved the success in visual recognition. In their methods, shared visual codebooks [10] are constructed based on either interest region detectors [11], [12] or dense sampling [13], [14]. However, the discriminative information of object parts, which can play an important role (verified in fine-grained visual recognition [2]), are missing due to failing to localise object parts explicitly. We propose a novel approach in this work: generative learning of (probabilistic) models of object parts represented by sparse landmarks, and their automatic detection (see Figure 1 for the illustration). Our method learns effective detectors from a small set of positive examples.

### A. Related work

**Implicit object parts detection** – The part-based approach for object or object category description, detection and classification, assumes that objects can be represented as rigid or deformable constellations of local parts. In the part-based approach, some distortion types, such as pose invariance (rotation, scale, translation), are easier to realise on the local

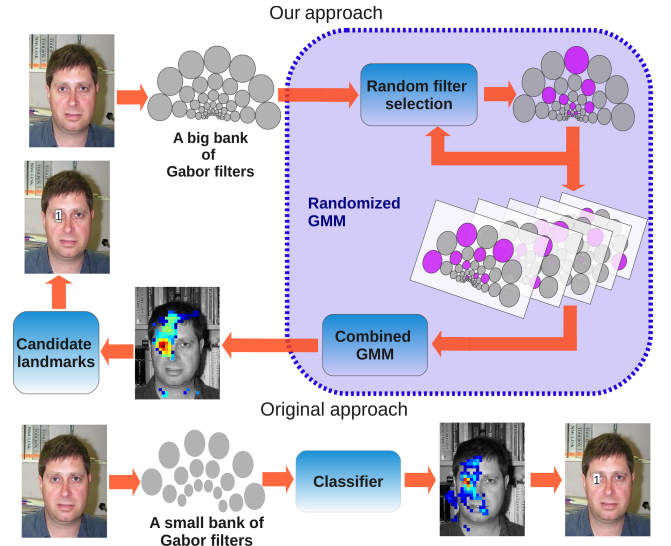


Figure 1. A flowchart showing the difference between our and the original approaches.

part detection level, and some other distortions, such as the occlusion, on the constellation matching level. In [15], similar to the concept of our work, Gaussian derivatives (steerable filters) were adopted and part pdf’s were estimated by a single diagonal Gaussian. Manually annotated landmarks inside objects are used in the method by Berthod et al. [16]. They use the scanning window approach, extract various features, and use randomised classification trees to classify each location into a specific part class. Tree outputs are also transformed to heuristic scores. However, the distinguished information provided by object parts in part-based models [1] cannot also explicitly localise the object parts having semantic meaning. In the light of this, our method can be viewed as a learning and detection pipeline to explicitly provide the locations of parts, where the multi-resolution Gabor features can be replaced with other local image descriptors, such as steerable filters [17] or local binary pattern (LBP) [18]. These were experimentally compared in our previous work [19], where the multi-resolution Gabor features obtained superior results. Moreover, the learning part could be replaced with any classifier, but in our experiments the proposed Gaussian mixture model pdf estimation outperformed, for example, support vector machines [20] and no negative examples are needed.

**Multi-resolution Gabor feature** – Our method utilises multi-resolution Gabor features which have been particularly suc-

successful in many computer vision and image processing applications, especially in biometrics: iris recognition [21], face recognition (two best in [22]) and fingerprint matching [23]. Traditionally, Gabor features have been considered as texture descriptors [24]–[26], but also encoding of local object parts was one of the first applications, e.g., in the dynamic link architecture by Lades et al. [27] and its extension to elastic bunch graph matching [28]. In this work, we utilise a matrix structure of Gabor responses referred to as multi-resolution Gabor feature or simple Gabor feature space [29], which enable efficient row- and column-wise shifts to search local parts over varying scale and orientation. The multi-resolution Gabor responses are transformed to object part conditional probability likelihoods using Gaussian mixture models [30].

**Contributions** – The main contributions and novelties of our work are four-fold:

- We propose an alternative approach to interest point and dense sampling based approaches – an explicit probabilistic “vocabulary” of object parts.
- We developed a feature selection procedure which uses much more Gabor filters than used in existing literature and thus exploits the power of Gabor space in full.
- To overcome the problem of rapidly increasing number of training examples due to a large number of Gabor features, we constructed a novel random forest inspired generative learning procedure which learns effectively from a small number of examples: *randomised Gaussian mixture model*.
- We designed a likelihood-driven part detection procedure with efficient non-maximum suppression.

The effectiveness of our approach is verified in the extensive experiments and full source codes, experiment scripts and data will be made publicly available.

## II. MULTI-RESOLUTION GABOR FEATURES

The core element is the 2D Gabor filter function [31]:

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

where  $f$  is the central frequency of the filter,  $\theta$  the rotation angle of the Gaussian major axis and the plane wave,  $\gamma$  the sharpness (bandwidth) along the major and  $\eta$  along the minor axis. In the given form, the aspect ratio of the Gaussian is  $\eta/\gamma$ . This function has the following analytical form in the frequency domain

$$\begin{aligned} \Psi(u, v) &= e^{-\frac{\pi^2}{f^2}(\gamma^2(u'-f)^2 + \eta^2 v'^2)} \\ u' &= u \cos \theta + v \sin \theta \\ v' &= -u \sin \theta + v \cos \theta. \end{aligned} \quad (2)$$

In the spatial domain in (1), the Gabor filter is a complex plane wave (a Fourier basis) multiplied by an origin-centred Gaussian, and in the frequency domain in (2), it is a single real-valued Gaussian centred at  $f$ . The multi-resolution forms and parametrisation in (1) and (2) are used in this work.

Multi-resolution Gabor features (Figure 2) are constructed from responses of the filters in (1) or (2) by using multiple filters on several frequencies  $f_m$  and orientations  $\theta_n$ . Frequency

corresponds to the scale and is thus drawn from [19]

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

where  $f_m$  is the  $m$ th frequency,  $f_0 = f_{max}$  is the highest frequency, and  $k > 1$  is the frequency scaling factor. The filter orientations are drawn from

$$\theta_n = \frac{n2\pi}{N}, \quad n = \{0, \dots, N-1\}.$$

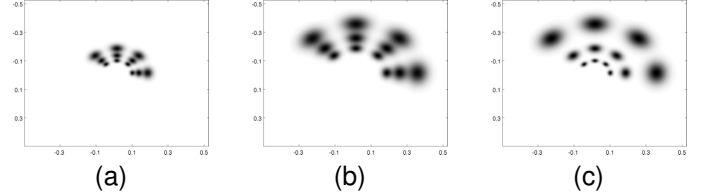


Figure 2. Multi-resolution Gabor feature: (a)  $M = 4$  orient. and  $N = 3$  freq., (b) the base frequency  $f_{max}$  increased, (c) the scaling factor  $k$  increased.

### A. Feature matrix

Instead of a patch of pixels used by the popular descriptors such as SIFT [32], we do not fix the patch size, but only fix the centroid of a part and describe the part by multi-resolution Gabor feature at the centroid point. This provides computational simplicity for processing.

The multi-resolution Gabor features at  $(x_0, y_0)$  can be arranged into a matrix form

$$\mathbf{F} = \begin{pmatrix} \psi(x_0, y_0; f_0, \theta_0) & \psi(x_0, y_0; f_0, \theta_1) & \dots & \psi(x_0, y_0; f_0, \theta_{n-1}) \\ \psi(x_0, y_0; f_1, \theta_0) & \psi(x_0, y_0; f_1, \theta_1) & \dots & \psi(x_0, y_0; f_1, \theta_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi(x_0, y_0; f_{m-1}, \theta_0) & \psi(x_0, y_0; f_{m-1}, \theta_1) & \dots & \psi(x_0, y_0; f_{m-1}, \theta_{n-1}) \end{pmatrix}$$

where rows correspond to responses on the same frequency ( $f_0 = f_{max}$ ) and columns correspond to responses at the same orientation ( $\theta_0 = 0^\circ$ ). Despite computation in a single location, every filter “sees” to its vicinity defined its effective bandwidth (the Gaussian envelope controlled by  $\gamma$  and  $\eta$ ). The property which makes multi-resolution Gabor features computationally attractive is the fact that simple row-wise and column-wise shifts of the response matrix correspond to scaling and rotation in the input space [29].

## III. LEARNING GENERATIVE PART MODELS

The key steps of our procedure depicted in Figure 1 are: i) transformation of all training images to the aligned object space (Sec. III-A), ii) extraction of Gabor features at part locations (Sec II), iii) probability density (pdf) estimation using the randomized Gaussian mixture model procedure (Sec. III-B). These steps will be described in the following.

### A. Aligned object space

In order to remove pose variation of objects in the training images (see Figure 3) we establish a normalised object space for each class and transform all training images to the space.

We could fix the coordinates of some landmarks similar to [33], but since that would re-distribute the spatial variation of the fixed landmarks, we adopt the mean shape concept proposed by Cootes et al. [34]. Our mean “shape” process using the parts is given in Algorithm 1 (see Figure 3 right-hand-side for the result).

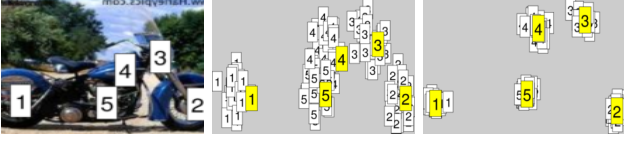


Figure 3. From left: a class example with annotated parts (Caltech-101 motorbike), all examples in the original coordinates (yellow: the example) and all examples in the aligned space (yellow tags).

---

**Algorithm 1** Aligned mean object space.

---

- 1: Select a random seed image and use its landmarks as the initial object space.
  - 2: **for all** images **do**
  - 3: Estimate homography to the object space (e.g., using Umeyama’s method [35]).
  - 4: Transform image landmarks to the object space.
  - 5: Refine the object space by computing average of transformed landmarks.
  - 6: **end for**
  - 7: Return the mean object space and transform all images and landmarks.
- 

**B. Randomised Gaussian Mixture Model PDF**

The standard workflow of Gabor features (“original approach” in Figure 1) is feature extraction using a fixed “bank” of Gabor filters, concatenation of responses to a feature vector  $\mathbf{f}$ , post-processing (e.g.  $\hat{\mathbf{f}} \leftarrow \text{abs}(\mathbf{f})$ ) and input features to a classifier.

We have chosen the statistical approach and model each part  $P_i$  as the probability density function (pdf)  $p(\mathbf{f}|P_i)$ . For efficiency we adopt parametric density estimation and in particular the Gaussian mixture model

$$p(\mathbf{x}; \theta) = \sum_{c=1}^C \alpha_c \mathcal{N}(\mathbf{x}; \mu_c, \Sigma_c),$$

where  $\alpha_c$  is the weight of the  $c$ th component and  $\sum_{c=1}^C \alpha_c = 1$ . It is noteworthy that we do not perform post-processing but utilise directly the complex-valued Gabor responses ( $\mathbf{f} \in \mathbb{C}^D$ ) and complex-valued version of the  $D$ -dimensional Gaussian [36]

$$\mathcal{N}^{\mathbb{C}}(\mathbf{x}; \mu, \Sigma) = \frac{1}{\pi^D |\Sigma|} \exp [-(\mathbf{x} - \mu)^* \Sigma^{-1} (\mathbf{x} - \mu)] .$$

The pdf parameters to be estimated are

$$\theta = \{\alpha_1, \mu_1, \Sigma_1, \dots, \alpha_C, \mu_C, \Sigma_C\} .$$

which has total of

$$C(D^2 + 3D) + C - 1$$

free parameters. In our previous works, we compared several different algorithms and found that the maximum-likelihood estimation by the standard expectation maximisation (EM) algorithm [37] is the best if the number of components  $C$  is known. That, however, is not the case in practise and thus unsupervised GMM methods need to be employed. We tested the two popular methods, the method by Figueiredo and Jain (FJ) [38] and greedy EM (GEM) by Verbeek et al. [39], and in our experiments the FJ algorithm generally performed better and was more stable with limited data.

**Randomisation procedure** – For some of the classes in publicly available benchmark datasets (Caltech-101, Caltech-256, Pascal VOC) the number of training images is small, for example, only tens of images. In that case, the maximum likelihood estimation fails or overfits. The overfitting occurs with any classifier and the standard procedure is to limit the size of the Gabor bank (number of frequencies/orientations). In the existing literature, we did not find any method with a flexible number of Gabor features, but always the number of orientations and frequencies [40]–[42] was constant. A fixed set, on the other hand, is always a compromise over all classes – the *bias-variance tradeoff*.

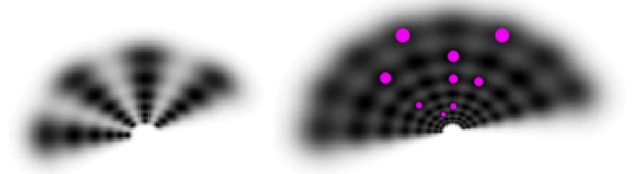


Figure 4. Left: a small fixed size Gabor feature bank used, e.g., in [19]; Right: a large set with the optimal combination of 9 by the proposed method.

We want to use a large number of Gabor features and at the same time avoid overfitting in the GMM estimation. With a small number of examples and a large number of features the problem becomes low bias and high variance. The standard tools to circumvent this issue are bagging, boosting and randomisation. Inspired by the success of similar extension to decision trees, random forests [43], we decided to apply bagging and randomisation to Gaussian mixture models: *randomised Gaussian mixture models*. Our randomised learning procedure is given in the pseudo code in Algorithm 2.

---

**Algorithm 2** Randomised GMM pdf of Gabor features.

---

- 1: Apply 4-fold randomisation of the training data
  - 2: Apply a big bank of Gabor filters to all annotated parts of the category  $C$
  - 3: **for**  $T$  iterations **do**
  - 4: Randomly select  $K$  Gabor filters and use them as features
  - 5: Estimate GMM with the selected features
  - 6: Evaluate and store landmark detection results over the 4-fold validation set
  - 7: **end for**
  - 8: Prune the same or too similar filter combinations
  - 9: Choose the  $B$  best filter sets and form their combination
- 

The big bank of Gabor filters allows full exploitation of the Gabor feature space. Nevertheless, the big pool of Gabor filters makes detections more complicated, as filters with neighbouring frequencies and orientations will give similar responses, lowering accuracy of detections. We allow each landmark of each object to have its unique frequency and orientation representation (Figure 4). Thus, to solve the problem of neighbouring Gabor filters, avoid deficiency in training data and provide the most discriminative landmark representation, we suggest to select  $K$  Gabor filters from the bank randomly. The procedure is repeated  $T$  times for each landmark of each object. Based on the 4-fold landmark detection evaluation of each trial on the training data, the combination of the  $B$  best sub-banks are selected to represent each landmark. The chosen Gabor filters are then applied

to each pixel of the image and transformed into likelihood maps with the corresponding GMM. In our implementation, FJ GMM algorithm is successfully trained for all of the categories with  $K = 9$  Gabor filter responses.

#### IV. PROBABILISTIC DETECTION OF OBJECT PARTS

Since the representation of each part is now encoded into  $B$  random Gaussian mixture models in the aligned object space, we need an efficient and effective procedures to combine the likelihoods of each GMM and non-maximum suppression to select the best candidates over the whole image. The pseudo code of our detection procedure is in Algorithm 3

---

**Algorithm 3** Detection using rand-GMM & Gabor features.

---

- 1: Apply  $B$  sets of  $K$  Gabor filters
  - 2: Compute  $B$  likelihood maps using the estimated GMMs
  - 3: Threshold each likelihood map to retain the proportion of  $P_1$  highest likelihoods
  - 4: Compute the product likelihood of the  $B$  thresholded maps
  - 5: Apply recursive global maximum search with suppression
- 

As each of the  $B$  ( $B = 5$  in our experiments)  $K$ -dimensional random feature GMMs miss some part of the landmark's description, the result classifier tend to overfit producing a lot of false positives (Figure 5 2nd row). Therefore, to suppress undesired false positives, we utilise the most strict combination rule of classifiers - the product rule [44]. Another

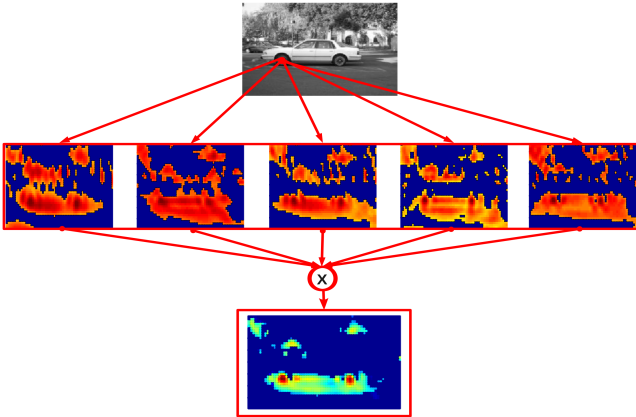


Figure 5. 1st row: Image of a car side with annotated part: front wheel. 2nd row: 5 thresholded likelihood maps, corresponding to the part. 3rd row: Final likelihood map from which candidates are sampled. Colours encode the likelihood values.

procedure used for suppressing false positives is thresholding ( $P_1 = 40\%$  in our experiments), applied before the product is calculated. Finally, the pixels with the highest product likelihood values correspond to the best candidates of each landmark. Hence, we should find peaks of the separate modes (local maximas). For that we use a simple but efficient iterative global maximum search with consecutive suppression. The suppression procedure is based on the discrimination ability of Gabor filters related to the lowest frequency of the bank. Our simple method returns candidates until the whole spatial space is covered. Consequently, for the very discriminative landmarks, we have just a few distinctive peaks leading to a few candidates, and for less discriminative landmarks, we may have a large amount of lower peaks and more candidates. As

a result, our method produces different amount of candidates for the landmarks based on their certainty, trying not to skip any correct location even if it leads to producing several outliers. In this sense, our approach deviates from all other works significantly, as the number of landmarks is not fixed. It is also noteworthy, that for some classes non-discriminative landmarks still produce the best detection.

#### V. EXPERIMENTAL RESULTS

**Data and Parameter Settings** – For experiments, we selected natural and divergent categories from the Caltech-101 dataset. Images of each category have been randomly assigned into approximately equal sized training and testing parts. Our method tolerates small amount of training data, thus in our experiments we have used categories with the number of training images from 28 to 406. Categories selected from Caltech 101 are: airplanes (406 train/394 test images), car side (58/65), dollar bill (28/24), faces easy (206/229), motorbikes (377/412), revolver (41/41), stop sign (30/34), watch (118/121), yin yang (30/30), menorah (43/44), grand piano (49/50) and dragonfly (34/34). Our method is also tested on BioID face database containing 1521 images with human faces (507/1024), recorded under natural conditions, i.e. varying illumination and complex background. Faces in the images vary in size, have different facial expressions, facial hair and glasses, belong to different gender and racial groups. We use FGnet Markup Scheme of the BioID Face Database, which have 20 landmarks useful for facial analysis and recognition [45]–[47].

**Performance Evaluation** – To the authors' best knowledge no performance evaluation method for local part detection has been proposed. Our method derives from the  $d_{eye}$ -measure which is used for measuring face localisation accuracy [48], [49]. Our metric computes distances from predicted coordinates of the object parts to the groundtruth coordinates that are normalised by the diagonal of the corresponding bounding box. For the object detection purposes, the normalised distance  $\leq 0.05$  is considered as excellent,  $\leq 0.10$  good and  $> 0.25$  as a failure (thresholds are adopted from face detection [48], [49]). Using these fixed thresholds, we can suggest a number of the best landmark candidates and test if any of them falls within the thresholds. The sooner the correct landmarks are fetched, the better. As a result, we can plot either a cumulative detection curve, or detection bars. The curve denotes how fast the correct landmarks are found by increasing the number of fetched landmarks. While the graph reports on landmark level, detection bars report on image level, i.e. for how many images at least some number of landmarks are detected. These performance graphics are illustrated in Figure 6.

**Results** – The detection bars for all Caltech-101 classes are shown in Figure 6 and some typical detection examples in Figure 7. From the results we see that the method provides almost perfect results for the faces class. The second best category is motorbikes and the third airplanes. These three are the classes with the largest amount of training images. The difference between faces (for 0.10 almost all landmarks were detected in all images), motorbikes (97%) and airplanes (95%), can be explained by the fact that the faces class is much more easier as compared to the motorbikes and airplanes which contain many sub-classes. Results for the best classes are followed by yin yang and stop sign, classes with the most simple structure. The worst performing class is dragonfly,



which have the biggest appearance variation with only few training examples.

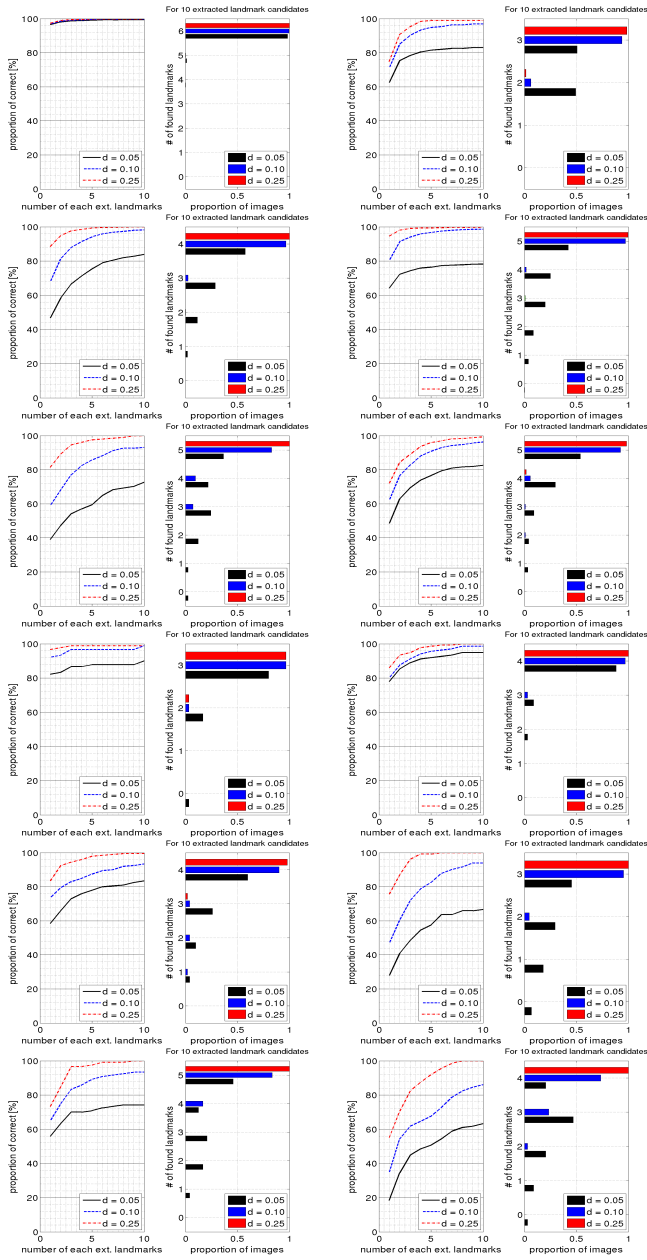


Figure 6. Catech-101 landmark detection (left: cumulative detection graph, right: detection bars for 10 best candidates). From top down, from left to right: faces, car side, airplanes, motorbikes, revolver, watch, yin yang, stop sign, grand piano, menorah, dollar bill, dragonfly.

BioID is a face database used for recognition of persons and facial expressions, thus requiring very precise landmark detection. In the case of BioID  $d_{eye}$ , distance between two eye centres, was used for normalising distances (instead of bounding box diagonal as in Catech-101). We also added an extra error threshold (0.15) to capture details of the detection curve in Figure 8. Even though only 23% of images have all 20 landmarks detected with precision 0.05, already 77% have detections with precision 0.15 which is already visually plausible.

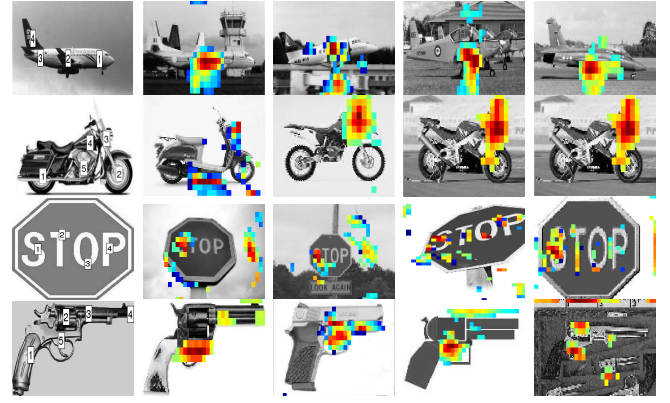


Figure 7. Catech-101 landmark detections. Left-most image shows all landmarks and on the right are examples for 10% best detection likelihoods of airplanes (lm #2), Motorbikes (lm #3), stop\_sign (lm #1) and revolver (lm #5).

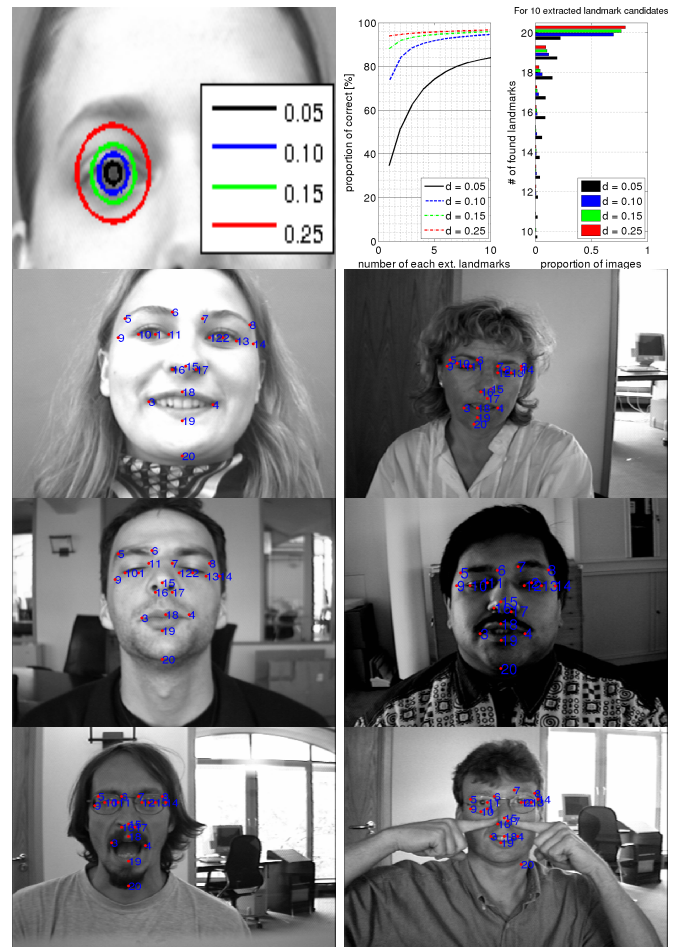


Figure 8. Top left: Detection precision regions. Top right: BioID faces landmark detection results. Bottom: Example detections in different scale and illumination conditions.

## VI. CONCLUSION

The bio-inspired Gabor features have been particularly successful in many computer vision problems. In this work, we adopted Gabor features and unsupervised Gaussian mixture models to define a probabilistic detector of local object parts.

Our work deviates rather strongly from other similar works since we do not limit the size of Gabor feature space, we learn the detector from positive examples only and provide both the full likelihood map and the best candidate points. To overcome the problems of curse of dimensionality due to a large number of Gabor filters and overfitting due to a small number of examples, we proposed effective and efficient learning and detection procedures utilising the novel concept of random Gaussian mixture models.

## REFERENCES

- [1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part based models, *IEEE Transactions on PAMI* 32 (9) (2010) 1627–1645.
- [2] K. Duan, D. Parikh, D. Crandall, K. Grauman, Discovering localized attributes for fine-grained recognition, in: *CVPR*, 2012.
- [3] H. Pirsiavash, D. Ramanan, Steerable part models, in: *CVPR*, 2012.
- [4] E. Riabchenko, J.-K. Kamarainen, K. Chen, Density-aware part-based object detection with positive examples, in: *ICPR*, 2014.
- [5] M. Sun, S. Savarese, Articulated part-based model for joint object detection and pose estimation, in: *ICCV*, 2011.
- [6] M. Andriluka, S. Roth, B. Schiele, Pictorial structures revisited: People detection and articulated pose estimation, in: *CVPR*, 2009.
- [7] Y. Yang, D. Ramanan, Articulated pose estimation with flexible mixtures-of-parts, in: *CVPR*, 2011.
- [8] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Transactions on PAMI* 29 (6) (2007) 915–928.
- [9] K. Chen, S. Gong, T. Xiang, C. C. Loy, Cumulative attribute space for age and crowd density estimation, in: *CVPR*, 2013.
- [10] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, Object retrieval with large vocabularies and fast spatial matching, in: *CVPR*, 2007.
- [11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. V. Gool, A comparison of affine region detectors, *Int J. of Computer Vision* 65 (1/2) (2005) 43–72.
- [12] M. Agrawal, K. Konolige, M. Blas, Censur: Center surround extremas for realtime feature detection and matching, in: *ECCV*, 2008.
- [13] T. Tuytelaars, C. Schmid, Vector quantizing feature space with a regular lattice, in: *ICCV*, 2007.
- [14] E. Tola, V. Lepetit, P. Fua, A fast local descriptor for dense matching, in: *CVPR*, 2008.
- [15] P. Felzenszwalb, D. Huttenlocher, Pictorial structures for object recognition, *Int J of Computer Vision* 61 (1) (2005) 55–79.
- [16] M. Bertholdt, J. Kappes, S. Schmidt, C. Schnör, A study of parts-based object class detection using complete graphs, *Int J Comput Vis* 87 (2010) 93–117.
- [17] E. Simoncelli, W. Freeman, E. Adelson, D. Heeger, Shiftable multiscale transforms, *IEEE Transactions on Information Theory* 38 (2) (1992) 587–607.
- [18] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Transactions on PAMI* 24 (7) (2002) 971–987.
- [19] J. Ilonen, J.-K. Kamarainen, P. Paalanen, M. Hamouz, J. Kittler, H. Kälviäinen, Image feature localization by multiple hypothesis testing of Gabor features, *IEEE Transactions on Image Processing* 17 (3) (2008) 311–325.
- [20] J. Ilonen, Supervised local image feature detection, Ph.D. thesis, Lappeenranta University of Technology (2007).
- [21] J. Daugman, High confidence visual recognition of persons by a test of statistical independence, *IEEE Transactions on PAMI* 15 (11) (1993) 1148–1161.
- [22] K. Messer, et al., Face authentication test on the BANCA database, in: *ICPR*, 2004.
- [23] A. Jain, Y. Chen, M. Demirkus, Pores and ridges: Fingerprint matching using level 3 features, *IEEE Transactions on PAMI* 29 (1) (2007) 15–27.
- [24] A. C. Bovik, M. Clark, W. S. Geisler, Multichannel texture analysis using localized spatial filters, *IEEE Transactions on PAMI* 12 (1) (1990) 55–73.
- [25] B. Manjunath, W. Ma, Texture features for browsing and retrieval of image data, *IEEE Transactions on PAMI* 18 (8) (1996) 837–842.
- [26] J. Han, K.-K. Ma, Rotation-invariant and scale-invariant gabor features for texture image retrieval, *Image and Vision Computing* 25 (9) (2007) 1474–1481.
- [27] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, *IEEE Transactions on Computers* 42 (1993) 300–311.
- [28] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, *IEEE Transactions on PAMI* 19 (7) (1997) 775–779.
- [29] V. Kyrki, J.-K. Kamarainen, H. Kälviäinen, Simple Gabor feature space for invariant object recognition, *Pattern Recognition Letters* 25 (3) (2003) 311–318.
- [30] P. Paalanen, J.-K. Kamarainen, J. Ilonen, H. Kälviäinen, Feature representation and discrimination based on Gaussian mixture model probability densities - practices and algorithms, *Pattern Recognition* 39 (7) (2006) 1346–1358.
- [31] J.-K. Kamarainen, V. Kyrki, H. Kälviäinen, Invariance properties of Gabor filter based features - overview and applications, *IEEE Transactions on Image Processing* 15 (5) (2006) 1088–1099.
- [32] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision* 60 (2) (2004) 91–110.
- [33] M.C. Burl, M. Weber, P. Perona, A probabilistic approach to object recognition using local photometry and global geometry, in: *ECCV*, 1998.
- [34] T. Coates, C. Taylor, D. Cooper, J. Graham, Active shape models – their training and application, *Computer Vision and Image Understanding* 61 (1) (1995) 38–59.
- [35] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, *IEEE Transactions on PAMI* 13 (4) (1991) 376–380.
- [36] N. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction), *The Annals of Mathematical Statistics* 34 (1).
- [37] J. Bilmes, A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models (1997).
- [38] M. Figueiredo, A. Jain, Unsupervised learning of finite mixture models, *IEEE Transactions on PAMI* 24 (3) (2002) 381–396.
- [39] J. J. Verbeek, N. Vlassis, B. Kröse, Efficient greedy learning of Gaussian mixture models, *Neural Computation* 15 (2) (2003) 469–485.
- [40] D. Vukadinovic, M. Pantic, Fully automatic facial feature point detection using Gabor feature based boosted classifiers, in: *ICSMC*, 2005.
- [41] G. Guo, G. Mu, Y. Fu, T. S. Huang, Human age estimation using bio-inspired features., in: *CVPR*, 2009.
- [42] J. Yang, Y. Shi, J. Yang, Finger-vein recognition based on a bank of Gabor filters., in: *ACCV*, 2009.
- [43] L. Breiman, Random forests, *Machine Learning* 45 (1) (2001) 5–32.
- [44] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas, On combining classifiers, *IEEE Transactions on PAMI* 20 (3) (1998) 226–239.
- [45] V. Rapp, T. Senechal, K. Bailly, L. Prevost, Multiple kernel learning svm and statistical validation for facial landmark detection, in: *Automatic Face Gesture Recognition and Workshops (FG)*, 2011.
- [46] H. Dibeklioglu, A. Salah, T. Gevers, A statistical method for 2-d facial landmarking, *IEEE Transactions on Image Processing* 21 (2) (2012) 844–858.
- [47] H. Yang, I. Patras, Face parts localization using structured-output regression forests, in: *ACCV*, 2013.
- [48] M. Hamouz, J. Kittler, J.-K. Kamarainen, P. Paalanen, H. Kalviainen, J. Matas, Feature-based affine-invariant localization of faces, *IEEE Transactions on PAMI* 27 (9) (2005) 1490–1495.
- [49] Y. Rodriguez, F. Cardinaux, S. Bengio, J. Mariéthoz, Measuring the performance of face localization systems, *Image and Vision Computing* 24 (2006) 882–893.