

# Gaussian refinements on Gabor filter based patch descriptor

László Lefkovits<sup>a</sup>, Szidónia Lefkovits<sup>b</sup>

<sup>a</sup>Sapientia - Hungarian University of Transylvania  
lefkolaci@ms.sapientia.ro

<sup>b</sup>”Petru Maior” University Tîrgu-Mureş  
szidonia.lefkovits@science.upm.ro

## Abstract

Recent research in object detection tends to put an accent not only on global object methods, but concentrates mostly on object parts and the relationship between them. One of the most widespread part-based object model was proposed by Felzenszwalb et al. [1]. Such systems can be divided into three main parts: the detection of interest points, the development of adequate local descriptors and the object model.

This article deals with the most important phase: the elaboration of local descriptors. We have therefore created a patch descriptor based on two-dimensional Gabor filters. The idea of the descriptor thus developed is to select only a few from the multitude of definable Gabor filters: those most adequate for a given object part. In our previous works, we designed a response-map that played the role of the local descriptor, based on the above-mentioned filters and the GentleBoost learning algorithm [2] or the SVM classification method [3].

In this paper we propose an improvement to the filter selection process which considers not only the magnitude of the complex Gabor filter responses, but the real and imaginary parts and their statistical distribution as well.

For this purpose, we have created an RBF Neural Network able to learn the statistical distribution of Gabor filter responses. This network improved the selection procedure of the most suitable filters for a given image patch. The idea of using the RBF NN has been suggested by several authors [4, 5, 6] whose systems are based on the Gaussian distribution of Gabor filters.

In conclusion, we have compared the above-mentioned three methods – GentleBoost, SVM and RBFNN – and have deduced that the combination of Gaussians characterizes the patch better than just the magnitude value of the complex responses.

*Keywords:* Gabor filter, local descriptor, Gaussian approximation, RBF Neural Networks

*MSC:* 68T45

## 1. Introduction

The goal of object detection is to automatically localize the object or object categories in 2D images. Humans solve this task very easily, without any effort. Despite extensive research, no application has been developed for all purposes of detection; there are some applications suited to specific tasks which can be used in a limited fashion. The most important problem of object detection resides in the fact that an object may appear on different occasions, in several aspects, dimensions, positions or viewpoints. A notable factor of this process is the diversity of backgrounds, which also makes detection more difficult.

There are two major ways to approach the problem of object detection: the global and local aspect-based methods. Global aspect-based systems [7, 8] present the object as whole, indivisible entity. In general, the detection process uses the image pyramid technique, i.e., the images are resized and scanned over the image using a standard dimensional window. Recent research into artificial vision tends to gloss over global object detection and concentrates mainly on part-based detection. Generally, in a system based on local aspect [1, 9], there are three distinguishable parts: interest points – these represent a set of points where local object parts are searched for; local descriptor – represents a formal description of image patches corresponding to different object parts; object model – it is the mathematical formalism based on which the detected object parts compose the target object.

This paper attempts to create a novel local descriptor based on two dimensional Gabor filters and to classify the object of interest with several classification methods. We make a comparison to our previous approaches where the GentleBoost algorithm [2] and Support Vector Machines [3] were used for classification purposes. Here, every Gabor filter response is approximated by a RBF Neural Network classifier. The output of the network is the value which is classified by a majority vote.

This paper is organized as follows: the first section is a short overview of local descriptors in the literature; the second consists of a theoretical review of the Gabor filters, the SVM classifier and the RBF Neural Networks; finally, we present the results our experiments, comparing them to other systems, draw certain conclusions, and refer to any work that may be done in the foreseeable future.

## 2. Theoretical background

In Computer Vision, there are a wide range of proposed local features for general detection purposes, namely: Scale Invariant Feature Transform (SIFT) [10], PCA-SIFT [11], Histogram of Oriented Gradients (HOG) [12], Gradient Orientation Location Histogram (GLOH) [13], Locally Binary Patterns (LBP) [13], Bag of Words [14].

It is known that none of these features can be applied on its own to obtain optimal detection performance, but their combination, used with the most adequate classification algorithm, can lead to the desired detection rates.

In our experiments we have chosen a local descriptor based on the 2D Gabor filters. It has been proven that Gabor wavelets work similarly to the mammalian visual cortical system [18]. Thus, cortical receptive fields which process visual information can be modeled fairly well by Gabor wavelets [15].

Similar systems which deal with the Gabor filter for object detection have been proposed by [16], who have used a set of Gabor filters to form a jet with different frequencies and orientations. In several cases, a similarity measure [6] is set up to detect a given discriminative feature. Other authors [5] define the similarity measure as a probabilistic likeliness.

In many applications, it is not sufficient to determine a similarity measure between Gabor filters; it is also desirable to compare the amplitudes and phases of complex responses based on their distribution. Generally, the distribution of responses is considered to be the normal distribution [4, 5, 6].

Using the Gaussian approximation of the Gabor filter responses is the main idea of this article. The RBF Neural Network, with a given number of seeds, approximates a filter response quite well, and the classifier created from a set of RBF Neural Network gives the desired detection results.

The main advantage of this approximation is that it takes into account the complex form of the Gabor filter response and makes a decision in a 2D space. This method considers the phase information as well, and not just the frequency or the magnitude value of the Gabor wavelet, as do the majority of approaches based on the idea of Gabor filters [17].

## 2.1. Gabor filters

In our system, the first step is to extract more Gabor filter responses from the facial region of interest. Gabor wavelets have a wide area of use, especially in bioinformatics. It has been physiologically proven that the working principle of the mammalian visual cortex can be modeled extremely well by these wavelets [15].

The Gabor wavelet is a sinusoidal wave modulated by a Gaussian. The mathematical definition of the Gabor wavelet is [18]

$$g(x, y) = \frac{1}{k} e^{-\pi \left[ \frac{(x - x_0)_r^2}{\alpha^2} + \frac{(y - y_0)_r^2}{\beta^2} \right]} e^{i[\xi_0(x - x_0) + \nu_0(y - y_0) + P]}, \quad (2.1)$$

where index  $r$  means the rotation of the envelope surface with  $\theta_0$  in trigonometric direction. The response of the filter in a given point  $(x_0, y_0)$  is the Gabor coefficient, which means the convolution of the image  $I(x, y)$  with the Gabor filter  $g(x, y)$

$$C(x_0, y_0) = \iint I(x, y) g(x - x_0, y - y_0) dx dy. \quad (2.2)$$

The Gabor wavelet is a plane wave modulated by a Gaussian envelope. This function is defined in a 9D parameter space, where  $\frac{1}{k}$  is the amplitude of the Gaussian envelope;  $\theta_0$  the rotation angle of the Gaussian and the plane wave;  $(\alpha, \beta)$  the standard deviation of the Gaussian in 2D;  $(x_0, y_0)$  the center of the Gaussian;

$(\xi_0, \nu_0)$  the spatial frequency of the sinusoidal wave; and  $P$  the phase of the wave.

These 9 independent parameters can be reduced to only 4, considering some useful relations between the frequencies, the orientation and the attenuation of the Gaussians. The relationships between the wavelength,  $\lambda$ , bandwidth,  $bw$  and the Gaussian attenuation,  $\alpha$ ; respectively between the angle of the wave,  $\theta_0$ , the bandwidth,  $bw$ , and the Gaussian attenuation in the perpendicular direction,  $\beta$ , lead to only 4 parameters.

The 4 degrees of freedom which remain are the wavelength, the angle of the wave, the bandwidth and the aspect ratio  $(\lambda, \theta, bw, S)$ . The definition domain of them is infinite, but for a given image, with limited number of pixels, these parameters have an upper limit. For more details see [2].

Taking into account the 4D space obtained, which is also limited at the same time, we define a considerable number of Gabor filters. Based on these, the system computes the filter response centered on the image patch. In order to choose only the most representative filters and the weight of each one in the final decision, a learning algorithm must be applied. In our last paper, we proposed the GentleBoost algorithm [2] or the Support Vector Machines [3] for this purpose.

## 2.2. Support Vector Machines Classification Method

The Support Vector Machines [19] are supervised learning machines for binary classification problems. They may be used if the input space is linearly separable or if the input space can be transformed into another, higher dimensional space, where it can be linearly classified. The main idea of these machines is to obtain the best separation hyperplane which maximizes the distance of the nearest entities from each class of training set.

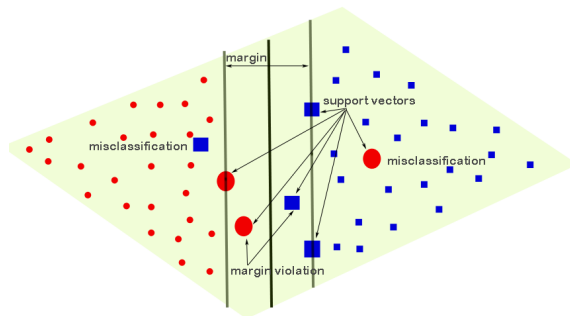


Figure 1: Linear separation by SVM

This separation can be formulated as an optimization problem. The equation of the hyperplane is

$$w^T \Phi(x) + b = 0, \text{ where } w \in \mathbb{R}^n, \text{ and } b \in \mathbb{R}. \quad (2.3)$$

In equation 2.3,  $w$  is the normal vector of the hyperplane,  $\Phi$  is the transformation of the inputs in a higher-dimensional space,  $b$  is the bias.

Given  $k$  inputs in the training data set,  $\mathcal{S} = \{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{\pm 1\}, i = \overline{1, k}\}$ . For one class the classification response can be greater or equal  $+1$  and for the other class less or equal  $-1$ . This separation can be formulated as

$$y(x_i) \cdot w^T \Phi(x_i) + b \geq 1, i = \overline{1, k}, \text{ where } y(x_i) = \begin{cases} +1 & \text{if } x_i \text{ is entity of class 1} \\ -1 & \text{if } x_i \text{ is entity of class 2.} \end{cases} \quad (2.4)$$

These restrictions will form the inequality constraints of the following optimization problem

$$\max \left( \frac{1}{\|w\|} \right), \text{ with the following constraints } y(x_i) \cdot w^T \Phi(x_i) + b \geq 1. \quad (2.5)$$

This optimization problem can be solved by the Lagrangian multipliers method and Karush–Kuhn–Tucker conditions.

In our previous experiments [2, 3], the SVM classifier had a double role: first, it had to extract the most appropriate  $n$  from the given number of filter responses in order to discriminate the target facial feature from other facial features; second, it had to obtain the optimal separating plane of the two classes.

The 2D responses of every Gabor filter are classified by their own SVM, determining the optimal separating curve between positive and negative responses. In that case, the goal is not to obtain the final decision of the responses with one classifier alone, but to compare the Gabor responses based on the performance of the classifiers obtained.

Thus, the appropriateness of Gabor filters can be determined depending on the number of resulting support vectors. Out of the best  $n$ , we have obtained an  $n$  dimensional feature vector. These feature vectors are separated by the final classifier, which is also based on the SVM learning algorithm. In the classification phase, input space is linearly separated into a higher dimensional space, after transforming it with the Gaussian RBF kernel and using the kernel trick [19].

The disadvantage of the approach presented above is that it considers only one dimensional input space of the complex Gabor filter response. Only one value is taken into account: either magnitude, or phase, or real part, or imaginary part. It creates the feature vector which is used for classification from the best responses. The main idea of this article is finding a method which does not eliminate the 2D real and imaginary part of the filter response, and somehow describes the input space in two-dimensions.

### 2.3. Radial Basis Function Neural Network

The Radial Basis Neural Network is a special feed-forward neural network composed of an input layer, a single hidden layer and an output layer. The activation function of the hidden layer is not the sigmoidal function or the hyperbolic tangent, but the

RBF (Radial Basis Function). The value of this function is dependent on the distance to a given center  $x_0$ , and is generally expressed as  $\Phi(x, x_0) = \|x - x_0\|$ . The best known RBF function is the Gaussian function with mean value  $\mu$  and variance  $\sigma^2$  :

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}. \quad (2.6)$$

The linear model for the neural network output is

$$f_{out} = \sum w_i \Phi_i(x) + b, \quad (2.7)$$

where  $w$  is the weight of each RBF response in the output and  $b$  is the bias.

In our experiments, a separate RBF network is trained for each Gabor filter. The goal of the RBF NN is to approximate the input space as well as possible.

The 2D responses of the Gabor wavelet can be easily considered as inputs of an RBF Neural Network. The output of the RBF NN will be the classification of an image patch in the two classes (background or facial feature). The architecture of the network created can be seen in figure 2 generated by the Matlab Neural Network Toolbox.

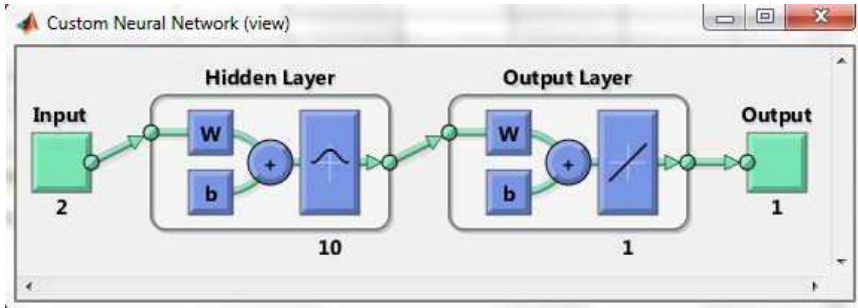


Figure 2: The architecture of the RBF NN for every feature

The main issue is choosing the optimal number of RBF in the hidden layer and the manner in which they are initialized. The adequate number of neurons was determined experimentally, and the selection of centers and spreads of the  $n$  Gaussians was suggested by [20], namely, the k-means clustering algorithm.

The algorithm assumes that the sum of the distances are minimized:

$$\operatorname{argmin} \sum_{i=1}^N \sum_{x_j \in Cl_i} \|x_j - \mu_i\|^2. \quad (2.8)$$

The aim of the k-means clustering algorithm is to separate the input points into  $n$  clusters and to place each data point in the cluster to the center of the one it is nearest to.

### 3. Results and experiments

In our experiments, we defined two-dimensional Gabor wavelets to compute the image features. The sets of features have been trained by the multiple RBF Neural Networks, and the classification of the image patches based on the response of the RBF approximation is a majority voting process. This method, compared to our previous efforts, deals with the 2D complex responses of the Gabor filter, not just the magnitude value, or separately with the real or imaginary part of the complex number. This two dimensional approximation considers the spacial distribution of the filter responses. Figure 3 shows the response of one filter for the training set; the red dots are the considered facial feature locations and the blue dots are the background responses of the same Gabor filter.

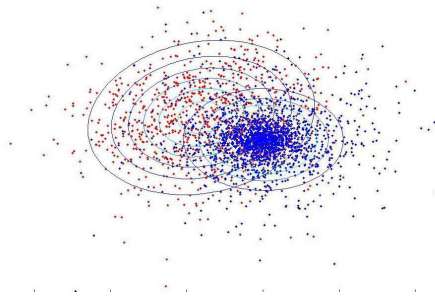


Figure 3: The responses of one filter for the training set

In our previous paper we had defined 3024 fine-tuned Gabor filters. These were built, as described before, in the 4 dimensional feature-space  $(\lambda, S, bw, \theta)$ . The most adequate filters had to be selected based on the training set. For the training and test set, images from the FERET [21] database were used.

The training set consists of 730 positive and 2000 negative examples, and the test set of 160 and 500 patches. The image patch used in the training phase is  $33 \times 33$  pixels centered on the eye and the negative images have been extracted randomly from the face, but not the eye.

The responses of each filter have to be approximated. Approximation was done by an RBF NN. Its architecture is the following: the input is the filter response, the output is a real value between  $[-1, 1]$ , and the hidden layer consists of 10 neurons. The adequate number of neurons was determined experimentally. For each Gaussian activation function in the network, we had to decide where to place the initial central point and how to calculate the initial spread of the surface. We have used the k-means clustering algorithms for this purpose. The cluster centers are also the centers of the chosen Gaussians, and their spread was computed by the average distance of every point in a given cluster to its center. Of course, the neural network adjusts not only the weights of each neuron, but also the initial

cluster centers and the initial spreads.

After the training process the centroids are illustrated for one RBFNN in figure 4(a) and the corresponding classification curve is shown in figure 4(b).

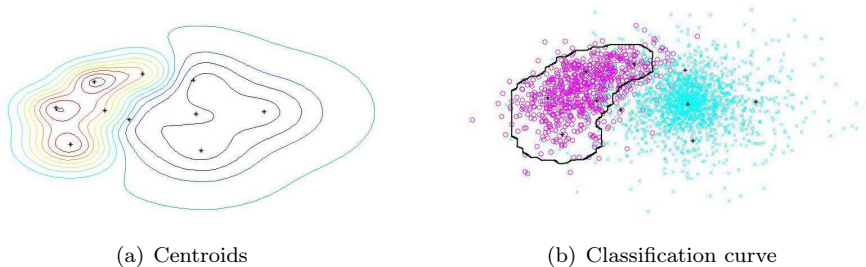


Figure 4: The approximation of the input space by RBFNN

The classification performance of this method, compared to our previous methods for the same train and test set, is shown below, table 1. The measurements present the detection error, the false positive error, and the false negative error. Three methods are compared the GentleBoost algorithm [2], the SVM classification method [3] and the RBF NN presented here.

	GentleBoost	SVM 32D	RBF NN
<b>Detection rate</b>	1.71%	3.22%	2.03%
<b>False Positive Error</b>	0.31%	1.98%	2.58%
<b>False Negative Error</b>	3.15%	4.5%	0.4%

Table 1: Classification performances of the three methods

We can see that the detection rate is slightly worse than that of GentleBoost: instead of 1.71% it is 2.03%. But the advantage is that false negatives have decreased considerably, from 3.05% to 0.4%. Compared to the SVM, all three values have been optimized. The classification performances are shown in figure 5, comparing all three methods. According to the experiments, the GentleBoost detects best, but the RBFNN eliminates false detections better than the GentleBoost.

The advantage of this method is that it takes the 2D input space into consideration, as well as the fact that the training and final decision is also made in 2D. So this improves the training of 1D inputs, even if the classification method is less strong than the Boosting algorithm. Once the approximation surface is obtained for each filter response, the final decision is made by a majority voting of the best surfaces.



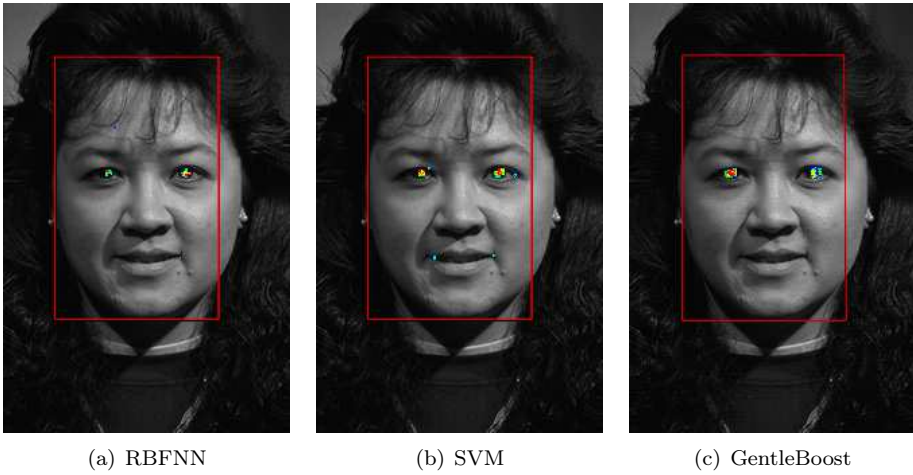


Figure 5: Detection example from FERET database

## 4. Conclusion and future work

This paper presents a facial feature detector based on Gabor filter responses and the RBF Neural Network classifier. It is placed alongside previously presented methods [2, 3] that use the same filters, but different classification algorithms. The classification performance of the method presented in this paper is better than that of previous results, mainly because it eliminates the high number of false detections of the GentleBoost. The goal of using this RBF network approximation is to create a classification surface in 2D from the complex responses of the Gabor filters. The other two methods in papers [2, 3] predict the final classification result based only on 1D input space. The 2D surface is trained from the responses of one Gabor filter; to make a final decision, more such surfaces have to be created based on several Gabor filters, not just one. This is one of the drawbacks of the approach presented. As to future work, we suggest combining the GentleBoost algorithm with the 2D approximation of filter responses. We also intend to apply our best algorithm similarly for more facial features in order to create a constellation of facial parts obtained from local descriptors.

## References

- [1] FELZENSZWALB, F.-P. , HUTTENLOCHER,P.-D. , Pictorial Structures for Object Recognition, *International Journal of Computer Vision* Vol. 61 No. 1(2005), 55–79.
- [2] LEFKOVITS, SZ., Novel Gabor Filter Based Patch Descriptor, *10th Jubilee International Symposium on Intelligent Systems and Informatics SISY, IEEE Conference*, Vol. 1 (2012).

- [3] LEFKOVITS, SZ., Enhanced Gabor Filter Based Facial Feature Detector, *The Proceedings of the European Integration - Between Tradition and Modernity, Computer Science Section*, Vol. 1, (2013).
- [4] ILONEN, J., KAMARAINEN, J.-K., PAALANEN, P., HAMOUZ, M., J. KITTLER AND KÄLVIÄINEN, H., Image feature localization by multiple hypothesis testing of Gabor features, *IEEE Transactions on Image Processing*, Vol. 7 No. 3, (2008), 311–325.
- [5] TOIVANEN, M. LAMPINEN, J., Incremental Bayesian learning of feature points from natural images, *Computer Vision and Pattern Recognition Workshop*, (2009), 39-46.
- [6] WISKOTT, L., FELLOUS, J.-M., KRÜGER, N., VON DER MALSBURG, CH., Face Recognition by Elastic Bunch Graph Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, (1997), 775-779.
- [7] VIOLA P., JONES M., Robust Real-Time Face Detection, *International Journal of Computer Vision*, Vol. 57, No. 2, (2004), 137 - 154
- [8] LIENHART R., KURANOV E., PISAREVSKY V., Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection, *In DAGM 25th Pattern Recognition Symposium*, (2004), 297–304
- [9] FERGUS R., PERONA P., ZISSERMAN A., Object class recognition by unsupervised scale-invariant learning, *In CVPR*, (2003), 264–271
- [10] LOWE, D. G., Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, Vol. 60, (2004), 91–110
- [11] KE Y., SUKTHANKAR R. PCA-SIFT: A more distinctive representation for local image descriptors, *In Computer Vision and Pattern Recognition CVPR*, (2004), 506–513
- [12] DALAL N., TRIGGS B. Histograms of oriented gradients for human detection, *In Computer Vision and Pattern Recognition CVPR*, (2005), 886–893
- [13] TUYTELAARS T., MIKOLAJCZYK K. Local Invariant Feature Detectors: A Survey, *Now Publishers Inc., Hanover, MA, USA*, (2008)
- [14] LEIBE B. Interleaved Object Categorization and Segmentation *PhD thesis*, 2004
- [15] SERRE T., WOLF L., BILESCHI S., RIESENHUBER M., POGGIO T. Robust object recognition with cortex-like mechanisms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, (2007), 411–426
- [16] TAMMINEN T., LAMPINEN J. Sequential Monte Carlo for bayesian matching of objects with occlusions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 6, (2007), 930–941
- [17] VUKADINOVIC D., PANTIC M. Fully automatic facial feature point detection using Gabor feature based boosted classifiers, *In IEEE International Conf. on Systems, Man and Cybernetics*, (2005), 1692–1698
- [18] MOVELLAN J. R. Tutorial on Gabor filters,
- [19] CORTES C. VAPNIK V. Support-Vector Networks, *Machine Learning*, (1995), 273–297
- [20] MCCORMICK CR. Radial Basis Function Network <http://chrisjmccormick.wordpress.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/>,
- [21] FERET DATABASE [http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html),