

IMPROVEMENTS ON GABOR DESCRIPTOR RETRIEVAL FOR PATCH DETECTION

Szidónia LEFKOVITS

*Department of Mathematics-Computer Science
“Petru Maior” University
Nicolae Iorga Street No. 1
540088 Tîrgu-Mureş, Romania
e-mail: szidonia.lefkovits@science.upm.ro*

Abstract. The localization of object parts in the component-based object detection is among the main tasks to solve. This paper presents several improvements of the proposed local image descriptor based on Gabor wavelets. Including these descriptors in the desired application is an ambitious challenge if we take into account the high number of parameters. Determining of parameters can be very hard because of their infinite definition range. Defining the filters is done in two stages: a theoretical consideration narrows the domain and the cardinality of parameters; this is followed by adequate experiments to select the most characteristic descriptor for a target image patch. The descriptor is created from a given number of 2D Gabor filters chosen by the GentleBoost learning algorithm. Comparing the proposed descriptor to those found in the state of the art, we can conclude that the selected filters are adaptable to any target object. In contrast to this, the majority of filter-based descriptors have fixed values for the parameters that do not allow to be ductile to the given object. Parameters fine-tuning allows the descriptor to be general, and discriminative at the same time. The effect of the following experiments has been analyzed during the investigation: elimination of redundancy between the weak classifiers, using the LoG interest points in the detection process. Finally, we propose an acceleration algorithm in order to determine the response map faster. By means of the descriptor, the response map is created, which accurately localizes the target object part and can easily be integrated in almost all detection systems.

Keywords: Local descriptor, 2D Gabor wavelets, fine-tuning parameters, response map, GentleBoost, part-based object model, acceleration algorithm, LoG interest points, mutual information

1 INTRODUCTION

In the area of object detection, the term feature is the image patch which differs from its neighboring pixels and captures a visual attention. In pattern recognition, features may be pixels, edges, corners, intensity regions or texture changes that describe only a part of an image; these are so-called local image features. The descriptor can also refer to the whole image using the information from all of its pixels; these are the global features. The part-based object model, which uses local features, has the advantage of handling occlusions or small deformations of the object.

The descriptor must not be very specific, but not too general either. Local image descriptors are numerical representations of object parts. They are usually computed in previously detected interest points. The aim of this is to shorten the computation time, but it also assumes that the set of interest points has not missed the target object part. Otherwise, if the interest point detection was left out, an exhaustive search would have to be used over the image. It would be ideal if detected features semantically corresponded to the object parts.

2 STATE OF THE ART

The most widespread local descriptors in the specialized literature are the following: intensity-based descriptors, scale-invariant descriptors, local binary patterns and filter-based descriptors [1]. The simplest descriptor uses the intensity values of pixels. In this case, the measure of similarity is the cross-correlation between two image patches. A disadvantage of this descriptor is a high dimensionality, which can be reduced, for example, with the PCA method. Another way to reduce the dimensionality of image patches is to use the so-called ‘bag of words’ [2].

The best-known histogram-based descriptor is HOG [3]. The histogram of gradients is computed in a dense grid, with a given step for each image. The descriptor is a histogram of 8 directions and it is also weighted by a Gaussian envelope in a pixel region. The descriptor contains 128 values: 16 histograms with 8 bins computed in a so-called pixel block. The SIFT (Scale Invariant Feature Transform) introduced by Lowe [4] is based on the HOG descriptor. This method has two parts: the interest point detector and the descriptor that is applied on the detected interest points for a single scale and orientation. These are determined by the properties of the interest points. Here, the necessity of dimensionality reduction appears as well, which is why the PCA-SIFT was introduced [5].

In the state of the art, several authors have used the Gabor wavelets as feature descriptors [10].

Wiscott et al. [6] or Tamminen and Lampinen [7] define a fixed jet form of a given number of Gabor filters without any connection to the descriptor and the target image patch. Vukadinovic and Pantic [8] define 48 filters that are optimized for surrounding interest points. They determine the descriptor for several well-

defined facial points. Shen and Bai [9] use the mutual information theory integrated in the AdaBoost algorithm as an optimized Gabor wavelet selection.

Our proposal is to use Gabor filter-based descriptors as local image descriptors. The selection of filters with GentleBoost is optimized in order to find the most appropriate filter for the target image patch.

The paper is organized as follows: the first section is a general presentation of the local descriptors; the second section describes the state of the art of local descriptors; in the third section, the Gabor filters are defined and the parameter space is reduced by using certain theoretical relations, while the fourth section experimentally determines the most appropriate descriptors for a given image patch. Finally, the fifth section presents the results of the experiments. The obtained descriptor is upgraded based on several improvements brought in this paper.

3 THEORETICAL STUDY

Filter-based descriptors extract certain components of the image. If the whole family of orthogonal filters is applied, a reconstructible decomposition of the images is obtained. This means that the image can be represented by its wavelet coefficients obtained from the decomposition. In this paper, the goal is not a decomposition or complete reconstruction, but to determine the most adequate set of filters for the analysed object part. The aim is to find the set of adequate filters that best represent the image or the target image patch. We chose the Gabor function as a filter descriptor. It has been neurophysiologically proven that the working principles of the visual receptive field of mammals may best be compared to Gabor filter responses. The 2D Gabor wavelets do not represent an orthonormal basis, but in some conditions, the image reconstruction is possible [11].

One possibility to characterize the neighborhood of a given pixel is to compute the filter responses of the image patch. The more different the applied filters are, the better accuracy of the descriptor is obtained.

The response of the filter is actually the wavelet coefficient, which may be obtained by convolving the image $I(x, y)$ with the Gabor filter $g(x, y)$ in a given point (x_0, y_0)

$$C(x_0, y_0) = \iint I(x_0, y_0)g(x_0 - x, y_0 - y) dx dy. \tag{1}$$

This coefficient describes the neighborhood of the (x_0, y_0) image point, according to the Gabor function:

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

where r means the rotation of the envelope surface with the θ_0 angle in trigonometric direction defined by

$$\begin{aligned} (x - x_0)_r &= (x - x_0) \cos \theta_0 + (y - y_0) \sin \theta_0, \\ (y - y_0)_r &= -(x - x_0) \sin \theta_0 + (y - y_0) \cos \theta_0. \end{aligned} \tag{3}$$

The 2D Gabor function is a plane wave modulated by a Gaussian envelope. Thus the 2D Gabor filter is determined in a high dimensional space depending on 9 different parameters, namely: $1/k$ the amplitude of the Gaussian envelope, (α, β) the standard deviations in both directions of the plane, the rotation angle θ_0 of the Gaussian, (x_0, y_0) the maximum point of the Gaussian, (ξ_0, ν_0) the spatial frequency of the sinusoidal wave, and it is the phase of the wave.

In order to reduce the high dimensionality of parameters, theoretical relations must be found between the parameters. The easiest way to limit the parameter domain is in transforming it into the frequency domain. The spatial frequencies (ξ_0, ν_0) are expressed in polar coordinates because these measures are much more intuitive. Thus F_0 is the frequency, respectively ω_0 is the direction of the sinusoidal wave

$$F_0 = \sqrt{\xi_0^2 + \nu_0^2}, \quad (4)$$

$$\omega_0 = \arctan \frac{\nu_0}{\xi_0}.$$

An interesting property of Gabor filters is the pure real form of their Fourier transform (if $x_0 = y_0 = 0$ and $P = 0$, it means that the envelope is centered on the origin of the coordinate system, respectively the wave starts in the same point)

$$G(\xi, \nu) = \exp(-\pi(\alpha^2(\xi - \xi_0)_r^2 + \beta^2(\nu - \nu_0)_r^2)). \quad (5)$$

From the half-magnitude response in the frequency domain, two interesting relations can be deduced [12]

$$\lambda = \sqrt{\frac{\pi}{\ln 2}} \alpha \frac{2^{bw} - 1}{2^{bw} + 1}, \quad (6)$$

$$\theta_0 = 2 \arctan \left(\frac{\lambda}{\beta} \cdot \sqrt{\frac{\ln 2}{\pi}} \right). \quad (7)$$

The first formula (6) is a relation between the central frequency $\lambda = \frac{1}{F_0}$ and the bandwidth bw , which indicates the number of useful wavelengths taken into consideration. The second formula (7) is the connection of the central frequency F_0 and the width of the frequency domain covered, indicating the minimal number of directions taken into account. The bandwidth is defined in terms of half-magnitude responses as the \log_2 difference of minimum and maximum frequencies. Usually, it is measured in octaves.

$$bw = \log_2 \frac{F_{max}}{F_{min}} \quad (8)$$

In order to limit the domain of parameters, we have accepted neurophysiological assumptions gained from the experiments [12]:

- the orientation θ_0 of the Gaussian envelope (the minor axis α orientation) and the wave orientation are almost equal $\omega_0 = \theta_0$;
- the biological orientation sensitivity is $10^\circ - 40^\circ$; as a consequence, we consider $L = 12$ discrete orientations with a step of 15° . The orientations are $\theta_l = l \cdot \theta_0$ where $\theta_0 = \frac{\pi}{L}$ and $l \in \{0, 1, \dots, L - 1\}$;
- the aspect ratio of the filter $S = \frac{\beta}{\alpha}$ is between $[1, 2]$, and thus, the envelope is mostly elliptical; notice that the direction of the plane wave is equal to the minor axis of the Gaussian;
- the biological half-magnitude frequency bandwidth is between 1 and 1.5, which is why we limit the bandwidth to a distinct range $[1, 2]$.

The central frequency of the family of filters is usually between $F_0 \in [F_{min}, 0.5]$. The $F_{max} < 0.5 \text{ pixel/cycle}$ is a consequence of the Nyquist sampling theorem. The minimum frequency has to be deduced from the dimension of the analyzed image patch. Let us take the patch dimension to be denoted by $2R$, this determines the used part of the infinity Gabor filter. The neglected part represents an intrinsic evaluation error ε .

The error is the ratio between the volume of the part used and the whole integral

$$\varepsilon = 1 - \frac{\int_{-R}^R \int_{-R}^R \exp\left(-2\pi \frac{x^2+y^2}{\beta^2}\right) dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-2\pi \frac{x^2+y^2}{\beta^2}\right) dx dy}. \tag{9}$$

Based on the definition of the error, we have deduced a useful relation between the maximum wavelength and the finite dimension of the filter

$$\frac{\lambda}{R} < \frac{\pi}{\sqrt{-\ln 2 \cdot \ln \varepsilon}} \cdot \frac{1}{S} \cdot \frac{2^{bw} - 1}{2^{bw} + 1}. \tag{10}$$

From relation (10) we can deduce the maximum of the usable wavelength λ_{max} , hence the minimum frequency F_{min} . Commonly the frequencies are chosen in geometric progression taking 2–3 frequencies in an octave. It is well known that every patch has a fundamental frequency in each direction; accordingly, we consider that several frequencies form the interval $[F_{min}, F_{max}]$. Taken the above into consideration, the 9-dimensional space could be reduced to 4 dimensions (λ, θ, bw, S) . With a different set of parameters and by varying the values of λ and θ , we can cover the frequency space in several ways (Figure 1).

We have proposed a descriptor that describes the patch computing the Gabor filter on its central point. For a given patch, we may define a huge number of Gabor filters with different parameters. In order to use only a few of them, the most representative have to be chosen. The GentleBoost algorithm is applied with this purpose in mind. The cyclic process selects the best weak classifier which minimizes the least weighted square error of the classification.

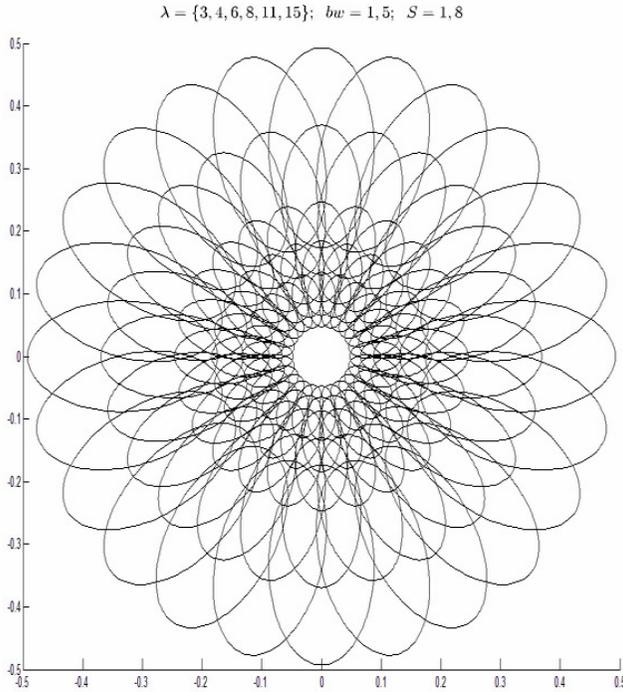


Figure 1. Covering the frequency domain with partial lapped filters

The weak classifier is defined as the difference of object probability and non-object probability. Its value represents the confidence value of the decision. A weak classifier can be implemented with a simple regression stump with 3 parameters (a, b, θ) .

$$\hat{y} = a \cdot (x > \theta) + b = \begin{cases} a + b, & \text{if } x > \theta \\ b, & \text{otherwise} \end{cases} \tag{11}$$

After the selection of the weak classifier, we finally get the strong classifier

$$\text{sign}H(x) = \text{sign} \left(\sum_{t=1}^T h_t(x) \right). \tag{12}$$

A fine-tuning of the parameters is done in the training process, where the Gentle-Boost algorithm selects the filters based on positive and negative examples. The performance of the selected filters is measured on a set of validation images. The descriptor contains as many selected filters as needed to achieve the expected performance.

3.1 LoG Laplacian of Gaussian Detector

This detector has been invented by Lindeberg [16] and is used for blob-like feature detection. The operator is obtained by the second order derivative filter, the Laplacian smoothed by a Gaussian. The analytical form of the 2D filter is

$$\nabla^2 G = \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^4} e^{-(x^2+y^2)/2\sigma^2}. \tag{13}$$

It can be observed that this is a negative kernel. It has a strong negative peak in the center and a positive ring surrounding it (Figure 2).

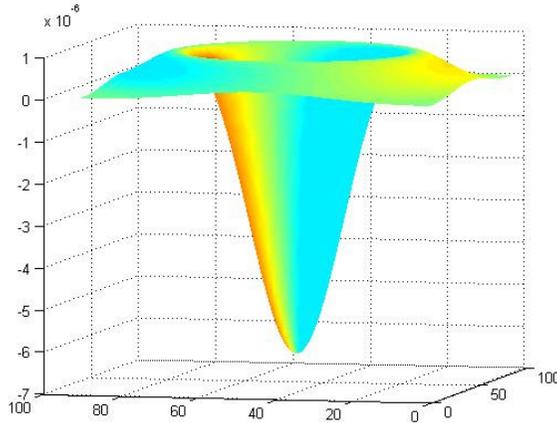


Figure 2. Bidimensional LoG filter

Modifying the scale variable σ , it searches for the maximum similarity of the image region on which it is applied. It is a scale-invariant blob detector because it detects the scale of the region by computing the scale space extrema of a certain point. The LoG responses are therefore given in a descending order.

3.2 Mutual Information

Mutual information measures the mutual dependence of two random variables, X and Y . It can be computed from the marginal probabilities $p(x)$, $p(y)$ of the two variables and the joint probability $p(x, y)$.

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \tag{14}$$

The mutual information can be expressed in terms of the marginal $H(X)$, $H(Y)$ and conditional $H(X, Y)$ and/or joint entropies $H(X|Y)$, $H(Y|X)$ as well:

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y). \end{aligned}$$

The probabilities in the case of discrete variables can be computed by the number of outcomes for a given value divided by the number of total samples in the training set. These can be easily obtained by computing the 1D histogram and 2D histogram of discrete values from the training set.

4 FINE-TUNING PARAMETERS

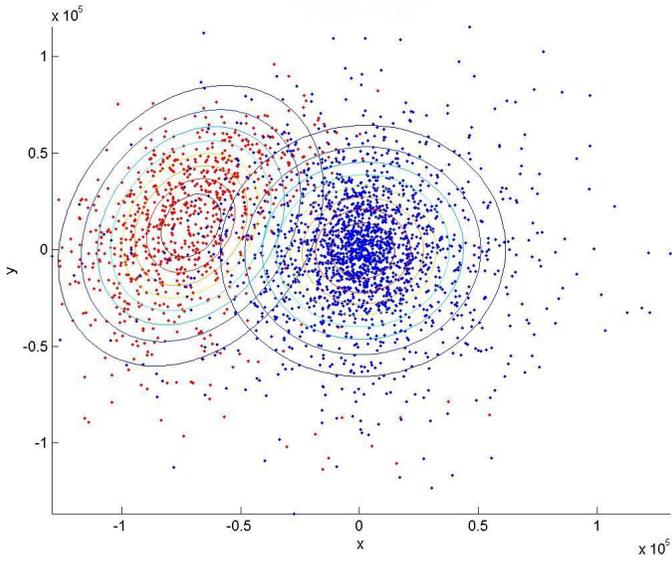
Due to biological and theoretical limitations, we have defined a number of 3024 Gabor filters. The main idea is to choose the most appropriate ones for a given image patch. To this end, we have used the supervised learning algorithm GentleBoost. The training data set consists of 730 positive images and 2000 negative ones. The test database contains 160 positive and 500 negative images. The eyes are extracted from the labeled face database, FERET [13]. The patches are cropped uniquely, obtaining a 33×33 sub-image centered on the pupil. The negative images are randomly acquired from the human face, except for the eye. This condition assures the ability to discriminate from other facial features.

The experiments concentrate on studying one parameter at a time, restricting the possible domain of parameter values. The problem is to choose from a huge number of defined filters those, which are most corresponding to the target object.

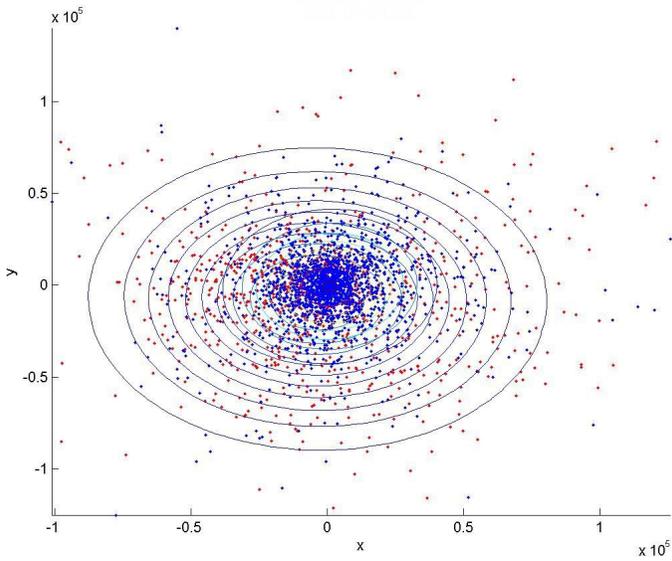
The convolution response of a Gabor filter with an image is a complex number. From a complex number, we may retrieve the amplitude, the argumentum, the real part and the imaginary part. The first group of experiments concerns the efficiency and effectiveness of these. In state of the art [6, 7, 9], the distribution of the module and phase are combined.

Figure 3 presents the distribution of filter responses for positive images and negative images. The mean value of the negative images is very close to the origin of the coordinate system. If the mean of positives is distant from the $(0, 0)$ point, it shows that the filter will be a weak classifier with a good performance (Figure 3 a)). If the mean value of positives is also $(0, 0)$, then it cannot be used for classification (Figure 3 b)). In our experiment, we have compared the classifiers based on the amplitude of responses, basing the responses on the normal distribution. Measuring the results on the validation set, it could be observed that the difference between detections parameters of two classifiers was not significant.

Table 1 measures global detection error, false positive detection error and false negative detection error for the two types of classifiers based on different informational values and using a given number of filters. Determining distribution param-



a)



b)

Figure 3. Separable and non-separable filter responses

Amplitude Response								
No. of cl.	4	8	12	16	20	24	28	32
ErD	10.68	3.43	1.88	1.42	0.97	0.76	0.56	0.36
ErFP	1.92	0.27	0.31	0.31	0.07	0.10	0.00	0.00
ErFN	20.28	6.89	3.60	2.64	1.96	1.48	1.18	0.75

Gaussian Response								
No. of cl.	4	8	12	16	20	24	28	32
ErD	7.20	4.97	4.41	4.03	3.57	3.27	3.04	2.89
ErFP	3.14	2.02	2.11	1.90	1.75	1.50	1.35	1.33
ErFN	15.27	10.83	8.98	8.26	7.18	6.78	6.38	5.97

Table 1. Amplitude and normal distribution error rates of filter responses

eters complicates calculations, and because of that, we shall only use amplitude values hereafter.

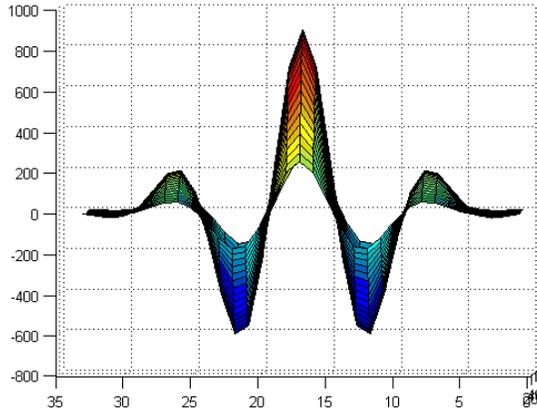
The aspect ratio defines the attenuation of the Gaussian envelope in both directions. Most of the authors [8, 14, 15] consider aspect ratio 1. In our experiments, we have compared several values for the aspect: $S \in \{0.5, 0.75, 1, 1.2, 1.5, 1.8\}$. The results clearly prove that $S > 1$. Intuitively, it would seem that the direction of propagation of the plane wave is equal to the major axis of the Gaussian, meaning that $S < 1$. But the result of the experiments was just contrariwise: an aspect value greater than 1. It follows that for a value of $1.8 - 2$, the attenuation of the Gaussian envelope is twice as large in the propagation direction of the wave than it is in the perpendicular direction. The best classification performances were obtained for $S = 1.8$.

The bandwidth defines the number of useful wavelengths taken into consideration. Considering the relation (6) for the interval $[-\alpha, +\alpha]$, we get, as wavelengths, 1.3λ and 2.4λ for $bw = 1$ and $bw = 2$, respectively (Figure 4).

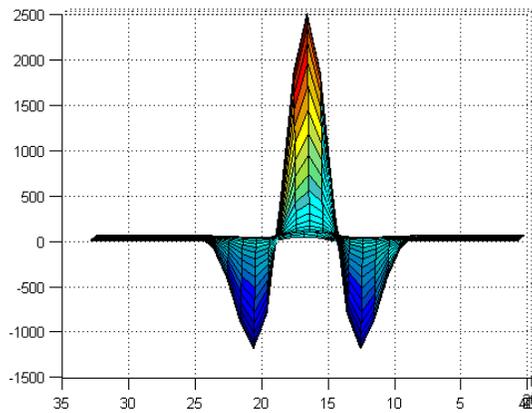
The goal in the detection process is to achieve a good detection rate; thus, the number of misses must be almost 0. Detection error is determined by the training and test samples from the image database used. At a given error rate, this factor can be changed by modifying image weights. The false negative error rate needs to be as small as possible with respect to the false positive rate. This ratio of positive image weights/negative image weights can be modified in the learning phase by enhancing positive image weights. The significance of this is to increase the importance of positive images in the training set.

5 RESULTS AND IMPROVEMENTS OF THE EXPERIMENTS

The best descriptor obtained from the 3024 predefined Gabor filters contains only 32 filters (Table 2), with its parameters presented in Table 3 and Figure 5. The abbreviations in this table refer to detection error, false negative and false positive detection error rates.



a)



b)

Figure 4. Influence of bandwidth on the defined filter; a) $bw = 1$, $\lambda/\alpha = 1.3$, b) $bw = 2$, $\lambda/\alpha = 2.4$

We can draw the following conclusion from the results obtained: the detection error becomes sufficiently stable and has a high performance at the same time with only 32 filters. If we analyze the selected filters, we may observe that the 8th and the 15th filters are the same, and the 12th and the 26th are very similar. This fact suggested to us that we should measure the amount of redundancy between filters. The results of the measurements are shown in Table 4.

Since 7 pairs have mutual information greater than 0.3, the elimination of these filters can presumably bring new improvements in the classifier. We have modified

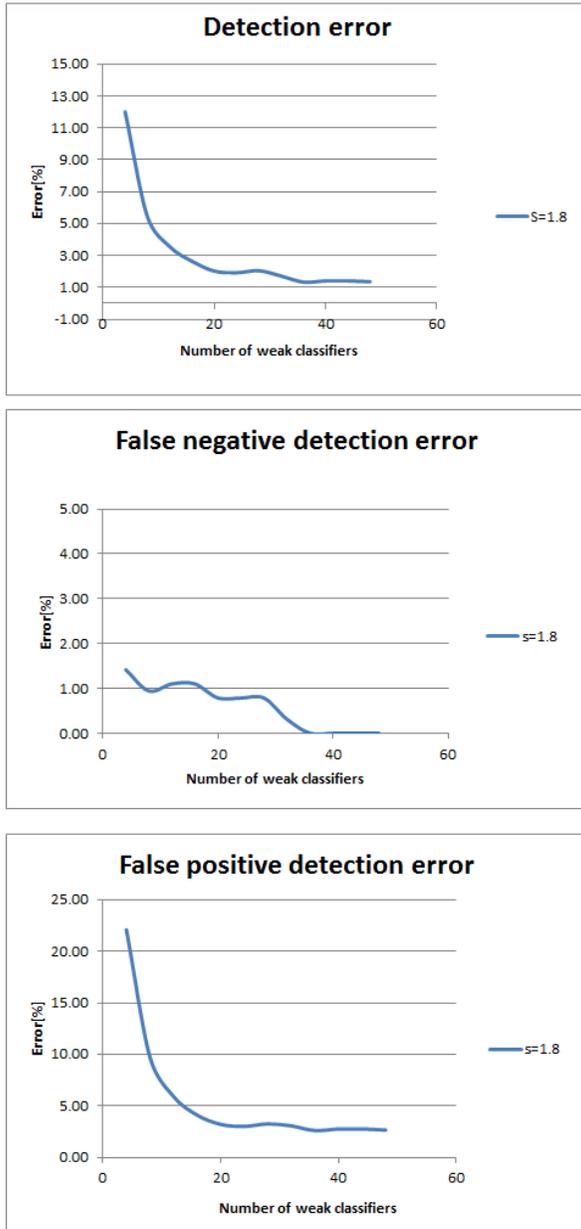


Figure 5. Descriptor performance

No. of cl.	λ	$\theta[^\circ]$	bw	No. of cl.	λ	$\theta[^\circ]$	bw
1.	18	15	1.5	17.	8	45	2.0
2.	10	135	1.0	18.	11	120	1.0
3.	4	90	2	19.	16	15	2.0
4.	5	15	1.0	20.	6	165	1.0
5.	14	0	1.5	21.	9	45	1.0
6.	4	60	1.0	22.	8	60	1.0
7.	11	60	1.5	23.	5	135	1.0
8.	22	15	1.0	24.	22	15	2.0
9.	6	150	1.0	25.	4	60	1.0
10.	6	60	2.0	26.	10	0	1.0
11.	4	120	2.0	27.	4	120	2.0
12.	11	0	1.0	28.	5	30	1.0
13.	4	165	1.0	29.	16	75	2.0
14.	12	90	1.0	30.	9	135	1.0
15.	22	15	1.0	31.	9	90	1.0
16.	18	45	1.0	32.	22	30	1.0

Table 2. Descriptor parameters

No. of cl.	4	8	12	16	20	24	28	32
ErrD	12.00	5.42	3.56	2.64	2.02	1.92	2.05	1.71
ErrFP	1.42	0.94	1.10	1.10	0.79	0.79	0.79	0.31
ErrFN	22.10	9.70	5.90	4.10	3.20	3.00	3.25	3.05

Table 3. Descriptor performance

the GentleBoost algorithm in order to get rid of the classifiers which have a high mutual information with the previously selected information. A similar algorithm had been used in a face recognition system by Shen and Bai [9]. The parameters of the resulting classifiers are illustrated in Table 5, and their performance in Table 6. The measurements show that in the case of a small number of filters, the elimination of redundancy does not bring a considerable improvement to performance.

The descriptor created does not only act as a descriptor, but it is used as a classifier as well. Applying it for each point of the image (Figure 6 a)), we get the response map (Figure 6 b)).

The positive values represent object detections, while the negative ones represent non-object detections. These values, computed with the GentleBoost learning algorithm, represent not only a classification decision, but also a confidence value regarding the decision made.

Figure 7 b) represents the confidence values around an object point. In order to create the final detector, we must use not only the sign, but the effective value as well, mostly for positive values. The oscillations that appear have to be eliminated

No. of cl.	λ	$\theta[^\circ]$	bw	No. of cl.	λ	$\theta[^\circ]$	bw
1.	18	15	1.5	17.	6	30	1.0
2.	10	135	1.0	18.	22	30	1.5
3.	4	90	2.0	19.	22	105	1.0
4.	5	15	1.0	20.	5	105	1.0
5.	14	0	1.5	21.	7	165	1.0
6.	4	60	1.0	22.	14	15	1.5
7.	11	60	1.5	23.	12	120	1.0
8.	22	15	1.0	24.	8	60	1.0
9.	6	150	1.0	25.	4	30	2.0
10.	6	60	2.0	26.	10	60	1.0
11.	4	120	2.0	27.	9	75	1.5
12.	11	0	1.0	28.	6	120	1.0
13.	4	165	1.0	29.	16	45	2.0
14.	12	90	1.0	30.	4	15	1.0
15.	22	0	2.0	31.	5	15	2.0
16.	20	45	1.0	32.	5	0	1.0

Table 5. Irredundant descriptor parameters

No. of cl.	4	8	12	16	20	24	28	32
ErrD	12.00	5.42	3.56	2.74	2.48	2.00	1.94	1.69
ErrFP	1.42	0.94	1.10	1.10	0.94	0.63	0.79	0.63
ErrFN	22.10	9.70	5.90	4.30	3.95	3.30	3.05	2.70

Table 6. Irredundant descriptor performance

in order to use the obtained map as a patch or object detector. For this purpose, mean filtering must be used so as to obtain a smooth surface 7b), the peak of which corresponds to the location of the target object part. The algorithm described assumes the evaluation of these responses in every point of the image. The proposed improvement consists of computing the responses only in a regular grid, omitting lots of uninteresting points. This elimination imposes analysis of the responses in the neighborhood of the object.

S=1.8										
Relative Position		-16	-14	-12	-10	-8	-6	-4	-2	0
Positive	Horizontal	-11.06	-8.77	-7.69	-6.85	-1.05	9.66	8.70	6.43	11.43
Positive	Vertical	-7.57	-5.39	-0.83	-0.56	1.60	2.41	8.35	13.98	11.43
False positive	Horizontal	-17.33	-15.54	-16.77	-8.40	-10.44	-4.10	-0.16	2.93	8.27
False positive	Vertical	-2.25	-0.95	-2.61	-6.84	-4.12	-1.28	4.20	5.31	8.27
Relative Position		0	2	4	6	8	10	12	14	16
Positive	Horizontal	11.43	4.35	8.42	6.46	11.96	4.89	-0.26	-2.97	-10.16
Positive	Vertical	11.43	1.13	-5.87	-11.11	-15.33	-17.37	-19.29	-16.58	-14.27
False positive	Horizontal	8.27	5.77	4.01	7.09	2.29	-4.70	-6.03	-4.60	-3.38
False positive	Vertical	8.27	5.41	-6.04	-11.88	-11.75	-14.06	-17.81	-15.69	-15.55

Table 7. Variation of classifier responses horizontally and vertically

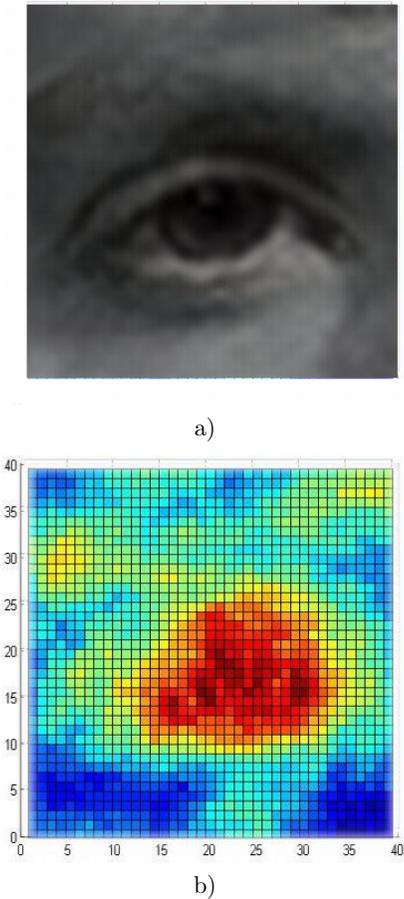
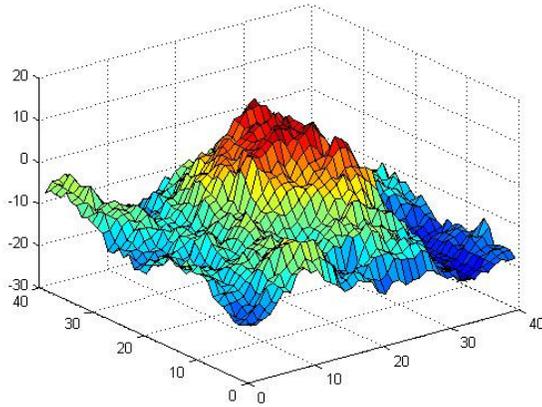


Figure 6. Response map in 2D; a) An example image, b) The corresponding descriptor

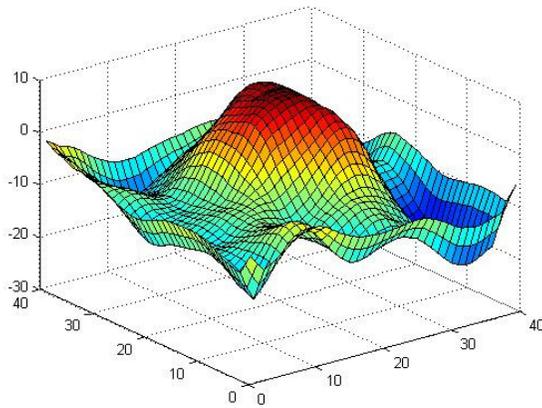
Figure 8 a) represents the variation of responses around the positive points horizontally and vertically. Figure 8 b) stands for the same, but around a false positive detection.

By comparing these charts, we can conclude the following:

- the response values for positive detections are high
- the area covered horizontally and vertically by positive responses is larger than those of false positives
- the number of points which indicate a detection in a given direction is greater than 10 pixels
- we define as interesting regions the image regions where the average value of responses for 9 (3×3) adjacent points is also positive



a)

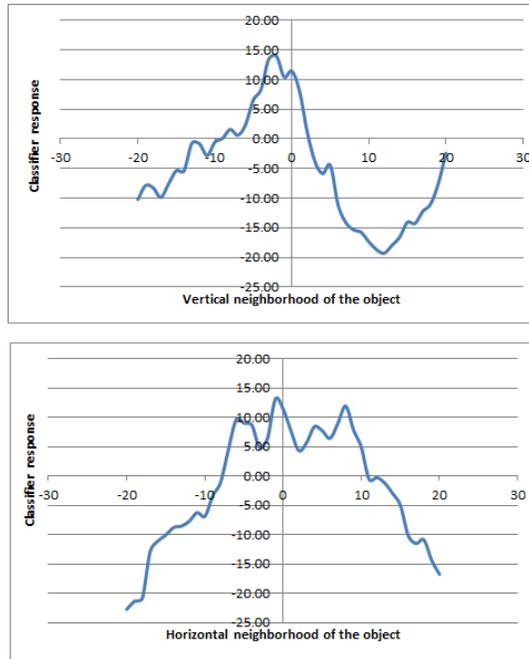


b)

Figure 7. Filtered response map

In order to create the response map we propose the following algorithm (Figure 9) which scans the image 4 by 4 pixels. The steps of the algorithm are:

- for the positive responses the average value of the 9 surrounding points is evaluated
- if this region is an interesting region, then the neighboring 16 points are computed too
- if this is not an interesting region, then the responses of the intermediate points are not computed by the descriptor, they are simply interpolated from the grid points

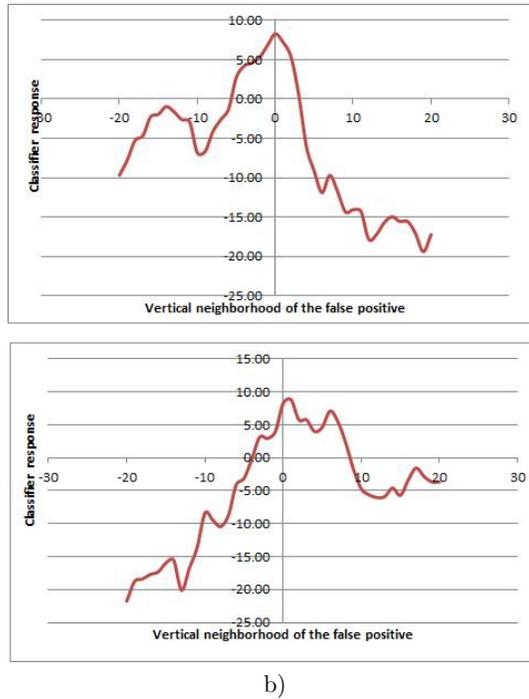


a)

- finally, a 5×5 mean filter is applied for smoothing the 3D response map (Figure 7).

The peak of the response map obtained in this manner represents the central point of the detected object. The algorithm described above has a detection rate of 96% and improves processing time, computing the response map 14 times faster than the non-enhanced algorithm.

Another way to reduce computation time is to use so-called ‘interest points’. Interest points are detected based on certain geometrical properties that present some kind of consistency, ex. scale or rotation invariance. The LoG (Laplacian of Gaussian) detector filters the image and selects the parts which have a circular aspect. Due to the circular aspect of the eye, we have decided to use the LoG operator for eye detection. The experiments done on the training and test sets show a detection rate of 95% in the first 200 LoG points per image. The average distance of the closest LoG to the marked pixel (correct location) is approximately 2.27 pixels. This property suggests the applicability of the LoG detector in the training process as well as the final detection process. We have created a separate image database using the LoG points. The patches for positive images were centered on the LoG point closest to the eye marked, while the strongest LoG regions of the face were considered negative image patches. Using the same fine-tuning process presented in the previous section, we have created another classifier and



b)

Figure 8. Variation of classifier responses; a) In the neighborhood of the object, b) In the neighborhood of the false detection

descriptor based on LoG points that is also made up of 32 filters. The comparative results of this descriptor for the training set and test set are presented in Tables 8 and 9. We can conclude that detection results are slightly better for the manually marked database. With 32 classifiers taken into account, the detection error is around 1.5% for marked points and 3.71% for LoG points. Detection performances in the test set are closer: detection error rates of 3.98% and 3.71% have been measured.

With this new descriptor based on LoG points, a response map can be created with little modification to the filtering algorithm in a similar way. At first, descriptor responses are computed only in the LoG points. Next, detection continues with the evaluation of interesting zones (9 adjacent points) where the average value is positive as well. The efficiency of the algorithm depends on the number of LoG points taken into consideration. The localization error, measured in pixels, is double compared to the more accurate descriptor that is based on marked points. The bottleneck of the LoG-Gabor descriptor is the detection of image patches that do not have a circular aspect. In these cases, finding and applying other interest point detectors becomes a necessity.

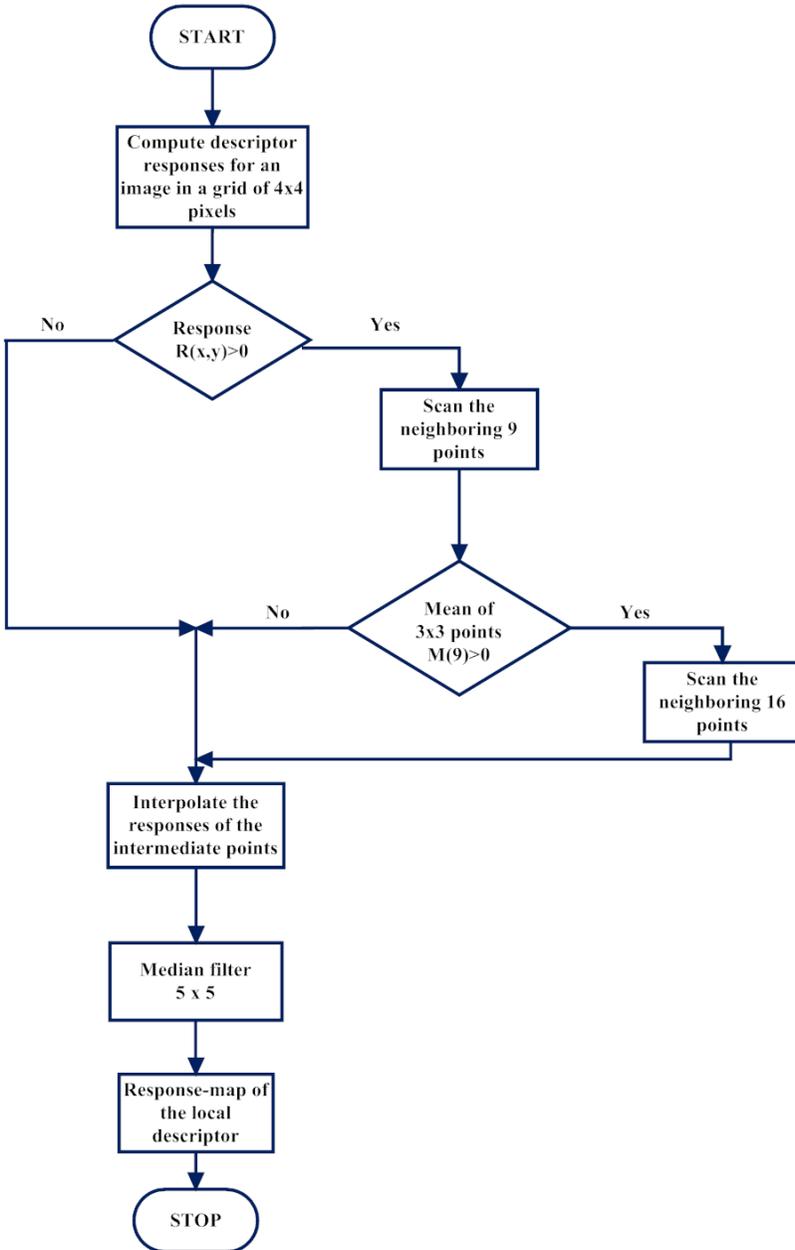


Figure 9. Accelerating the response map computation

LoG								
No. of. cl.	4	8	12	16	20	24	28	32
ErrD	11.20	8.83	6.98	5.75	5.26	4.61	4.22	3.71
ErrFN	3.72	2.97	1.41	1.44	1.41	1.03	1.03	0.72
ErrFP	20.18	15.86	13.68	10.93	9.89	8.90	8.05	7.30

Marked Position								
No. of. cl.	4	8	12	16	20	24	28	32
ErrD	12.33	5.87	4.08	3.21	2.59	2.06	1.72	1.49
ErrFN	2.29	2.19	1.23	0.86	0.27	0.21	0.03	0.14
ErrFP	23.33	9.90	7.19	5.79	5.13	4.09	3.58	2.98

Table 8. Detection performances of LoG points and marked points on the training set

LoG								
No. of. cl.	4	8	12	16	20	24	28	32
ErrD	12.37	7.38	6.00	5.67	4.68	4.17	4.48	3.98
ErrFN	3.63	3.21	3.07	2.23	2.51	2.65	2.37	2.93
ErrFP	21.75	11.85	9.15	9.35	7.00	5.80	6.75	5.10

Marked Position								
No. of. cl.	4	8	12	16	20	24	28	32
ErrD	12.64	6.88	5.40	4.61	3.71	3.94	3.71	3.71
ErrFN	3.14	3.62	2.52	2.20	1.57	2.52	2.52	2.36
ErrFP	21.70	10.00	8.15	6.90	5.75	5.30	4.85	5.00

Table 9. Detection performances of LoG points and marked points on the test set

6 CONCLUSION AND FUTURE WORK

This paper presents a local image descriptor based on Gabor wavelets. It presents several improvements in order to reduce the parameter space of the used filters. Comparing the proposed descriptor to those found in the state of the art, we can conclude that the selected filters are adaptable to any target object.

This paper also analyses the redundancy between selected filters. Therefore, we implement an updated version of the GentleBoost algorithm which eliminates the filters which have high mutual information with the previously selected ones. To reduce the computation time we compute the response map only in the so called LoG interest points. Finally, for more accurate detection, we suggest an acceleration algorithm in order to determine the response map faster.

Accordingly, the detector obtained is very robust and detects the most appropriate instance of the object in a given image with high accuracy (96%). The computational efficiency of the training and detection process is very time-consuming because of the convolutions, hence the need for using the presented improvements in order to reduce the high computational complexity of the system. An important

contribution is the implementation of these expansions by a cluster of computers working in parallel.

As to the future we propose to apply the presented descriptor in the same way for several parts of the object to be detected. For each of these components, we get a response map that localizes each part separately. The obtained maps have to be processed using the detection model. We propose the deformable object model [17] for the most effective application. With this model, false detections can easily be eliminated using not only this descriptor, but also the special relations between the parts.

Acknowledgements

Financial support for the work of Szidónia Lefkovits was provided from programs co-financed by the European Social Fund under the responsibility of the Managing Authority for the Sectoral Operational Program of Human Resources Development, Investing in People! Key Area of Intervention 1.5 “Doctoral and Postdoctoral Scholarship in Support of Research”; Project Title: “Integrated System for Quality Improvement of the Doctoral and Post-Doctoral Research in Romania and Promoting the Role of Science in Society”, Contract: POSDRU/159/1.5/S/133652.

REFERENCES

- [1] LU, Y.—ZHOU, J.—YU, S.: A Survey of Face Detection, Extraction and Recognition. *Computing and Informatics*, Vol. 22, 2003, No. 2, pp. 163–195.
- [2] LEIBE, B.—LEONARDIS, A.—SCHIELE, B.: Robust Object Detection with Interleaved Categorization and Segmentation. *International Journal of Computer Vision*, Vol. 77, 2008, pp. 259–289.
- [3] DALAL, N.—TRIGGS, B.: Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 2005, Vol. 1, pp. 886–893.
- [4] LOWE, D: Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, 2004, No. 2, pp. 91–110.
- [5] KE, Y.—SUKTHANKAR, R.: PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, 2004, Vol. 2, pp. 506–513.
- [6] WISKOTT, L.—FELLOUS, J. M.—KRÜGER N.—VON DER MALSBERG, CH.: Face Recognition by Elastic Bunch Graph Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, No. 7, pp. 775–779.
- [7] TAMMINEN, T.—LAMPINEN, J.: Sequential Monte Carlo for Bayesian Matching of Objects with Occlusions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, 2004, No. 6, pp. 930–941.

- [8] VUKADINOVIC, D.—PANTIC, M: Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers. IEEE International Conference on Systems, Man and Cybernetics, 2005, Vol. 2, pp. 1692–1698.
- [9] SHEN, L.—BAI, L.—BARDSLEY, D.—WANG, Y: Gabor Feature Selection for Face Recognition using Improved Adaboost Learning. Proceedings of International Workshop on Biometric Recognition System (IWBR 2005), LNCS, 2005, Vol. 3781, pp. 39–49.
- [10] SHEN, L.—BAI, L.: A Review on Gabor Wavelets for Face Recognition. Pattern Analysis and Applications, Vol. 9, 2006, No. 2, pp. 273–292.
- [11] LEE, T. S.: Image Representation Using 2D Gabor Wavelets. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, 1996, No. 10, pp. 959–971.
- [12] MOVELLAN, J.: Tutorial on Gabor Filters. <http://mplab.ucsd.edu/tutorials/gabor.pdf>.
- [13] FERET Database. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.
- [14] ILONEN, J.—KAMARAINEN, J.K.—PAALANEN, P.—HAMOUZ, M.—KITTLER, J.—KÄLVIÄINEN, H.: Image Feature Localization by Multiple hypothesis Testing of Gabor Features. IEEE Transactions on Image Processing, Vol. 17, 2008, No. 3, pp. 311–325.
- [15] CAMPADELLI, P.—LANZAROTTI, R.—LIPORI, G: Precise Eye Localization Through a General-to-Specific Model Definition. Proceedings of the 17th British Machine Vision Conference (BMVC), 2006, Vol. 1, pp. 187–196.
- [16] LINDBERG, T.: Feature Detection with Automatic Scale Selection. International Journal of Computer Vision, Vol. 30, 1998, No. 2, pp. 79–116.
- [17] FELZENSZWALB, P. F.—HUTTENLOCHER, D. P.: Pictorial Structures for Object Recognition. International Journal of Computer Vision, Vol. 61, 2005, No. 1, pp. 55–79.



Szidónia LEFKOVITS received her Ph.D. degree in computer science from the Babes-Bolyai University Cluj-Napoca (Romania) in 2012. Currently she works as Assistant Lecturer at the Department of Mathematics-Computer Science of the “Petru Maior” University Țirgu-Mureș (Romania) since 2008. She received her electrical engineering degree in computer science in 2009 from the Sapientia University Țirgu-Mureș (Romania) and at the same time graduated in mathematics and computer science in 2008 from the “Petru Maior” University Țirgu-Mureș. She is the author of over 10 scientific papers. Her topics of interest

include, but are not limited to, computer vision, object detection, image processing and pattern recognition, databases, programming techniques, data structures.