Babeș-Bolyai University Cluj-Napoca

Faculty of Mathematics and Computer Science

Robust Object Detection Methods with Applications in Face Detection

PhD Thesis Extended Abstract

Scientific advisor:
Prof. univ. dr. Horia F. Pop

PhD candidate:
Lefkovits Szidónia

2012

Contents of the Thesis

Contents of the Abstract

Acknowledgements

I would like to express my sincere gratitude to professor dr. Horia F. Pop, my scientific advisor, for his help during this scientific activity.

I am grateful to my parents as well for the help they have offered me. The passion for research, encouragements and lots of ideas have come from my father, and my mother has morally supported me during the entire time I have been preparing this thesis.

Introduction

The human visual system has the extraordinary capacity of recognizing a wide variety of objects or object categories from only two- or three-dimensional visual information. Solving this task of detection or recognition without any effort happens mainly due to the capacity for extremely quick parallel processing, the ability for self-development, being able to learn from experience and the exceptional structure of the human brain.

The aim of artificial vision is not to understand human perception, but to model and create robust automatic detection systems. Artificial vision is a branch of general object detection that processes two-dimensional images as a projection of three-dimensional space.

The most up-to-date research has not led to a general system that could be useful for solving all practical applications. Each of the existing systems is created with a specific aim and work in certain given conditions.

Creating such an automatic detection system is a challenge for every researcher in this field. The main difficulty comes from the extreme variety in which all objects appear. This diversity is induced by the huge set of existing objects, and each of the these may appear in an infinite number of aspects: dimension, form, color, position, illumination, shadow, occlusion.

At the same time, geometric forms may also present ambiguity owing to the projection of the 3D space in the image plane. Significant disturbance factors in the aspect of the objects are the way of illumination, the shadows and surfaces that absorb or reflect light. Another bottleneck arises from the great quantity of information that an image contains. The appearance of the same object in a multitude of environments, circumstances or contexts is also unfavorable [35].

The object detection from two-dimensional images plays an important part in several different areas of activity. The most widespread applications are in the fields of medicine, traffic supervision, access control, identification and authentication systems, industries using intelligent robots, cinematographic productions etc.

List of Publications

ISI Journals

- **Sz. Lefkovits** [57] ”Improvements on Gabor Descriptor Retrieval for Patch Detection” *Computing and Informatics* (ISI IF=0.66, SRI = 0.15) [under review]

ISI Proceedings

- **Sz. Lefkovits** [54] ”Numerical Computation Method of the General Distance Transform” *KEPT 2011 International Conference Knowledge Engineering Principles and Techniques*, Babeș-Bolyai University, 2011. ISSN: 2067-1180, pp. 143–152, indexed MathSciNet, ZMath.

- **Sz. Lefkovits** [48] ”Performance Analysis of Face Detection Based on Haar-features” *Proceedings of the International Conference on Complexity and Intelligence of the Artificial and Natural Complex Systems: Medical Applications of the Complex Systems: Biomedical Computing*, Petru Maior University, 2008, ISBN: 978-973-7794-76-5, pp.184–192.

International indexed BDI Journals

- **Sz. Lefkovits** [58] ”Novel Gabor Filter Based Patch Descriptor” *10th Jubilee International Symposium on Intelligent Systems and Informatics, IEEE Conference*, Serbia, 2012 (Australian Research Council class C) [accepted]

- **Sz. Lefkovits** [55] ”Numerical Computation Method of the General Distance Transform” *Studia Informatica Universitatis Babeș-Bolyai,* vol. 56, no. 2, 2011, pp. 68–74. indexed MathSciNet, ZMath.

- **Sz. Lefkovits** [49] ”Assessment of Building Classifiers for Face Detection” *Acta Universitatis Sapientiae, Electrical and Mechanical Engineering*, no. 1, 2009, ISSN: 2065-5916, pp. 175–186, indexed EBCSO Databases, EZB − Karlsruhe Institute of Technology

International Journals acknowledged by the National University Research Council in Higher Education CNCSIS, B category

- **Sz. Lefkovits** [53] ”Combining Boosted Global- and Part-aspect Face Detectors” *Scientific Bulletin of the Petru Maior University*, vol.8, no. 1, 2011, ISSN: 1841-9267, pp. 35–40.

International Conferences

- **Sz. Lefkovits** [56] "Hybrid Face Detector Based on Boosted Classifiers" *The Eighth Conference of PhD Students in Computer Science*, Optimization session, University of Szeged, Hungary, Institute of Informatics, 2012 **„Best talk of the Session"** award

- **Sz. Lefkovits** [50] "Teaching Improvements on Haar-based Classifiers" *International Conference IETM-3*, Mathematics and Computer Science section, Petru Maior University 2009, ISSN: 1844-2048, pp. 1050–1060.

- **Sz. Lefkovits**, C. Enăchescu [59] "Face Detection System Based on Artificial Intelligence", I*nternational Conference ICELM*, Decision Systems, Mathematical Modelling and Statistics section, Petru Maior University, 2008, ISBN: 973-7794-00-10, pp. 112–121

National Conferences

- **Sz. Lefkovits** [52] "Classification Mixture for Object Detection" *Cluj Academic Days*, 2010, ISSN: 2066-5768, pp. 5–10

Original Contributions

- detection in a two-phase approach: tapering the domain of interest with a rapid detector based on global aspect, followed by detailed processing with a local aspect-based detector [51]
- creating the database for global object detection (section 3.4, [48])
- enhancing the cascade classifier algorithm, considerable decrease of false detections due to the selection of significant negative images (section 3.4, [59], [49])
- building the proprietary face detector and measuring its detection performances (section 3.4, [49], [50])
- defining a local descriptor based on 2D Gabor filters (section 4.2, [56])
- study, analysis and theoretical design of the 2D Gabor filters (section 4.2.1.1, [52])
- experimenting and evaluating the 2D Gabor filter parameters that define the local descriptor (section 4.2.3, [56])
- narrowing the domain of values useful for the 9 parameters of the 2D Gabor filters (section 4.2.3, [56])
- designing the selection system and evaluation of the filters characteristic for the detectable object (section 4.2.3, [58])
- creation, determination and evaluation via experiments of the proposed local descriptor (section 4.2.3, [58])
- developing an original algorithm for computing the general distance transform (section 4.1.4.1, [54], [55])
- creating the object detection system and implementing it with parallel processing on a distributed system (section 4.2.4, [52], [53], [56], [58])

Thesis Structure

1 Introduction. In this chapter, the goal, the task, the problematics and the main applications of object detection are presented. It continues with the list of publications related to the thesis and the original contributions made by this dissertation.

2 Detection Systems. General Context. This chapter contains an overview and the general structure of object detection systems. The most important methods and system of the domain are presented in short. The techniques used for each part of the detection systems are mentioned. At the end of this chapter, the proposed detection system is described, which will be studied and analyzed in detail in the course of the thesis.

3 Global Aspect-based Detection System. In this chapter a global aspect-based system is presented. Sections 3.1, 3.2 and 3.3 contain the succinct theoretical basis. Section 3.1 describes the AdaBoost algorithm invented by Y. Freud and R. Schapire [32, 78]. In section 3.2, there is an analysis of the construction model of weak classifiers with Haar functions created by P. Viola and M. Jones [92]. Section 3.3 presents the idea of R. Lienhart et al. [61] of enhancing the performance of the system with the cascade classifiers.

Section 3.4 discusses the hidden problems of these systems, which need to be solved in order to create such a functional system. Here, we describe the method of creating databases and the proprietary algorithm as well, using which we have managed to create proprietary classifiers [48, 59] with performance comparable to existing classifiers. This chapter ends with presentation of our own experiments with the created classifier [49, 50] and draws conclusions about the advantages and drawbacks of such a system, pointing towards the next chapter, in which we create the second part of our detector.

4 Local Aspect-based Detection System. This chapter details the theoretical and practical aspects of a part-based detection system. Section 4.1 presents a deformable object model [28]. In section 4.1.3, we provide full details for the detection algorithm by defining the state table that facilitates the effective implementation of the proposed model. In section 4.1.4, we suggest and implement an original algorithm [54, 55] for the effective evaluation of the general distance transform that is critical to the evaluation of the cost function in the deformable model. This algorithm is compared with the lower envelope algorithm [70]. The experiments performed outline the advantages of the proposed algorithm. If the lower envelope can be computed analytically then the processing times of the two algorithms are alike, but if the lower envelope has to be determined numerically, then the proposed algorithm is lot more efficient.

Section 4.2 outlines the theoretical foundations of 2D Gabor filters [72]. In section 4.2.1.1 we propose a method of designing two-dimensional Gabor filters through which the domain of values of the 9 parameters is reduced. Using theoretical and experimental restrictions, the domain has been narrowed down, but not sufficiently to use them as local descriptors. Another method is required to decrease the number of filters with help of the Gentle AdaBoost algorithm [33].

In section 4.2.2 we present the implementation of this algorithm [56, 58]. In section 4.2.3 by using the Gentle AdaBoost algorithm and by evaluating the results of the experiments, on a validation set, most characteristic filters could be selected for the image patch analysed. Section 4.2.4 illustrates the performances of the local descriptor created [53, 56, 58]. The practical results are concentrating on human eye detection. Taking into account the necessity to decrease processing time, we present a parallel distributed implementation of the training process and detection process for this descriptor.

Keywords

object detection, global aspect, local aspect, Haar functions, two-dimensional Gabor wavelets, frequency domain analysis, filter design, parameter optimization, labeled databases, supervised learning, AdaBoost, Gentle AdaBoost, background generator, classifier, false detection decrease, local descriptor, filter selection, response map, general distance transform, parallel implementation.

Global Aspect-based Detection System

Global aspect-based systems present the object as a whole, indivisible entities. In general, the detection process uses the image pyramid technique, i. e. the images are redimensioned and scanned over the image with a standard dimensional window. In this chapter, we present such a system [48, 49, 50, 53, 59]. Sections 3.1, 3.2 and 3.3 describe the theoretical foundations which constitute the system.

Section 3.1 presents the AdaBoost algorithm elaborated by Y. Freud and R. Schapire [32, 78]. The algorithm builds a strong classifier from a of set of weak classifiers, a series of classifiers with well-determined weights are adapted to the problem of decision. Section 3.2 analyzes the method of building weak classifiers by means of Haar functions. Using the AdaBoost algorithm and the Haar functions as weak classifiers, P. Viola and M. Jones [92] create a detection system with extraordinary performance in real-time.

Section 3.3 presents the idea of R. Lienhart et al. [61] for enhancing the performance of the system by transforming the monolithic classifier to a cascade classifier.

We have not found any reference about the methodology classifier creation in the state of the art. Researchers only publish the results of their experiments, keeping the databases and techniques used in strict confidence. Section 3.4 discusses the hidden problems [50], which must be solved in order to create a functional system. Here we describe method for creating the database [49] and the proprietary algorithm as well, by means of which we have managed to create proprietary classifiers [48] that have results comparable to existing classifiers.

Building the proprietary Classifier

The experiments presented hereafter were carried out by the Haartraining application that is part of the OpenCV program collection [11]. Our aim was to create our own classifier for detecting every type of object.

To facilitate our work, we have created auxiliary applications for labeling and cropping different types of images with various extensions by using the Boost C++ and OpenCV program suites. This auxiliary application enables us to label interest points on the face and to crop them to specific sizes. The points selected are stored in a text file and are used in the learning process. The database comprises 2893 images of distinct human faces.

We have started our experiments with a series questions in mind that we have answered by creating our own classifier:

1. Which is the best size for the input images? The best results were obtained from the measurements carried out for a size of 18×24; these had less false detection at the same detection rate.

2. How to crop the faces? The more significant details an image contains, the larger its size is. Considerably better results have been obtained by cropping not only the face, but the outline of the head as well.

3. Which images are the significant ones? We define facial images that have not yet been detected as significant-positive images. These images must be reintroduced as an expansion to positive images from the training database. We must find the types of faces that address the deficiencies of the training set. This process is laborious because, after modifying the training set, the entire training process must be reconsidered and executed again.

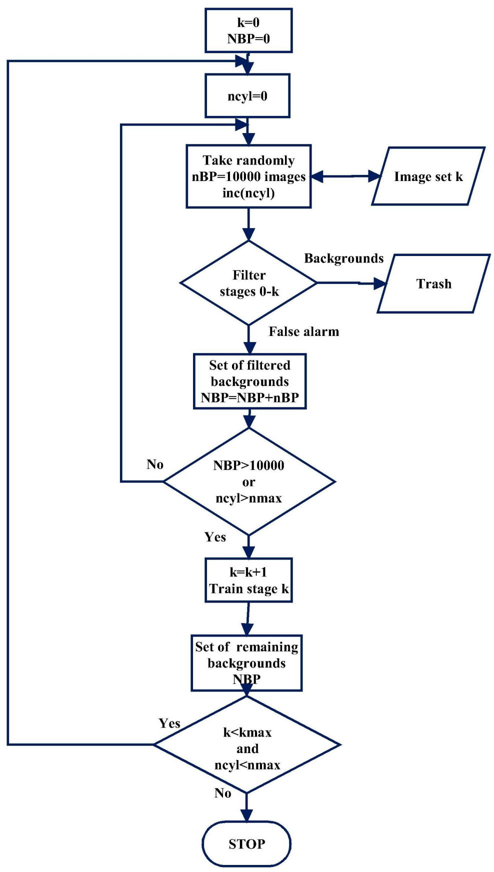

Figure 1: Significant Background Generator
© Sz. Lefkovits

4. What is the optimal size of the training set? The results of the experiments have shown that the more diverse the training set is, the better the performances are, but the running time of the training process increases at the same time. In the end, the training set consists of 2893 facial images and millions of background images.

5. How can we achieve a lower false detection rate than $5 \cdot 10^{-6}$? The detection process uses the image pyramid technique. The images of the pyramid (the same images at different scales) are scanned by a sliding window at a given dimension (18×24 pixels). The number of images to be evaluated is at the order of magnitude of several hundreds of thousands or even millions. This requires for the rate of false positive detections to be under $5 \cdot 10^{-6}$, a performance that can only be accomplished if the database consisted of millions of backgrounds. A large number of negative images would extend training time and would require massive hardware storage. We have reached the conclusion that a stage must to be trained using approximately 3000-5000 facial images and 10000 significant background images.

   We define images which do not look like faces but are indicated as such as negative-significant images, or, in other words, false positive images. For labeling and cropping significant-negative images we have created the significant background generator algorithm (figure 1). This generator is used before creating a new stage in order to provide sufficient number of negative images. The goal of the background generator is to find over 10000 false positive images to train the next stage with. Our experiments have shown that the time needed to generate significant-negative images for 15-20 stages is much longer than the effective training time with the AdaBoost learning algorithm.

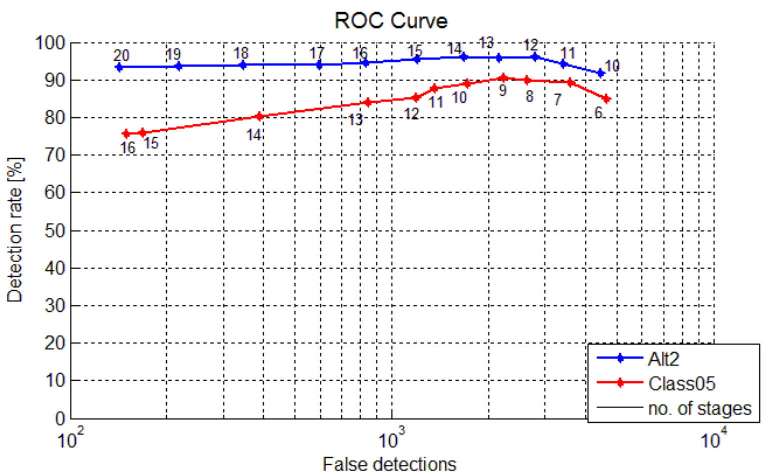

Figure 2: Stage by Stage ROC Curve
© Sz. Lefkovits

Performances of the Proposed Classifier

After obtaining the classifier that satisfies the above-mentioned conditions, we have compared its performance to existing classifiers using the CMU [94] test set, a reference database in the field. For the sake of comparison, we have plotted the detection rate depending on the number of false detections, namely the ROC curve of the classifier. An important result was a significant decrease of false detections. Based on the ROC curve, we could see that along with the decease of false detections, the detection rate also decreased considerably. At the same time, the number of weak classifiers increased substantially with the addition of new stages. At the same time, the number of weak classifiers increased substantially with the addition of new stages. The $Clas05$ classifier was created using a database consisting of young European faces without beards, mustaches or glasses. We have compared our classifier to frontalface_alt2 [11] using a series of images from the internet and those assured by FotoVision Studio [1]. We may state that we have obtained better results, i.e. the number of detections were similar; however, the number of false positive detections had been reduced considerably. The performances of the classifier created and frontalface_alt2 could be compared visually on the CMU [94] test set and a number of images that we have provided (figure 3).

Several global aspect-based detection systems had been created by H. Rowley et al. [80], H. Schneiderman [83], P. Viola and M. Jones [92] and R. Lienhart et al. [63]. The crucial disadvantage of these systems is the necessity for an image database that covers the aspects of the target object's appearances as well as possible. In general, they concentrate on detecting only one object class for a single spatial orientation. In order to handle several orientations, creating distinct databases for each orientation separately is required, and to distinguish orientations, it is necessary to specialize the classifier via parallel structures or by a tree-like structure with specific decision nodes.

The significant advantage of these systems is the high detection rate and the detection time, which corresponds to real-time applications as well. Based on the experiments we have carried out, we were able to draw the conclusion that the detection which uses the global aspect is based on the ability of the system to eliminate negative images as quickly as possible without losing any positive images. The disadvantage of these systems is that they detect a relatively high amount of false positives that do not resemble faces at all. In order to reduce the number of false detections, we have proposed to build a system which, in the first phase, quickly detects image regions that may contain faces, and in the second phase, detects facial features. In this way, spatial configuration of the facial features detected would lessen the probability of false positive detections. Certainly, the detection performance increases at the expense of processing time. For detection of human faces, we have used the system proposed by Lienhart et al. [61] and for facial feature detection we use the deformable model based on 2D Gabor filters as local descriptors.

The own classifier *Clas05* | *frontalface_alt2* classifier

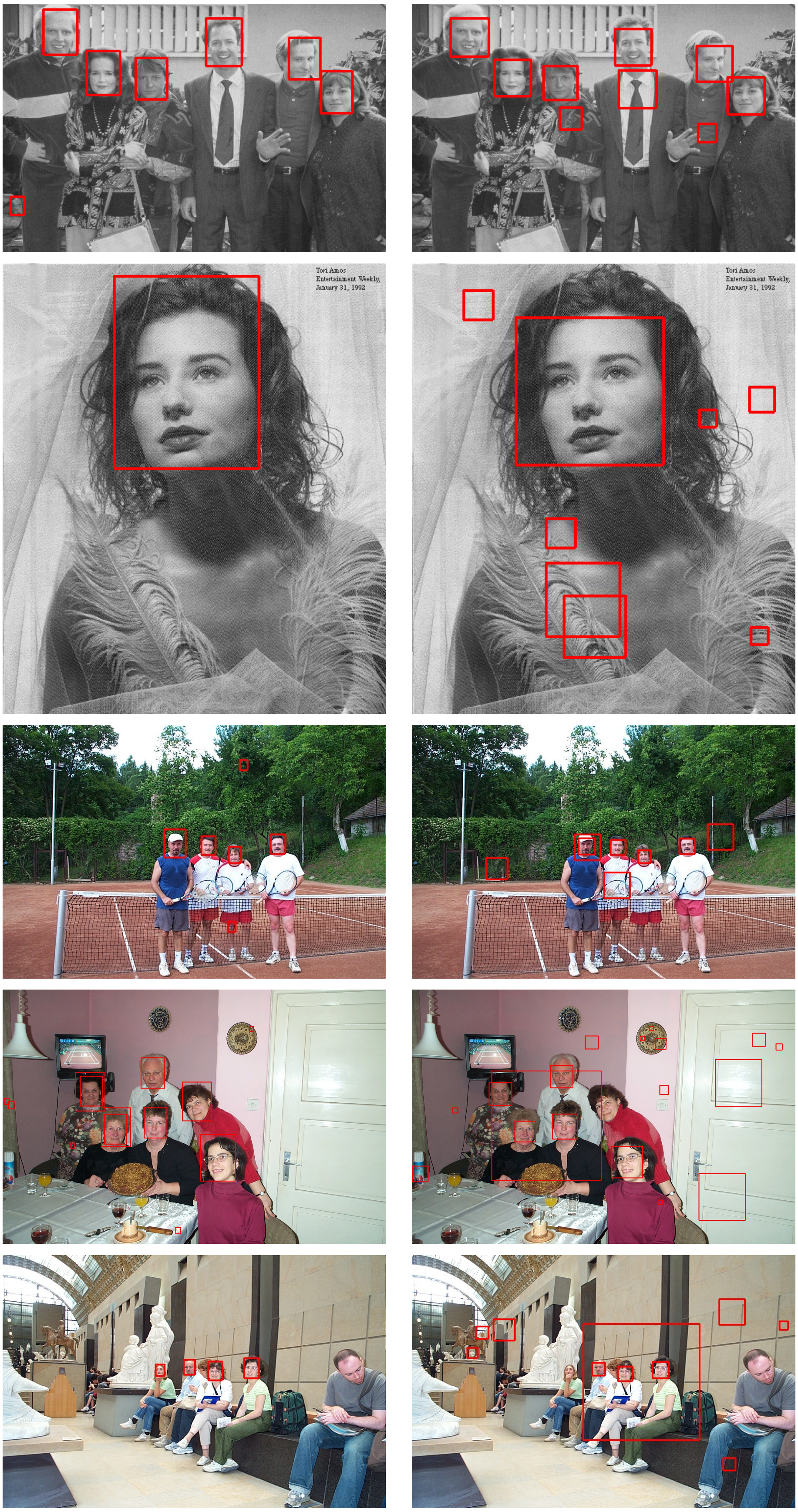

Figure 3: Detection Examples

Local Aspect-based Detection System

Recent research in the area of artificial vision tends to gloss over individual object detection and concentrates mainly on establishing new methods for object class detection. It is necessary to create generic models based on object parts and the relationship between them. One of the possible solutions is deformable object models. Generally, in a system based on local aspect, three parts may be distinguished:

- 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 – represents the mathematical formalism based on which the detected object parts compose the target object.

This chapter details the theoretical and practical aspects of a part-based detection system. Section 4.1 presents a deformable object model [28]. The theoretical part of the definition and determination of model parameters is inspired by the model used by Felzenszwalb et al. [28]. In section 4.1.3 we give details of the detection algorithm define the state table that facilitates the effective implementation of the proposed model. A determining factor for the complexity of the algorithm is the solution of the distance transform.

In section 4.1.4 we suggest and implement an original algorithm [54, 55] for the effective evaluation of the general distance transform which is determinant in the evaluation of the cost function in the deformable model. This algorithm is compared with the lower envelope algorithm [70]. If the lower envelope can be computed analytically then the processing times of the two algorithms are alike, but if the lower envelope has to be determined numerically then the proposed algorithm is lot more efficient.

Section 4.2 outlines the theoretical foundations of 2D Gabor filters [72]. A 2D Gabor filter is defined by 9 parameters, hence the filters used are extremely numerous. It is necessary to restrict the domain in which the filter parameters can change. The results of biological and physical research reduce this space. We propose a method of designing two-dimensional filters [58]. Using theoretical and experimental restrictions, the domain has been narrowed down, but not sufficiently to use them as local descriptors In section 4.2.3, by using the Gentle AdaBoost algorithm and by evaluating the results of the experiments, on a validation set, most characteristic filters could be selected for the image patch analyzed. The aim is to

determine a restricted set of filters that can be used efficiently in localization and detection of object parts.
Section 4.2.4 illustrates the performance of the local descriptor created [53, 56, 58] and the possibility for its use in human eye detection. Taking into account the necessity to decrease processing time, we present a parallel distributed implementation of the training process and detection process for this descriptor.

Deformable Object Model

Deformable object models are coded, in general, by the local visual properties of the object parts and their connection data. In this case an object is defined as a set of parts and a set of interconnections between certain parts. Such a model can be defined by an undirected graph $G = (V, M)$, where $V = \{v_1, v_2, \cdots, v_n\}$ represent the set of vertexes corresponding to the object parts and $M = \{(v_i, v_j)|v_i, v_j \in V\}$ represent the set of edges, i.e. the interconnection between the parts.
In these methods, the solution for the detection is generally the minimization of an energy function [31]. The energy or the cost of a particular configuration depends, on the one hand, on the similarity measure $\mathbf{a}_i$ of the pats v_i and the image patches in corresponding locations x_i, and on the other hand, on the relative distances $\mathbf{d_{ij}}$ of the connected parts.
Using these functions, the dissimilarity measure of the object model and the object images may be defined, thus the cost function is

$$C(X) = \left(\sum_{i=1}^{n} \mathbf{a}_i(x_i) + \sum_{(v_i,v_j)\in M} \mathbf{d}_{ij}(x_i, x_j) \right). \tag{1}$$

The minimization of this function determines the configuration of the object that best fits to the appearance of the object.

Felzenszwalb et al. [28] propose the limitation of the graph structure G for the effectiveness of the calculations, an adequate model for the $\mathbf{d}_{ij}$ functions and an algorithm that takes the $\mathbf{d}_{ij}$ and $\mathbf{a}_i$ functions into consideration, ensuring an effective minimization of the cost function. One possible way to determine the parameters of the model is to formulate it statistically. Statistical formalism facilitates learning model parameters by examples. Basically, all parameters can be determined from the training set using the expectation maximization algorithm.
At the same time, the minimization of the cost functions is equivalent to the maximization of the aposteriori probability of the object configuration in the image match.

After determining the configuration, detection consists of minimizing the cost function defined by the relation (1) and determining its minimum at the same time. This task can be solved by the technique of dynamic programming, in which we have detailed the use of the state tables and the algorithms used.

The complexity of the algorithm is $\mathcal{O}(h^2 n)$. The most significant bottleneck of this algorithm is the minimization process which has to be applied for every possible position h, which can be the order of magnitude of tens of thousands. The minimization function is, in fact, a general distance transform. To compute this we present an original algorithm in section 4.1.2. This algorithm has a $\mathcal{O}(h \log h)$ complexity. It has to be executed for each node or object part separately. Thus the global complexity resulting for the cost minimization is $\mathcal{O}(n \cdot h \log h)$.

Distance Transform

Distance transform is a mathematical operator that computes the distance map of the image. This map is an image as well.

The transform is defined on a set of points $\mathcal{P}$, $D_{\mathbf{a}} : \mathcal{P} \rightarrow \mathbb{R}_+$ as

$$D_{\mathbf{a}}(y) = \min_{x \in \mathcal{P}} M(x,y) = \min_{x \in \mathcal{P}} \left(\mathbf{d}\left(x,y\right) + \mathbf{a}\left(x\right)\right). \tag{2}$$

Proposed Algorithm

In this chapter we suggest an original algorithm for the computation of the general distance transform according to the definition (2). Contrary to known algorithm, this method can be applied in the most general case, taking into account only two restrictions defined by Paglieroni [75]

$$\begin{gathered} \mathbf{d}(x,y) = f(|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|) \\ |x_i - y_i| < |z_i - t_i| \Rightarrow \\ \Rightarrow f(m_1, m_2, \ldots, |x_i - y_i|, \ldots, m_n) < f(m_1, m_2, \ldots, |z_i - t_i|, \ldots, m_n). \end{gathered} \tag{3}$$

The idea of the algorithm is based on the following observation. In order to find the minimum of the measure function (2) in a given point y we have to compute the distance $\mathbf{d}(x,y)$ for each point x. Supposing that in the minimization process, an actual minimum point x_0 has been already found, the points x for which the effective calculation of the measure function is not necessary, are the points where the dissimilarity function $\mathbf{a}(x)$ is greater than the actual value and are located at a greater distance than the actual point

$$\mathbf{a}\left(x\right) > \mathbf{a}\left(x_0\right), \tag{4}$$

$$\mathbf{d}(y, x) > \mathbf{d}(y, x_0), \forall y. \tag{5}$$

The sum of the two functions may in no way be greater than the value of the measure function in the actual point.

The algorithm is based on the order relation of the dissimilarities and distances, and it defines the potential minimum points. Potential minimum points are those in which evaluating the measure function makes sense. The distance transform is going to be the smallest value of the measure function out of the potential minimum points.

This algorithm suggests a recursive implementation. The method presented can be conceived with a binary search tree. In this tree the nodes are introduced in the increasing order of dissimilarities (increasing order of $\mathbf{a}$). The position of the nodes in the binary tree is determined by their position relative to the position of their parents. If all the dissimilarities $\mathbf{a}(1), \ldots \mathbf{a}(n)$ are introduced, then the distance transform is solved through a preorder traversal. The algorithm is implemented according to the pseudocode (algorithm 1)

Algorithm 1 Distance Transform(a,d) – © Sz. Lefkovits

Input: a, d
Output: C_a, argmin(C_a)

```
1: SORT([a,d]) {sort in ascending order by a}
2: tree:=CREATE([a,d](1))
3: for i := 2..n execute
4:    INSERT(tree,[a,d](i)) {insert in ascending order of a,
      with the ordering relation of the nodes according to d }
5: end for
6: (argmin, min):=MINIMTD(tree_root, 1, dim([a,d])) {preorder traversal}
7: return C_a, argmin(C_a)
```

We have compared the proposed algorithm (1) to the algorithm presented by Meijster [70] and implemented by Felzenszwalb et al. [28]. The comparative experiments argue for the algorithm of parabolas in case of distances l_1 and l_2.

Howerver, for a large number of points/nodes ($n > 5000$) the algorithm proposed finds the distance transform faster. The experiment contradicts the "almost linear" [27] execution time $\theta(n)$ of the parabola algorithm which becomes slower than the proposed algorithm $\theta(n \log n)$. Meijster's algorithm supposes the determination of the lower envelope, which is an important factor in the duration of the algorithm. Our algorithm eliminates this inconvenience, reducing the execution time by 1-2 orders of magnitude.

Local Descriptors using Gabor Filters

Rephrasing the ideas presented in the previous chapter, a local descriptor that may be used in a robust detection process is necessary. As we have said the detection process is supervised with the significant object parts labeled. In the first phase, we create a robust classifier based on the global aspect of the object, and in the second phase, we create a classifier based on the local aspect, which serves to eliminate false detections. We intend to create a local descriptor based on the two-dimensional Gabor filter responses.

This section is divided into three parts. The first part presents the theory of the two-dimensional Gabor filters [72] emphasizing the theoretical relations that are at the foundation of filter design. The second part presents the Gentle AdaBoost classifying, highlighting the reason for choosing it and the method it is implemented by. In the last part, based on the theory exposed, a local descriptor resulting from the experiments is determined; we analyze its behavior as a descriptor and we propose possibilities for its application in the deformable object model proposed.

2D Gabor Filters

Representing the images through the decomposition coefficients of the 2D Gabor wavelets suggests filtering the images with the two-dimensional Gabor filter in a well-defined series of points. In fact, the determination of the decomposition coefficients is a result of a convolution process between the image $I(x, y)$ and the Gabor filter $g(x, y)$. The general formula of a two-dimensional Gabor filter is identical to the spatial wave proposed by Daugman [19, 20]

$$g(x,y) = \frac{1}{k} e^{-\pi\left[\frac{(x-x_0)_r^2}{\alpha^2} + \frac{(y-y_0)_r^2}{\beta^2}\right]} e^{i[(\xi_0(x-x_0)+\nu_0(y-y_0))+P]}, \tag{6}$$

where

$$(x - x_0)_r = (x - x_0)\cos\theta + (y - y_0)\sin\theta, \tag{7}$$

$$(y - y_0)_r = -(x - x_0)\sin\theta + (y - y_0)\cos\theta. \tag{8}$$

Thus the two-dimensional Gabor filter is defined by 9 parameters.

- $\frac{1}{k}$: the amplitude of the Gaussian envelope;
- (α, β): the dimension of the axis of the Gaussian envelope;
- θ: the rotation angle of the envelope of the Gaussian;
- (x_0, y_0): the maximum point of the envelope of the Gaussian;

- (ξ_0, ν_0): the spatial frequency of the sinusoidal plane wave;

- P: the phase of the sinusoid.

(9)

While choosing Gabor filters, we have 9 grades of freedom, which makes it difficult to choose the filters to be used. A rational decrease of the parameters' space is necessary in order to effectively use the filter proposed. The theoretical analysis of the frequency domain presented in section 4.2 was made with the goal of finding certain relations and restrictions between the parameters, which leads implicitly to the decrease in dimension of the definition domain and to the theoretical limitation of the possible values for each filter. The relations taken in consideration are:

1. $\frac{1}{k}$ the amplitude of the Gaussian envelope is a multiplication factor, doesn't appear to have any significant role. An arbitrary constant value may be empirically chosen. Because these wavelets are used with the aim of image filtering and used as classification functions, the comparison of their responses becomes necessary. In order to make a realistic comparison, the maximum energy going through the filters needs to be identical. Based on the theory of filtering, it was considered useful that the integral of the Gaussian envelope be equal to the unity, implying $\frac{1}{k} = \frac{1}{\alpha\beta}$.

2. Based on neurobiological assessments, the orientation of the Gaussian envelope and the direction of the major axis θ are considered to be identical to the direction of propagation of the plane wave $\theta = \omega_0$.

3. The ratio of the two axis α and β determines the aspect $S = \frac{\beta}{\alpha}$. Generally, the aspect is considered to be $S \geq 1$, which means that the major axis is smaller than the minor axis of the envelope in the spatial domain.

4. Point (x_0, y_0) is considered to be zero because it is the central point of the Gabor wavelet. This point is taken as the reference point in the expression of the filter. The reference point is going to coincide with the image point where the filter is applied.

5. Phase P of the wavelet can be considered zero, $P = 0$ because the reference point (x_0, y_0) traverses all interest points on the image.

6. The connection relationship of the filter parameters (bandwidth bw, wavelength λ_0 and axis α) is expressed as

$$K_\alpha = \frac{2^{bw} - 1}{2^{bw} + 1}, \tag{10}$$

$$\frac{K_\alpha}{\lambda_0} = \frac{C}{\alpha}. \tag{11}$$

Obtaining

$$\frac{\alpha}{\lambda_0} = \frac{C}{K_\alpha}. \tag{12}$$

This relation shows the fact that for a given bandwidth, the ratio of the major axis and the wavelength λ is constant.

Considering the previous theoretical relations, the number of independent parameters can be reduced to four, namely:

- $\lambda_0 = \dfrac{1}{F_0}$ the wavelength
- S the aspect of the envelope
- bw the bandwidth
- θ the propagation direction of the plane wave.

Experimental Creation of the Local Descriptor

The majority of applications based on Gabor filters use a set of empirically chosen parameters whose scientific argumentation is insufficient. Each author describes the method of using the filters without giving any tangible, detailed data about the parameter domain or the size of the filters used. Commonly, 40 filters are used with 8 orientations and 5 frequencies. This information are not sufficient to reproduce or develop the theoretical analysis or even apply them in our own research. Optimizing of the number of Gabor filters and, implicitly, determining of their parameters is a field researched by several authors [40, 45, 85, 93]. These optimizations consist of reducing the number of filters, choosing a minimal set of parameters and determining the significant locations where applying them is worth it.

In the following section, we shall elaborate on the method of building a local aspect-based descriptor, taking into consideration a certain part of the image. The creation process of the local classifier implies two distinct stages. In the first stage, a database has to be created that contains the images of the local part, and in the second stage we develop a classifier for the corresponding local aspect.

The supervised learning process implies a huge number of object images that must cover almost all appearances of the object as well as possible. In order to create the local descriptor, 730 object images and 2000 background images are used in the training set and 159 object and 500 non-object images in the test set. In practical experiments, we have to affix particular values to each of the parameters. Neurological research limits the domain of values

as follows: aspect S takes values between 1 and 2 and the bandwidth bw between 1.2 and 1.7 octaves.
The propagation angle of the Gabor wavelet is determined from the relations that establish the possibilities for image reconstruction based on filter responses. The maximum number of orientations that are worth taking into account is 24. In this way the orientation θ_l is defined as an integer multiplicative of the minimal rotation angle

$$\theta_0 = \frac{2\pi}{2L} = \frac{\pi}{L}, \text{ where } L \in \{6, 8, 12\}, \tag{13}$$

obtaining $\theta_l = l \cdot \theta_0$ a space of maximum 12 distinct values. The orientations considered are only in the 1st and 2nd quarter because the filters having orientations in the 3rd and 4th quarter are the complex conjugates of these.
For the wavelength the entire discrete space must be considered, with its upper limit established by the filter dimension $R = 16$ and with an accepted discretization error

$$\frac{\lambda}{R} < \frac{\pi}{\sqrt{-\ln 2 \cdot \ln(\varepsilon)}} \cdot \frac{1}{S} \cdot \frac{2^{bw} - 1}{2^{bw} + 1}, \tag{14}$$

while its lower limit is established by the sampling theorem, namely

$$F_{max} < \frac{F_s}{2} = \frac{1}{2} = 0.5\, pixel/cycle, \text{ thus } \lambda > 2. \tag{15}$$

In the experiments determining the local aspect-based classifier, the size the Gabor filter was considered to be 33×33 pixels, and the domain of parameter values

$$\begin{aligned} \lambda &\in \{4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 22\} && \text{- 14 values} \\ S &\in \{0.5; 0.66; 1; 1.2; 1.5; 1.8\} && \text{- 6 values} \\ bw &\in \{1; 1.5; 2\} && \text{- 3 values} \\ \theta &\in \left\{ \frac{l\pi}{12} \middle| l = \overline{0, 11} \right\} && \text{- 12 values.} \end{aligned}$$

Hence, for a single point we have defined a number of $14 \times 6 \times 3 \times 12 = 3024$ filters, and the response of each filter is a complex number. The domain of values is that of the set of complex numbers, where a relation of order does not exist. It is necessary to analyze the type of the value, which must be used for the creation of classifiers based on Gabor filter responses. In our experiments, we have considered it useful to analyze the informational content of 5 value-types of the filter responses: real part, imaginary part, module, argument and statistical distribution.

In conclusion, the space analysed for a single image point has a dimension of $15120 = 3024 \times 5$. The main purpose of the experiments was to decrease this space by choosing only the most significant characteristics.

Next to parameter space, there are several factors which determine the learning process and the preprocessing of the images. The results of the experiments shall coherently delineate the influence of each factor defined in the processes of selecting and deciding on the optimal local descriptor.

The GentleBoost [33] learning algorithm has been used for the analysis and the selection of the most appropriate characteristics. This algorithm, besides the choice of significant characteristics, also provides us with the confidence level of the decision used in the evaluation of the classifier obtained.

In order to decide what the influence of each factor on the performances of the local descriptor is, the experiments were performed under the same conditions, modifying only a specific analyzed factor. The criterion under which the factors are evaluated is the evolution of errors based on the number of classifiers. Error evaluation was done not only on the training set, but on the test set as well.

During the experiments, we have evaluated the following errors:

- The detection error is the ratio of the false detections and the total number of detections, given in percentage.
- The false negative detection rate is the ratio of the falsely evaluated positive images and the total number of positive images, given in percentage
- The false positive detection rate is the ratio of the falsely evaluated negative images and the total number of negative images, given in percentage

Detailed experiments have been carried out for the following factors:

a) informational analysis of the filter responses: These experiments have been done to be able to decide about the method of using filter responses. The real part, the imaginary part, the module and the argument are the values analyzed in the complex Gabor filter responses.

b) aspect of the filters S: The aspect is defined by the ratio of the axis of the Gaussian envelope. Most of the specialized literature considers the aspect to be 1. In order to choose the optimal value for the aspect $S = \frac{\beta}{\alpha}$ (α is the major axis equal to the propagation direction, β is the perpendicular axis on this direction), we have carried out several experiments on the training set and on the test set as well, comparing the decision power of the final classifier.

c) bandwidth bw: The wavelength in the propagation direction relative to filter length is inversely proportional to the filter width; in other words, if the filter width increases the wavelength is decreases. At the same wavelength and bandwidth, the growth of the aspect increases with the filter length.

d) wavelength λ and number of orientations θ: Several 2D Gabor filters may be considered as weak classifiers, with more wavelengths λ and propagation directions θ than the number of parameters computed from the covering condition of the frequency domain with tangential filters. This means that we take in account more wavelengths λ and more propagation directions θ than we have obtained in the theoretical filter design using tangential filters. In the experiments, we have considered more values for these two parameters (wavelength, orientation), which means that the filters are overlapping in the frequency domain. The learning process using the Gentle AdaBoost algorithm has the task to choose the most significant classifiers through which the given local aspect can be detected efficiently.

e) parameters of the learning process: If we take the ratio of positive images and negative images to be 3:1 in the learning process, we shall notice the effect of the inequality of the weights; namely, the false negative detection rate is being considerably reduced in spite of the growth of the false positive rate. The same phenomenon can be observed in the testing phase as well, but the effect of the weights is diminished.

f) histogram equalization and image normalization: The results presented indicate that the equalization function has a significant role. The normalization function has a less effective influence on the process of classifier creation, because the images from the test set present a relatively small dispersion of the intensity value of the points. Nevertheless, for the creation of a detector that is used in real life, these functions contribute to maintaining detection performance.

g) selecting specific interest points (LoG): The number of points in which the filter responses have to be evaluated is a decisive factor on the performance of the object detector. The processing time is directly proportional to the number of points considered. One way of reducing the cardinality of this set of points is using a series of interest points only. We have noticed that in case of labeled points, the learning process as well as the test process becomes more effective. By using the LoG interest point detector, the detection rate increases, yet detection time decreases dramatically.
   Detection performance becomes considerably weaker, but the significant decrease of the detection time suggests that this procedure is amenable to image sequences.

Performance of the Local Descriptor Created

Using the methodology described in the previous chapter, we have obtained the most effective local descriptor for the training and test set from the database of images created. The learning process uses positive images with a weight 3 times greater than that of the negative image weights. The best results for the labeled points have been obtained with image patches that are uniformized by the normalization and equalization processes. A classifier with a high performance is composed of 40 filters with the aspect of S=1.8 and the configuration of the parameters given in table 1.

In the learning process, the false detection rate became 0 just after 28 weak classifiers and the positive images of the training set were complectly separated from the negatives with only 68 classifiers (figure 4). In case of test images, the false positive error rate became zero, and the global detection error was reduced to under 2% with only 32 classifiers (figure 5).

	Classifier	λ	$\theta[°]$	bw
1.	GF_02718	18	15	1.5
2.	GF_02248	10	135	1.0
3.	GF_01574	4	90	2.0
4.	GF_01636	5	15	1.0
5.	GF_02493	14	0	1.5
6.	GF_01555	4	60	1.0
7.	GF_02313	11	60	1.5
8.	GF_02932	22	15	1.0
9.	GF_01825	6	150	1.0
10.	GF_01772	6	60	2.0
11.	GF_01592	4	120	2.0
12.	GF_02275	11	0	1.0
13.	GF_01618	4	165	1.0
14.	GF_02437	12	90	1.0
15.	GF_02932	22	15	1.0
16.	GF_02734	18	45	1.0
17.	GF_01979	8	45	2.0
18.	GF_02347	11	120	1.0
19.	GF_02609	16	15	2.0
20.	GF_01834	6	165	1.0

	Classifier	λ	$\theta[°]$	bw
21.	GF_02086	9	45	1.0
22.	GF_01987	8	60	1.0
23.	GF_01708	5	135	1.0
24.	GF_02933	22	15	2.0
25.	GF_01555	4	60	1.0
26.	GF_02167	10	0	1.0
27.	GF_01592	4	120	2.0
28.	GF_01645	5	30	1.0
29.	GF_02645	16	75	2.0
30.	GF_02140	9	135	1.0
31.	GF_02113	9	90	1.0
32.	GF_02941	22	30	1.0
33.	GF_01681	5	90	1.0
34.	GF_02626	16	45	1.0
35.	GF_02185	10	30	1.0
36.	GF_02815	20	0	1.0
37.	GF_02338	11	105	1.0
38.	GF_02176	10	15	1.0
39.	GF_01915	7	120	1.0
40.	GF_02078	9	30	2.0

Table 1: The most Performant Classifier $S = 1.8$
© Sz. Lefkovits

The descriptor thus created does not only act as a descriptor, but is used as a local patch detector as well. By applying the descriptor in all points of the image, we can obtained the responses map of the descriptor.

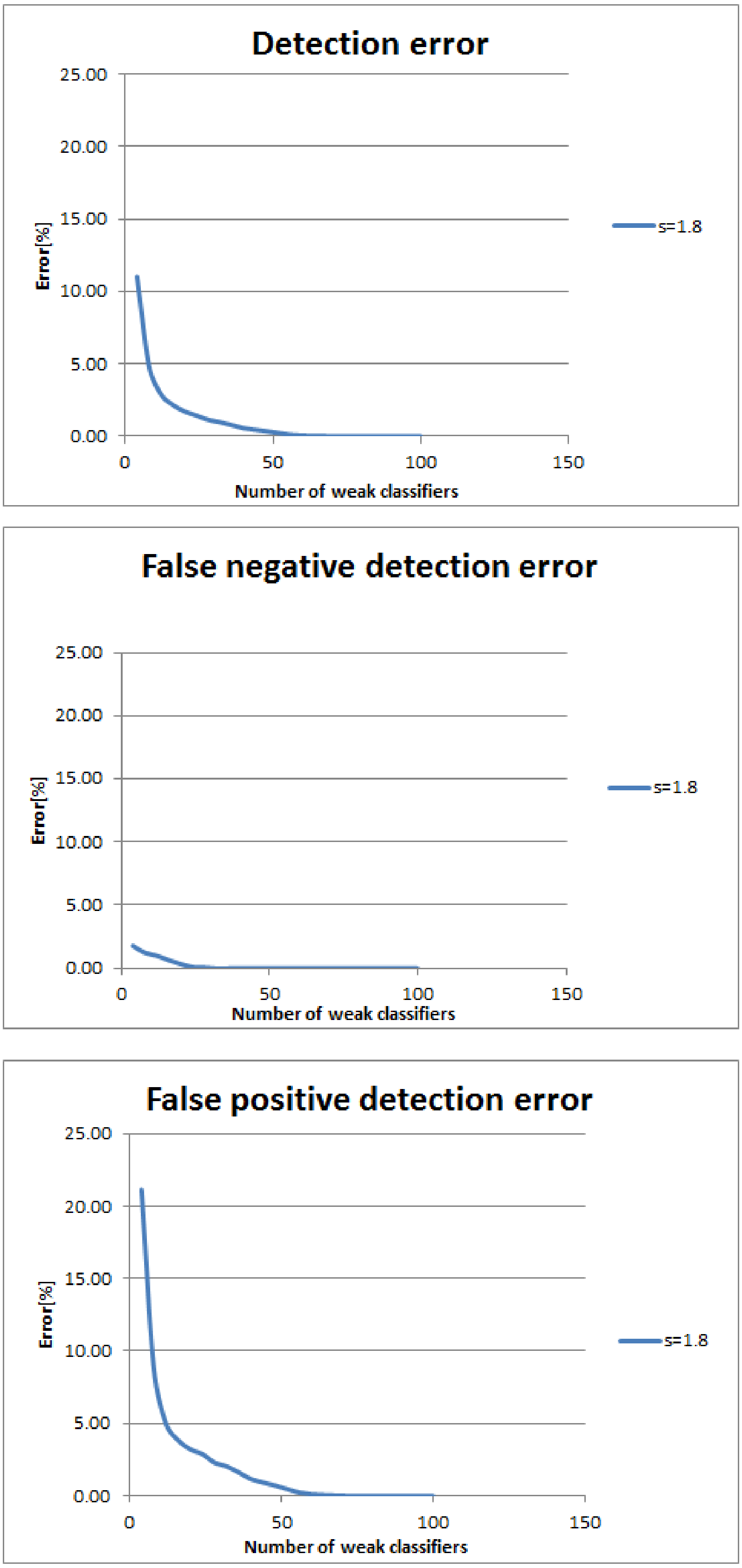

Figure 4: Detection Error Rates for the Training Set of the most Performant Classifier
© Sz. Lefkovits

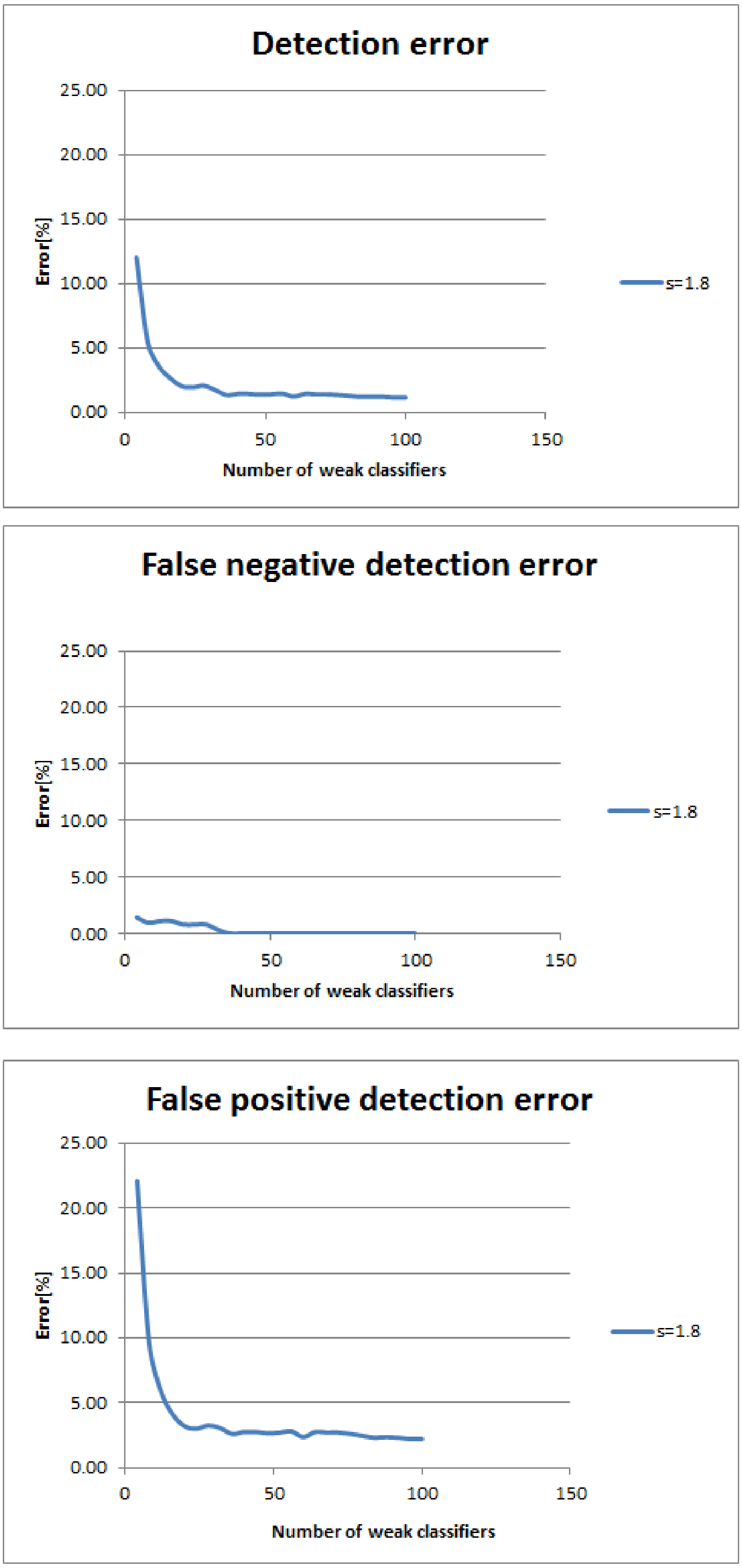

Figure 5: Detection Error Rates for the Test Set of the most Performant Classifier
© Sz. Lefkovits

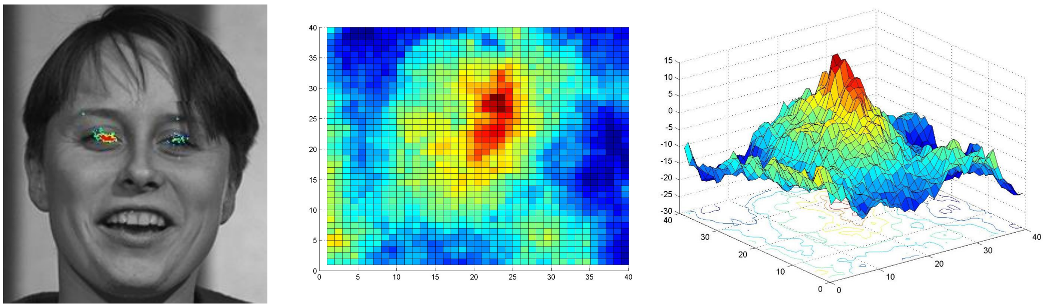

Figure 6: Response Map in the Neighbourhood of the Object
© Sz. Lefkovits

In the first phase, the response map is a surface of the real values provided by the Gentle AdaBoost classification function [33] ($R > 0$ object, $R < 0$ non-object). Applying the algorithm presented in section 4.2.4 to an image we obtain the response map of the descriptor. The filtering process at 4×4 pixels is sufficient for the patch detection and is necessary to decrease detection time by a factor of almost 16. This map is going to be used in the object model with the goal of object detection. The object model determines the possible configurations based on the classifier response eliminating the positive detections that cannot be assigned to the target object. This means that the response map has to be evaluated according to the performance of the local descriptor.

One of the major bottleneck of the descriptor proposed is the long processing time, because in the learning process, every filter in each of the image points form the training set must be evaluated. In the detection process, the number of convolutions depends on the number of filters which form the strong classifier and number of interest points from an image. We have shown that for a correct detection, evaluating the descriptor in a grid of points with 4-pixel steps is enough. With the aim of decreasing processing time, we have implemented the descriptor on a parallel multiprocessor system using a cluster of computers. The application created uses the Matlab [2] program suite. This suit enables parallel processing via the MDCS (Matlab Distributed Computing Server). Figure 7 shows some examples of the most effective descriptor created (table 1 and figures 4, 5) using this system in order to detect the human eye.

The detection system defined in this chapter is a novelty because it uses the 2D Gabor filters as local descriptors. Thorough study shows that a well-chosen set of filters based on informational criteria is able to form a sufficiently general – and in the same time, specific – descriptor.

However, by judiciously delimiting the domain of filter parameter values, the descriptor creation process can be automated without evaluating intermediate performances. During the descriptor creation process, a database of labeled images of the target object parts is necessary. In the learning process, image patches corresponding to the parts have to be rescaled to a standard dimension. In the local aspect base detection process, the images used have the same size. This problem can be solved with multiresolution processing or with a global aspect detector. In the first phase, the system proposed uses a fast global aspect-based detector, thus a simple resizing of the image is enough for the detection process.

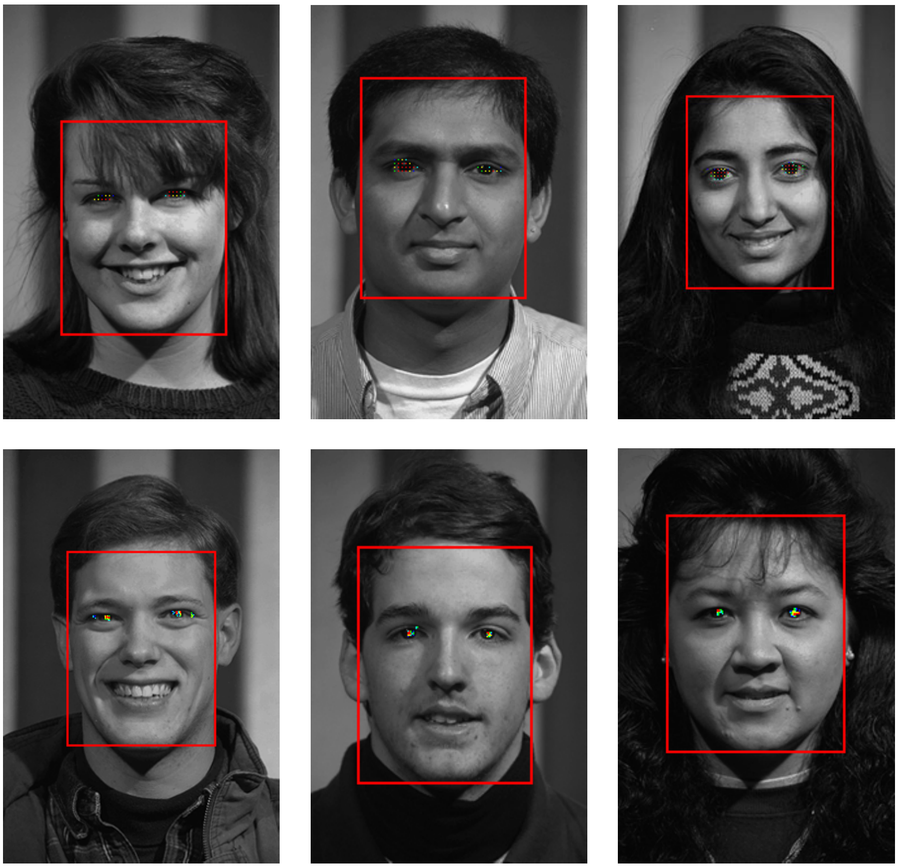

Figure 7: Test Image Examples for the FERET database
© Sz. Lefkovits

The running time of the detection process is determined by the number of filters used and the number of points in which the descriptor is evaluated. For the detection process, we have proposed and implemented a parallel processing system which computes the filter responses. By this parallel processing, detection time can be reduced to the desired time limits. The local descriptor created shows the same properties as the Gabor filters: invariance to translation, rotation and scaling.

The experiments performed in detecting the human eye with the descriptor created show remarkable performance in detection. Hence, such a descriptor is adequate to be integrated into any part-based object detection model.

Conclusions and Feature Work

The goal of this PhD thesis has been to study and create a robust object detection system. The robustness of a detector refers to its performance: the fact that the detected object is present beyond all doubt, and all its instances have been detected. In technical terms, this means a high detection rate and a low false detection rate, i.e. almost zero. For this purpose, a two-component system has been created.

The first component is a global aspect-based detection system that scans the image of interest and uses the image pyramid technique. The traits of this system are high-speed detection and high detection rate, but with a large number of false detections. The detection rate could only be reached via a newly-created positive image database (section 3.4), which takes into account almost all instances of the target object. By using the methodology presented in section 3.4, a proprietary classifier has been created with a higher performance than existing public classifiers. We have obtained a very low false detection rate of of $5 \cdot 10^{-6}$, which is nevertheless insufficient for robust detection. There are still several false zones that are being detected. Analyzing the false images, we can conclude that these images do not resemble the object of interest at all. This is the motivation for the second detection phase, a slower detector that would analyze the details.

The second component is a local aspect-based detection system. This is a slower system, but it processes the images in detail. The system analyzed is made up of two parts: the object model and the local descriptor. The object model is the well known deformable model (section 4.1), however it relies on the information provided by the local descriptor. The deformable object model has an efficient solution using the proprietary algorithm proposed in section 4.1.3.

The effectiveness of the algorithm consists of correctly applying the dynamic programming technique and the rapid evaluation of the general distance transform using an original algorithm (section 4.1.4.1).

The most important factor that determines the performances of a local aspect-based detection system is the local descriptor and the way in which interest points are chosen. The starting idea came from the principles of receptive fields in biological systems. Hence we have suggested a local descriptor based on 2D Gabor filters. In section 4.2.1, we have analyzed the creation of such a descriptor in theory and through experiments. The problem of defining the filter limitations, the method of evaluation, and of effectively creating the local descriptor has been addressed a series of rigorous experiments. The methodology of the approach, the systematic and scrupulous work materialized in a series of successful experiments presented in section 4.2.3. The persuasive results of the experiments outline the possible applicability of the suggested descriptor in the deformable model.

The created detection system is a supervised one that requires an image database covering all the instances of the detectable object. The method of the approach is that the system tries to find an optimum compromise between the global and local aspects, the global and detailed processing, speed and detection performance. The theoretical and experimental analysis presented in the thesis, is a solid foundation for creating an effective object detection system.

The parameters of the learning and detection process can be enhanced by subsequent development. In order to shorten processing time, the execution time of the large number of convolutions that it is determined by must be decreased. This can be done by finding the optimal dimension of the filters or by using only integer coefficients that have an optimized quantization level. Simultaneously, the necessity of using a parallel processing multisystem with judiciously distributed computational tasks appears. Having solved the efficiency of the computations, the creation of a more complex system that could handle object in different points of view, with partial occlusions appearing in different aspects and contexts could be undertaken.

Next to straightforward developments of this thesis, several further research subjects may be brought in as future works:

- creation of detection models which rely on semisupervised or unsupervised learning systems;
- building local descriptors made up of visual information from patches;
- complex study of 2D Gabor filters in the vicinity of interest points;
- creating recognition systems applicable for identification or authentication.

References

[1] www.fotovision.ro. [Online; accessed 2012]. (document)

[2] http://www.mathworks.com/. [Online; accessed 2012]. (document)

[3] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html. [Online; accessed 2012].

[4] http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html. [Online; accessed 2012].

[5] http://coreldraw.com/media/. [Online; accessed 2012].

[6] Bioid database. https://support.bioid.com/downloads/facedb/index.php. [Online; accessed 2012].

[7] P. Addison, J. Walker, and R. Guido. Time–frequency analysis of biosignals. *Engineering in Medicine and Biology Magazine, IEEE*, 28(5):14 –29, september-october 2009.

[8] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition, October 2007.

[9] G. Borgefors. Hierarchical Chamfer matching: A parametric edge matching algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10:849–865, November 1988.

[10] G. Borgefors and I. Nyström. Efficient shape representation by minimizing the set of centres of maximal discs/spheres. *Pattern Recognition Letters*, 18(5):465 – 471, 1997.

[11] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA, 2008. (document)

[12] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167, 1998.

[13] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, June 1986.

[14] O. Cuisenaire. *Distance Transformations: Fast Algorithms and Applications to Medical Image Processing*. PhD thesis, Universite Chatolic de Louvain, 1999.

[15] M. N. Dailey, G. W. Cottrell, C. Padgett, and R. Adolphs. Empath: A neural network that categorizes facial expressions. *J. Cognitive Neuroscience*, 14:1158–1173, November 2002.

[16] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893, 2005.

[17] P. E. Daniellson. Euclidian distance mapping. *Computer Vision Graphics and Image Processing*, 14:227–248, 1980.

[18] J. Daugman. Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94(11):1927 –1935, nov. 2006.

[19] J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847 – 856, 1980. (document)

[20] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 2(7):1160–1169, 1985. (document)

[21] J. G. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(7):1169–1179, 1988.

[22] R. L. De Valois and K. K. De Valois. Spatial vision. *Annual Review of Psychology*, 31(1):309–341, 1980.

[23] H. Eggers. Two fast euclidean distance transformations in z2based on sufficient propagation. *Computer Vision and Image Understanding*, 69(1):106 – 116, 1998.

[24] R. Fabbri, L. D. F. Costa, J. C. Torelli, and O. M. Bruno. 2D euclidean distance transform algorithms: A comparative survey. *ACM Comput. Surv.*, 40:2:1–2:44, February 2008.

[25] H. Farzin, H. A. Moghaddam, and M.-S. Moin. A novel retinal identification system. *EURASIP J. Adv. Sig. Proc.*, 2008, 2008.

[26] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(PrePrints), 2009.

[27] P. F. Felzenszwalb and D. P. Huttenlocher. Distance transforms of sampled functions. Technical report, Cornell Computing and Information Science, September 2004. (document)

[28] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *Int. J. Comput. Vision*, 61(1):55–79, 2005. (document)

[29] R. Fergus. Object class recognition by unsupervised scale-invariant learning. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, 2003.

[30] R. Fergus. *Visual Object Category Recognition*. PhD thesis, Robotics Research Group Departament of Engineering Science, University of Oxford, 2005.

[31] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. In *Computers, IEEE Transactions on*, volume C-22, pages 67–92, 1973. (document)

[32] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, vol. 55, pp. 119-139, Art. No. SS971504,*, 1997. (document)

[33] J. Friedman, T. Hastie, and R. Tibshirani. Additive Logistic Regression: a Statistical View of Boosting. *The Annals of Statistics*, 38(2), 2000. (document)

[34] D. Gabor. Theory of communication. part 1: The analysis of information. *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429 –441, 1946.

[35] C. Galleguillos and S. Belongie. Context based object categorization: A critical survey. Technical Report CS2008-092, 08/2008 2008. (document)

[36] D. Gao. *A Discriminant Hypothesis for Visual Saliency: Computational Principles, Biological Plausibility and Applications in Computer Vision*. PhD thesis, University of California, 2008.

[37] C. Harris and M. Stephens. A Combined Corner and Edge Detection. In *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, 1988.

[38] C. Huang, B. Wu, H. Ai, and S. Lao. Omni-directional face detection based on real adaboost. In *In International Conference on Image Processing*, 2004.

[39] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9):850 –863, sep 1993.

[40] J. Ilonen, J.-K. Kamarainen, P. Paalanen, M. Hamouz, J. Kittler, and H. Kälviäinen. Image feature localization by multiple hypothesis testing of Gabor features. *IEEE Transactions on Image Processing*, 17(3):311–325, 2008. (document)

[41] Y. Jie, Y. Y. fang, Z. Renjie, and S. Qifa. Fingerprint minutiae matching algorithm for real time system. *Pattern Recognition*, 39(1):143 – 146, 2006.

[42] J. Jones and L. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in the cat striate cortex. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 58:1,233–1,258, 1987.

[43] A. Jorgensen. Adaboost and histograms for fast face detection. Master's thesis, University of Stockholm, 2006.

[44] T. Kadir and M. Brady. Saliency, scale and image description. *Int. J. Comput. Vision*, 45(2):83–105, 2001.

[45] P. Kalocsai. Face recognition by statistical analysis of feature detectors. *Image and Vision Computing*, 18(4):273–278, 2000. (document)

[46] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, CVPR*, pages 506–513, 2004.

[47] T. S. Lee. Image representation using 2D Gabor wavelets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(10):959 –971, Oct. 1996.

[48] S. Lefkovits. Performance analysis of face detection systems based on haar features. In *Complexity and Intelligence of the Artifical and Neural Complex Systems*, volume 1, pages 184–192, 2008. (document)

[49] S. Lefkovits. Assessments of building classifiers for face detection. In *International Conference On Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics MACRO*, pages 175–186, 2009. (document)

[50] S. Lefkovits. Teaching improvements on Haar-based classifiers. In *International Conference IETM-3, Petru Maior University, Tîrgu Mureș*, pages 1050–1060, 2009. (document)

[51] S. Lefkovits. Classification mixture for object detection. In *National Conference ZAC Cluj Academic Days*, pages 5–10, 2010. (document)

[52] S. Lefkovits. Algritmi de detecție facială – studiu comparativ. Technical report, Babeș-Bolyai Cluj-Napoca, 2011. (document)

[53] S. Lefkovits. Combining boosted global- and part-aspect face detectors. *Scientific Bulletin of the Petru Maior University*, 8(1):35–40, 2011. (document)

[54] S. Lefkovits. Numerical computation method of the general distance transform. *KEPT 2011 International Conference Knowledge Engineering Principles and Techniques*, 1(1):143–152, 2011. (document)

[55] S. Lefkovits. Numerical computation method of the general distance transform. *Studia Informatica Universitatis Babeș-Bolyai*, 53(2):68–74, 2011. (document)

[56] S. Lefkovits. Hybrid face detector based on boosted classifiers. In *The Eighth Conference of PhD Students in Computer Science, University of Szeged*, volume 1, 2012. (document)

[57] S. Lefkovits. Improvements on Gabor descriptor retrieval for patch detection. *Computing and Informatics*, 2012. (document)

[58] S. Lefkovits. Novel Gabor filter based patch desciptor. In *10th Jubilee International Symposium on Intelligent Systems and Informatics SISY, IEEE Conference*, volume 1, 2012. (document)

[59] S. Lefkovits and C. Enăchescu. Face detection system based on artificial intelligence. In *International Conference ICELM,Petru Maior University Tîrgu-Mureș Computer Science section, Decision Systems, Mathematical Modelling and Statistics*, pages 112–121, 2008. (document)

[60] B. Leibe. *Interleaved Object Categorization and Segmentation*. PhD thesis, Swiss Feredral Instituite of Technology Zurich, 2004.

[61] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. pages 297–304. 2003. (document)

[62] R. Lienhart, L. Liang, and E. Kuranov. A detector tree of boosted classifiers for real-time object detection and tracking. In *IEEE ICME2003*, pages 277–280, 2003.

[63] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, pages 900–903, 2002. (document)

[64] C. Liu. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(5):572 –581, may 2004.

[65] P. M. Lopez. *Object Component Models using Gaobr Filters for Visual Recognition*. PhD thesis, Tehnical University Lisabon, 2008.

[66] R. A. Lotufo and F. A. Zampirolli. Fast multidimensional parallel euclidean distance transform based on mathematical morphology. *Graphics, Patterns and Images, SIBGRAPI Conference on*, 0:100, 2001.

[67] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[68] S. Marcelja. Mathematical description of the responses of simple cortical cells. *J Opt Soc Am*, 70(11):1297–1300, Nov. 1980.

[69] C. R. Maurer, Jr., R. Qi, and V. Raghavan. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25:265–270, February 2003.

[70] A. Meijster, J. B. T. M. Roerdink, and W. H. Hesselink. A general algorithm for computing distance transforms in linear time. In J. Goutsias, L. Vincent, and D. S. Bloomberg, editors, *Mathematical Morphology and its Applications to Image and Signal Processing*, volume 18 of *Computational Imaging and Vision*, pages 331–340. Springer US, 2002. (document)

[71] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, pages 128–142. Springer, 2002. Copenhagen.

[72] J. R. Movellan. Tutorial on Gabor filters. (document)

[73] J. Mundy. Object recognition in the geometric era: A retrospective. pages 3–28. 2006.

[74] M. J. P. Viola. Fast multi-view face detection. Technical Report TR2003-096, Mitsubishi Electric Research Laboratories, Cambridge, July 15, 2003.

[75] D. W. Paglieroni. Distance transforms: Properties and machine vision applications. *CVGIP: Graphical Models and Image Processing*, 54(1):56 – 74, 1992. (document)

[76] C. P. Papageorgiou, M. Oren, and T. Poggio. 1998, a general framework for object detection. *Computer Vision, 1998. Sixth International Conference on*, pages 555–562, 1998.

[77] D. Pollen and S. Ronner. Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411, 1981.

[78] E. Robert. A Brief Introduction to Boosting. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999. (document)

[79] A. Rosenfeld and J. Pfaltz. Distance functions on digital pictures. *Pattern Recognition*, 1(1):33 – 61, 1968.

[80] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. pages 38–44, 1998. (document)

[81] Y. Sato. A Gabor wavelet pyramid based object detection algorithm. In *International Symposium on Neural Networks ISNN*, 2011.

[82] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Mach. Learn.*, 37:297–336, December 1999.

[83] H. Schneiderman. *A Statistical Approach to 3D Object Detection Applied to Faces and Cars*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2000. (document)

[84] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:411–426, 2007.

[85] L. Shen, L. Bai, D. Bardsley, and Y. Wang. Gabor feature selection for face recognition using improved Adaboost learning. In *Proceedings of International Workshop on Biometric Recognition System, in conjunction with ICCV'05*, pages 39–49, 2005. (document)

[86] F. Shih and Y.-T. Wu. The efficient algorithms for achieving euclidean distance transformation. *Image Processing, IEEE Transactions on*, 13(8):1078 –1091, 2004.

[87] T. Tamminen and J. Lampinen. Sequential Monte Carlo for bayesian matching of objects with occlusions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(6):930 –941, june 2006.

[88] M. Toews and T. Arbel. Detection over viewpoint via the object class invariant. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 765–768, Washington, DC, USA, 2006. IEEE Computer Society.

[89] M. Toivanen and J. Lampinen. Incremental bayesian learning of feature points from natural images. *Computer Vision and Pattern Recognition Workshop*, 0:39–46, 2009.

[90] T. Tuytelaars and K. Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Now Publishers Inc., Hanover, MA, USA, 2008.

[91] M. Vidal-Naquet and S. Ullman. Object recognition with informative features and linear classification. In *ICCV*, pages 281–288, 2003.

[92] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision - to appear*, 2002. (document)

[93] D. Vukadinovic and M. Pantic. Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In *IEEE International Conf. on Systems, Man and Cybernetics 2005*, pages 1692–1698, September 2005. (document)

[94] B. Wang. CMU VASC image database. http://vasc.ri.cmu.edu/idb/images/face/frontal_images/. [Online; accessed 2012]. (document)

[95] X. Wang and H. Qi. Face recognition using optimal non-orthogonal wavelet basis evaluated by information complexity. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 164 – 167 vol.1, 2002.

[96] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[97] Y. Zhou, Y. Li, Z. Wu, and M. Ge. Robust facial feature points extraction in color images. *Eng. Appl. Artif. Intell.*, 24(1):195–200, Feb. 2011.