ENHANCED GABOR FILTER BASED FACIAL FEATURE DETECTOR

LEFKOVITS Szidónia, Teaching Assistant,
"Petru Maior" University of Tîrgu Mureş
LEFKOVITS László,
Sapientia University of Tîrgu Mureş

Abstract: The object detection from two-dimensional images plays an important role in several different areas of activity. 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. Locating interest points in images is the most determinative part of object detection. This paper discusses selection methods for choosing the most characteristic set of Gabor filters for facial feature detection. Choosing the most adequate parameters and the contribution of every used filter in the feature response is a high computational complex problem. This paper presents a new descriptor based on 2D Gabor filters and SVM learning algorithm and compares it to the local image descriptor built from the same filters, but using GentleBoost algorithm for classification [1].

Keywords: local descriptor, 2D Gabor wavelets, Gentle Boost, soft margin SVM

Introduction

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 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 [2].

The three main parts of such a system are: the interest points, the local descriptor and the object model. The interest points represent set of points which stand in a way out of their environment; in other words they capture the visual attention. The local descriptor represents a formal description of an image region in the neighborhood of the interest point. The local descriptors, applied in deformable object models, have the advantage of handling small deformation and partial occlusions. Based on the physical structure of the object, the model separates the target objects from other objects.

In this paper the interest and non-interest points are numerically represented using a large set of Gabor filters. From these Gabor responses a descriptor is created based on filters selected by the GentleBoost learning algorithm, on one hand, and by the Support Vector Machines classification method, on the other hand. The classification performance of these systems is measured and compared.

Similar system was proposed in [3] that uses the HOG descriptor and SVM classification for human detection. Other authors define a jet of Gabor filters [4] to obtain the local features. The method proposed [5] defines 48 Gabor filters 6 frequencies and 8 directions for an image patch of 13×13 around the interest point. In [6] the author define a set of Gabor filters optimized for facial feature recognition, using only 4 frequencies and 6 orientations. The selection of filters can also be done by genetic algorithms as proposed in [7]. 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. In our paper we clearly specify the defined Gabor filters. This time the same set of filters is classified using the SVM classification algorithm.

The paper is organized as follows: The first section presents an introduction to local descriptors, the second consists of a theoretical review of the Gabor wavelets and SVM learning algorithm and the third section presents the proposed system and the obtained experimental results.

Theoretical review

Gabor wavelets

In order to determine an adequate image descriptor the first step is to define the interest points of an object or object part. The Gabor wavelets have a wide area of use, especially in bioinformatics systems, because they work similar to the mammalian visual receptive field. Every image can be decomposed using a family of orthogonal wavelets. The Gabor wavelets do not form an orthogonal basis, but in some conditions the reconstruction is possible [8] Nevertheless, the decomposition and reconstructions is computationally very time-consuming. In this paper the aim is not to decompose the image-patch in its wavelet coefficients, but to determine the most suitable Gabor filters which characterize a given image patch. Thus, the Gabor filter describes not only the interest point, corresponding to the center of the image patch, but creates a local image descriptor as well, covering the area of interest.

The 2D Gabor functions are defined as follows [9]

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

where r means the rotation of the envelope surface with θ_0 in trigonometric direction.

The response of the filter in a given point (x_0, y_0) is the Gabor coefficient, which means the convolution of the image $I(x, y)$ with the Gabor filter $g(x, y)$

$$C(x_0,y_0)=\iint I(x_0,y_0)\,g(x_0-x,y_0-y)\,dxdy. \quad (2)$$

The Gabor wavelet is a plane wave modulated by a Gaussian envelope. This function is defined in a 9D parameter space, where $\frac{1}{k}$ the amplitude of the Gaussian envelope; θ_0 the rotation angle of the Gaussian and the plane wave; (α,β) the standard deviation of the Gaussian in 2D; (x_0, y_0) the center of the Gaussian; (ξ_0,ν_0) the spatial frequency of the sinusoidal wave; and P the phase of the wave.

In practice this high dimensionality of parameters can hardly be handled. Due to several theoretical observations in the frequency space the 9 parameters $(k,\theta,\alpha,\beta,x_0,y_0,\xi_0,\nu_0,P)$ can be reduced to only 4, namely (λ,θ,bw,S) , where $\lambda=\frac{1}{F_0}=\sqrt{(\xi_0^2+\nu_0^2)}$ is the wavelength in the frequency domain, the orientation of the wave is $\omega_0=\arctan\left(\frac{\nu_0}{\xi_0}\right)$ and the aspect ratio of the Gaussian $S=\frac{\beta}{\alpha}$.Here we consider that the orientation of the wave and the orientation of the Gaussian are equal $\omega_0=\theta$. Furthermore it is assumed, that the envelope is centered on the coordinate system $(x_0=y_0=0)$ and there is no phase $(P=0)$ between the wave and the envelope.

The relation between λ,α and the bandwidth bw is

$$\lambda=\sqrt{\frac{\pi}{ln2}\alpha\frac{2^{bw}-1}{2^{bw}+1}} \tag{3}$$

and the relation between θ,β and bw is

$$\theta=2\arctan\left(\frac{\lambda}{\beta}\sqrt{\frac{ln2}{\pi}}\right). \tag{4}$$

The above mentioned equations were deduced form the half-magnitude profile in frequency domain.

The domain of each parameter was restricted taking the neurophysical observations described in [9] into account, which considers the orientation sensitivity between 10° and 40°; the aspect ratio grater or equal to 1 and the half-magnitude frequency bandwidth between 1 and 2.

Taking the obtained 4D space we define a considerable number of Gabor filters. Based on these, the system computes the filter response centered on the image patch. In order to choose only the most representative filters and the weight of each one in the final decision a learning algorithm has to be applied. In our last paper we proposed the GentleBoost algorithm for this purpose [1], but if the data is considered to be linearly separable in a given space, then the SVM classification method could be applied as well.

The Support Vector Machines

The Support Vector Machines [10] are supervised learning machines for binary classification problems. They are able to separate the inputs linearly, or if they are not linearly separable they can map them in a higher dimensional feature space, where the linear separation is possible. The learning algorithm finds the best hyperplane, which separates the entities included in the training set. In other words, it finds the hyperplane which maximizes the distance to the nearest entities in each class. The larger the separating margin between the classes the lower is the generalization error of the obtained classifier [11]. The optimization problem consists of maximizing the distance between the closest data points.

Given k $k$inputs in the training data set $\mathrm{S}=\left\{\left(x_i, y_i\right) | x_i \in \mathrm{R}^n,\ y_i \in \left\{\pm 1\right\}\right\}$ for $i=1,2,\ldots,k$.

The classification is achieved by a hyperplane of this form:

$$w^T\Phi\left(x\right)+b=0,\ \text{where}\, w\in\mathrm{R}^n, b\in\mathrm{R} \tag{5}$$

where Φ is the transformation of the inputs in a higher-dimensional space, b is the bias, the translation of the hyperplane from origin and the w is the normal vector of the hyperplane. For each point we obtain a classification value of ± 1. Thus, $w^T\Phi\left(x_i\right)+b\geq +1$, if i is object and $w^T\Phi\left(x_i\right)+b\leq -1$, if i is non-object. This can be reformulated like $y\left(x_i\right)\cdot w^T\Phi\left(x_i\right)+b\geq 1,\ i=1,2,\ldots,k$. The inequality constraints on the margin becomes equality $\left|w^T\Phi\left(x_i\right)+b\right|=1$.

The distance from a point x_i to the hyperplane $\left(\mathrm{P}\right)$ is the length of the perpendicular segment from the point x_i . Or it can be computed as the distance of the projection of x_i x_i on the normal vector of the plane and any point on the plane $\left(x\in\mathrm{P}\right)$. $dist=\left|w^T\left(\Phi\left(x_i\right)-\Phi\left(x\right)\right)\right|=\dfrac{1}{\|w\|}$.

The optimization problem becomes the maximization of the distance

$$\max\left(\frac{1}{\|w\|}\right) \text{ with the following constraint } y_i\left(w^T\Phi\left(x_i\right)+b\right)\geq 1, \tag{6}$$

where $\|w\|^2=w^T\cdot w$ The solution is obtained from the Lagrangian of the problem with respect to k inequality constraints.

$$\mathrm{L}\left(w,b,\lambda\right)=\frac{1}{2}w^T w-\sum_{i=1}^{k}\lambda_i\left[y_i\left(w^T\Phi\left(x_i\right)+b\right)-1\right]. \tag{7}$$

The solution will be given by the quadratic programming.

$$w=\sum_{sp\in SV}\lambda_i y_i\Phi(x_i) \ \text{ and } b=\frac{1}{No_{sp}}\sum_{sp_i}\left(y_{sp_i}-\sum_{sp_j}\left(\lambda_{sp_j}y_{sp_j}\Phi\left(x_{sp_i}\right)^T\Phi\left(x_{sp_j}\right)\right)\right). \tag{8}$$

It can be observed that very few λ are different from 0. The x_i corresponding to $\lambda_i\neq 0$ are the support vectors. The bias b $b$is the average value of the biases, obtained for each support vector from the condition $y_i\left(w^T\Phi\left(x_i\right)+b\right)=1$.

In general, the training data set is not linearly separable; in this case, a transformation function has to be defined, $\Phi:\mathrm{S}\rightarrow\mathrm{S}$. The classification is made in a higher dimensional space. Thus, the separation hyperplane and the support vectors are obtained in space S . It can be observed that the images of the support vectors in the input space S are not necessarily the closest points to the image of the separation plane.

We applied the Gaussian radial basis function kernel [11] is the transformation from the input space to the feature space. The corresponding kernel is

$$\mathrm{K}(x_{sp_i}, x_{sp_j}) = \Phi(x_{sp_i})^T \Phi(x_{(sp_j)}) = exp\left(-\frac{1}{2\sigma^2}\left\|x_{sp_i} - x_{sp_j}\right\|\right). \tag{9}$$

The inputs are not linearly separable, but by allowing some errors (outliers) is also gives a good solution. This type of SVM is called the soft-margin SVM. In this case the margin is violated; there can be two forms of error

- the entity is correctly classified, but it is between the margin and the separation plane;
- the entity from the object class will be on the other side of the separation plane and vice-versa.

The optimization function includes, in this case, the sum of the slack variables and the constraints become less restrictive.

$$\min\left(\frac{1}{2}w^T w + C\sum_{i=1}^{k}\xi_i\right);\ \text{subject to}\ y_i\left(w^T\Phi\left(x_i\right)+b\right) \geq 1-\xi_i\ \text{ and }\ \xi_i \geq 0;\ i = 1,2,\ldots,k\ . \tag{10}$$

C is a constant that represents the weight of the violation regarding the original objective function.

The solution of the new Lagrangian is the same as (8).

$$\mathrm{L}(w,b,\xi,\lambda,?\nu) = \frac{1}{2}w^T w + C\sum_{i=1}^{k}\xi_i - \sum_{i=1}^{k}\lambda_i \cdot \left[y_i(w^T\Phi(x_i)+b)-1+\xi_i\right] - \sum_{i=1}^{k}\mu_i\xi_i\ . \tag{11}$$

Differentiating by ξ the upper bound of the λ_i Lagrange multipliers are determined

$$\frac{\partial \mathrm{L}}{\partial \xi_i} = C - \lambda_i - \mu_i = 0 \text{and} \mu_i \geq 0 \Rightarrow 0 \leq \lambda_i \leq C\ . \tag{12}$$

In case of the soft-margin the support vectors are not only those which support the plane and define the margin, but also the misclassified entities are called support vectors, because they satisfy the constraint in (10). For the marginal support vectors the Lagrange multipliers are $0 \leq \lambda_{sp_i} < C$ and $\xi_{spi} = 0$; and for the non-marginal support vectors (the misclassified entities) $\lambda_{sp_i} = C$ and $\xi_{sp_i} > 0$ (13). This property was used to evaluate the goodness of the obtained separation plane. The number of support vectors for which λ was equal indicated the number of misclassification and the other values, which were distinct, pointed to the actual support vectors of the separation hyperplane.

Experimental results

In our experiments the Gabor wavelets have been used to compute the image features and the SVM algorithm to train and classify the image patches. The parameters of the four dimensional feature-space $\left(\lambda, S, bw, \theta\right)$ have been fine-tuned in the training process in order to find the most adequate filters based on the positive and negative training examples. Similar to our previous paper [1], we have defined 3024 Gabor filters: 14 frequencies, 6 aspect ratios, 3 bandwidths and 12 orientations taking in account the real part, imaginary part, magnitude and the distribution of the complex responses. The experiments have been carried out for the eye extracted from the FERET [12] database. The training set consists of 730 positive and

2000 negative examples and the test set of 160 and 500 patches. The image patch used in the training phase is 33×33 pixels centered on the eye and the negative images have been extracted randomly from the face, but not the eye. In order to compare the performance of the GentleBoost classifier [1] and the SVM the same training and validation conditions have been ensured. In these experiments the SVM classifier has a double role: first it has to extract from the given number of filter responses the most appropriate n, where $<< 3024$, in order to discriminate the eye from other facial features; secondly it has to obtain the optimal separating plane of the two classes.

The 2D responses of every Gabor filter is classified by its own SVM, determining the optimal separating line between positive and negative responses. We can observe that the negative responses are concentrated around the origin, and the positives are situated further. In 2D space an errorless linear classification of these responses is not possible, thus we applied the soft-margin version of the SVM. In the first step the aim is not to obtain the final decision of the responses with only classifier, but to compare the Gabor responses and the performance of the obtained classifiers.

No cl.	λ	θ_0	bw
1	18	15	1.5
2	10	135	1.0
3	4	90	2.0
4	5	15	1.0
5	14	0	1.5
6	4	60	1.0
7	11	60	1.5
8	22	15	1.0
9	6	150	1.0
10	6	60	2.0
11	4	120	2.0
12	11	0	1.0
13	4	165	1.0
14	12	90	1.0
15	22	15	1.0
16	18	45	1.0
17	8	45	2.0
18	11	120	1.0
19	16	15	2.0
20	6	165	1.0
21	9	45	1.0
22	8	60	1.0
23	5	135	1.0
24	22	15	2.0
25	4	60	1.0
26	10	0	1.0
27	4	120	2.0
28	5	30	1.0
29	16	75	2.0
30	9	135	1.0
31	9	90	1.0
32	22	30	1.0

Table 1 Weak classifiers GentleBoost

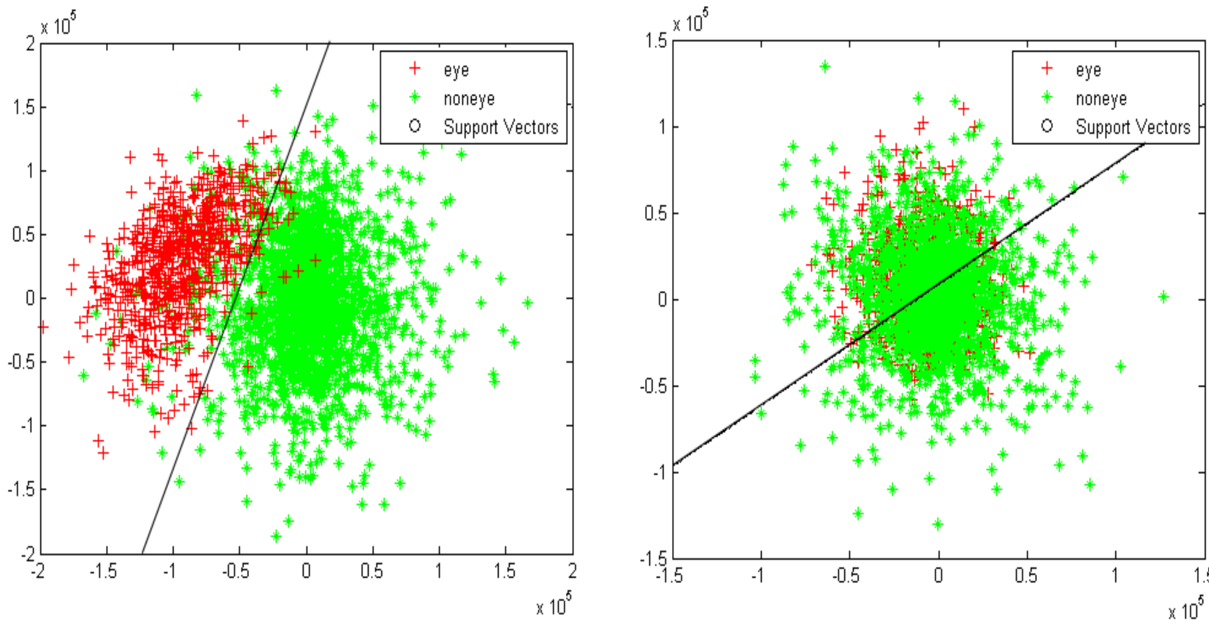
No cl.	λ	θ_0	bw
1	18	15	2.0
2	20	15	2.0
3	20	15	2.0
4	18	15	2.0
5	20	15	2.0
6	18	15	2.0
7	22	15	2.0
8	22	15	2.0
9	22	15	2.0
10	18	0	2.0
11	18	15	1.5
12	18	15	1.5
13	18	15	1.5
14	16	0	2.0
15	18	0	2.0
16	20	15	1.5
17	20	15	1.5
18	16	0	2.0
19	18	0	2.0
20	20	15	1.5
21	16	0	2.0
22	16	15	1.5
23	16	15	1.5
24	16	15	2.0
25	20	0	2.0
26	16	0	1.5
27	16	0	1.5
28	22	15	1.5
29	16	15	2.0
30	22	15	1.5
31	16	0	1.5
32	20	0	2.0

Table 2 Weak classifiers SVM

The generalization error can be bounded upperly by the number of support vectors [13].

Thus, the goodness of Gabor filters can be determined based on the number of resulting support vectors

Taking relation (13) this into consideration we computed the number of support vectors for all the filters. The most adequate 32 weak classifiers were selected to create a classifier formed of them. Their parameters are presented in Table 2. These these can be compared to the best 32 weak classifiers obtained in our previous work [1] using the



a. the most separable filter responses **b. linearly not separable filter responses**

Figure 7

GenteBoost algorithm (Table 1).

Figure 1 shows a separable filter response and its optimal separation line versus an inseparable feature. We can observe that the two classes cannot be 100% linearly separated, but if we admit some outliers, then the SVM can compute the optimal separation line.

Taking this into account we are able to evaluate the classification error. This error is related to the number of support vectors. In Figure 1a there are very few support vectors and in Figure 1b the number of support vectors is huge.

In order to achieve a good detection rate, with as few misses as possible we have to create a descriptor formed of more than one filter. Based on the goodness of filters, measured with the obtained number of support vectors, a jet of n Gabor filters is built. The n Gabor responses are separated with the SVM as well, but in a high order space. Here, instead of the real and imaginary part of the filter response, the value of the magnitude is the input.

Thus, the positive and negative response magnitudes are separated in a n dimensional space.

Due to the measurements we can conclude that the linear separation in this high dimensional space is less performant than using the Gaussian RBF kernel function – relation

(9). Table 3 presents the obtained detection error, false positive error and false negative error for linear hyperplane and RBF kernel, using $n = 32$ dimensions.

	Linear SVM in 32D space	RBF kernel SVM in 32D space
ErrD[%]	5.83	3.22
ErrFP[%]	4.17	1.98
ErrFN[%]	7.65	4.5

Table 3 Comparison of linear and non-linear kernels

We can observe that if we use a complex kernel, which separates the inputs of the training set more accurately. If the separation surface overfits then the generalization error will increase. It is much better to admit outliers and try to obtain the separation plane minimizing the distance and the errors.

Comparing the order of the space, in other words the number of classifiers included in the SVM, we can draw the following conclusions: the error rates are decreasing with the increase of the filters (Table 4). But after only 32 filters the detection error becomes sufficiently stable.

No. of cl.	16	32	48
ErrD[%]	4.16	3.22	2.12
ErrFP[%]	2.6	1.98	1.23
ErrFN[%]	5.87	4.5	3.1

Table 4 Performance of SVM depending on the no. of classifiers

The binary classifier obtained in this way can be compared to our previous work (Table 5) where the selection and classification of the Gabor filters was made by the GentleBoost algorithm. As you can observe the error rates in this case are almost the same.

No. of cl.	16	32	48
ErrD[%]	2.64	1.71	1.36
ErrFP[%]	1.10	0.31	0.23
ErrFN[%]	4.10	3.05	2.61

Table 5 Performance of GenteBoost depending on the no. of classifiers

In the space where inputs are the linearly separable the distance of one entity to the hyperplane can be considered as a similarity measure. As far it is form the margin as better the classification is. Appling the same classifier to every point of the image a response map can be created, which localizes the target object part more accurately.

The advantage of the SVM algorithm is the faster training and especially validation process. In the training phase the ranking of the obtained features has to be made only once, not in each loop as for the GentleBoost. The validation process is only a simple evaluation, verifying if the entity is on the left or right side of the hyperplane. The only two bottlenecks of the SVM algorithm is assuming that the training data is linearly separable. But if it is not the case, we can admit some outliers. The other is finding the minimum of the quadratic programming problem. Generally the minimum is computed iteratively until it converges. The iterative algorithms not necessarily find the global minimum. If a separating hyperplane is found and the detection rate is high enough then the hyperplane can be considered optimal. Figure 2 illustrates the detection comparing the two methods on a FERET [12] database image.

a. Example image GentleBoost Algorithm

b. Example image SVM Algorithm

Figure 8 Detection example

Conclusion and future work

This paper presents a facial feature detector based on Gabor filter responses and SVM classifier. It is put in parallel with a previously presented method [1] using the same filters but another classification algorithm. The classification performance of the two algorithms is approximately the same. But the advantage of SVM is the less computational complexity and a shorter processing time for training but especially for validation. Accordingly, the detector obtained is very robust and presents high detection accuracy. As for the future we propose to compare several kernels for the nonlinear transformation of input entities in a higher dimensional feature space. Besides we intend to apply the same algorithm for several object parts in the same time in order to obtain a constellation of parts in the deformable object model.

References

[1] Szidonia Lefkovits, "Novel Gabor filter-based patch descriptor," *IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics*, vol. 1, no. 1, pp. 373-378, 2012.

[2] Szidonia Lefkovits, "Robust Object Detection Methods with Applications in Face Detection," *PhD Thesis*, 2012.

[3] Navneet Dalal and Bill Triggs, "Histograms of Oriented Gradients for Human Detection," pp. 886-893, 2005.

[4] Toni Tamminen and Jouko Lampinen, "Sequential Monte Carlo for Bayesian Matching of Objects with Occlusions," vol. 28, no. 6, pp. 930-941, 2006.

[5] Danijela Vukadinovic and Maja Pantic, "Fully automatic facial feature point detection using Gabor feature based boosted classifiers," vol. 2, pp. 1692-1698, 2005.

[6] J. Ilonen, J. -K Kamarainen, P. Paalanen, M. Hamouz, and J. and Kalviainen, H. Kittler, "Image Feature Localization by Multiple Hypothesis Testing of Gabor Features," vol. 17, no. 3, pp. 311-325, 2008.

[7] Xiaoling Wang and Hairong Qi, "Face Recognition Using Optimal Non-Orthogonal Wavelet Basis Evaluated by Information Complexity," pp. 10164-10168, 2002.

[8] Tai Sing Lee, "Image Representation Using 2D Gabor Wavelets," pp. 959-971, 1996.

[9] Javier R. Movellan, "Tutorial on Gabor filters".

[10] Vladimir Vapnik and Corinna Cortes, "Support-Vector Networks," vol. 20, no. 3, pp. 273-297, 1995.

[11] Nello Cristianini and John Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*.: Cambridge University Press, 2000.

[12] *FERET Database*., http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.

[13] Christopher J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," vol. 2, no. 2, pp. 121-167, 1998.