FACE DETECTION AND TRACKING IN DYNAMIC BACKGROUND OF STREET

Jacek Naruniec, Władysław Skarbek
Faculty of Electronics and Information Technology, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warszawa, Poland

Antonio Rama
Dept. Teoria del Senyal i Comunicacions, Universitat Politécnica de Catalunya (UPC), Barcelona, Spain

Keywords: Face detection, face tracking, Gabor filter, Linear Discriminant Analysis, Dual Linear Discriminant Analysis, reference graph.

Abstract: The paper presents a novel face detection and tracking algorithm which could be part of human-machine interaction in applications such as intelligent cash machine. The facial feature extraction algorithm is based on discrete approximation of Gabor Transform, called Discrete Gabor Jets (DGJ), evaluated in edge points. DGJ is computed using integral image for fast summations in arbitrary windows and by FFT operations on short contrast signals. Contrasting is performed along radial directions while frequency analysis along angular directions. Fourier coefficients for a small number of rings create a feature vector which is next reduced to few LDA components and then compared to the reference facial feature vector. Detected eyes and nose corners are chosen to fit reference face by spatial relationships. Tracking is based on the same rule, but the corners are searched only within already detected facial features neighborhood. Optionally for face normalization eyes centers are found as centers of outer and inner eye corners. Comparison of manual and automatic eye center detection shows still significant advantage of manual approach, measured in terms of accuracy in face recognition by Linear Discriminant Analysis (LDA) and Dual Linear Discriminant Analysis (DLDA) algorithms.

1 INTRODUCTION

Face detection is important preprocessing task in biometric systems based on facial images. The result of detection gives the localization parameters and it could be required in various forms, for instance:

- a rectangle covering the central part of face;
- a larger rectangle including forehead and chin;
- eyes centers (the choice of MPEG-7 (MPEG-7, 2004));
- numerous landmarks including eyes, nose and mouths corners, eyebrows, nostrils;
- two eyes inner corners and two nose corners (the choice of this paper).

While from human point of view the area parameters are more convincing (cf. Fig. 1), for face recognition system fiducial points are more important since they allow to perform facial image normalization – the crucial task before facial features extraction and face matching (Beumer et al., 2006).

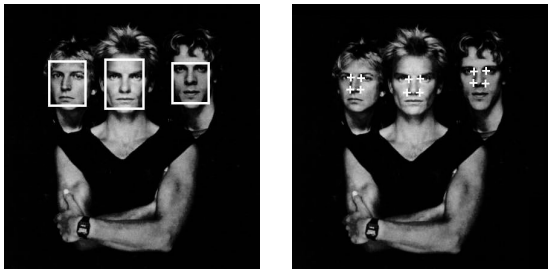


Figure 1: Area versus fiducial points parameters for face localization.

We may observe for each face detector the following design scheme (Yang et al., 2002):

1. *Define admissible pixels and their local neighborhoods of analysis.* Admissible pixels could cover the whole image area or its sparse subset, for instance edge or corner points. The neighborhoods could consist of one pixel only or even thousands of them forming rectangular windows of analysis or for instance rings of small rectangles (cf. Fig.

Naruniec J., Skarbek W. and Rama A. (2007).
FACE DETECTION AND TRACKING IN DYNAMIC BACKGROUND OF STREET.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 404-410
DOI: 10.5220/0002142404040410

2).

2. *Design a feature extractor* which produces a collection of features for each admissible local neighborhood. It may be as simple as admissible pixel color frequency or as complex as long vector of 100 Angular Radial Transformation (ART) coefficients.
3. *Design a classifier* which decides whether the collection of features extracted from the given neighborhood of analysis could be face relevant. If so the admissible pixel becomes *face relevant point.* It could be a simple classifier based on comparison of feature with a threshold or more complex Support Vector Machine (SVM classifier) using Gaussian kernel.
4. *Define a post-processing scheme* which selects representative face relevant points defining face locations. The representatives could be obtained as centroids of connected components in the set of all face relevant points or results of more complex clustering scheme combined with graph matching to reject inconsistent ensembles of face relevant points.

On top of the above scheme each detector includes a multi-resolution mechanism to deal with face size. It is implemented either through analysis in image pyramid or by scaling the local neighborhood of analysis together with relevant parameters.

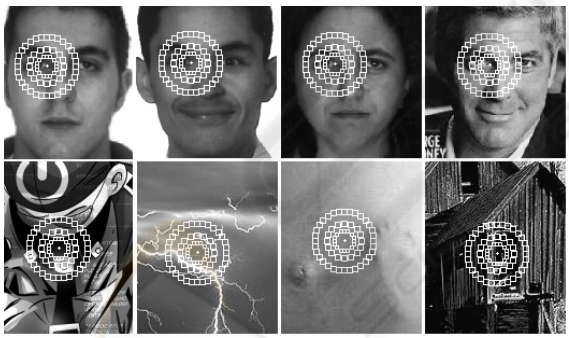


Figure 2: Rings of small squares as neighborhoods of analysis in our method.

One of the most known face detectors is based on AdaBoost classifier. It was introduced by Viola and Jones in 2001 (Viola and Jones, 2001). Let us trace the design scheme of this prominent method:

1. The local neighborhood of analysis is a small window of size 20×20 scaled up by 20%. The pixel is admissible if and only if it is upper left corner of the analysis window completely included in image domain.
2. In analysis window at fixed positions small contrasting filters are defined of specific size and type. The filter returns the contrast between white and black region defined as the difference between total intensities in the regions.
3. The regional contrast is compared with filter specific threshold giving a weak classifier. The weak decisions are linearly combined using cost coefficients elaborated according the AdaBoost machine learning scheme. The AdaBoost is a multi-classifier well known from the late 1980s which due a special weighting scheme of training examples ensures the high performance of strong classifier providing that weak classifiers have the success rate about 0.5. Authors of (Viola and Jones, 2001) applied an early and suboptimal heuristics given for AdaBoost training algorithm in (Freund and Schapire, 1997). However, their face recognition system described in (Jones and Viola, 2003) which also used the AdaBoost concept, contained the optimal training procedure which is methodologically sound. The algorithm proposed by them is a generalization of one described in (Shapire, 2002).
4. In post-processing stage the centroid of enough large connected components of face relevant window corners represents the detected face window.

While AdaBoost is satisfactory solution for facial window detection, its extensions to detect fiducial points, for instance eye centers, are not equally effective. The normalization of facial image based on AdaBoost is not accurate and it results in poor face recognition and verification. In this paper we develop a novel method for detection of face fiducial points which is based on very rough discrete approximation of Gabor transform called here Discrete Gabor Jet (DGJ). The method gives very good results for detection of frontal face views with almost perfect false acceptance rate.

In practice when we deal with a temporal sequence of images face detection cooperates with face tracking. There are many techniques for object tracking in video. However, having robust face detector the simple practical approach is its local use for each image frame interleaved by global face detector called with the frequency proportional to motion activity of tracked objects.

2 DGJ FACE DETECTOR

The whole process of face detection consists of several steps illustrated in Fig. 3. The fiducial points are

searched only within edge points in the image. Canny edge detector's low and high thresholds must be set to a very small value, that even with poor lighting conditions, no feature's edges could be discarded.

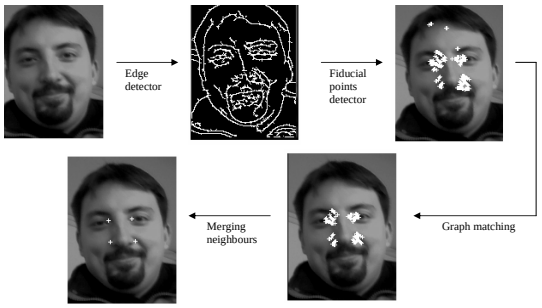


Figure 3: Illustration of all steps for our face detector.

The feature extraction in local neighborhood of analysis is performed in two stages. Firstly the frequency analysis is performed on selected rings (first type coefficients) and on contrasts between pairs of rings (second type coefficients). In the second stage a modified LDA analysis produces many-dimensional features discriminating face and non-face fiducial points. Each type of fiducial point has its specific LDA matrix.

The classifier for the particular fiducial point is based on distance to the centroid of fiducial feature vectors. The standard ROC is built up to tune the distance threshold to the required false acceptance rate.

2.1 Extractor I: Discrete Gabor Jet

The kernel of Gabor filter (Gabor, 1946) in spatial image domain is a Gaussian modulated 2D sine wave grating with parameters controlling wave front spatial orientation, wave frequency and rate of attenuation. While Gabor wavelet is accurate tool to represent local forms with complex textures its computation excludes real time applications in pure software implementations.

Therefore we seek for the representation which can describe changes of local image contrasts around the given pixel in both angular and radial direction. To this goal we design rings of small squares and evaluate frequency of luminance changes on such rings (cf. Fig. 2).

There are two types of Gabor jets. The first type detects angular frequency on selected rings while the second type represents angular frequencies for radial contrast between two selected rings.

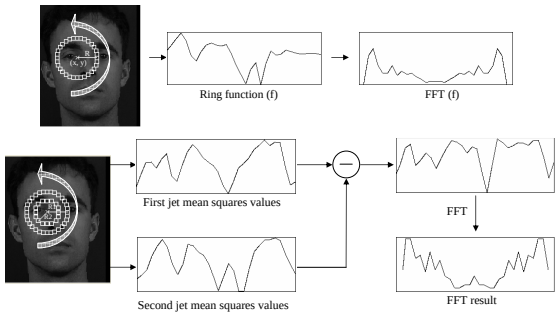


Figure 4: First type jet (up) – frequency features on single ring of squares, and second type jet – frequency features of double ring of squares.

First type jets. The concept is illustrated in Fig. 4 in upper part.

Each jet of the first type is characterized by radius r of the ring, the number of squares $n = 2^k$, the center (anchoring) point (x,y). The sizes of all squares on the ring are equal. The size is the maximum possible providing that squares do not intersect each other (except of intersection perhaps by one pair).

The sum of pixel values in each square is computed using the integral image like in AdaBoost detector (Viola and Jones, 2001). The sequence of n such values is normalized to be included in the unit interval $[0,1]$. Finally the obtained sequence f is transformed by FFT. Only the first $n/2$ of DFT complex coefficients are joined to the output feature vector.

Second type jets. The idea is shown in Fig. 4 in bottom part.

Now the jet consists of two rings with radii $r_1 < r_2$ with the same center (x,y) and with the same number $n = 2^k$ of equal size squares.

Like for the first type jets the sum of pixel values in each square is computed using the integral image, but now the mean value of each square is computed. Differences between mean values of each square in the first ring and the corresponding mean values in second ring are taken. Next the obtained differential signal is normalized to the unit interval and then transformed by FFT. Again only the first $n/2$ of DFT complex coefficients are joined to the output feature vector.

In the final design of Gabor jets for base points detection and tracking (two inner eye corners and two nostril corners) we take four jets of the first kind and two jets of the second with parameters defined in the following table:

After the first stage of feature extraction, the feature vector has $3*16+3*32 = 144$ of real compo-

Table 1: Jets for base points detection.

type	1	1	1	1	2	2
n	16	16	32	32	16	32
r_1	16	24	12	19	16	12
r_2	–	–	–	–	24	19

nents.

For outer eyes corners we cannot use whole rings, which would cover also area outside the face. In this case we use only right and left halves of the jets to analyze only the desired face region. Because the information is now smaller, we have to use more rings to achieve good detection results.

2.2 Extractor II: Modified Linear Discriminant Analysis

Having DFT description over local rings as feature vector of length 240 we discriminate them using a Modified Linear Discriminant Analysis (MLDA).

In case of face detection when we deal with two classes only, i.e. with *facial descriptions* and *non-facial descriptions*, the classical LDA enables only scalar discriminative feature. It makes harder separation of two classes by linear approach. More information about classified region is required. Therefore we modify the concepts of within and between-variances and related scatter matrices, in order to get vectorial discriminative features.

Namely, the classical LDA maximizes the Fisher ratio of between-class variance over within-class variance defined as follows ((Fisher, 1936),(Fukunaga, 1992)):

$$\begin{aligned} f_X &:= \frac{\mathrm{var}_b(X)}{\mathrm{var}_w(X)} \\ \mathrm{var}_w(X) &:= \frac{1}{|I_f|}\sum_{i\in I_f}\|x_i-\bar{x}^f\|^2+\frac{1}{|I_{\bar{f}}|}\sum_{i\in I_{\bar{f}}}\|x_i-\bar{x}^{\bar{f}}\|^2 \\ \mathrm{var}_b(X) &:= \|\bar{x}^f-\bar{x}\|^2+\|\bar{x}^{\bar{f}}-\bar{x}\|^2 \end{aligned} \tag{1}$$

where the training set X of feature vectors is divided into the facial part indexed by I_f and the non-facial part with remaining indices $I_{\bar{f}}$.

It appears that we obtain better discrimination results with the following class separation measure:

$$\begin{aligned} m_X &:= \frac{\mathrm{mvar}_b(X)}{\mathrm{mvar}_w(X)} \\ \mathrm{mvar}_w(X) &:= \frac{1}{|I_f|}\sum_{i\in I_f}\|x_i-\bar{x}^f\|^2 \\ \mathrm{mvar}_b(X) &:= \|\bar{x}^f-\bar{x}\|^2+\frac{1}{|I_{\bar{f}}|}\sum_{i\in I_{\bar{f}}}\|x_i-\bar{x}^f\|^2 \end{aligned} \tag{2}$$

Like in classical case, the optimization procedure requires replacing variances by traces of scatter matrices:

$$\begin{aligned} m_X &:= \frac{\mathrm{trace}(S_{mb}(X))}{\mathrm{trace}(S_{mw}(X))} \\ S_{mw}(X) &:= \frac{1}{|I_f|}\sum_{i\in I_f}(x_i-\bar{x}^f)(x_i-\bar{x}^f)^t \\ S_{mb}(X) &:= (\bar{x}^f-\bar{x})(\bar{x}^f-\bar{x})^t+\frac{1}{|I_{\bar{f}}|}\sum_{i\in I_{\bar{f}}}(x_i-\bar{x}^f)(x_i-\bar{x}^f)^t \end{aligned} \tag{3}$$

Since for the large number of positive training examples the scatter matrix S_{mw} is of full rank then we can optimize $m(W) := m_{W^tX}$ w.r.t. the training set X by Cholesky factorization ((Golub and Loan, 1989)) of $S_{mw} = C_{mw}C_{mw}^t$ and solving the following EVD problem for the symmetric, semi-definite matrix $S'_{mb} := C_{mw}^{-1}S_{mb}C_{mw}^{-t}$:

$$\begin{aligned} &S_{mb}W = \lambda S_{mw}W,\ S_{mw} = C_{mw}C_{mw}^t,\ W' = C_{mw}^tW \\ &S'_{mb} = W'\Lambda(W')^t \\ &W = C_{mw}^{-t}W' \end{aligned} \tag{4}$$

If columns of W' are sorted by decreasing eigenvalues λ_i and we want to have $n \geq 2$ LDA features then we select from W the first n columns as the optimal solution.

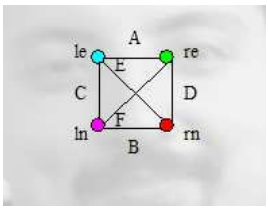
2.3 Post-processing: Graph Matching

Let fiducial facial points be depicted according the notation from the Fig. 5. All detected fiducial points to be preserved must pass, for at least one scale, through the graph matching procedure with the following pseudocode.

```
forall left eyes le do
  forall right eyes re do
    if distance(le.Y,re.Y)<15*scale
       and 30*scale>(re.X-le.X)>20*scale
       forall left nostrils ln do
         if distance(le.X,ln.X)<30*scale
         and 35*scale>(ln.Y-le.Y)>20*scale
           forall right nostrils rn do
             if distance(re.X,rn.X)<30*scale
             and 35*scale>(rn.Y-re.Y)>20*scale
               set distance(le,re):=norm;
               normalize other distances:
                 distancen:=distance/norm;
               get total_distance as the sum of distances
                 between actual graph and reference graph;
               if total_distance<threshold
                  consider points as detected face;
           endfor
       endfor
  endfor
endfor
```

This pseudocode could be shortly explained by the statement, that we look for the left eye corner to the left from the right eye corner, the nose corner at the

bottom of eyes and so on. All the points that fulfill these constraints are then normalized by the distance of the eye corners, and then compared to the reference graph. If the Euclidean distance of both graphs is below specified threshold, points are considered as face features.



edge	length
A	1.00
B	1.05
C	0.99
D	0.98
E	1.41
F	1.43

Figure 5: Fiducial points and their symbols used in graph matching algorithm: le - left eye, re - right eye, ln - left nostril, rn - right nostrils corners.

Reference graph is trained by averaging normalized distances in the set of many faces. It is interesting that these distances between fiducial points indicate their displacement in the corners of a square.

2.4 Face Tracking

Because of face detection takes larger amount of time (about 80 ms for CIF image, depending on number of detected edges) it cannot be directly used in real time applications. To achieve detection at about 25 frames per second a tracking algorithm has been proposed. Instead of continuous detecting facial features, we can run face detection once per larger period of time (ie. one second), and in the meantime search for facial features only in the closest neighborhood of points already found, as shown in Fig. 6. Tracking is identical as detection, but because of much smaller search region and known face scale, it is much faster.

Figure 6: Search areas for tracking of two nose and two eyes inner corners.

In our final implementation there are two coexisting threads shown in Fig. 7. Detection thread is responsible for detection of the new faces in the image. Tracking algorithm can discard and modify previously found fiducial points. It also has an option to find eyes centers in the middle of both eye corners.

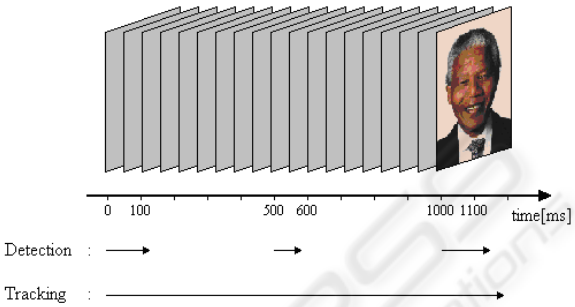


Figure 7: Detection and tracking threads.

3 EXPERIMENTAL RESULTS

All experiments have been performed in Visual C++ environment, running on 64bit Athlon 3500+ processor. ROC for individual base points (inner eyes corners and nose corners) have been made for Mpeg, Banca and BioID database and presented in Fig. 8. As we mentioned before these features are next applied to reference graph and false acceptance and false rejection rates of the final detector are much lower than in separate facial points detectors.

We have tested detection algorithm for face recognition by Linear Discriminant Analysis (Skarbek et al., 2004) and Dual Linear Discriminant Analysis (Kucharski, 2006). Database contained nearly frontal pictures of three databases:

- Mpeg (3175 pictures, 635 people),
- AltKom (1200 pictures, 80 people),
- Yale (165 pictures, 15 people).

One test was performed on manually, and the second on automatically selected eyes centers. Result of such analysis is presented in Fig. 9. Image ROC was created for face recognition based on single image. In person ROC, for every recognition, a set of pictures have been taken. Equal error rate pointed in figures determines threshold at which false acceptance and false rejection rate becomes equal.

We have also verified the algorithm in various environments that simulate real cash machine conditions. It appears, that DGJ algorithm has almost perfect false acceptance rate while achieving very good

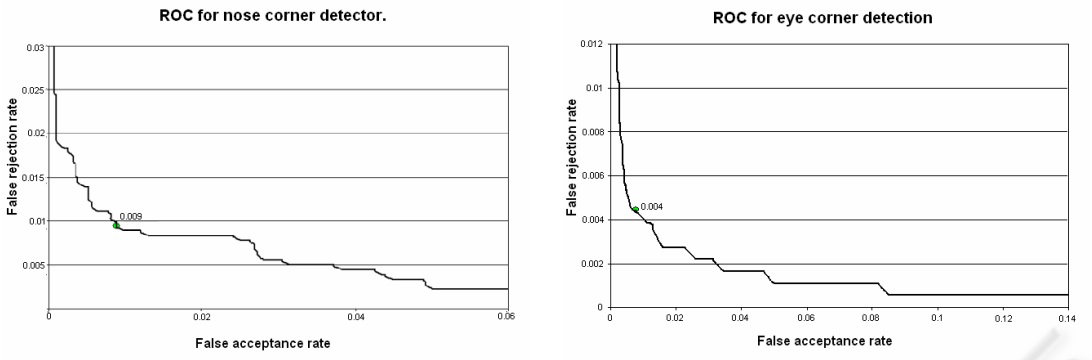


Figure 8: ROC for eye and nose corner detectors.

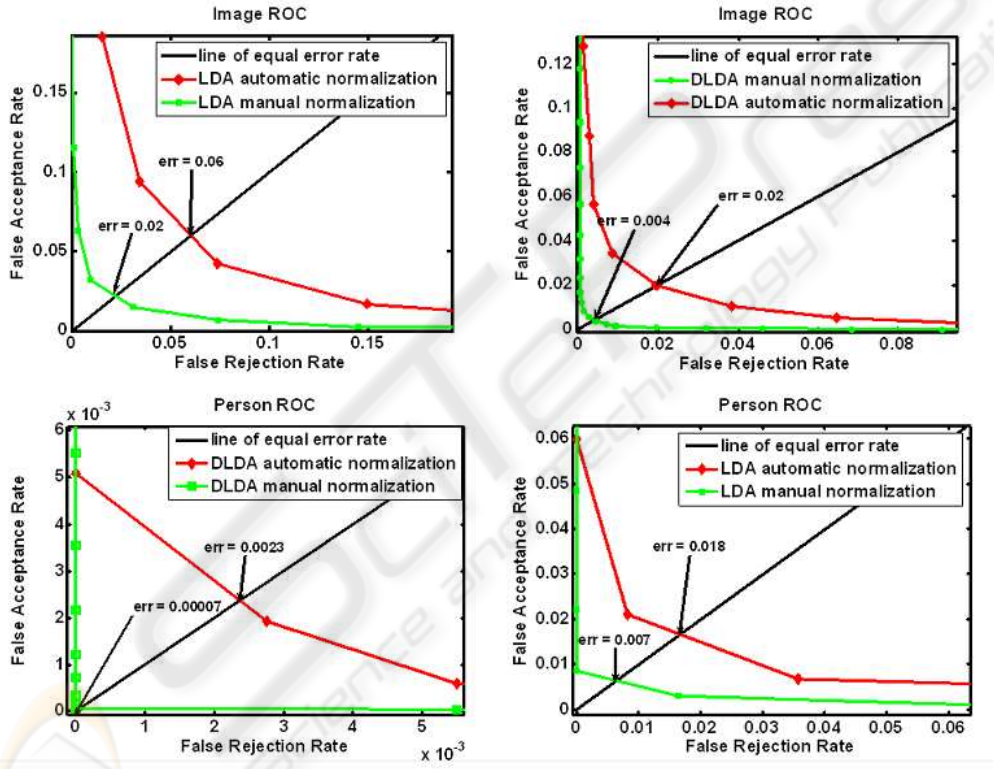


Figure 9: Recognition results for manual and automatic eye center detection.

detection results. However, usually faces that weren't frontal or near frontal have been missed. Also very dark conditions sometimes caused errors. Some of the results are shown in Fig. 10. Detector itself, for CIF resolution, achieves up to 13 frames per second, but in combination with tracking - 25 frames per second.

4 CONLUSIONS

The paper presents a novel face detection and tracking algorithm which could be part of human-machine interaction in applications such as intelligent cash machine. The facial feature extraction algorithm is based on discrete approximation of Gabor Transform, called Discrete Gabor Jets (DGJ), evaluated in edge points. DGJ is computed using integral image for fast summations in arbitrary windows and by FFT operations

Figure 10: Detection results for manual (upper row) and automatic eye center detection (lower row).

on short contrast signals. Contrasting is performed along radial directions while frequency analysis along angular directions. Fourier coefficients for a small number of rings create a feature vector which is next reduced to few LDA components and then compared to the reference facial feature vector. Detected eyes and nose corners are chosen to fit reference face by spatial relationships. Tracking is based on the same rule, but the corners are searched only within already detected facial features neighborhood. Optionally for face normalization eyes centers are found as centers of outer and inner eye corners. Comparison of manual and automatic eye center detection shows still significant advantage of manual approach, measured in terms of accuracy in face recognition by Linear Discriminant Analysis (LDA) and Dual Linear Discriminant Analysis (DLDA) algorithms.

ACKNOWLEDGEMENTS

The work presented was developed within VISNET 2, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 Programme.

REFERENCES

Beumer, G., Tao, Q., Bazen, A., and Veldhuis, R. N. J. (2006). A landmark paper in face recognition. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 73–78. IEEE Computer Society.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Freund, Y. and Schapire, R. E. (1997). A decision theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and Systems Sciences*, 55(1):119–139.

Fukunaga, K. (1992). *Introduction to Statistical Pattern Recognition*. Academic Press.

Gabor, D. (1946). Theory of communication. *Proc. of IEE*, 93(26):429–441.

Golub, G. and Loan, C. (1989). *Matrix Computations*. The Johns Hopkins University Press.

Jones, M. and Viola, P. (2003). Face recognition using boosted local features. Technical Report TR20003-25, Mitsubishi Electric Research Laboratories.

Kucharski, K. (2006). *Face Indexing by Image Components Method*. PhD thesis, Warsaw University of Technology.

MPEG-7, editor (2004). *Multimedia content description interface. Visual Descriptor Extensions.*, volume 15938-3:2002/Amd.1:2004. ISO/IEC.

Shapire, R. E. (2002). The boosting approach to machine learning – an overview. In *MSRI Workshop on Nonlinear Estimation and Classification*.

Skarbek, W., Kucharski, K., and Bober, M. (2004). Dual lda for face recognition. *Fundamenta Informaticae*, 61:303–334.

Viola, P. and Jones, M. (2001). Robust real-time object detection. *Second Int'l Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling*.

Yang, M. H., Kriegman, D. J., and Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58.