

Object Localization Using Linear Adaptive Filters

Ben-Zion Shaick and Leonid Yaroslavsky

Tel-Aviv University, Faculty of Engineering
Dept. of Interdisciplinary Studies
Tel Aviv 69978, Israel
Email: zion@eng.tau.ac.il

Abstract

We present a novel approach to localization of objects in clutter images with the use of linear adaptive filters in a two-object classifier: target object versus clutter object. An automatic optimized feature extraction processing is suggested to generate two pair of models: “target” and “clutter” models from training image databases, and “clutter-like-target” and “target-like-clutter” models from positive and negative detection errors examples respectively.

Experimental results obtained on testing database of known acquisition system containing “face” and “non-face” objects show that the proposed approach outperforms other literature reported results both in term of detection rate and false alarm rate.

1 Introduction

Object localization is a fundamental problem in computer vision. One of the typical and challenging applications in this field that has been extensively investigated is the problem of human face detection and localization.

A variety of methods have been developed for solving this problem: Top-Down approaches [1,2], Bottom-Up approaches [3,4], Statistical approaches [5,6], Neural network approaches [7,8], Fast algorithms [9,10], Template matching by means of matched filter [11], Template matching by means of linear adaptive filters [12], and others methods [13,14].

In this paper, we present an approach to localization of objects in clutter images with the use of linear adaptive filters in a two-object classifier: target object versus clutter object. We will show that one can optimize linear filters in terms maxi-

mization of the ratio of the filter response to the target object to standard deviation of its respond to “clutter objects”. We will refer to this ratio as to Signal to Noise Ratio (SNR).

As a practical application, localization of frontal views of human faces in gray-scale images with complex background is selected for testing the proposed object localization method.

In most object localization application, the input images are obtained with a fixed acquisition system. The results of experimental testing the developed method on a database for a fixed acquisition system show its superior discrimination capabilities in face detection in clutter images.

2 Optimal localization of an exactly known object target in additive noise

Let $u(x - x_0, y - y_0)$ be the target object located in coordinates $\langle x_0, y_0 \rangle$ and x and y denote the spatial two-dimensional image indices. Let also assume that the target object is corrupted in the imaging system by additive noise $n(x, y)$ such that imaging system output image can be modeled as

$$v(x, y) = u(x - x_0, y - y_0) + n(x, y) \quad (1)$$

Target coordinates $\langle x_0, y_0 \rangle$ are to be estimated from the observed input image $v(x, y)$. The optimal filter frequency response for this case is given by [15,16]:

$$H(f_x, f_y) = \frac{U^*(f_x, f_y)}{|N(f_x, f_y)|^2} \quad (2)$$

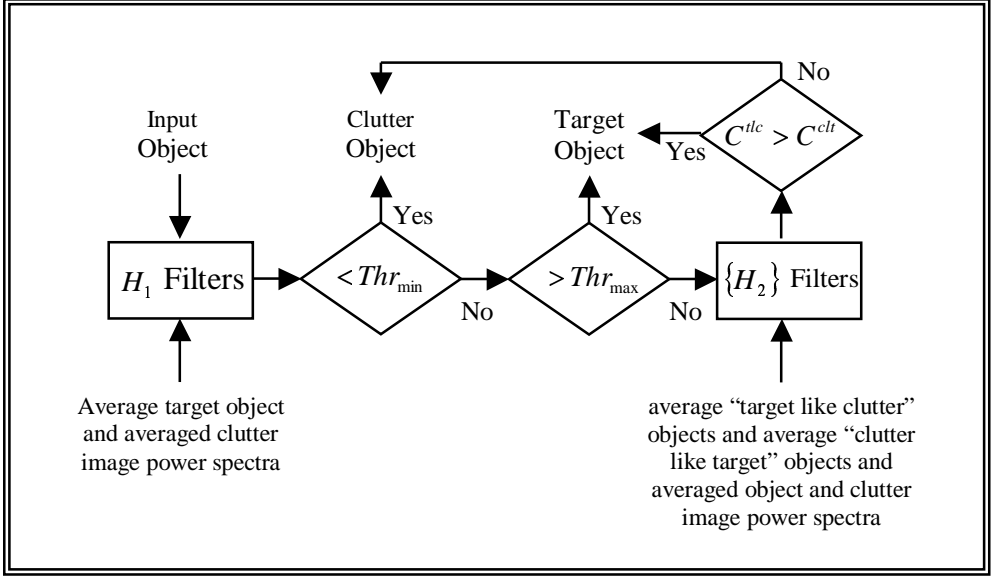


Figure 1: The classification algorithm

where N and U are the Fourier transform of the noise and target object respectively, $*$ denotes the complex conjugate operation, f_x and f_y are the two-dimensional transform domain indices. The output correlation image is optimal in terms of the ratio of target signal peak height to standard deviation of output noise (Signal to Noise Ratio - SNR).

3 Optimal localization of an exactly known object target in cluttered background

The optimal filter frequency response for the task of localizing an object $u(x, y)$ in a cluttered image $v(x, y)$ is given by [15,16]:

$$H(f_x, f_y) = \frac{U^*(f_x, f_y)}{AV_{x_0, y_0} AV_{bg} |V_{bg}(f_x, f_y)|^2} \quad (3)$$

where $|V_{bg}(f_x, f_y)|^2$ is power spectrum of the background component of the image scene and AV_{x_0, y_0} and AV_{bg} denote averaging over unknown coordinates $\langle x_0, y_0 \rangle$ of the target object and over possible realizations of the background

component of the image respectively. The output correlation image is optimal in terms of the ratio of target signal peak height to standard deviation of clutter component of the image (SNR).

4 Optimal classification of an exactly known target object and a clutter objects class

Let $u(x, y)$ be the target object and $\{v(x, y)\}$ be the clutter object class. We want to find the filter that will maximize the SNR at its output:

$$\max_H \{SNR\} = \max_H \left\{ \frac{|C_{tr}|^2}{\text{var}\{\{C_{cl}\}\}} \right\} \quad (4)$$

where $|C_{tr}|^2$ is the squared filter response to the target object and $\text{var}\{\{C_{cl}\}\}$ is the variance at the filter output when it is applied to the clutter class members.

We can rewrite Equation 4 by applying inverse Fourier transform with $x = y = 0$ to the numerator and Parseval's theorem to the denominator:



Figure 2: Examples of faces after illumination correction and contrast normalization.

$$\max_H \left\{ \frac{\left| \sum_{f_x} \sum_{f_y} H(f_x, f_y) U(f_x, f_y) \right|^2}{\sum_{f_x} \sum_{f_y} |H(f_x, f_y) V(f_x, f_y)|^2} \right\} \quad (5)$$

Using Schwartz inequality, obtain

$$\frac{\left| \sum_x a(x)b(x) \right|^2}{\sum_x |a(x)|^2} \leq \sum_x |b(x)|^2$$

$$\max_H \{SNR\} \leq \max_H \left\{ \sum_{f_x} \sum_{f_y} \left| \frac{U(f_x, f_y)}{V(f_x, f_y)} \right|^2 \right\} \quad (6)$$

and the equality holds when

$$H(f_x, f_y) = \frac{U^*(f_x, f_y)}{AV_{cl} |V(f_x, f_y)|^2} \quad (7)$$

where AV_{cl} denotes averaging over the power spectrum of the clutter class members.

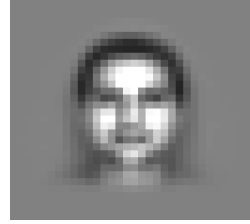


Figure 3: The Average Face

5 Optimal classification of target objects class versus clutter objects class

If target objects have certain variability, two options are possible: either to build optimal filter for every individual representative of the class of the target objects or to build a filter that will be optimal on average for this class. In the latter case, one should use the average target object $\bar{U}(f_x, f_y)$ in Equation 7 such that

$$H(f_x, f_y) = \frac{\bar{U}^*(f_x, f_y)}{AV_{cl} |V(f_x, f_y)|^2} \quad (8)$$

6 Optimal classification of two “similar” object classes

It is not possible to determine the threshold that discriminates the target object from the clutter object in the case of two “similar” class objects since we are using only the averaged object image in the filter. To overcome this difficulty we suggest using two classification filters as follows:

$$H_{opt}^U(f_x, f_y) = \frac{\bar{U}^*(f_x, f_y)}{AV_{obj} |V(f_x, f_y)|^2} \quad (9)$$

for the first class and:

$$H_{opt}^V(f_x, f_y) = \frac{\bar{V}^*(f_x, f_y)}{AV_{obj} |U(f_x, f_y)|^2} \quad (10)$$

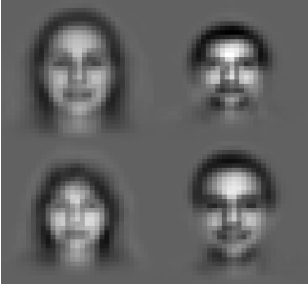


Figure 4: “Face-like –Non-Face” clusters.



Figure 5: “Non-Face-like-Face” clusters.

for the second class, where $U(f_x, f_y)$ and $V(f_x, f_y)$ are the Fourier spectra of the two similar classes $u(x, y)$ and $v(x, y)$ respectively.

Let C^U and C^V be the output signal of $H_{opt}^U(f_x, f_y)$ and $H_{opt}^V(f_x, f_y)$ filters respectively. The classification criterion is then given by:

$$Input\ Object = \begin{cases} C^U > C^V & U \\ else & V \end{cases} \quad (11)$$

7 The classification algorithm

When the input object is to be classified (see Figure 1), the first H_1 filter will be applied.

The object is classified as “clutter” if the filter output is smaller than a predetermined threshold Thr_{min} , while it is classified as “target” if the filter output exceeds Thr_{max} . Intermediate values imply that the filter cannot reliably classify the input image. In this case, the second group of $\{H_2\}$ filters will be applied. The first filter reliably classifies the input object for the case when the “target” and “clutter” objects differ substantially, while the second filter group reliably classifies the input object for the case of “target-like-clutter” and “clutter-like-target” similar objects. The frequency response of the first filter is given by:

$$H_1(f_x, f_y) = \frac{\bar{U}_{Target}^*(f_x, f_y)}{AV_{obj} |V_{Clutter}(f_x, f_y)|^2} \quad (12)$$

The $\{H_2\}$ filters group is prepared from two sets of filters:

$$\left. \begin{cases} H_2^{tlc}(f_x, f_y) = \\ \frac{\bar{U}_{Target-like-Clutter}^*(f_x, f_y)}{AV_{obj} |V_{Clutter-like-Target}(f_x, f_y)|^2} \end{cases} \right\} \quad (13)$$

$$\left. \begin{cases} H_2^{clt}(f_x, f_y) = \\ \frac{\bar{V}_{Clutter-like-Target}^*(f_x, f_y)}{AV_{obj} |U_{Target-like-Clutter}(f_x, f_y)|^2} \end{cases} \right\}$$

If C^{tlc} and C^{clt} are the maximal output signals of $\{H_2^{tlc}\}$ and $\{H_2^{clt}\}$ filters groups respectively then the classification procedure according to Equation 11 is given as

$$Input\ Object = \begin{cases} C^{tlc} > C^{clt} & Target \\ else & Clutter \end{cases} \quad (14)$$

8 Experimental implementation

We test the described classification algorithm in the application to detection of human faces in clutter images.

The averaged “face” image was created from the database of 124 individual faces (see Figure 2) taken from the Computer Vision Center, Purdue University [17]. Each image in the database was subjected to the following preprocessing steps:

- Cropping the image to exclude empty region.
- Converting color images to monochrome ones with 256 intensity levels.

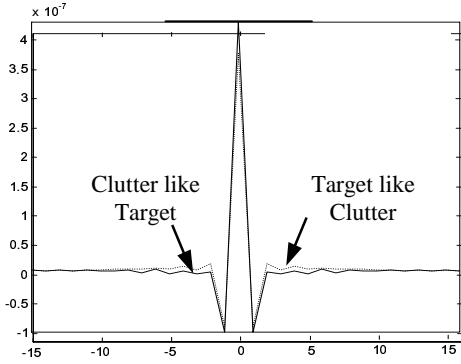


Figure 6: Impulse response of the whitening filter

- Aligning faces to be vertical and fixing the location of both the center of each eye and center of the mouth.
- Masking the face to include facial feature, hair and upper part of the shoulders.
- Normalizing image intensity by locally applying the histogram equalization process in a running window of size smaller than the face size.
- Resampling obtained images to a 32×32 pixels size.
- Rotating ($\pm 5^\circ$) and magnifying (1:1.3) the face width and height to imitate variation in size and to produce a new artificial set of training images. This step will make the classifier tuned to those geometrical distortions.

The resulting database was as large as 9168 images. All the database images were accumulated to produce an average “face” image (see Figure 3). An averaged “non-face” image was also created from the database of images that do not contain faces. The intensity of each image in the database was locally normalized in the same way as in the preprocessing step for the “face” database. Then, the square of the power spectrum of each image was found and finally the spectra of images of entire database were accumulated to produce the average “non-face” image power spectrum. A database of “face-like-non-face” and “non-face-like-face” images was prepared by applying H_I filter of Equation 11 to the entire “face” and “non-face” database and picking up the objects with lowest correlations (950 from 9168) and those with highest correlations (650 from 10^7 objects)

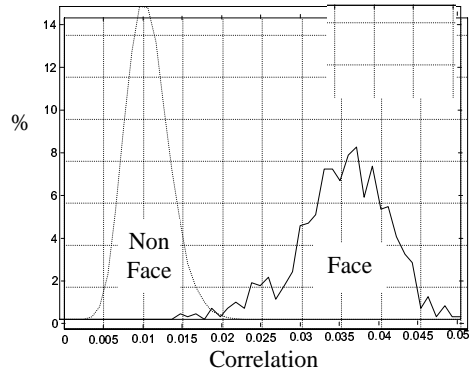


Figure 7: Correlation histogram graph at the output of H_1 filter.

respectively. The new databases were then separated into clusters by the K nearest mean algorithm (see Figures 4.5).

The square of the power spectrum of each “face like non-face” and “non-face like face” image was found and the spectra of images of the two classes were accumulated to produce the average “face like non-face” and “non-face like face” power spectrum image respectively. A graph of one-dimension cross-section of corresponding “whitening” filter (the denominator part of equations 13) is shown in Figure 6.

9 Experimental Results

Figure 7 shows solid and dotted histogram curves of the H_I filter output when applied on the training databases of “faces” (9168 objects) and “non-faces” (10^7 objects) respectively. The two curve (Gaussian shape) imply that our training database is sufficiently large for reliable testing the filter performance and for reliable selection parameters Thr_{min} and Thr_{max} . Figure 8 shows ROC curves obtained by testing the described detection algorithm on the training databases. The solid curve indicates the detection results that were obtained by applying the H_I filter on the “non-face” database and on the “face” database (9168 objects) separately.

The detection rate indicates the percentage of correct “face” classification, whereas the false

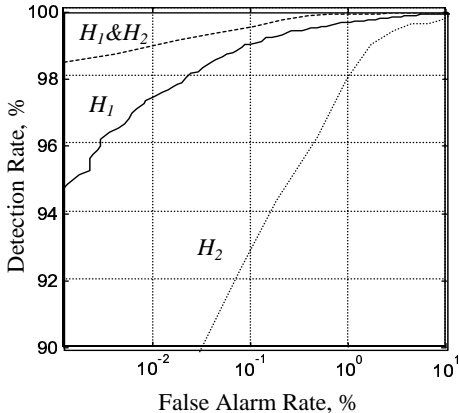


Figure 8: ROC graphs obtained from the training set

alarm rate indicates the percentage of wrong “non-face” classification. The dotted curve indicates the detection results that were obtained by applying the H_2 filter on the “face-like-non-face” and “non-face-like-face” objects separately. These objects were not reliably classified and thus were rejected by the H_1 filter. They are approximately 20% of the “non-face” database and 10% of the “face” database. The dashed curve represents the best detection results that were obtained by applying both H_1 and H_2 filters. These results indicate that the detection rate higher than 99% can be achieved with the probability of false alarms less than 0.01 of percent.

We selected a testing database (different from the training set) with 100 faces and 8000 non-face images of 32×32 pixels size. The faces in the testing database are in the range of tilt angle of $\pm 5^\circ$ and scale of 1:1.3 magnification orders. Testing the algorithm on this testing set resulted in a correct detection of 99 from 100 faces (99% detection rate), and one wrong detection from 8000 (0.013% false alarm rate).

10 Conclusions

An algorithm for reliable localization of inexactly known objects in complex images was described. The algorithm assumes generating, from a training database, average “target” object and “clutter” object images and forming from them the optimal

linear adaptive filters implemented in Fourier transform domain. To further increase the correlator discrimination capability it is suggested to additionally to generate, from the database, subsets of “target-like-clutter” images and “clutter-like-target” images. The algorithm is capable of automatic feature extraction by the use of the developed whitening filter. Experimental verification carried out over database of known acquisition system has shown 99% detection rate and 0.01% false alarms rate, which is superior to other known methods [1-11,13].

References

- [1] G. Yang, and T.S. Huang, “Human face detection in a complex background”. *Pattern recognition*, 27, pp. 53-63, 1994.
- [2] C. Kotropoulos and I. Pitas, “Rule-based face detection in frontal views”. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99)*, Phoenix, Arizona, pp. 2537-2540, 1997.
- [3] T.K. Leung, M.C. Burl and P. Perona, “Finding faces in cluttered scenes using random labeled graph matching” *Proceedings. Fifth International Conference on Computer Vision (ICCV'95)*, Cambridge, Massachusetts, pp. 637-644, 1995.
- [4] K.C. Yow and R. Cipolla, “Feature-based human face detection”. *Image and vision computing*, 15, pp. 713-735, 1997.
- [5] B. Moghaddam and A. Pentland, “Probabilistic visual learning for object representation”, *IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI)*, 19, pp. 696-710, 1997.
- [6] H. Schneiderman and T. Kanade, “A statistical method for 3D object detection applied to faces and cars”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2000)*, Hilton Head Island, Carolina, pp. 746-751, 2000.
- [7] K.-K. Sung and T. Poggio, “Example-based learning for view-based human face detection”. *IEEE trans. on PAMI*, 20, pp. 39-50, 1998.

- [8] H.A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection". *IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI)*, 20, pp. 23-38, 1998.
- [9] D. Maio and D. Maltoni, "Real-time face location on gray-scale static images", *Pattern Recognition*, 33, pp. 1525-1539, 2000.
- [10] H. Wang and S.-F. Chang, "A highly efficient system for automatic face region detection in MPEG video", *IEEE Transaction on circuits and systems for video technology*, 7(4), pp. 615-628, 1997.
- [11] R. Brunelli and T. Poggio, "Template matching: matched spatial filters and beyond". *Pattern Recognition*, 30(5), pp. 751-68, 1997.
- [12] L.P. Yaroslavsky and B.-Z. Shaick, "Transform Oriented Image Processing Technology for Quantitative Analysis of Fetal Movements in Ultrasound Image Sequences". *Proceedings of the IX European Signal Processing Conference (EUSIPCO'1998)*, Rhodes, Greece, pp. 1745-1748, 1998.
- [13] E. Osuna, R. Freund and F. Girosi, "Training support vector machines: An application to face detection" *Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, Sun Juan, Puerto Rico, pp. 130-136, 1997.
- [14] A.L. Yuille, P.W. Hallinan and D.S. Cohen, "Feature extraction from faces using deformable templates". *International Journal of Computer Vision*, 2, pp. 99-111, 1992.
- [15] L. Yaroslavsky and M. Eden, *Fundamentals of Digital Optics*, Birkhauser, Boston, 1996.
- [16] L. Yaroslavsky, "The theory of optimal methods for localization of objects in pictures", *Progress in Optics*, XXXII, pp. 145-201, 1993.
- [17] http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html