Face Detection by Cascade of Gaussian Derivates Classifiers Calculated With a Half-Octave Pyramid

John A. Ruiz - Hernandez, Augustin Lux and James L. Crowley
Grenoble INP, INRIA-Grenoble Rhone-Alpes Research Center, France
{John-Alexander.Ruiz-Hernandez, Augustin.Lux, James.Crowley}@inrialpes.fr

Abstract

This paper presents a method for object detection based on a cascade of scale and orientation normalized Gaussian derivative classifiers learnt with Adaboost. Normalized Gaussian derivatives provide a small but powerful feature set for rapid learning using Adaboost. Real time detection is made possible by use of a fast integer coefficient algorithm that computes a half-octave Gaussian pyramid with linear algorithmic complexity using a cascade of binomial kernel filters. The method is demonstrated by training a boosted classifier for frontal face detection using standard data sets. Experiments demonstrate that this approach can provide detection rates that are comparable or superior to those obtained with integral images while dramatically reducing the required training effort.

1. Introduction

Viola and Jones [24] have demonstrated that a rapid and robust object detector can be constructed using the adaboost algorithm to train a cascade of linear classifiers from image descriptors provided by integral images. Integral images have the interesting property that a very large number of descriptors (180 000 for a 24 x 24 image), can be computed by a very fast integer algorithm, leading to real time algorithms suitable for use in embedded systems. Unfortunately, integral images also have drawbacks. As an image descriptor they are not invariant to image scale or rotation. Furthermore the very large space of detectors can lead to extremely long learning times (on the order of weeks or months) for training a robust classifier.

In this paper we propose an alternative image descriptor using scale normalized Gaussian derivatives. We describe a fast (integer coefficient) pyramid algorithm for computing Gaussian derivatives in real time, and show that this algorithm can be used to provide Gaussian derivatives that are normalized in scale and orientation in linear time. We demonstrate that such normalized Gaussian derivatives can be used to learn a boosted object detector that is comparable to integral images in robustness and computation time, while derived from a much shorter training effort.

As with Viola and Jones, we demonstrate our approach using the problem of frontal face detection from static images. Face detection is an important problem in computer vision, with numerous practical applications including image stabilization for photography and video telephones, mug-shot recognition, facial expression recognition, and visual surveillance.

A variety of statistical learning techniques have been demonstrated for appearance based face detection [26]. For example, Garcia and Delakis describe a convolutional neural network architecture for frontal face detection with a high variability in position [8]. Osadchy et al. used similar neural-network architecture for simultaneous multi-view face detection and facial pose estimation [18]. While it is possible to use such neural networks to build detectors that are robust to pose changes, there is no systematic method for training such detectors. Determining the appropriate parameters such as number of layers remains a difficult problem.

Support vector machines (SVM) have recently been used for a number of object detectors. Heisele et *al* has proposed at hierarchy of SVM's to reduces the dimension of a feature space and thus optimize detection in images [12]. While SVM methods are known for generalization, a number of open problems remain including choice of kernel. They are also known to provide detection algorithms that have a high computational cost. Neither Neural network approaches nor support vector machine methods have surpassed the use of adaboost and integral images for building robust real time object detectors.

Many authors have explored extensions to Viola and Jones' method. Meynet and *al* have demonstrated detection using a cascade of boosted anisotropic Gaussian classifiers [17]. Other recent works have explored use of extended feature sets [15] or modifications to the learning algorithm or the cascade structure [2, 19, 22] for such problems as pedestrian detection [25]. The difficulty with these approaches has been the high dimensionality of the feature space, resulting in learning times that can span several weeks.

In this paper, we show that similar results can be obtained with a greatly reduced feature space provided by normalized Gaussian derivatives features. Gaussian derivatives have been known to provide a scale and

978-1-4244-2154-1/08/$25.00 ©2008 IEEE

rotation invariant image description [3] since the 1980s. Freeman and Adelson [6] demonstrated that Gaussian derivatives are steerable, providing a means to compute derivatives in arbitrary directions from linear sums of canonical derivatives computed from separable filters. Lowe demonstrated that local orientation statistics of scale normalized gradients (SIFT) [16] could provide a robust and stable image descriptor for tracking and matching interest points. Delal and Triggs [5] demonstrated that such detectors could be used for learning to detect object categories. Yokono and Poggio have recently demonstrated the use of Gaussian derivatives filters for object recognition [27]. Hall and Crowley used the Gaussian derivatives to generate a face template with log polar histograms [11]. This work was extended by Gourier and *al* who used chromatic Gaussian derivatives features for face detection and pose estimation [9].

A critical problem with Gaussian derivatives is computational cost. Direct computation of derivatives using FIR filters for scale normalization can give rise to algorithms with quadratic complexity. Even with recursive filters, algorithmic complexity and computational cost can make real time computation unfeasible. In this paper we propose the use of a half octave binomial pyramid to provide a fast algorithm for scale normalized Gaussian derivatives. This algorithm has been shown to have O(N) complexity, and to be computable using only integer operations [4].

2. Gaussian Derivatives as Image Features

This section describes the use of a cascaded binomial kernel to obtain a Gaussian pyramid in which each layer has an identical ratio of impulse response to sample rate

2.1. The Gaussian Jet

Koenderink has described how the visual appearance of a neighborhood can be represented by a local Taylor series expansion [13]. The coefficients of this Taylor series constitute a feature vector, referred to as the "Local Jet" that compactly represents the neighborhood appearance for indexing, matching and recognition. Ter Haar Romeny and others [23], have shown that invariance to scale and orientation can be obtained when the local jet is computed using Gaussian Derivatives. The basis functions for this expansion are local derivatives computing using a Gaussian support:

$$G(x,y;\sigma) = e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where σ is a scale factor that expresses the size of the support in terms of the second moment (or variance) Schiele [21] has referred to this as a Gaussian Receptive Field. The Gaussian support measures the average intensity of the neighborhood. This information does not contribute to the identification of the neighborhood, and can therefore be omitted in detection and recognition.

The first and second derivatives of the Gaussian support are:

$$G_x(x,y;\sigma) = \frac{\partial G(x,y;\sigma)}{\partial x} = -\frac{x}{\sigma^2} G(x,y;\sigma)$$

$$G_y(x,y;\sigma) = \frac{\partial G(x,y;\sigma)}{\partial y} = -\frac{y}{\sigma^2} G(x,y;\sigma)$$

These capture information about changes of the surface normal and measure the intensity of edges. The second order derivates are given by:

$$G_{xx}(x,y;\sigma) = \frac{\partial^2 G(x,y;\sigma)}{\partial x^2} = (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2}) G(x,y;\sigma)$$

$$G_{yy}(x,y;\sigma) = \frac{\partial^2 G(x,y;\sigma)}{\partial y^2} = (\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2}) G(x,y;\sigma)$$

$$G_{xy}(x,y;\sigma) = \frac{\partial^2 G(x,y;\sigma)}{\partial x \partial y} = \frac{xy}{\sigma^4} G(x,y;\sigma)$$

The second order derivatives are good descriptors for image features such as bars, blobs and corners. Higher order Gaussian derivatives are more sensitive to the image noise and only provide useful information in cases where the second order derivatives are strong.

A local jet based on Gaussian derivatives has been proposed as a model for receptive fields in the primary visual cortex [28]. The use of Gaussian derivatives is motivated by their capacity of describe image neighborhoods and also by their representation of image in specific orientations and frequencies. This representation allows an excellent support in object detection and learning processes.

A key problem in computing the local jet is determining the scale at which to evaluate the derivatives. Lindeberg [14] has described scale invariant features based on profiles of Gaussian derivatives across scales. In particular, the maximun of the Laplacian, evaluated over a range of scales at an image point is known to be "equivariant" to scale. Lowe has referred to this as the "natural" scale and has used this scale to construct a detector for scale normalized interest points [16] for the SIFT descriptor.

Just as the local jet can be normalized in scale, it can also be normalized in orientation. Since the early years of computer vision, the arc-tangent of the ratio of first derivatives (the angle of the gradient) has been used to

estimate the orientation of the gradient at any image point. Freeman and Adelson [6] have shown how a basis set of Gaussian derivatives can be "steered" to a desired orientation by weighting the derivative terms with the appropriate sine and cosines terms. Using their technique it is relatively easy to define a steerable basis for the local jet. Such a basis can be normalized to the intrinsic scale using the Laplacian profile and then normalized in orientation to the intrinsic orientation.

In order for normalized Gaussian derivatives to provide an effective means for real time object detection, we need an algorithm for computing such features that is competitive with Integral Images. Such an algorithm is provided by computing a half-octave Gaussian pyramid using a Binomial filter as kernel.

2.2. Half-Octave Gaussian Pyramid

Computing a scale invariant local Jet for an NxN image requires computing second order derivatives (Laplacian of the Gaussian) of the image at Log(N) scales. A linear time pyramid algorithm for this calculation has been known since the 1980's [1, 3]. The result of this algorithm is a Half-Octave Gaussian Pyramid. An integer coefficient version of this algorithm [4] has been demonstrated using repeated convolutions of the binomial kernel (1,2,1). Implementations that compute such pyramids on PAL sized images at video rates exist for the current generation of computer work-stations.

The pyramid algorithm is composed of an initial convolution with a Gaussian kernel filter $G(x,y,\sigma_o)$ followed by a series of processing stages, k=1 to K, as shown in figure 1. In our experiments we used $\sigma_0=1$. For each stage k, the pyramid is composed of three images $p_0(x, y, k)$, $p_1(x, y, k)$, $p_2(x, y, k)$. At the k^{th} stage, the image $p_0(x, y, k)$ is produced by resampling the image $p_2(x, y, k-1)$ by a factor of 2 in the x and y directions, using a re-sampling operator, $S_2\{\cdot\}$.

$$p_0(x, y, k) = S_2\{p_2(x, y, k-1)\}$$

The image $p_1(x, y, k)$ is the result of

$$p_1(x, y, k) = p_0(x, y, k) * G(x, y, \sigma_o)$$

where "$*$" is the convolution operator. The image $p_2(x, y, k)$ is produced by

$$p_2(x, y, k) = p_1(x, y, k) * G(x, y, \sqrt{2}\sigma_o)$$

This scaled copy can be obtained by cascaded convolution with the Gaussian kernel filter:

$$p_2(x, y, k) = p_1(x, y, k) * G(x, y, \sigma_o) * G(x, y, \sigma_o)$$

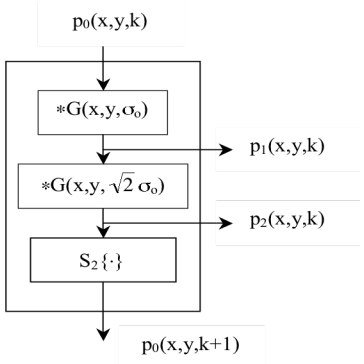

Figure 1. Schema for the k^{th} level in the pyramid.

The result of repeating these stages is a sequence of images in which the scale factor grows a power of 2. The process can be further accelerated by using a separable form of binomial approximation for the Gaussian filter, as shown by Riff [4].

$$G(x,y;\sqrt{2}) \approx \begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix} = \begin{bmatrix}1\\ 2\\ 1\end{bmatrix} * \begin{bmatrix}1 & 2 & 1\end{bmatrix}$$

Because each image in the pyramid has been smoothed by a Gaussian impulse response of $\sigma_k=2^k\sigma_o$, differences of adjacent image pixels in the row and column directions are equivalent to convolution with Gaussian derivatives.

$$\frac{\partial p(x,y,k)}{\partial x} = p * G_x(x,y;2^k\sigma_0) \approx p(x+1,y,k) - p(x-1,y,k)$$

$$\frac{\partial p(x,y,k)}{\partial y} = p * G_y(x,y;2^k\sigma_0) \approx p(x,y+1,k) - p(x,y-1,k)$$

$$\frac{\partial^2 p(x,y,k)}{\partial x^2} = p * G_{xx}(x,y;2^k\sigma_0) \approx p(x+1,y,k) - 2p(x,y,k) + p(x-1,y,k)$$

$$\frac{\partial^2 p(x,y,k)}{\partial y^2} = p * G_{yy}(x,y;2^k\sigma_0) \approx p(x,y+1,k) - 2p(x,y,k) + p(x,y-1,k)$$

$$\frac{\partial^2 p(x,y,k)}{\partial y \partial x} = p * G_{xy}(x,y;2^k\sigma_0) \approx p(x+1,y+1,k) - p(x+1,y-1,k) - p(x-1,y+1,k) + p(x-1,y-1,k)$$

Derivative values may be easily determined for image positions between image samples using bilinear interpolation. Derivatives for scale values between the pyramid levels can be computed using quadratic interpolation between adjacent levels in the pyramid. In

this way, Gaussian derivatives may be determined for any required value of x, y, σ.

3. Adaboost with Gaussian Derivatives

Adaboost has been proposed by Freund and Schapire [7] as an efficient learning algorithm for constructing a strong classifier as an additive combination of boosted weak classifiers that are individually only slightly better than random. At each iteration, adaboost determines a new weak classifier relative of the lowest value in the weight distribution in the training set. The principal advantage of Adaboost is that the training error converges exponentially towards zero and the generalization performance grows at each iteration when the null training error is reached by the algorithm [7].

3.1. Gaussian Derivative Feature Vector

In our experiments we used a 24 x 24 window size. A half-octave pyramid for a 24 by 24 image window gives five pyramid levels with a total of 10 images, providing 11140 Gaussian derivative features. These features are calculated in ten sub-levels between the first and fourth levels in the pyramid. In this feature set we ignore the initial stage of the pyramid, as well as regions within four pixels of the image boundaries. Figure 2 illustrates the feature set, by showing the derivative impulse responses at different levels of a pyramid computed for an image window that contained a single impulse pixel surrounded by zeros.

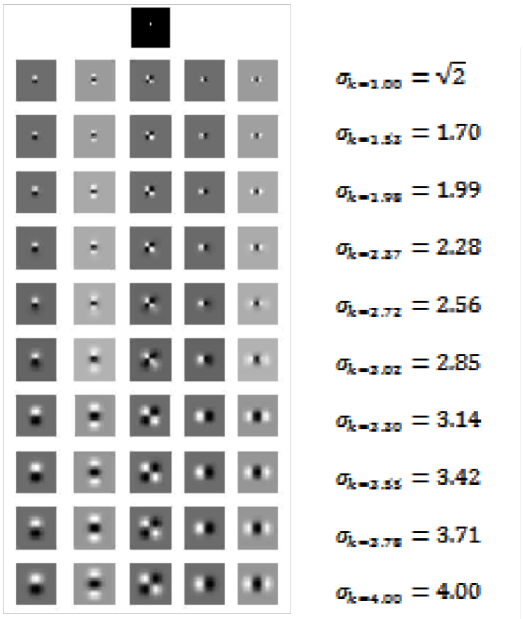

Figure 2. Impulse responses for derivative features ($p_y(x,y,1)$, $p_{yy}(x,y,1)$, $p_{xy}(x,y,1)$, $p_x(x,y,1)$, and $p_{xx}(x,y,1)$, calculated at steps of $\Delta\sigma$=0.3 by interpolation.

3.2. Cascade of Weak Classifiers

The cascade of weak classifiers is successively applied to all image sub-windows. The first layer has a small number of weak classifiers that reject a pre-defined percentage of negative sub-windows and detect nearly 100% of the positive sub-windows in the image. The next layer is then trained to reject the same percentage of negative sub-windows and detect nearly 100% of positive sub-windows using the sub-windows that were improperly classified by the previous layer. This procedure is repeated to provide a cascade of classifiers that increasingly concentrate on a reduced number of difficult sub-windows. This technique allows an improvement in the detection speed with excellent detection and false positives rates. Viola and Jones [24] have shown that adaboost could be used to train a cascade of boosted classifiers using integral image features as simple classifiers.

4. Experimental Results

In this section we report on our experiments comparing scale and orientation normalized Gaussian derivative features with integral image features for face detection by cascade of classifiers.

4.1. Performance of Gaussian Derivatives Features vs. Integral Image Features

To compare Gaussian derivative features with integral image features, we used Adaboost to train 400 weak classifiers for face detection in 24 x 24 pixel image windows with Gaussian Derivative features and with the Haar-like features provided by integral images. The training set consisted of 1000 face images and 1000 non-faces images, while the test set consisted of 1000 face images and 30000 non-face images. Figure 3 compares the performance of the two detectors. In this figure we see that Gaussian derivative features systematically outperformed integral images for all 400 weak classifiers.

The error rate of detection with Gaussian derivatives features decreases quickly with the number of Adaboost iterations. For the first ten iterations the error rate with Gaussian derivatives is less than half of the error rate for integral images, in the consecutive iterations the error rate detection for the Gaussian Features is less than one fourth of the error rate detection for integral images. The false positive rate is also slightly better with Gaussian derivatives features gaining about 1%.

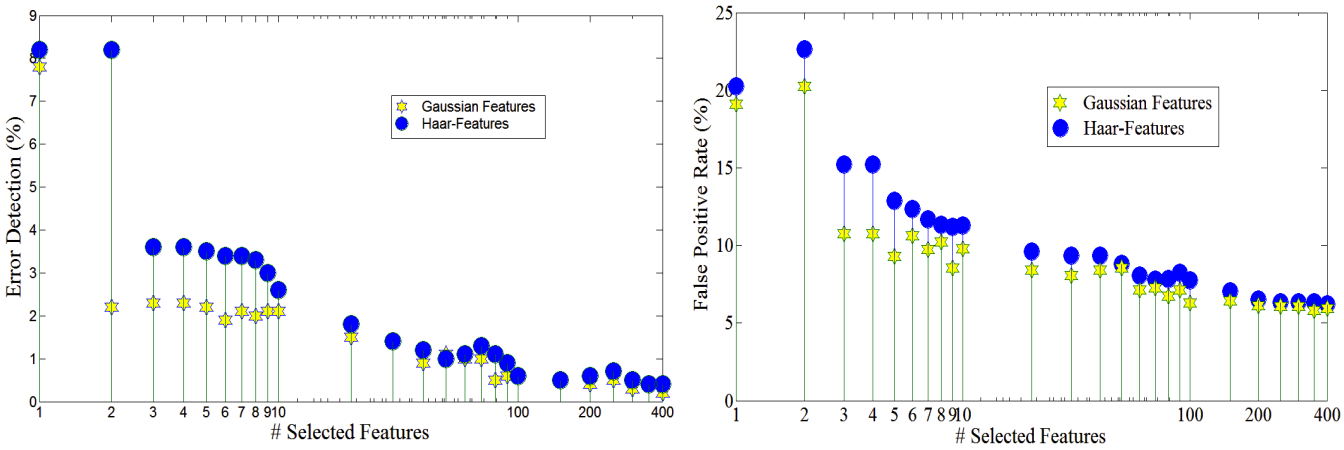

Figure 3. Error Detection and false positive rate values for a 400 weak classifiers trained with Gaussian features (yellow star) and Haar Features (blue circle) using adaboost. (Notice: for best visualization we used a log scale for # selected features axis)

4.2. Cascade of Gaussian Derivatives Features

To evaluate our approach, we have constructed a cascade of Gaussian classifiers with 15 layers. During the learning process we used a training set with 2000 face images. A training set of 4000 non-face images was used for the first layer; for each subsequent layer the non-face datasets contained 4000 non-face sub-windows bootstrapped from 3500 images without faces. The training process has been performed on a typical desktop 2.0 Ghz Pentium-4 computer. The total training time was about 2 days. In comparison, Viola and Jones have used training data with 10000 face images, and 10000 non-face images with a feature set of 180000, for each layer, their training time seems to be in the order of weeks [24].

When applying the classifier for detection on arbitrary images, the pixels classified as "face" are clustered in order to obtain a single response for adjacent pixels. For detection, we retain a median window, supported by a sufficient number of face pixels.

We tested the learned cascade on the MIT+CMU frontal face database [20]. Figure 4 shows the global result represented by a ROC curve. The test data set is composed of 130 images with 507 labeled faces (about 2% of the faces are cartoons, which we do not want our system to detect. Nevertheless, we include them in the results [15]). For this data set, we achieved a detection rate of 92% with a total of 350 false positives (operating point in the ROC curve).

5. Conclusions

The primary advantage of the integral images computation for Haar-like features is that it provides a very fast calculation for a very large set of image features. The resulting features are essentially binary coefficient image filters that tend to be sensitive to small changes in image position, scale and rotation. A cascade of boosted classifiers overcomes these sensitivities by brute force. However, the very large number of features used results in prohibitively long training times, making construction of new classifiers difficult. Furthermore, the resulting detector is only robust to changes in scale and orientation if the training data set contains positive examples at different scales and orientations.

Gaussian Derivative features provide a natural alternative to integral images. As with integral images, a fast O(N) integer coefficient algorithm exists to provide a rich feature space, whit the advantage that the detectors in this feature space are much less sensitive to changes in scale and orientation. More importantly, the resulting detectors can easily be made invariant to scale and orientation effects by adapting the feature set to a reference scale and orientation.

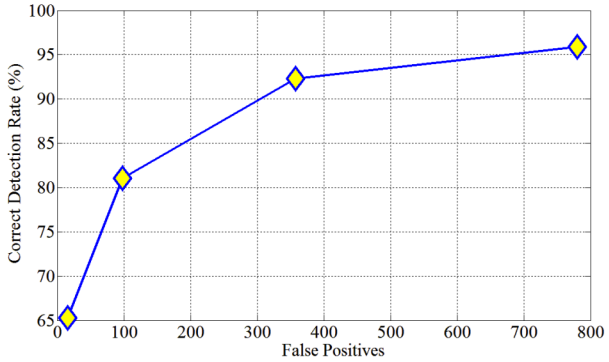

Figure 4. ROC curve for our face detector on the MIT+CMU test set. The detector was run once using a step size of 1.1 and starting scale of 1.25 (73,130,500 sub-windows scanned)

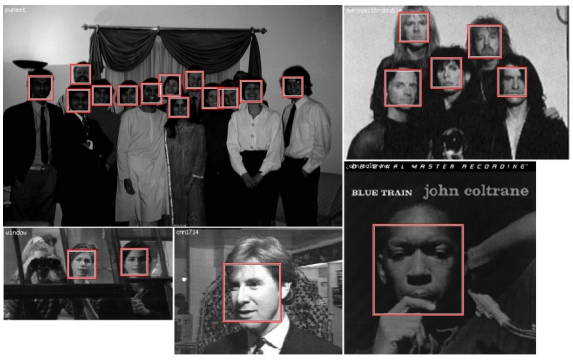

Figure 5. Output of our face detector on a number of test images from the MIT+CMU test set.

References

[1] P.J. Burt and E.H. Adelson. The Laplacian pyramid as a compact image code. In IEEE Communications Transactions, vol. 31, p.532-540. 1983

[2] H. Chang, A. Haizhou, L. Yuan and L. Shihong. High-Performance rotation invariant multiview face detection. In transactions IEEE on Pattern Analysis and Machine Intelligence, vol 29, p.671-686, 2007.

[3] J. L. Crowley and A.C Parker. A representation for shape based on peaks and ridges in the difference of low pass transform. In transactions IEEE on PAMI, 6(2), p. 156-170, 1984.

[4] J. L. Crowley and O. Riff. Fast computation of scale normalised Gaussian receptive fields. Proceedings Scale-Space, Isle of Skye, Scotland, Springer Lecture Notes in Computer Science, volume 2695, 2003.

[5] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2005, pp 886- 893 2005.

[6] W. T. Freeman and E. H. Adelson. The design and use of stereable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(9):891-906, 1991.

[7] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, pp. 119-139, 1997.

[8] C. Garcia and M. Delakis. Convolutional face finder: a neural architecture for fast and robust face detection. In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 26, issue 11, p. 1408-1423, 2004.

[9] N. Gourier, D. Hall and J. Crowley. Facial features detection robust to pose, illumination and identity. In Proceedings IEEE of POINTING '04, Visual Observation of Deictic Gestures, In association with ICPR 04, 2004.

[10] D. Hall. Viewpoint Independent Recognition of Objects from Local Appearance. Thesis dissertation at Institute National Polytechnique de Grenoble, 2004.

[11] D. Hall and J. Crowley. Face detection by robust generic features computed from luminance. in “Reconnaissance des Formes et Intelligence Artificiel (RFIA)”, 2004.

[12] B. Heisele, T. Serre, S. Prentice and T. Poggio. Hierarchical classification and feature reduction for fast face detection with support vector machines. In Pattern Recognition, p.2007-2017, 2003.

[13] J. J. Koenderink, A. J. van Doorn, "Representation of Local Geometry in the Visual System", Biological Cybernetics, pages 367-375, 1987.

[14] T. Lindeberg. Scale-Space Theory in Computer Vision, Kluwer Academic Publishers, 1994.

[15] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, volume 2, p.53-60, 2004.

[16] David G. Lowe. Object recognition from local scale-invariant features. In International Conference on Computer Vision, p.1150-1157, 1999.

[17] J. Meynet , V. Popovici and J.P. Thiran. Face detection with boosted Gaussian features. In Pattern Recognition, v.40 n.8, p.2283-2291, 2007.

[18] M. Osadchy, M. L. Miller and Y. Le Cun. Synergistic face detection and pose estimation with energy-based methods. In Advances in Neural Information Processing System (NIPS), 2006.

[19] Z. Qiang, Y. Mei-Chen and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In IEEE Computer Vision and Pattern Recognition Conference, Volume 2, p.1491-1498, 2006.

[20] Henry A. Rowley, S. Baluja and T. Kanade. Rotation invariant Neural Network-based face detetction. In IEEE Conference in Computer Vision and Pattern Recognition, p.963, 1998.

[21] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. In International Journal of Computer Vision, p.31-50, 2000.

[22] Z. Li Stan and Z. ZhenQiu. FloatBoost learning and statistical face detection. In transactions IEEE on Pattern Analysis and Machine Intelligence, vol 26, p.1-12, 2004.

[23] B. M. ter haar Romeny, L. M. J. Florak A. H. Salden, and M. A Viergever, "Higher Order Differential Structure of Images", Image and Vision Computing, Vol 12 No. 6, pp 317-325, 1994

[24] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In IEEE Computer Vision and Pattern Recognition, 2001.

[25] Y.W. Xu, X.B. Cao, H. Qiao and F.Y.Wang. A cascaded classifier for pedestrian detection. In IEEE Intelligent Vehicles Symposium, 2006.

[26] M-H. Yang, D.J. Kriegman and N. Ahuja, Detecting faces in images: a survey, in IEEE Transactions, Pattern Analysis and Machine Intelligence, vol 24, no. 1,2002.

[27] J.J. Yokono and T. Poggio. Oriented filters for object recognition: an empirical study. In IEEE conference in Automatic Face and Gesture Recognition, Issue, 17-19, p.755– 760, 2004.

[28] R. A. Young, "The Gausssian Derivative Theory of Spatial Vision: Analysis of Cortical cell receptive field line-weighting profiles", Technical Report, General Motors Research Labs, 1985.