International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.4, pp. 122-126, 2013

LDA based Reduced Joint Integral Histogram for Feature Extraction Case of study: Face Detection

Ameni YANGUI JAMMOUSSI, Sameh FAKHFAKH GHRIBI and Dorra SELLAMI MASMOUDI
Computers Imaging and Electronic Systems Group
CEMlab research Laboratory ENIS
University of Sfax, Sfax Engineering School, POB, 3038 Sfax, Tunisia

***Abstract*—The face pattern is described by extracted features using the new Reduced Joint Integral Histogram (RJIH) data structure. Extending the classical representations of integral images and integral histograms, it joins the global information of two images. Then, we turn to Linear Discriminant Analysis (LDA) to project the obtained Joint Integral Histogram from $d-$dimensional subspace to one dimensional subspace. Best features are selected by Adaboost learning framework. The experimental results demonstrate that our proposed method RJIH under Adaboost training process further improve the performance of the JIH in terms of detection rate and false positive rate.**

I. Introduction

Appearance based approach had shown good performances in face detection applications [1]. The whole concept is to collect a large set of face and non-face images, and adopt machine learning algorithms to learn a face pattern to perform classification [2]. The type of features contributes well on the performance of these approaches. Generally, in these methods, there are two key questions that can be asked: what the accurate feature to extract and what the convenient learning algorithm to apply? As the feature extraction process play a crucial role in these methods, we review in this brief the recent advances in feature extraction process [3].

The Haar-like rectangular features of the Viola and Jones framework [4] are very easy to compute due to the integral image representation, and yield good performances for building frontal face detectors.
In a number of follow-up works, a number of researchers noted the limitation of the conventional Haar-like features set especially for multi-view face detection, and proposed to generalize the feature set.
For instance, Lienhart and Maydt [5] extended the feature set by introducing 45 degree rotated rectangular features and center-surround features.

Nonetheless, Haar-like features and its variants are not the only effective extracted features that has shown tremendous success. Many features, have performed well in face detection systems. Another well-known feature characterized by its robustness to illumination variations called Local Binary Patterns (LBP) have been shown good performances for face recognition tasks. Inspired by LBP, many researchers have proposed other effective features like Multi-block Local Binary Patterns (MLBP)[6], Locally Assembled Binary features (LAB)[7] and HLBP[8] which combines the advantage of both LBP and Haar-like features.

To further improve the performance of the system, more complex features were proposed in the literature and various efforts have also been made to improve the feature computation speed and so the final detector's speed. Inspired by the integral image representation used by Viola and Jones for the Haar-like features computation, the Integral Histogram (IH) technique was adopted for faster histogram computation [15]. In fact, histograms are among the most prominent features used in many computer vision applications from object based retrieval, to segmentation, to object detection, to object tracking. A histogram can be defined as an array of numbers in which each element (bin) corresponds to the frequency of values in a given data. For example, each bin counts the number of pixels values having the same color values of an image. The computational complexity of such method is one major bottleneck of the histogram extraction. There are some attempts to deliver accelerated alternatives techniques to the basic exhaustive search. Computer vision problems that depend on the optimal solutions, such as detection and tracking, remain demand a theoretical breakthrough in histogram extraction. For this purpose, the integral histogram [15] is proposed which converts into the extraction of rectangular region histograms, which are computed by intersection of the integral histogram at the four corner references using simple arithmetic operations.

These integral graphs have been proved efficiency for computation in software and hardware in terms of speed and facilitate the computation of different values under a square region in a constant time.
In our previous work [9], taking into consideration the advantage of integral representations, we make use of both integral image and integral histogram by the way of an image data structure referred to as Joint Integral Histogram (JIH). We have demonstrate the effectiveness of this method in feature extraction process. However, JIH is a three dimensional array which increase the computation complexity of feature

extraction process and slow down the final face detector. On the one hand, with JIH based feature, the number of selected weak classifiers by Adaboost is reduced considerably. On the other hand, the complexity is augmented which explain the need to reduce the JIH dimensionality to further improve the performance of such proposed technique.

The remaining of this paper is organized as follows. Section II introduces the JIH data structure. Section III presents the use of JIH for feature extraction process. In section IV, our proposed RJIH method. Section V presents experimental results and discussion. Section VI presents a brief description of the basic Adaboost. Finally, section VII summarizes our work and draws some conclusions.

II. Feature Extraction using Reduced Joint Integral Histogram (RJIH)

A. JIH basic principle

The JIH is inspired by both the integral images and the integral histograms. In a JIH, the value at each bin is determined by two images. The equation of JIH is constructed as follows[10][11]:

$$\begin{aligned} JIH(x,y,bi) &= \sum_{x'\le x, y'\le y} \delta(x',y') f(x',y') where \\ & \quad \delta(x',y') = 1 \quad if \quad g(x',y') = bi \end{aligned} \tag{1}$$

As the equation shows, the JIH is composed of a combination of an integral image and an integral histogram. The contribution of (x, y) to JIH is jointly determined by two functions $f(.)$ and $g(.)$. The function $g(x, y)$ determines which bin to increase and $f(x, y)$ determines the value to increase at that bin. In a JIH, instead of remembering bin occurrences, the value at each bin indicates an integral defined by two signals.

B. JIH based Features

In our previous work[9], we made use of the JIH for efficient feature extraction. The main purpose behind the exploitation of the JIH is to take advantages of two different images and so their complementary benefits. We joined the discriminating information of Haar-like features and the invariance to illumination of LBP image. The integral image is applied to the pixel gray level values and the integral histogram to the LBP image which has been proved by experimental results. Thus, we used both the pixel gray level values image and the LBP patterns extracted to jointly represent the target and embed it into the Adaboost framework. We recall in what follows the basic principles of LBP and Haar-like features.

1) Local Binary Patterns (LBP): Local Binary Patterns (LBP) was first proposed as a gray level invariant texture primitive extractor [12]. LBP operator describes each feature by its relative gray level to its neighboring pixels. If the gray level of the neighboring pixels is higher or equal, the value is set to one, otherwise is set to zero.

$$LBP_{P,R} = \sum_{k=0}^{P-1} s(gk - gc)2^k \tag{2}$$

The operator $LBP_{P,R}$ refers to a neighborhood size of P equally spaced pixels on a circle of radius R that form a circularly symmetric neighbor set.

2) Haar-like masks: The rectangular Haar features are reminiscent of Haar basis functions which have been used by Papageorgiou et al at 1998 [13].
More specifically, they use three kinds of features. A two-rectangle feature is the difference between the sum of pixels within two rectangular regions. Selected regions have the same size and shape and are horizontally or vertically adjacent. A three rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle.

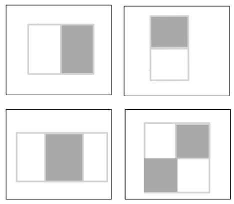

Fig. 1. Rectangle features

C. Reduced Joint Integral Histogram (RJIH)

The obtained JIH of an $p \times q$ image is a three dimensional array as $JIH[p+1][q+1][d]$, where d is the number of bins. In order to reduce the JIH complexity, we adopt the following process.
First, we try to reduce the number of bins of the histograms. There is no best number of bins, and with varying bin sizes we can reveal different feature values. Depending on the data and its analysis goals, different bin widths may be appropriate. Thus, experimentation is usually needed to determine the adequate width. The number of bins k can be calculated starting from a suggested bin width h as:

$$k = \left[\frac{\max x - \min x}{h}\right]. \tag{3}$$

Second, according to the literature review, there are some techniques for dimension reduction. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are both statistical techniques used to dimensionality reduction purposes. In [14], Fisher Linear function was used to find a projection orientation of histograms by which two classes are well separated. Thus, we adopt this method in our work to reduce JIH dimensionality.

D. Linear Discriminant Analysis

Starting with PCA which is a statistical linear technique for dimensionality reduction. PCA performs dimensionality reduction by projecting the data into a subspace of lower dimension[16]. Although there exist various techniques for dimension reduction, the PCA is the most popular linear technique. Different from PCA, which seeks efficient directions for representation, Fisher linear discriminant seeks directions efficient for discrimination by yielding the maximum ratio of between-class scatter to within-class scatter.

Assuming that we have a set of D-dimensional samples $\{X_1, X_2, ..., X_N\}$, N_1 of which belong to class ω_1, and N_2 to class ω_2. We seek for a scalar Y by projecting the samples X onto a line.

$$Y = W^T X \tag{4}$$

Of all the possible lines, we would like to select the one that maximizes the separability of the scalars.

- The mean values of the X and Y examples are:

$$\mu_1 = \frac{1}{N_i} \sum_{X \in \omega_i} X$$
$$and \quad \widetilde{\mu_1} = \frac{1}{N_i} \sum_{Y \in \omega_i} Y =$$
$$\frac{1}{N_i} \sum_{X \in \omega_i} W^T X = W^T \mu_1$$

- For each class the scatter is defined as an equivalent of the variance:

$$\widetilde{S_i^2} = \sum_{Y \in \omega i} (Y - \widetilde{\mu_i})^2 \tag{5}$$

and the quantity $(\widetilde{S_1^2} + \widetilde{S_2^2})$ is called the **within-class scatter** of the projected examples. Thus the Fisher linear discriminant is defined as the linear function $W^T X$ that maximizes the following criterion function:

$$J(W) = \frac{\mid \widetilde{\mu_1} - \widetilde{\mu_2} \mid}{\widetilde{S_1^2} + \widetilde{S_2^2}} \tag{6}$$

- Therefore, we will be seeking for a projection where examples from the same class are projected very close to each other and, at the same time, the projected means are as farther apart as possible.
- For this purpose, we need to express $J(W)$ as an explicit function of W in order to find W^*. For this reason we define the equivalent of the scatter in the projection which, in multivariate feature space become scatter matrices.

$$S_i = \sum_{X \in \omega_i} (X - \mu_i)(X - \mu_i)^T \tag{7}$$

$$S_1 + S_2 = S_w \tag{8}$$

The matrix S_w is called the within-class scatter matrix and is proportional to the sample covariance matrix.

- The scatter of the projection can be expressed as a function of the scatter matrix in the X feature space.

$$\widetilde{S_i^2} = \sum_{Y \in \omega i} (Y - \widetilde{\mu_i})^2$$
$$= \sum_{X \in \omega_i} (W^T X - W^T \mu_i)^2$$
$$= \sum_{X \in \omega_i} W^T (X - \mu_i)(X - \mu_i)^T W$$
$$= W^T S_i W$$

$$\widetilde{S_1}^2 + \widetilde{S_2^2} = W^T S_w W \tag{9}$$

$$(\widetilde{\mu_1} - \widetilde{\mu_2}^2) = W^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T W$$
$$= W^T S_B W$$

The matrix S_B is called the **between-class scatter**.

- Finally, the Fisher criterion can be expressed in terms of S_B and S_w:

$$J(W) = \frac{W^T S_B W}{W^T S_w W} \tag{10}$$

- Solving the generalized eigenvalue problem $(S_w^{-1} S_B W = JW)$ yields

$$W^* = argmax_w \left\{ \frac{W^T S_B W}{W^T S_w W} \right\} = S_w^{-1}(\mu_1 - \mu_2) \tag{11}$$

This is known as Fisher's Linear Discriminant, although it is not a discriminant but rather a specific choice of direction for the projection of the data down to one dimension.

After computing the RJIH, we make use of the Adaboost for feature selection. Its basic principle is detailed in the following section.

III. Basic Adaboost

The Adaboost technique (Adaptive Boosting) [17] presents a popular technique for selecting a set of more performing weak classifiers from a pool of over complete weak classifiers. A boosting algorithm is able to construct a strong classifier by a linear combination of weak classifiers chosen from a huge amount of set. The single strong classifier obtained is much more reliable than the weaker ones. The algorithm is described in the table (Table I).

IV. Experimental Results

A. Choice of the parameters

The CMU-MIT data set consists of 130 images with 507 labeled frontal faces. The system was trained using 500 faces and 1000 non faces. For the validation set, we have used 100 faces and 300 non faces. Faces were cropped to images of size 19×19 pixels. The number of non-faces is higher than the number of faces in order to represent the disparity of existing patterns on real images. In fact, in real images there are much more non-face patterns than face patterns. The Haar-like features as image features are good references in the context of face detection. So, we compare the performance of our proposed features to the Haar-like features.

TABLE I
ADABOOST ALGORITHM

Given sample images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive samples respectively.

Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives samples.

For $t = 1 \ldots T$:

1. **Normalize the weights**

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$

so that w_t is a probability distribution.

For each feature, j, train a classifier h_j is restricted to using a single feature.

The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i |h_j(x_i) - yi|$.

2. **Choose the classifier**, h_t, with the lowest error ϵ_t.
3. **Update the weights:**

$w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$
where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$

The final strong classifier is:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

where $\alpha_t = log(1/\beta_t)$

TABLE II
COMPARISON OF DIFFERENT METHODS FOR JIH DIMENSION REDUCTION ON THE TEST SET USING A SINGLE STAGE COMPOSED BY 20 FEATURES

	JIH	JIH 32 bins	JIH 16 bins	RJIH
DR	31%	17%	28%	66%
FPR	11.66%	13.66 %	17.33%	8.7%

B. Comparative study

Firstly, we conduct experiment to compare between the dimension reduction methods. JIH, reducing the number of bins and RJIH. For a preliminary comparison of different methods, we compare the response of 20 features in terms of detection rate and false positive rate.

Experiments have been made for histograms with 256, 32 and 16 bins. The reduction of the number of bins reduces the complexity of JIH data structure. In fact, this leads to shorter classifiers and hence to faster classification. However, the obtained results show that the more bins used in the histograms, the better the detection rate and false positive rate. Thus, we intend to keep more number of bins and adopt a dimensionality reduction method. The experiments show that the use of the RJIH further improve the performance results. This can be illustrated by the ROC curves for the first stage comparing the three following methods: JIH, LDA based RJIH and Viola and Jones as a reference method in face detection.

In our previous work [9], we have proved that features

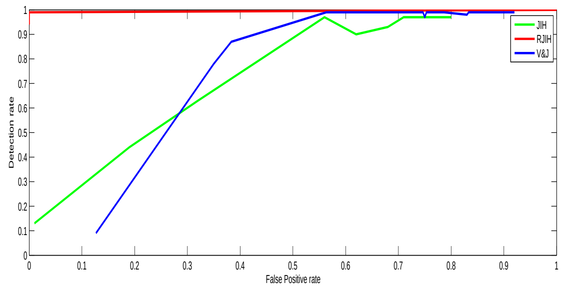

Fig. 2. The ROC curves for the first stage: JIH 256 bins, RJIH and Viola and Jones

based JIH are more effective than the conventional Haar-like features. However, the computational time of these features is higher. On the one hand, JIH based feature extraction increases the computational complexity. On the other hand, the reduction of the number of weak classifiers can be very important for the speedup of the detector. Thus, there is a compromise between the number of weak classifiers and their complexity computation. To further demonstrate the effectiveness of the RJIH, we should train a robust cascade structure and compare the speed of the final detector to the Viola an Jones detector.

V. CONCLUSION

In this paper, we proposed a new method called (RJIH) for feature extraction. Hence, the face patterns are presented by pairs of JIH based features and Fisher projection orientation under the Adaboost framework. Using the RJIH based features, we facilitate the computation of weak classifiers. This improvement leads to cost minimization when implemented on FPGAs architecture.

REFERENCES

[1] M. Yang, D. Kriegman and N. Ahuja, Detecting Faces in Images:survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.1, pp. 34-58, January 2002.
[2] A. Ferreira, Survey on Boosting Algorithms for Supervised and Semi-supervised Learning, Instituo de Telecomunicaoes, october 2007.
[3] C. Zhang and Z. Zhang, A survey of Recent Advances in Face Detection, Technical report, June 2010.
[4] P. Viola and M. Jones, Robust Real-time Object Detection, Second International Workshop on Statistical and Computational theories of vision modeling, learning, computing and sampling, Vancouver, Canada, July. 2001.
[5] R. Lienhart and J. Maydt, An Extended set of Haar-like Features for Rapid Object Detection, 2004.
[6] L. Zhang, R. Chu, S. Xiang and S.Z. Li, Face Detection based on Multi-block LBP representation, 2007.
[7] S. Yan, S. Shan, X. Chen, and W. Gao, Locally assembled binary (LAB) feature with feature-centric cascade for fast accurate face detection, In Proc. of CVPR, 2008.
[8] A. Roy, and S. Marcel, Haar Local Binary Pattern Feature for Fast Illumination Invariant Face Detection, Lausane, Switzerland, 2009.
[9] Joint Integral Histogram based Adaboost for Face Detection System, International Journal of Computer Applications, 2011.
[10] K. Zhang, G. Lafruit, R. Lauwereins and L.V. Gool, Joint Integral Histogram and its Application in Stereo Matching, International Conference on Image Processing, Hong Kong, 2010.
[11] K. Zhang, G. Lafruit, R. Lauwereins and L. Van Gool, Constant Time Joint Bilateral Filtering Using Joint Integral Histograms, IEEE Transactions On Image Processing, Vol.21, NO.9, September, 2012.

[12] T. Ojala, M. Pietikainen and T. Maenpaa, Multi-resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patternes, IEEE Trans. on PAMI, 2002.
[13] C. Papageorgiou, M. Oren, and T. Poggio, A general framework for object detection, IEEE International Conference on Computer Vision, pp. 555562, Janvier 1998.
[14] H. Wang, P. Li, and T. Zhang, Histogram Feature based Fisher linear discriminant for face detection, 2007.
[15] F. Porikli, Integral Histogram: A fast way to extract histograms in cartesian spaces, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
[16] L. van der Maaten and E. Postma, Dimensionality Reduction: A comparative Review, Tilburg center of Ceative Computing, Tilburg University, October 2009.
[17] Y. Freund and R.E. Shapire, A decision-thoretic generalization of on-line learning and an application to boosting, European Conference on Computational Learning Theory, pp. 23-27, 1995.