Enhanced Weakly Trained Frontal Face Detector for Surveillance Purposes

Wael Louis, K.N. Plataniotis, and Yong Man Ro

Abstract— **Face detection is becoming popular in surveillance applications; however, the need of enormous size face/non-face dataset, large number of features, and long training time are persistent problems. This paper claims that only a subset of the total number of features conserves the major power to detect faces; hence, this subset is capable to detect faces with high detection rate. The proposed detector fuses the results of two classifiers where one is trained with only 40 Haar-like features and the other is trained with only 50 LBP Histogram features. A pre-processing stage of skin-tone detection is applied to reduce the false positive rate. The detector is examined on real-life low-resolution surveillance sequence. Conducted experiments show that the proposed detector can achieve a high detection rate and a low false positive rate. Also, it outperforms Lienhart detector and tolerates wide range of illumination and blurring changes.**

I. Introduction

Extensive research for face detection has been conducted since the 1970's as face detection can be used for surveillance purposes, human tracking, human-computer interaction, and many other applications. Many face detection techniques are presented in Yang et al. [1]. Image-based face detectors (i.e. detectors that consider the face detection problem as two-class pattern recognition problem) require a tremendous amount of time in the designing stage [2][3] in order to achieve the desired outcomes. The long training time is a result of the increased system complexity due to using large number of training face/non-face dataset, extracting large number of features, using different classification schemes, or other factors. Also, large amount of time is required in the pre-training stage where much time is to be spent collecting billions of non-face images in addition to collecting the face dataset then modify them by cropping and aligning them. Most of the conducted research tries to solve the designing stage issue by improving the classification schemes that can reduce some of the discussed factors; however, comparably less research is conducted to find features with high discriminative power, and even lesser research is conducted trying to implement a detector that requires small face/non-face dataset which will reduce both the pre-training and training stages.

One of the most famous image-based face detection techniques is the Viola and Jones [2] which uses the Haar-like features [4]. This technique is considered as a breakthrough in the face detection due to its high detection rate and high speed of processing using the AdaBoost algorithm with cascade of classifiers. However, the Haar-like features have a limited discriminative power; therefore, in order to achieve the desired outcomes, Viola and Jones used $6,060$ Haar-like features distributed on 38 layers of cascade of classifiers. These $6,060$ features are chosen from over $\approx$ $83,000$ features in the case of Viola and Jones where 24×24 pixel windows were used, also billions of non-face images were needed. Hence, weeks of training time was required. Furthermore, higher discriminative power feature called Improved Local Binary Patterns is used in Rodriguez [5] where 450 LBP features was distributed on 3 cascade of classifiers. Even though this method succeeded in reducing the number of cascade of classifiers but it required large face and non-face images. Again much time was required to establish this detector. Another approach was introduced by Fröba et al. [6], where Modified Census Transform (MCT) features were used. These features proved the capability to extract face discriminative features; however, the number of features is limited to the number of pixels, so two classification schemes were required. Another approach is also introduced by Chen et al. [7], instead of targeting the type of features, it tries to reduce the number of training dataset by obtaining an optimal subset of the full dataset. A large faces dataset was required in order to choose the optimum subset from. Shen et al. [8] targets the training and pre-training stages where a high power covariance features as well as small training dataset was used; however, extracting the covariance features is computationally expensive.

The method we are proposing is to implement a frontal face detector that is trained with few number of features extracted from a small face/non-face dataset. Hence, this detector can be trained within a short period of time, and its small face/non-face dataset can be collected within reasonable time (i.e. collecting thousands in comparison to billions in [2][5]). The detector is to be used for surveillance purposes using a 2D information from static camera mounted in a position where mostly frontal faces are captured. Two contributions are presented in this paper: first contribution lies in proving that only small subset of the total features contributes to the major part of the results; hence, we only use these few features to implement what we call Enhanced Parallel Detector (EPD). Second contribution lies in using a pre-processing stage using skin-tone detector to eliminate the pixels that belong with a high probability to non-skin elements. The EPD detector extracts two types of highly proven face discriminative features to train two classifiers

Wael Louis and K.N. Plataniotis are with the Multimedia Lab, The Edward S. Rogers Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Canada (e-mail: wlouis@comm.utoronto.ca).

Yong Man Ro is with the Image and Video System Laboratory, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon, 305-701, Korea.

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

cascaded in parallel. The decision is made by an equal weighted fused decision from both classifiers. Therefore, this method is designed for finding the faces instead of discarding the non-faces. Finally, unlike other explained techniques that consists of only one stage for face detection and non-face rejection, EPS works in two stages: first stage concentrates on non-face windows reduction and the second stage concentrates on face windows detection.

The paper is organized such that section II describes the technical details including the pre-processing stage, classification scheme, feature extraction, and multi-detection merging. Followed by section III where the conducted experiments as well as the training and evaluation phases are discussed. Finally the conclusion is on section IV.

II. Technical Details

Enhanced Parallel Detector (EPD) is designed such that it uses two different types of features to train two classifiers cascaded in parallel. Each of the classifiers is to be trained with few features that span the highest discriminative power. Each type of features targets different image structure. One is the Haar-like features [4], and the other is the LBP Histogram (Local Binary Pattern Histogram) features [3]. These features are specifically chosen as they can complete each other from an aspect that one type is proven to extract high face discriminative features which is the LBP Histogram, but on the other hand it's extraction is computationally expensive. The second type of features, Haar-like features, can be extracted very quickly, but has less face discriminative power than the LBP Histogram features. Each of the features will have different criterion in detecting the face as each type targets different image structure. LBP Histogram features gives an excellent description for corners, edges, spots and flat ends whereas Haar-like features produces excellent description for small details structure, edges and bars.

In image-based approach detector, the classifier is usually trained with many discriminative features in order to reduce the number of false positives while preserving the true positive. Hence, if the classifier is trained with few number of feature, then it will be capable to detect faces but also will have many false positive detections. Hence fusing the decisions of each classifier, which are trained with few number or features, will preserve the true positive detection and drastically reduces the false positive. Even with the reduction of false positives, applying this detector in real-life scenario where thousands of subwindows are examined will make it face a challenging situation.

Therefore, the EPD detector optimizes the idea of fusing the weakly trained detector [9] by adding a pre-processing step which drastically reduces the number of subwindows to be examined. So the EPD detector works as two stage detector: first stage is the pre-processing stage that concentrates on non-face windows rejection while the second stage, fused detector, concentrates on face window detection. By doing so the system will: (1) be trained with small number of features from small dataset, (2) preserve a high detection rate, (3) drastically reduce the false positive detections. The complete diagram for the EPD detector is as shown in Figure 1.

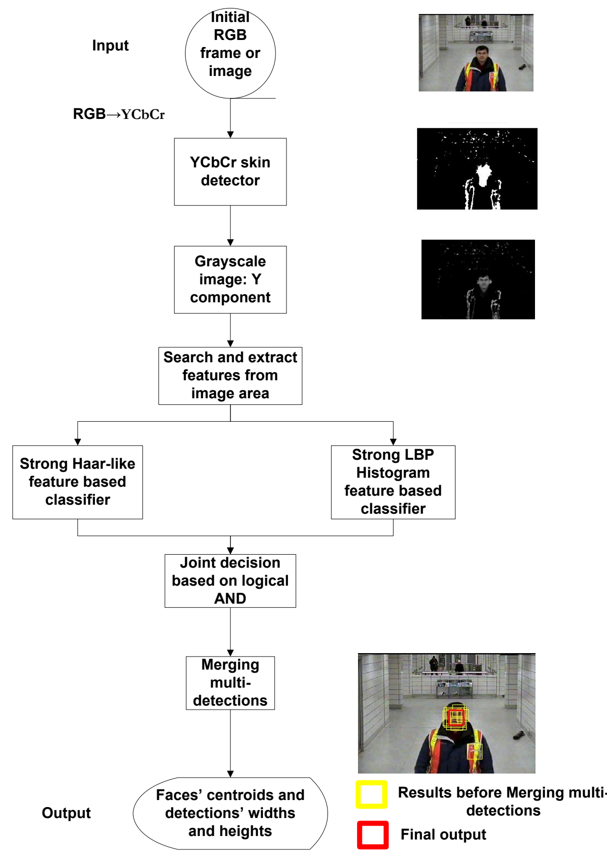


Fig. 1. Enhanced Parallel Detector diagram

A. Pre-processing stage using skin-tone detection

Fusing the results of the weakly trained detectors preserves the performance of the detection rate as well as keeps the false positive in a low level; however, applying the fused detector on a difficult dataset such as surveillance video will produce a high number of false positive detections. In order to keep the high detection rate and reduce the false positive, a skin detection technique is used as it is invariant to skin orientation and size. In a survey by Kakumanu et al. [10] it can be seen that many colorspaces were used for skin-tone detection. The main point of this pre-processing step is to discard as many non-skin pixels as possible so the following classifier will have less number of subwindows to examine. Therefore, this step is not a decision maker so a simple to implement skin-tone detector with small false negative rate (i.e. deciding a skin pixel as a non-skin pixel) is used. Despite there are many models in the literature; however, the YCbCr colorspace is assumed that is capable to separate the skin from non-skin color range by reducing the overlapping after transformation from the RGB colorspace. Chai et al.[11] implemented a simple technique using YCbCr for skin detection based on video sequence; this technique is used as a pre-processing step.

Chai et al. uses the Cb and Cr component to depict that a pixel is considered as skin pixel when its Cb and Cr values lie in the range

$$77 \leq Cb \leq 127, 133 \leq Cr \leq 173$$

.

B. Feature extraction

1) LBP Histogram: Simple LBP feature extraction algorithm operates by considering the value of the center pixel in a 3×3 pixels in a grayscale window as a threshold. It then assigns 1 to the neighboring pixels with a value greater than the threshold else it assigns it to 0. The resulted binary value is converted to decimal. This decimal value preserves the texture for this 3×3 pixel window. LBP features can be extracted with a circular neighbor ($LBP_{P,R}$) [12] (Circular LBP), with different radii (R) and points (P) as in Figure 2. There are 2^P binary words for each $LBP_{P,R}$. Also, there is subset of the 2^P words that spans most of the texture descriptor called uniform LBP $LBP^{u2}_{P,R}$ [12]. $LBP^{u2}_{P,R}$ are the words that have only two flipping bits from 0 to 1 and 1 to 0 (e.g. 01110000).

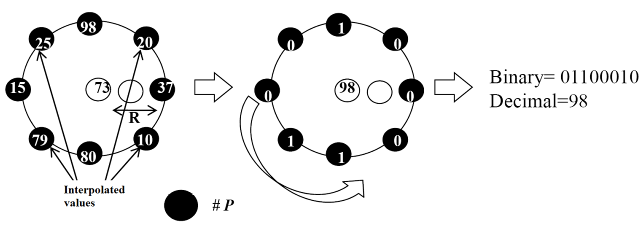


Fig. 2. LBP feature extraction

The LBP Histogram feature proved the ability to capture facial representation [3]. The image is divided into regions, and each region's $LBP_{P,R}$ is extracted, then the histogram for the region is calculated. The histograms for all regions are concatenated, and each bin is considered as a feature. The histogram of the LBP values from the image works as a descriptor for the image.

The LBP Histogram feature extraction in the proposed detector is designed to extract features in two phases, first phase is to extract the features based on the entire face image to obtain the overall description. For this purpose two LBPs feature extraction are used, one is the $LBP^{u2}_{8,1}$ and the other is $LBP^{u2}_{12,2}$. The second phase is to detect the smaller face description; therefore, $LBP_{4,1}$ is used with overlapping windows to target many places of the face. The LBP Histogram features extraction procedure is illustrated in Figure 3.

In Hadid et al. [3] only two LBPs were used so almost same preference is given for information obtained from within the face region (i.e. using small overlapping windows) and information obtained from the entire face (i.e. no overlapping window). However, to get more distinctive

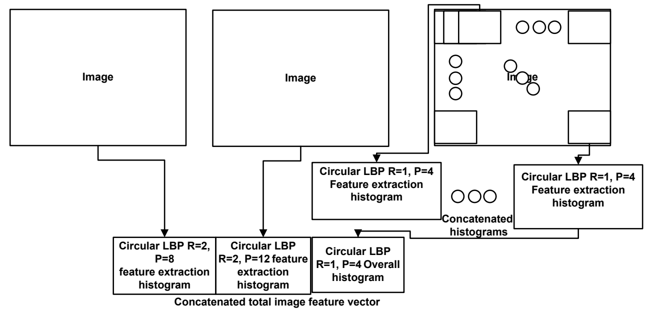


Fig. 3. LBP Histogram feature extraction

features, three LBPs features extraction combination are used, where two of them are for the overall face. The explained combination gives the best results on the tested data. Having higher number of combination more than three increases the computation complexity and wouldn't increase the performance significantly.

2) Haar-like features: The Haar-like features extraction is based on the work of Viola and Jones [2] which subtracts the sum of pixels of grayscale image in two adjacent rectangles. These two rectangles, one is considered as black region and the other is white region. The final feature result is the subtraction of the sum of the white region from the sum of the black region. These adjacent rectangles can be in different templates such as being two horizontally or vertically adjacent, or three adjacent rectangles in horizontal or vertical template where the black region is in the middle, or four adjacent regions in a square shape, where the white region in the main diagonal and the black region in the secondary diagonal. The Haar-like features works well in places where there are differences in brightness in the object as in the human face such as the area between the eyes and the forehead or the area between the eyes and the bridge of the nose. These two places are considered as the most discriminative Haar-like features in the human's face [2]. The feature extraction starts by a feature of smallest size (i.e. 2 pixels in the 2 rectangle template) and exhaust the image many times by keep increasing the size of feature until it reaches one of the dimensions of the image. This search is performed quickly using the integral image.

C. Classification

The AdaBoost [13] method uses weak classifiers $h_i(X)$, where each $h_i(X)$ is a single feature. The AdaBoost algorithm weighs and select the best n weak classifiers, where each weak classifier minimizes the classification error to construct a strong classifier $H(X) = \sum_{i=1}^{n} \alpha_i h_i(X)$. At each iteration of the boosting, the best weak classifier $h_i(X)$ is chosen, and the weight α_i is increased to the wrongly classified samples for the next iteration.

Many versions of AdaBoost are explained in the literature where all of them have the same concept but might differ in one or more of either error calculation, weight update or feature selections criterion. Based on Lienhart et al. [14], experiments conducted on the Discrete AdaBoost, Real

AdaBoost, and GentleBoost showed that GentleBoost gives the best results in the face detection problem. Therefore, GentleBoost [15] algorithm is used to construct the strong classifier.

D. Multi-detection merging

Multi-detections issue is considered as a drawback in many appearance based face detector due to the fact that the features are insensitive to small error changes [2][5]. While the scanning window scans the image, many detections might be located for the same detection. However, the number of detections in false positive regions are significantly less than those multi-detections in the faces regions.

The algorithm we use is based on only one parameter. It finds the centroid position of each detection, then cluster these positions with mode finding using the Euclidean distance from the points to cluster's centroid regions as a similarity measure.

The threshold β is considered as the minimum number of particles in each cluster to be considered as a detection. All the clusters that don't pass this test get deleted. The remaining detections' mean values are taken as the centroid of that detection. The size of the detected face is calculated as the average size of all remaining centroids in each cluster.

There is no rule of thumb to choose a generic β that works for all images. However it can be an application dependant since the bigger β the less the true positives and less false positives and vice versa when small β is used.

III. Experiments

A. Training dataset

The EPD detector is trained using Viola and Jones faces dataset, it consists of $4,916$ grayscale face images of size 24×24 pixels. No extra cropping, resizing and aligning are performed on the dataset. Also only $7,872$ grayscale non-face images of size 24×24 pixels are used in the training.

B. Evaluation datasets

Two datasets are used for evaluation purposes, first dataset is the Ole Jensen dataset [16]. Ole Jensen dataset contains $5,000$ grayscale face images and $10,000$ of grayscale non-face images of 24×24 pixels.

The other dataset is a real-life footage from a realistic environment where data became available to the University of Toronto team for research purposes. The footage is an RGB colorspace sequence taped by a camera mounted on the ceiling in vantage to capture frontal faces. The footage is of Codec Video 1 format with video rate of 5 frames/second. The sequence is of low resolution of 360×243 pixels. Faces appear in different sizes up to 60×80 pixels. 107 frames were examined that contained 89 faces. 59 frames are for single frontal face in different positions in order to examine the performance with different face sizes. 15 frames for two people appearing in the screen, to illustrate the ability of detecting more than one face, and finally 33 scenarios are for images where either a vacant place or non-frontal faces in the scene to inspect the false positive detections. Some of the real-life frames are shown in Figure 4

Fig. 4. Some of the examined frames

C. Method of evaluation

The evaluation is conducted based on the detection rate (DR) and number of False Positive detections (FP) when evaluated on the real-life frames. FP is the number of wrongly classifying a non-face region as a face region

$$DR = \frac{\text{TP}}{SF}$$

where TP (True Positive) is the number of correctly classified faces as faces and SF is the number of faces in the sequence.

Another evaluation was using the performance rate (PR) when the evaluation is conducted on the Ole Jensen face/non-face dataset.

$$PR = \frac{TP + TN}{T}$$

where TN (True Negative) is the number of correctly classified non-face as a non-face, and T is the total number of images in the dataset.

The decision whether the detection is considered as a face or not is conducted automatically using Lienhart et al. [17] approach. The face is considered as correct detection when the Euclidean distance between the face and the detector's face is less than 30% of the width of the detection's size, also the detection's height and width are within $\pm 50\%$ of the actual face width. The actual face location and size is extracted manually from the real-life sequence. The face region is considered as the region from above the eyebrow to the end of the chin. This region is considered since the dataset that trained the system (Viola-Jones dataset) was originally cropped such that the bounding box from above the eyebrow to below the mouth and is increased by 50%. Hence the regions from the top of the forehead to the bottom of the chin is considered.

D. Highest discriminative feature selection

Following the main objective of this paper which is to train each classifier, LBP Histogram classifier as well as the Haar-like classifier, with small number of features claiming only few of the features contribute the major results and the rest of the features contribute little information.

The feature extraction is conducted as explained in section II-B that two types of features are to be extracted the LBP Histogram features and the Haar-like features. From section II-B.1, the LBP Histogram features of $LBP^{u2}_{8,1}$ and $LBP^{u2}_{12,2}$ are used for full face description; hence, window of size

24×24 pixels is used whereas for the smaller $LBP_{4,1}$ feature, a scanning window of size 12×12 pixels and shifting of 2 pixels are used. For each image there are 59 features from the $LBP^{u2}_{8,1}$, 135 features from $LBP^{u2}_{12,2}$. 784 features from $LBP_{4,1}$. The total number of features extracted from each 24×24 pixel image is 978 features. The 784 features of $LBP_{4,1}$ is calculated as there are $LBP_{4,1} = 2^4 = 16$ features in each 12×12 pixel window; therefore, the 24×24 pixel image is subdivided into 49 windows each of 12×12 pixels overlapped by 10 pixels. As a results, the number of features is $49\times7 = 784$ features. Since $LBP^{u2}_{8,1}$ and $LBP^{u2}_{12,2}$ features don't have overlapping windows, so they are based only on extracting the uniform features. The smaller the scanning window and smaller the shifting the more the features to be extracted hence more computation is required. The Haar-like features extraction is the same used in Viola and Jones[2]

The LBP Histogram and Haar-like classifiers were each trained with 10,20,30,..,100 features, the GentleBoost algorithm was used to train the detector and to choose the high discriminative features. The performance rate (PR) is measured on the Ole Jensen face/non-face dataset. The time for each training session is recorded and tabulated in Table I. Figure 5 shows the PR versus number of features the detector is trained with.

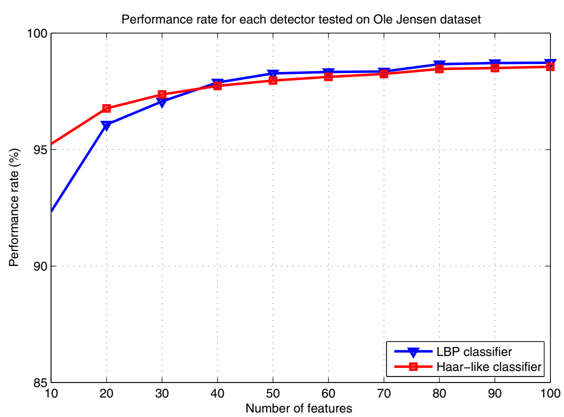


Fig. 5. Number of features trained the classifier versus performance rate for the LBP and Haar-like classifiers

It can be seen from Table I that the training time for the LBP Histogram classifier is significantly lower than the Haar-like classifier. This is expected since there are only 978 LBP Histogram features extracted from each 24×24 window in comparison to $\approx 83,000$ Haar-like features. Hence, the GentleBoost classifier in the case of the Haar-like classifier runs many folds over the LBP Histogram classifier to find the best features.

From Figure 5, it can be seen that for the LBP Histogram features, the system performance rate settles down after 50 features. Hence, the first 50 features contributed to 98.27% performance rate before reaching a plateau. Therefore, only the first 50 features were used to train our LBP Histogram classifier. On the other hand, same concept was adopted to decide the best number to train the Haar-like classifier. It can be observed that after the ≈ 40 features, the classifier starts reaching a plateau. On the other hand, taking under consideration the amount of time between two consecutive trained feature versus amount of improvement achieved, 40 features were chosen. Also, these 40 features are enough to provide 97.73% performance rate; whereas 50 features results a performance rate of 97.97%, and it requires extra 4 hours of training. Therefore, in the training session of the EPD detector, 50 features were used to train the LBP Histogram classifier while 40 features were used to train the Haar-like classifier.

Therefore we can conclude that only 50 out of 978 features of LBP Histogram features and 40 out of $\approx 83,000$ of Haar-like features conserve high discriminative power. This claim doesn't negate the fact that more features will result better detection (of course before over-fitting the classifier) or mention that the rest of the features are redundant; however, it proves that the rest of the feature contributes significantly lower than the first 40 or 50 features in our case.

E. Enhanced Parallel Detector results

Following the block diagram in Figure 1, the trained individual detector results are fused then merged as in section II-D, and the experiments are conducted on the real-life scenario. The effect of using the fused detector is examined by implementing the EPD detector diagram in Figure 1 without the pre-processing step in three different schemes: once using only the LBP Histogram classifier trained with 50 features, using only the Haar-like classifier using 40 features, and finally using the EPD detector where both Haar-like and LBP classifiers are trained with 50 and 40 features respectively and their fused result is obtained. The ROC is plotted for all operating points where β explained in section II-D is the parameter tweaked, the ROC is as seen in Figure 6. When evaluating the image, detector's window is scaled by 1.25 each time and its content is downsampled to 24×24 pixels.

Even though as shown from Figure 5, that when the individual face detector are applied on a face/non-face dataset, they prove the ability of distinguishing between face from non-face image with a high performance rate. However, applying the individual detectors on the real-life frames may result high false positive detections and low detection rate.

This issue lies behind two reasons:
(1) Multi-detection merging algorithm problem: the multi-detection algorithm is based on clustering, so having many false positive detections in the image will affect and move the centroid and size of the correct detections in the region, which leads to the second point.
(2) Method of evaluation problem: since an automatic decision is made based on Lienhart method of evaluation, then even if the individual detectors are able to detect the face (assuming only one face in the image), but the detection's centroid is moved due to the clustering method; then the tight boundary used by this method of evaluation will not meet. Hence, a misdetection will be encountered.

TABLE I

TIME IN HOURS REQUIRED TO TRAIN THE CLASSIFIERS USING INTEL(R) XEON(R) CPU X5355 2.66GHZ

Detector	10 Features	20 Features	30 Features	40 Features	50 Features
Haar-like (hr)	5.08	10.01	13.18	17.72	21.25
LBP Histogram ($\times 10^{-3}$ hr)	2.60	4.86	7.12	18.52	21.71
Detector	**60 Features**	**70 Features**	**80 Features**	**90 Features**	**100 Features**
Haar-like (hr)	23.96	26.69	30.32	31.94	34.53
LBP Histogram ($\times 10^{-3}$ hr)	26.93	17.62	21.34	26.82	38.43

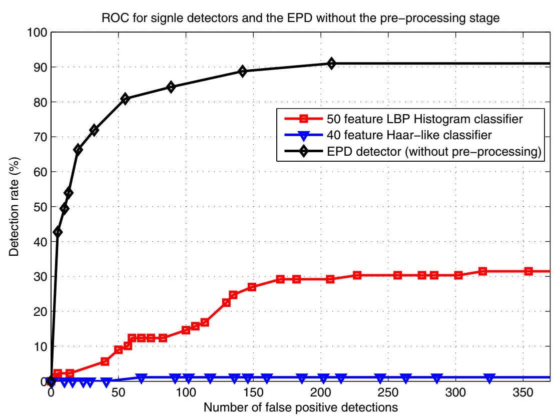


Fig. 6. Receiver Operating Characteristic for the EPD and the individual detectors

Figure 6 illustrates the detection rate versus the number of false positive detections, it proves the concept of the ability to achieve a significant improvement in the detection rate by fusing the detectors in comparison to individual detectors.

Furthermore, the complete EPD detector (with pre-processing stage) results is compared to Lienhart detector [14]. The GentleBoost algorithm model *haarcascade_frontalface_alt* is used since it uses same algorithm as the EPD detector. Lienhart detector is considered since it is a very close implementation to Viola and Jones detector, and it is available on OpenCV. The results are shown in Table II.

TABLE II

DETECTION RATE (%) COMPARISON BETWEEN THE EPD DETECTOR AND THE LIENHART DETECTOR

Detector	0 FP	1 FP	2 FP	6 FP	142 FP	208 FP
EPD	69.66	89.88	94.38	-	-	-
Lienhart	73.03	75.28	76.220	77.52	88.76	91.01

Therefore, it can be seen that the EPD detector worst case scenario is having 2 false positive detections and it is able to achieve a high detection rate while it is only trained with few number of features within $\approx$ 18 hrs using CPU of $2.66GHz$ in comparison to weeks of training in Viola and Jones with CPU of $400MHz$. Additionally, the EPD detector is trained with a total of 90 features in comparison to $6,060$ in Viola and Jones. Finally, it could outperform Lienhart detector which is trained with over 20 stages of classifiers.

F. Sensitivity analysis

The EPD detector robustness is examined in various artificial added illuminating and blurring scenarios. The robustness of the system as well as its comparison criterion to the Lienhart detector are explained. The β value was fixed for the highest detection rate for both EPD and Lienhart detectors of 94.38% and 91.01% respectively. For fair comparison, the degrading detection rate ΔDR percentage rate is measured. Where

$$\Delta DR = DR_0 - DR_n$$

Also the increase of the number of false positive ΔFP is calculated. Where

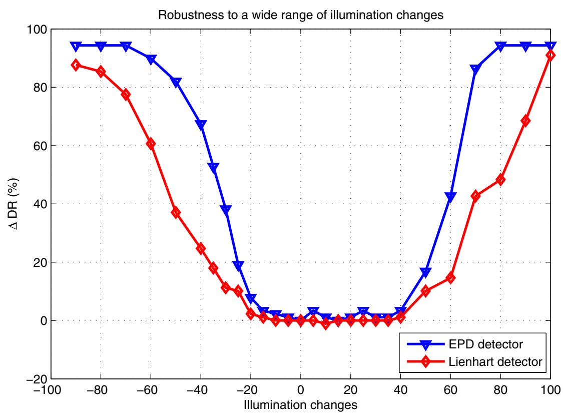
$$\Delta FP = FP_0 - DR_n$$

where DR_0 and FP_0 are the detection rate and the number of false positive detections for the non-noisy images respectively while DR_n and FP_n are the detection rate and number false positive detections for the noisy images respectively.

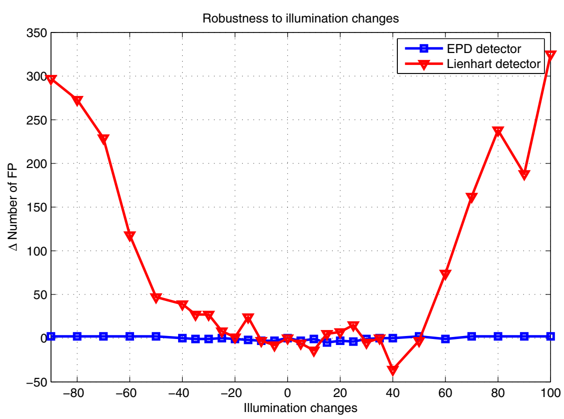
1) Illumination: The real-life examined frames were brightened and dimmed in the range of -100% to $+100\%$, this range of illumination is as seen in Figure 7.

Fig. 7. Sample of the brightening and dimming range

The robustness of the added brightness percentage versus the detection rate changes is plotted for both the EPD and Lienhart detectors to show the tolerance of the EPD detector to illumination changes, and to compare the EPD detector's robustness to that of Lienhart's detector. The EPD detector and Lienhart detector detection rate changes are shown in Figure 8 (a), also the detectors' false positive changes are shown in Figure 8 (b).



(a) Changes in the detection rate



(b) Changes in the number of false positive detections

Fig. 8. Robustness to illumination changes (a) changes in the detection rate (b) changes in the number of false positive detections

2) Blur: A gaussian filter of variance 1, 1.4 and 2 pixels were used to examine the EPD detector and Lienhart detector robustness to blurring. The blurred images look as in Figure 9.

ΔDR and ΔFP for the EPD detector and Lienhart detector are tabulated in Tables III and IV respectively.

TABLE III

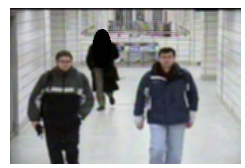
DETECTION RATE CHANGES, COMPARISON BETWEEN THE EPD DETECTOR AND LIENHART DETECTOR IN BLURRED FRAMES

Noise	EPD $\Delta DR\%$	Lienhart $\Delta DR\%$
1 pixel gaussian	7.87	2.24
1.4 pixel gaussian	15.73	0
2 pixel gaussian	30.34	14.6

(a) Original image

(b) 1 pixel gaussian blur

(c) 1.4 pixel gaussian blur



(d) 2 pixel gaussian blur

Fig. 9. The different blurring effect

TABLE IV

FALSE POSITIVE DETECTION CHANGES, COMPARISON BETWEEN THE EPD DETECTOR AND LIENHART DETECTOR IN BLURRED FRAMES

Noise	EPD $\Delta FP\%$	Lienhart $\Delta FP\%$
1 pixel gaussian	-3	-39
1.4 pixel gaussian	-1	-78
2 pixel gaussian	0	$+9$

IV. CONCLUSION

A frontal face detector that can be used for surveillance purposes is proposed. The Enhanced Parallel Detector (EPD) is trained with only few number of features extracted from a small face/non-face dataset within short training time. The EPD detector uses two weakly trained classifiers cascaded in parallel where each is trained with a subset of few number of high discriminative face features. The EPD detector is trained using only 90 features with the 40 and 50 highest discriminative power Haar-like and LBP Histogram features respectively. Hence, it is trained using $\approx \frac{1}{68}$ of the number of features used in Viola and Jones, and trained within ≈ 18 hrs.

Fusing the results of two weakly trained classifiers using only few number of two type of feature succeeded in achieving high performance results since each type of features targets different image structure; therefore, both features will agree on the face image and rarely agree on the non-face image. Logical AND is used to preserve the face image and discard the non-face image. Furthermore, a preprocessing step of skin-tone detector using YCbCr colospace is implemented to drastically decrease the number of false positive detections. The EPD detector has a different strategy for detection which is depicted by two stage strategy, where the first stage concentrates on rejecting the non-face images while the second stage highly concentrates on detecting the faces with less concentration on non-face rejection.

The EPD detector is applied on a real-life scenario and compared to the state-of-the-are Lienhart detector. With only the few features, the EPD detector could outperform Lienhart detector. Finally the robustness of the system was examined by artificially adding noise to the frames. One noise was by changing the illumination from -100% to $+100\%$ on the examined frames. Another noise is added by applying gaussian filter to blur the frames. The EPD detector proved its ability to handle a high range of illumination changes.

Acknowledgment

This work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The work is also partially supported by ORF-RE program of the Ontario Ministry of Research & Innovation through Multimodal SurvEillance System for SECurity-RElaTed Applications (MUSES_SECRET) project.

References

[1] M. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern analysis and Machine intelligence*, pp. 34–58, 2002.

[2] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[3] A. Hadid, M. Pietikainen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in *IEEE Proc. of CVPR*, vol. 2, 2004.

[4] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *IEEE Proc. of ICCV*, 1998, pp. 555–562.

[5] Y. Rodriguez, "Face detection and verification using local binary patterns," IDIAP-RR, École Polytechnique Fédérale de Lausanne, 2006, phD Thesis #3681 at the École Polytechnique Fédérale de Lausanne.

[6] B. Fröba and A. Ernst, "Face detection with the modified census transform," in *IEEE Proc. of FGR*, 2004, pp. 91–96.

[7] J. Chen, X. Chen, J. Yang, S. Shan, R. Wang, and W. Gao, "Optimization of a training set for more robust face detection," *Pattern Recognition*, 2009.

[8] C. Shen, S. Paisitkriangkrai, and J. Zhang, "Face detection from few training examples," in *IEEE Proc. of ICIP*, 2008, pp. 2764–2767.

[9] W. Louis and K. Plataniotis, "Weakly Trained Dual Features Extraction Based Detector for Frontal Face Detection," *To appear in IEEE Proc. ICASSP 2010.*

[10] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[11] D. Chai and K. Ngan, "Face Segmentation Using Skin-Color Map in Videophone Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 551–564, 1999.

[12] O. T., P. M., and M. T., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 871–987, 2002.

[13] Y. Freund and R. Schapire, "Experiments with a new boosting algorithm," in *IEEE Proc. of ICML*, 1996, pp. 148–156.

[14] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *IEEE Proc. of ICIP*, vol. 1, no. 1, 2002, pp. 900–903.

[15] J. Friedman, T. Hastie, and R. Tibshirani, "Special invited paper. additive logistic regression: A statistical view of boosting," *The annals of statistics*, vol. 28, no. 2, pp. 337–374, 2000.

[16] O. Jensen and R. Larsen, "Implementing the Viola-Jones Face Detection Algorithm," Master's thesis, Technical University of Denmark, 2008.

[17] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," *Lecture Notes in Computer Science*, pp. 297–304, 2003.