

Research Article

Co-Occurrence of Local Binary Patterns Features for Frontal Face Detection in Surveillance Applications

Wael Louis and K. N. Plataniotis

*Multimedia Laboratory, The Edward S. Rogers Department of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, Canada M5S 3G4*

Correspondence should be addressed to Wael Louis, wlouis@comm.utoronto.ca

Received 4 May 2010; Revised 16 September 2010; Accepted 9 December 2010

Academic Editor: Luigi Di Stefano

Copyright © 2011 W. Louis and K. N. Plataniotis. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Face detection in video sequence is becoming popular in surveillance applications. The tradeoff between obtaining discriminative features to achieve accurate detection versus computational overhead of extracting these features, which affects the classification speed, is a persistent problem. This paper proposes to use multiple instances of rotational Local Binary Patterns (LBP) of pixels as features instead of using the histogram bins of the LBP of pixels. The multiple features are selected using the sequential forward selection algorithm we called Co-occurrence of LBP (CoLBP). CoLBP feature extraction is computationally efficient and produces a high-performance rate. CoLBP features are used to implement a frontal face detector applied on a 2D low-resolution surveillance sequence. Experiments show that the CoLBP face features outperform state-of-the-art Haar-like features and various other LBP features extensions. Also, the CoLBP features can tolerate a wide range of illumination and blurring changes.

1. Introduction

Recently, surveillance cameras and Closed-Circuit Television (CCTV) are available in places that are highly occupied with people such as subway stations, airports, universities, and casinos. Increasing the number of cameras makes it a difficult task for the human to effectively monitor many cameras for suspicious activities simultaneously. As a result, much research is conducted to implement an intelligent system that mimics the human brain and achieves the monitoring job automatically with minimum human intervention. Monitoring the people is one task where finding the location of the face is used for surveillance applications; indeed, locating the place of the face is a preliminary step for other surveillance-related applications such as face recognition (i.e., find the identity of the person), emotion recognition, face tracking, and many other applications.

Humans can effortlessly detect and locate faces in real-life; however, locating faces using computer vision technology is not an easy task. Here we come to the problem of face detection for computer vision which can be interpreted as follows: given a still image or a video sequence frames, find

the *location(s)* and *size(s)* of the *face(s)* in the frame or image if *exist(s)* [1].

The face detection problem is tackled by numerous different methods since the 1970s. Face detection surveys in [1, 2] illustrate a very comprehensive study on face detection that covers most aspects of face detection techniques up to 2002. Despite the many approaches and techniques in these surveys, none of these techniques was adequate to perform in real-time basis. The term real-time, in this paper, is interpreted as the ability to process frames with a rate close to the examined sequence frame rate, under the condition that the frame rate ≥ 15 frames/second, that is, 15 frames/second in [3]. The first technique that could process in real-time basis was introduced in [3] in which it was published after these surveys.

Face detectors techniques can be divided into two main streams [1]: *featurebased approach* where human knowledge is used to extract explicit face features such as nose, mouth, and ears. then geometry and distances are used to decide on the existence of the face [4–6]. The advantages of this approach are the simple implementation, the high detection rate in simple uncluttered images, and the high tolerance to

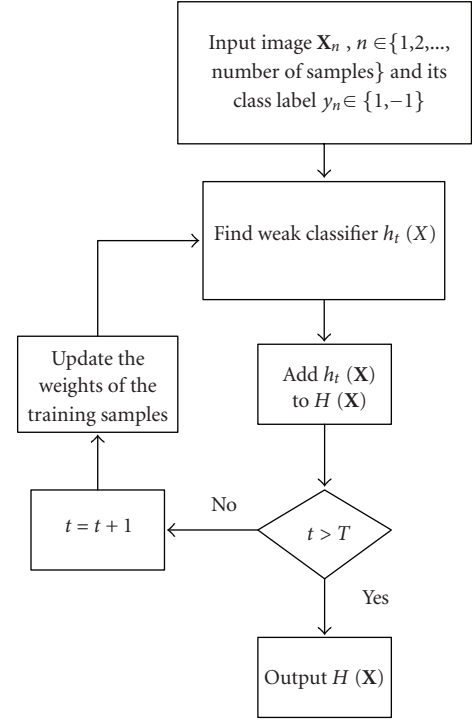
illumination changes. The disadvantages of this approach are the dramatic performance failure in cluttered and difficult, nonfrontal face, multiple faces, and low-resolution images. The second stream is the *image-based approach*; in this approach, the face detection problem is treated as a binary pattern recognition problem to distinguish between face and nonface images. This approach is a holistic approach that uses machine learning to capture unique and implicit face features. Most of the works in the last decade, including this work, follow the image-based approach due to the superiority over the featurebased approach from an aspect of the capability to handle low-resolution images, nonfrontal face images, and the possibility to process images in real-time. However, tremendous amount of time (i.e., weeks in [7]) is required to train the detector due to several problems, some of these problems will be explained briefly.

Since the image-based approach is the core of this paper, then it will be discussed. Image-based approaches can be categorized, based on the classification strategy used in the design process, into two subcategories: *appearance-based approach* and *boosting-based approach* [8]. Appearance-based approach category is considered as any image-based approach face detector that does not employ the boosting classification methods in its classification stage. However, other classification schemes are used such as neural networks [9, 10], Support Vector Machines (SVM) [11], Bayesian classifiers [12, 13], and so forth. All techniques in the appearance-based approach lack the ability to perform in real-time, and it takes an order of seconds to process an image [8]. The other image-based approach subcategory is the boosting-based approach, this approach started after the successful work of Viola and Jones [3] where high detection rate and high speed of processing (15 frames/second) using the AdaBoost (Adaptive Boosting) algorithm [14] and cascade of classifiers were used. Boosting-based approach is considered as any image-based approach that uses the boosting algorithm in the classification stage.

There are some problems associated with using the boosting-based approach to come about. Simple AdaBoost algorithm mechanism used in Viola and Jones is illustrated in Figure 1. AdaBoost uses a voting scheme from weak classifiers, $h_t(X)$, where X is the input image, t is the iteration number. These weak classifiers are used to construct a strong classifier, $H(X)$. The problem of the boosting algorithm in the face detection context is that each weak classifier, $h_t(X)$, is trained with a *single feature* as in [7, 11, 15, 16] following the pattern of the successful work of [3]. However, in earlier iterations, these single feature weak classifiers are capable of achieving a low classification error rate of 10%–30% [7] while in later iterations, $h_t(X)$ cannot achieve less than 40%–50% error rate.

This problem prevents the face detector from achieving a very highly accurate detection (i.e., $\approx 100\%$ accuracy), also it increases the number of small contributors $h_t(X)$ to achieve the desired accuracy in which will correspondingly increase the training time.

Although progresses are made to solve the explained face detection problem, the long training time and the insufficient number of discriminative features remain as challenging



t : boosting iteration number
 T : desired number of boosting iterations

FIGURE 1: AdaBoost training mechanism.

issues. The approaches followed in the literature to tackle this problem are either by focusing on improving the type of features or by improving the boosting algorithm, or a combination of both approaches. Haar-like features are used in [3], and extension to the Haar-like features have been proposed in [16–19]; however, Haar-like features have small tolerance to illumination changes [3]. Hence, the Local Binary Patterns Histogram (LBP Histogram) features which proved to have high tolerance to illumination [20] were used in [11], but LBP Histogram features are more computationally expensive than Haar-like features. LBP Histogram features [11] are not meant for real-time applications; they favor high discriminative power over speed. Hence, using these features prevent the face detector from processing images in real-time. Improved Local Binary Patterns (ILBP) features [8, 21] are less computationally expensive than LBP Histogram features; however, the number of ILBP features is limited to the number of pixels of the scanning window so a high detection rate, in comparison to LBP Histogram, cannot be achieved. Further extension to the LBP features is proposed in [22] in which Multi-Block LBP (MB-LBP) features are introduced. It is claimed in [22] that MB-LBP features are more informative than the Haar-like features, and they also have smaller feature vector length; hence, these advantages result to a faster training stage. Multidimensional covariance features in [23] are another type of features, but extracting these features is computationally expensive.

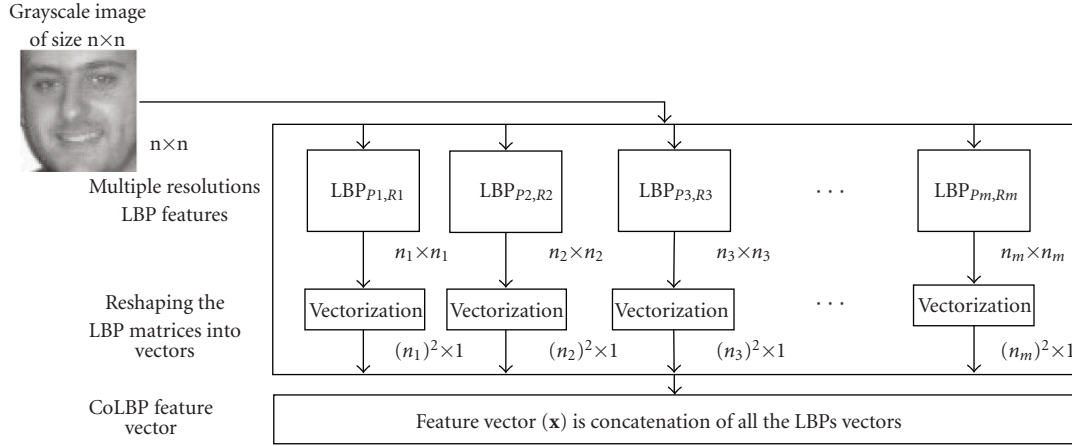


FIGURE 2: Single CoLBP feature vector extraction without co-occurrence of features.

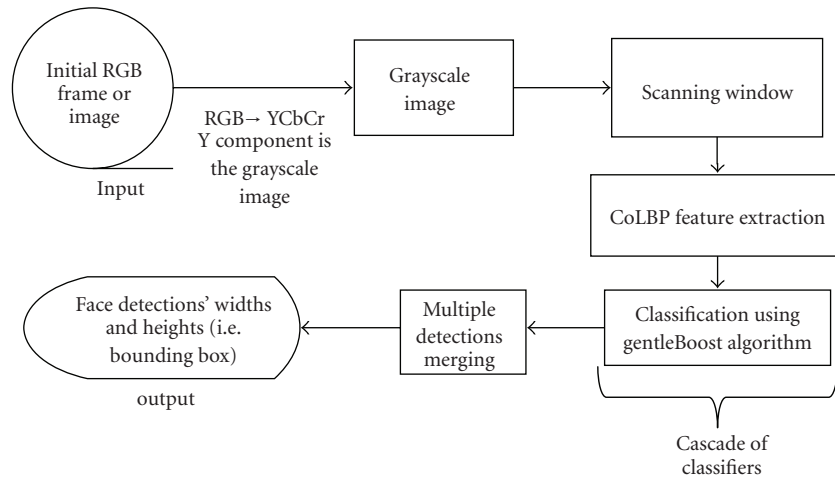


FIGURE 3: Illustration of face detection system using CoLBP features.

In this work, a type of features called Co-occurrence of Local Binary Patterns features (CoLBP) is proposed. The CoLBP features are used to implement a frontal face detector that is capable to achieve a high-performance rate. This face detector is used for surveillance purposes; it is applied on a low-resolution 2D information from a static camera mounted in a position where mostly frontal faces are captured. The proposed CoLBP features are based on the rotational LBP features [20]. This paper uses the rotational LBP with all possible resolutions in the examined scanning window to capture the maximum possible structure of the window that can be obtained using the rotational LBP operator. Unlike most of the known LBP features extensions in [11, 21, 22] where the pixels of the examined scanning window are transformed to LBP values, then the features are the histogram bins of the LBP values; in this work, the features are the LBP values of the pixels, as explained in Figure 2. Hence, extracting a feature of the proposed features (CoLBP) requires computing one pixel's rotational LBP value whereas in the histogram based LBP features as in [11, 21, 22], it requires to compute all of the examined

scanning window pixels LBP values in order to find their histogram bin. Therefore, the CoLBP features have less feature extraction computational overhead than histogram based LBP features. The main contribution of this paper is using the co-occurrence of multiple features to increase the feature's discriminative power. The multiple features are selected using the Sequential Forward Selection (SFS) algorithm. CoLBP features are not only computationally efficient but also provide high discriminative power capable to achieve a high detection rate.

The rest of the paper is organized such that Section 2 introduces the proposed CoLBP features; this section also gives a brief explanation about the classification scheme used to train the face detector and the post-processing step. The conducted experiments are in Section 3. Finally, the conclusion is given in Section 4.

2. Methodology

Figure 3 illustrates the CoLBP face detector that is based on the proposed CoLBP features. The Co-occurrence of the



FIGURE 4: Multiple resolutions $LBP_{P,R}$ on a gray-scale image to illustrate the ability to capture different image structure.

LBP features (CoLBP) features are built upon the rotational LBP features (see the Appendix for explanation about the rotational LBP features). The cascade of classifiers [3] and GentleBoost algorithm [24] are used to train the CoLBP detector. The multiple detections of the same face are merged using post-processing stage. Subsequent sections will explain in details the proposed CoLBP features as well as the method used to train the face detector using these features.

2.1. Co-Occurrence of Local Binary Patterns (CoLBP) Feature Extraction. LBP features drew much of the attention in object detection in general and face detection specifically due to its discriminative power as well as its high tolerance to illumination changes [11]. Detailed explanation about the LBP features and its feature extraction can be found in the Appendix. The main problem of the simple LBP features is that despite its capability to extract high discriminative face features [11], the number of features is limited to the number of pixels. This issue makes the simple LBP features insufficient to achieve a high-performance rate face detector. Various extensions are presented in the literature to solve this problem such as Sobel-LBP [25], CS-LBP [26], MB-LBP [22].

Despite the high discriminative power of the extended features [11, 22, 25, 26] in comparison to the rotational LBP feature, the *features* are the histogram bins of a region. Therefore, in the *classification stage*, in order to transfer the image from the pixel space into the feature space to be classified, all histograms' regions LBP features have to be computed to obtain the regions' distributions.

The proposed CoLBP features have the following advantages.

- (1) Less computational overhead than the extended LBP features that uses histogram bins as features as in [11, 22, 25, 26].
- (2) Overcomplete set of *discriminative* face features to achieve an accurate face detector; hence, it solves the problem of insufficient amount of information obtained by the simple LBP features. Overcomplete feature vector, in this paper, is a vector that its length exceeds the number of pixels of the examined window.

The CoLBP features proposed in this paper tackles the *computation overhead problem* by using the rotational $LBP_{P,R}$, where P, R correspond to the number of points and radius, respectively. Therefore, the features are the $LBP_{P,R}$ value of the pixels. Hence, in the *classification stage*, only

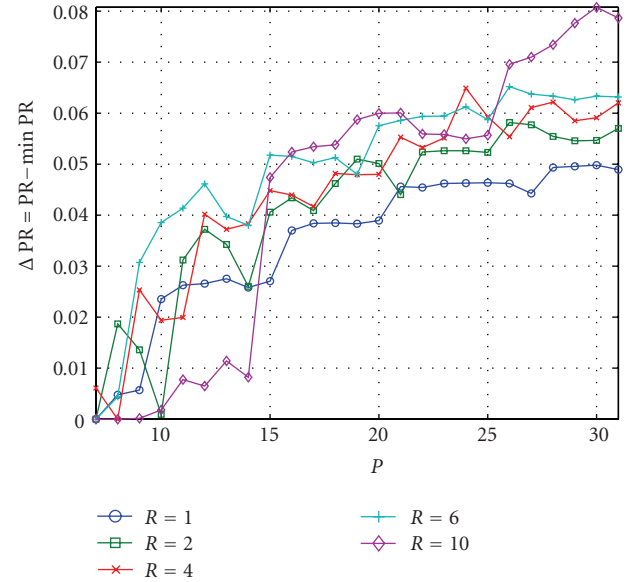


FIGURE 5: The relation between P and R in $LBP_{P,R}$.

the desired pixel's $LBP_{P,R}$ features are extracted. Therefore, CoLBP features are more computationally effective than histogram bins features.

Moreover, in order to overcome the problem of *limited features number* that prevents the system from providing enough information to achieve high-performance rate; the CoLBP feature vector is constructed by an exhaustive extraction for $LBP_{P,R}$ for all possible R s with different P values as illustrated in Figure 2. Each $LBP_{P,R}$ is considered as a resolution to capture the image structure [11]; hence, having different $LBP_{P,R}$ with different R s and P s capture the image structure with different resolutions [11] as it can be visualized through Figure 4. To be consistent throughout the paper, the name CoLBP, which stands for Co-occurrence of LBP features, is called on the feature vector that consists of multiple resolutions $LBP_{P,R}$ features. Hence, when Section 2.1.1 explains the co-occurrence of features, then it flows as “ c ” multiple features from the same vector CoLBP.

In order to choose P s for each R , an experiment is conducted that examines the relation of each possible R with a wide range of P s. P s that achieved the highest PR for each R are selected for the CoLBP feature vector. An example of this experiment is shown in Figure 5.

Hence, from Figure 5, the CoLBP feature vector consists of 3,164 rotational LBP features extracted from multiple $LBP_{P,R}$ resolutions. The rotational LBP features used are

Input: $(X_n, y_n, w_n)_{n=1}^N$, X_n is the training image and $y_n \in \{-1, +1\}$ is the class label, w_n is the weight, c is the number of co-occurred features;

Output: $z(X)$;

Initialize: $z(X) = \emptyset$;

(1) Retrieve the calculated CoLBP feature vectors $(x_n)_{n=1}^N$, $x_n \in \mathbb{R}^k$;

(2) Use decision stump to find all features $f(X) \in \mathbb{R}^k$ threshold values $\theta \in \mathbb{R}^k$ and parity values $p \in \mathbb{R}^k$;

for $i = 1, 2, \dots, c$ **do**

for $l = 1, 2, \dots, k$ **do**

$z_i^*(X)$ is concatenation of $z(X)$ with $s_i(X)$ that is trained with the l th feature, where $z_i^*(X)$ is a temporary $z(X) \triangleq (s_1(X), s_2(X), \dots, s_i(X))$;

$z_i^*(X_n)$ binarizes all examples $(X_n)_{n=1}^N$ using (2);

 Find the least weighted squared error of adding feature $l, J(l)$, using $z_i^*(X_n) = j$, where $j \in A$, $A = \{0, 1\}^i$, $\{0, 1\}^i$ is the Cartesian product of i terms;

for $n = 1, \dots, N$ **do**

 For all $j \in A$, find the estimated class label $\hat{y}_j(n)$;

$$\hat{y}_j(n) = \begin{cases} 1 & \text{if } P(y_n = +1 \mid z_i^*(X_n) = j) \geq P(y_n = -1 \mid z_i^*(X_n) = j) \\ -1 & \text{otherwise} \end{cases}$$

end

end

 Select $s_i(X)$ with the l th feature that makes $\arg \min_l J(l)$;

 Update $z(X)$ with $s_i(X)$;

end

ALGORITHM 1: SFS algorithm for selecting multiple CoLBP features.

LBP_{8,1}, LBP_{9,2}, LBP_{12,1}, LBP_{12,2}, LBP_{16,3}, LBP_{18,4}, LBP_{24,4}, LBP_{24,5}, LBP_{26,6}, LBP_{24,7}, LBP_{24,8}, LBP_{24,9}, LBP_{32,10}, and LBP_{32,11}.

To this end, it is shown how the CoLBP features can have less computational overhead than the histogram bins LBP features. Also, CoLBP has an overcomplete set of carefully selected discriminative features.

2.1.1. Co-Occurrence of Multiple CoLBP Features. The co-occurrence of features task can be defined as finding the joint probability of multiple features occurred simultaneously. Similar approach was recently proposed in [17, 18]. The objective of feature co-occurrence is claiming that a higher discriminative power can be achieved using the co-occurrence of multiple CoLBP features than taking same number of features separately. Therefore, in order to find the joint probability among multiple features, feature binarization is carried on as in [3]. Each feature $f(X)$ for the image X of the CoLBP has a threshold θ and parity $p \in \{1, -1\}$ calculated using a degenerative decision stump [27] from the training data such that the minimum number of examples are misclassified.

Having the parameters $(f(X), \theta, p)$, then given an input image X , $s(X)$ binarizes the input to 1 as being a face detection or 0 as being a nonface detection as in

$$s(X) = \begin{cases} 1 & \text{if } pf(X) > p\theta, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$f(X)$ is a single feature value, θ is a threshold values, and $p \in \{1, -1\}$ is the party that indicates the direction of the inequality.

This is a single feature binarization as in [3]. It is a specific case of a generalized case where more than one feature occur. Equation (2) shows the generalized form, $z(X)$, where it binarizes multiple features using(1)

$$z(X) \triangleq (s_1(X), s_2(X), \dots, s_c(X)), \quad (2)$$

where c is the number of co-occurred features, each $s_i(X)$ has $(f_i(X), \theta_i, p_i)$, where $i \in \{1, 2, \dots, c\}$.

Therefore, $z(X)$ is a vector which has c of the highest contributing $s_i(X)$ selected as in the following section. $z(X)$ has 2^c possible outcomes.

If $z(X)$ is used as a weak classifier for boosting-based approach detector with number of features co-occurrence of $c = 1$, then this $z(X)$ is similar to train each weak classifier with *single feature* as in many boosting-based approaches for the face detection problem in the literature including, but not limited to, [3, 11, 15, 16].

2.1.2. CoLBP Feature Selection. The combinations of $(s_1(X), s_2(X), \dots, s_c(X))$ in (2) are selected such that a minimum cost is achieved by $z(X)$. If $c = 1$, then selecting $z(X)$ for the minimum error is trivial since it is based on selecting one feature $f(X)$, and this $f(X)$ corresponds to the minimum error resulted using the decision stump. However, if $c > 1$ then finding $z(X)$ which achieves the minimum error is not an easy task. The optimal solution would be using the exhaustive search, where the solution is considered optimal from an aspect that the selected c co-occurred features achieve the minimum error. However, in the face detection problem, the feature vector dimension is usually in thousands (i.e., Viola and Jones had feature

vector \mathbf{x} , $\mathbf{x} \in \mathbb{R}^k$, where $k = 160,000$ features). Therefore, there are $\binom{k}{c}$ possible number of combinations for selecting $(s_1(\mathbf{X}), s_2(\mathbf{X}), \dots, s_c(\mathbf{X}))$ in $\mathbf{z}(\mathbf{X})$. As a result, a large number of combinations is possible; hence, many feature selection techniques have been proposed throughout the time, and below are some of them.

Sequential Backward Selection (SBS) [28] is a top-down approach that starts by a set that comprises all features then deletes features on one-by-one basis, where each deleted feature is the one that has the least contribution to minimize the error. On the other hand, the Sequential Forward Selection (SFS) [29] starts from a set of zero features and adds the feature that leads to minimum error. Both methods are suboptimal feature selection in comparison to the exhaustive search; however, they have the advantage of being less computational expensive than the exhaustive search method, and they are simple to implement. The disadvantage of SFS and SBS methods is the nesting effect in which the addition of the feature in the case of SFS or deletion of the feature in the case of SBS cannot be redone. For instance, in SFS, if a feature is added, then this feature will not be checked again whether it still has a high contribution to minimize the error; therefore, some features might lose their contribution after some iterations while they will still be considered. To solve the nesting issue, a method called Plus-l-Minus-r is introduced in [30] in which it adds l features using SFS and deletes r features using SBS; however, its main problem is the lack of theoretical approach to choose l and r . Also, it is more computationally expensive than SBS or SFS. An optimal decision is made by using the branch and bound method [31]; however, this method's complexity increases exponentially with the required number of features to be selected. The Sequential Forward Floating Selection (SFFS) and Sequential Backward Floating Selection (SBFS) [32] are similar to Plus-l-Minus-r in some sense but instead of being tied up by l and r , they keep adding and deleting features until minima is achieved. SFFS and SBFS methods are proved to outperform SFS and SBS [32].

From all these techniques a tradeoff between computation feasibility versus optimality is claimed in [32]. In this paper, the SFS is considered for the following reasons.

- (i) Despite that SFFS method proved the superiority over SFS [32]; however, same results were obtained when small number of features selection were needed (i.e., $c \leq 4$) [32]. Also, as will be proven experimentally in subsequent sections that the CoLBP face features tend to perform better when ($c < 4$). Therefore, using SFS or SFFS would give same results. One of the reasons that make SFS and SFFS perform similarly when c is small (i.e., 4) is that the nesting problem associated with SFS will not affect the system in the same manner as in the situation when several features are used.
- (ii) SFFS is more computationally expensive than SFS [32].

The implemented SFS is a modification to the original SFS [29] as it binarizes the input feature vector and finds

the highest contributing c co-occurred features. The SFS algorithm is illustrated in Algorithm 1.

$P(y = +1 \mid \mathbf{z}(\mathbf{X}))$, $P(y = -1 \mid \mathbf{z}(\mathbf{X}))$ are the conditional joint probability. Hence, for the input image \mathbf{X} , the binarization function $\mathbf{z}(\mathbf{X}) \in \{0, 1\}^c$, $\{0, 1\}^c$ is the Cartesian product of c terms, c is the number of co-occurred features, and $y \in \{-1, +1\}$ is the class label such that $y = 1$ is a face object and $y = -1$ is a nonface.

Therefore, $P(y = +1 \mid \mathbf{z}(\mathbf{X}))$, $P(y = -1 \mid \mathbf{z}(\mathbf{X}))$ are computed such that

$$\begin{aligned} P(y = +1 \mid \mathbf{z}(\mathbf{X})) &= \sum_{i \in \{n \mid y_n = +1\}} w_i, \\ P(y = -1 \mid \mathbf{z}(\mathbf{X})) &= \sum_{i \in \{n \mid y_n = -1\}} w_i, \end{aligned} \quad (3)$$

where w_i is the sample weight.

2.2. Classification. The proposed CoLBP features are used to implement a face detector (CoLBP detector). The CoLBP detector is trained using the GentleBoost algorithm [24] and cascade of classifiers [3]. The weak classifiers for the GentleBoost algorithm is $\mathbf{z}(\mathbf{X})$ obtained using c co-occurred features as explained in the previous sections. Therefore, for each boosting iteration, t , iteration's specific weak classifier, $\mathbf{z}_t(\mathbf{X})$, with the minimum error, J_t , is selected. Also, the weight of each sample w_n , where $n = 1, 2, \dots, N$, N is the number of samples, is increased for the misclassified samples and decreased for the correctly classified samples in each iteration of the boosting. Therefore, extra attention is given to the wrongly classified samples. In GentleBoost algorithms, the weighted squared error is used as an error measure. The final strong classifier is $H(\mathbf{X}) = \sum_{t=1}^T h_t(\mathbf{z}_t(\mathbf{X}))$, where $h(\cdot)$ is a confidence function that optimizes the GentleBoost's cost function. The overall training stage mechanism is as seen in Figure 1, the weak classifier step in Figure 1 is constructed by running Algorithm 1.

Many variations of boosting algorithms are explained in the literature [14, 19, 24, 27] where all of them have the same explained mechanism but might differ in one or more of either error calculation, weight update and/or feature selection criterion. The GentleBoost [24] was chosen for reasons such as GentleBoost proved the ability to outperform the Discrete AdaBoost and Real AdaBoost in face detection experiments [16], it is numerically stable, and it is simple to implement. Complete discussion about the GentleBoost algorithm can be found in [24].

2.3. Multiple Detections Merging. Applying image-based approach face detectors on an image will cause multiple detections for the same face as seen in Figure 6. The multiple detections occur for two reasons: first, due to the nature of the detection criterion where overlapping scanning window exhaustively search the image with different sizes and locations. Hence, there are windows where the difference in their content is small. Second, the classifier is trained to be insensitive to small localization error [3, 8, 9] so the classifier can handle different face variations. Despite

the multiple detections problem, trivially, the number of detections in nonface regions is significantly less than that in the face regions since the classifier is trained to achieve a high accuracy. Therefore, the algorithm used in this paper finds the centroid position of each detection, and then cluster these positions.

Furthermore, the multiple detection algorithm is based on a threshold β which decides the minimum number of detections in each cluster to be considered as a detection. All the clusters that do not pass this test are deleted. Afterwards, the detection within each cluster with the highest confidence values is only considered. The confidence of the detection is the value of the strong classifier $H(\mathbf{X})$.

3. Experiments

Previous sections have introduced the new CoLBP features with its method of extraction and properties studied. Also, the possibility of using the CoLBP features to train a classifier for object detection and a method to merge multiple detections are explained.

In this section, the explained CoLBP features will be applied for the face detection problem, and its performance will be evaluated. Specifically, it will be tested on a real-life 2D surveillance data, BioID dataset, as well as a face/nonface Ole Jensen and Viola and Jones datasets to investigate whether the proposed solution can achieve better detection results than the existing solutions.

The following experiments have been designed for the performance evaluation.

- (1) Evaluate the discriminative power of CoLBP features and observe the performance of co-occurrence of multiple features versus the separate ones.
- (2) Compare the CoLBP features to other various LBP features extensions presented in the literature.
- (3) Evaluate the performance of the CoLBP features-based face detector and compare it to the Haar-like features-based face detector.
- (4) Open the area to compare the CoLBP features face detector to the state-of-art face detectors by applying it on the BioID dataset.
- (5) Observe the robustness of the CoLBP features towards different illumination and camera blurring noise.

There are several terms that are usually used for face detection evaluation such as Detection Rate (DR), Performance Rate (PR), True Positive (TP), and False Positive (FP). DR is the ratio between the number of correctly classified faces to the total number of examined faces. PR is the ratio of the number of correctly classifying a face image as a face image and nonface image as nonface image to the total number of images evaluated. TP indicates correctly classifying a face image as a face. FP detection indicates incorrectly classifying a nonface image as a face.

3.1. CoLBP Features As Face Discriminative Features. The CoLBP features explained in Section 2.1 is examined in this



FIGURE 6: Multiple detections merging example.

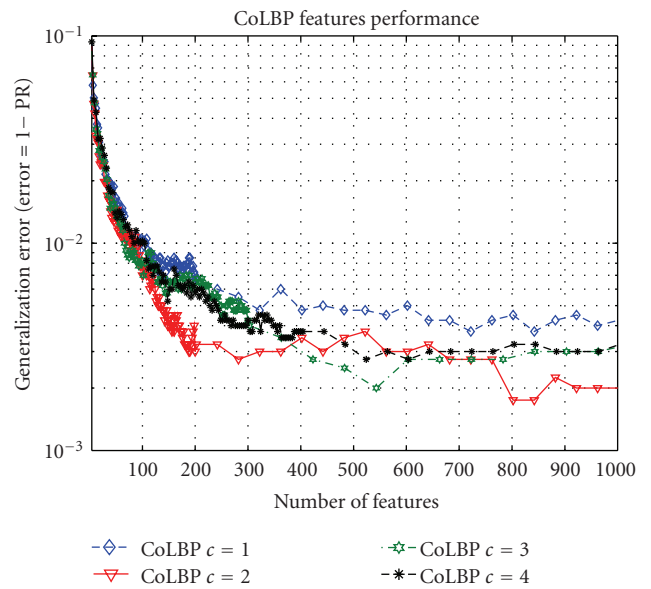


FIGURE 7: Generalization error for different number of co-occurred features.

experiment to prove the feasibility of providing face discriminative features. This experiment also compares different number of co-occurred features to prove the claim that having a co-occurrence of features produces higher performance rate than considering same number of feature separately.

The GentleBoost algorithm is used to train a classifier using different number of co-occurred features, c . The training and evaluation stages are performed using the Ole Jensen dataset [33] and its mirror images. Ole Jensen dataset consists of 15,000 gray-scale images of size 24×24 pixels, where 5,000 images are for frontal faces and 10,000 images correspond to nonface images. No extra cropping, resizing and aligning are performed on the dataset; hence, the dataset is used as it was provided in [33]. The training set contained 22,000 images, where 4,000 images and their mirror images correspond to face images and 7,000 images and their mirror images correspond to nonface images. Furthermore, 1,000 and 3,000 face and nonface images, respectively, and their mirror images are used as evaluation set. The evaluation is based on the performance rate (PR) measure.

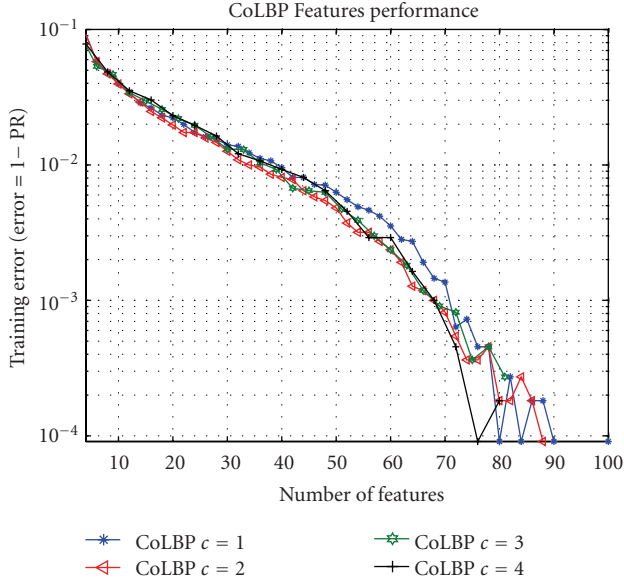


FIGURE 8: Training error for different number of co-occurred features.

The feature extraction of the CoLBP features follows the explanation in Section 2.1 such that all $LBP_{P,R}$ radii are extracted. Hence, there are a maximum of 11 possible R s for each window of size 24×24 . Having $R = 12$ will require a radius of 12 pixel radius excluding the center point; therefore, a diameter of 25 pixels is needed. Based on the different P s used, which are mentioned in Section 2.1, an overall number of features of 3,164 features are extracted from each window.

A comparison is made when $c = 1$, $c = 2$, $c = 3$, and $c = 4$ where the comparison is based on training the classifier with ≈ 1000 features. Figure 7 shows the generalization error whereas Figure 8 illustrates the training error for the same experiment. The error is calculated as $\text{error} = 1 - \text{PR}$. The number of features in Figures 7 and 8 are chosen such that it is unrelated to the number of iterations used to train the classifier. For example, following Section 2.2, the samples weights are updated in every iteration where the iteration consists of 1 CoLBP feature if $c = 1$ or 2 CoLBP features if $c = 2$, and so forth. Therefore, for fair comparison and to investigate the higher discriminative feature power using the co-occurrence of features then Figures 7 and 8 show the number of features the GentleBoost classifier is trained with.

Despite that all the c co-occurred features have almost same number of features when the training error converges to zero, all the co-occurred features outperformed the single CoLBP (i.e., CoLBP $c = 1$) features in the generalization error. The least generalization error resulted using $c = 2$, especially in early iterations. However, due to the fact that SFS reliability to select the best features decreases when the number of selected features increases, then it can be observed from Figures 7 that the performance of the system degrades when it is trained with $c > 2$. Especially if the comparison is conducted in early iterations.

To understand the significance achieved using $c = 2$ in comparison to $c = 1$ obtained in Figure 7, then taking an

arbitrary number of features, for example 50 features. The difference in PR between using $c = 1$ and $c = 2$ with 50 features is 0.004. So if an image of size 490×330 pixels is examined, and exhaustive search scanning window of size 24×24 with 1 pixel step size is used, then 143,369 windows is the total number of windows. Therefore, using $c = 1$ with 50 features will lead to an average of 573 wrongly classified windows more than that using same number of features with $c = 2$.

Co-occurrence of other types of features are introduced in the literature such as the co-occurrence of Haar-like features in [17, 18]. Despite it was claimed that the co-occurrence of Haar-like features increased their discriminative power however it was observed that this co-occurrence is prone to overfitting.

From this experiment, the following can be concluded:

- (i) CoLBP features are capable of extracting face discriminative features.
- (ii) Co-occurred features have higher discriminative power than separate features.

3.2. CoLBP Features versus Various Types of LBP Features.

Various extensions to the rotational LBP features are proposed in the literature including, but not limited to, the explained ones in Section 1. Therefore, to further prove the viability of the proposed CoLBP features, comparisons are made with Sobel-LBP [25], MB-LBP [22], and LBP Histogram [11] features. The comparisons are conducted in identical environment to examine the following: performance, training time, which consists of the time consumed to extract the feature vector and the time required to train a model, and classification time, which consists of the time required to extract the trained model features.

In addition to Ole Jensen dataset explained in Section 3.1; Viola and Jones dataset [3, 34] which consists of 4,916 gray-scale frontal face images of size 24×24 pixels and 7,972 gray-scale nonface images of size 24×24 is used. Both datasets' mirror images are obtained; hence, a total of 19,832 frontal face images and 35,744 nonface images are available. The datasets are divided into two halves; one half is used for training and the other half is used for evaluation. Therefore, the training dataset consisted of 9,916 face images and 17,872 nonface images of 24×24 pixels, and same number, but nonoverlapping, face and nonface images are used for the evaluation stage. The GentleBoost algorithm is used as a classifier. Hence, this experiment does not involve β explained in Section 2.3, and the decision is restricted to whether the detection is a face or a nonface image.

It is explained in [35, 36] that LBP Histogram features setup with the combination $(LBP_{12,2}^{u2}, LBP_{8,1}^{u2}, LBP_{4,1})$ gives higher performance rate than the LBP Histogram combination in [11] especially on the examined dataset; hence, this combination is used in this experiment. Furthermore, two setups for the Sobel-LBP, which is explained in [25], with 24×24 and 12×12 subregions are examined.

Figure 9 shows PR for different types of features using different number of features. It is clear from Figure 9, that CoLBP $c = 2$ outperforms all other types of features

TABLE 1: Comparisons between different types of LBP extensions. The times calculations are based on training and evaluating 27,788 images of size 24×24 pixels.

Feature type	Feature vector length	Training		Classification	
		Feature vector extraction time (sec)	Model training time (sec)	Trained model number of features	Model's feature extraction time (sec)
LBP Histogram	978	10.779	43.927	22	9.117
MB-LBP	8464	31.814	761.491	41	0.368
Sobel-LBP 12×12	2049	4.042	298.221	75	2.026
CoLBP $c = 1$	3164	25.245	219.693	38	0.296
CoLBP $c = 2$	3164	25.245	169.158	18	0.280

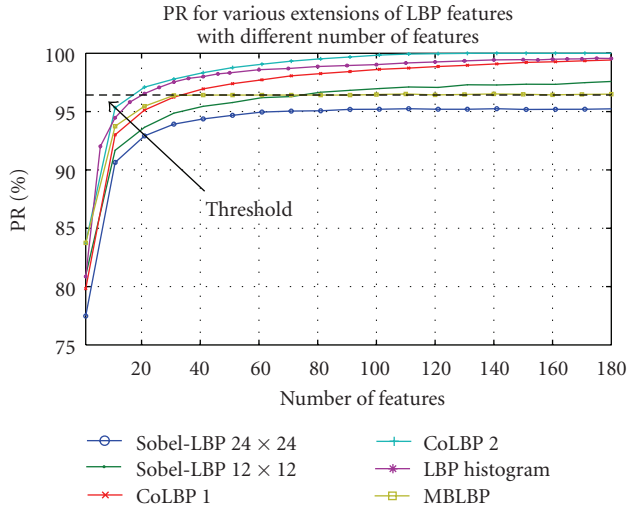


FIGURE 9: Comparison between the CoLBP features and various other LBP features extensions.

including the CoLBP $c = 1$. To have a fair comparison between these features, a common PR target threshold is chosen, and the corresponding model that achieves that threshold is considered. Hence, the comparison is conducted under identical environment from an aspect that all models are trained with same dataset, are evaluated on same dataset, used same classifier, and are capable of achieving an arbitrary PR of $\approx 96\%$. This arbitrary threshold was chosen since it is the maximum PR could be achieved with MB-LBP features, and it was desired that the MB-LBP features to undergo the comparison. Sobel-LBP with 24×24 pixels subregion was not considered in the comparison as it could not achieve the target PR. Having this threshold, the comparison results are tabulated in Table 1.

From Figure 9, CoLBP $c = 2$ trend outperforms all other experimented LBP extensions; however, judgement of the performance based on the number of features used to train the model is not meaningful. An argument can be raised to defeat such comparisons such that arguing: other LBP features might outperform CoLBP $c = 2$ when larger number of features are used, yet might be able to classify the images faster if the feature extraction time is faster than CoLBP. Hence, Table 1 proves that CoLBP $c = 2$ not only achieves a higher PR than the other examined LBP features but also

extracts a trained model feature faster. The reason behind this result is as illustrated in Section 2.1 that the CoLBP features are based on pixels rather than regions. Hence, only the model's specific LBP_{P,R} values are extracted rather than extracting all the LBP values for all examined window pixels, which is required in LBP extensions that use the histogram bins as features.

Furthermore, as explained in Section 3.1 that the co-occurrence of features not only increases the discriminative power of the features but also significantly reduces the training time by reducing the number of weak classifiers to half. For this reason, it can be seen from Table 1, that CoLBP $c = 2$ training time is less than CoLBP $c = 1$ since CoLBP $c = 1$ required 38 weak classifiers while CoLBP $c = 2$ required 18 weak classifiers. Also, CoLBP $c = 1$ required 38 features to achieve the target PR in comparison to 36 features in the case of CoLBP $c = 2$, since each weak classifier is trained with 2 features. Moreover, even though half the number of iterations are required in the CoLBP $c = 2$ in comparison to CoLBP $c = 1$, but it can be noticed that the training time is not reduced to half the time. The reason behind it is because of the overhead of using the SFS algorithm in CoLBP $c = 2$. Another observation from Table 1 is that having the CoLBP $c = 2$ require less number of features extraction in contrast to the CoLBP $c = 1$ (i.e., 38 in CoLBP $c = 1$ versus 36 in CoLBP $c = 2$) leads to a faster classification time.

Therefore, the following can be concluded from this experiment.

- (i) CoLBP $c = 2$ features outperform the LBP Histogram, Sobel-LBP, and MB-LBP features.
- (ii) CoLBP $c = 2$ features require less execution time to extract the trained model features. Hence, faster face detection algorithm can be achieved.
- (iii) Further proved Section 2.1 that CoLBP $c = 2$ not only outperforms CoLBP $c = 1$ but also requires less training time and has faster classification time.

3.3. CoLBP Face Detector. The CoLBP features are used to train a face detector using the cascade of classifiers technique in [3]. This experiment aims to prove that the CoLBP features are capable to achieve a face detector with a low FP of ≈ 1 in 10^6 examined window, this number is chosen as it is considered to what a face detector should have to be

TABLE 2: CoLBP detector cascade of classifiers training stages.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Number of features	6	7	26	36	61
TP	0.9900	0.9907	0.9907	0.9900	0.9900
FP	0.4761	0.3946	0.2516	0.02003	0.2344
	Stage 6	Stage 7	Stage 8	Stage 9	Total
Number of features	116	121	226	451	1050
TP	0.9900	0.9900	0.9900	0.9900	0.9135
FP	0.1604	0.1433	0.1881	0.1257	5.1453×10^{-6}

considered for practical applications [7], and TP of >90%. FP and TP values are chosen to coincide with the Viola and Jones detector [3] in order to achieve a fair comparison between CoLBP detector and Haar-like feature based detector.

Despite the fact that we cannot judge which number of co-occurred features of $c > 1$ performs better; however, it is clear from Figure 7 that CoLBP features with $c = 2$ has the least generalization error in earlier iterations. This is preferred since less number of features is required to reach the desired detection accuracy; hence, a faster classification speed is resulted. Therefore, $c = 2$ is used to train the cascade of classifiers.

The 19,832 frontal face images explained in Section 3.2 are used in this experiment. On the other hand, $\approx 20,000$ nonface images are downloaded from the World Wide Web; these images were manually investigated to ensure that they do not contain any faces. The images were downsampled with different ratios to increase their number. A total of $\approx 120,000$ nonface images were obtained. These images are of bigger resolution than the scanning windows. For example if an image is of size 490×330 pixels, and we use exhaustive search scanning window of size 24×24 with one pixel step, then 143,369 nonface windows are obtained. Therefore based on this example, the large number of nonface images can be imagined.

On each stage in the cascade of classifiers, a randomly selected 6,500 face images are used for training and randomly selected 1,500 face images are used for validation, under the condition that the training and validation set have no image in common. Also, 10,000 nonface images are used for each stage using the bootstrap strategy [3, 11] such that each step is trained with the misclassified nonface images by all previous stages. Furthermore, each stage is designed to achieve a minimum of 99% TP and a maximum of stage's dependant FP. The stage's FP can be found from Table 2 by taking the ceiling of each [FP] to nearest one decimal place (i.e., $0.4761 \rightarrow 0.5$).

After 9 stages, the minimum achieved FP using the CoLBP features is 5.1453×10^{-6} before depleting all nonface images dataset. The nonface dataset is depleted since the bootstrap strategy is used in training the cascade of classifiers. Table 2 illustrates the number of features, FP, and TP of each stage.

Comparison between Table 2 and Viola and Jones detector, where both have same objective of achieving total FP = 10^{-6} , and knowing that CoLBP detector is trained with

less number of face training dataset than Viola and Jones detector (i.e., 6,500 faces in the CoLBP detector versus 9,832 faces in Viola and Jones detector), shows that the CoLBP detector requires only 1,050 CoLBP $c = 2$ features while Viola and Jones requires 6,060 Haar-like features. Also the CoLBP features are selected from 3,164 features whereas the Haar-like features are selected from 160,000 features. Hence, it can be concluded that the CoLBP features cannot only extract discriminative face features to achieve a low total FP of 5.1453×10^{-6} but also require less number of features than Haar-like features in Viola and Jones detector, which indicates that the CoLBP features have higher discriminative power than the Haar-like features.

3.4. CoLBP Detector for Surveillance Application and Comparison to the Haar-Like Detector. The implemented CoLBP detector performance is evaluated on a real-life scenario as well as its performance is compared to the state-of-the-art Haar-like detector. The evaluation 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 taped by a camera mounted on the ceiling in vantage to capture frontal faces. The footage is an RGB colorspace sequence and is of Codec Video 1 format with video rate of 5 frames/second. Also, the sequence is of low-resolution of size 360×243 pixels, with decent illumination, and has multiple faces but noncrowded area. Faces appear in different sizes up to 80×60 pixels.

A total of 171 frames are extracted from the sequence where these frames contain all the frontal faces in the sequence in addition to some frames for vacant area. 105 frames are for single frontal face in different positions in order to examine the performance with different face sizes. 21 frames are for two people appearing in the screen to illustrate the ability of detecting more than one face, and finally 45 frames are for images where either a vacant place or nonfrontal faces appear in the scene to inspect the false positive detection tolerance. Some of the real-life frames are shown in Figure 10

For comparison purposes, the CoLBP detector is compared to the state-of-the-art Lienhart detector [15] (hereafter Haar-like detector). *haarcascade_frontalface_alt* model is used. This model is chosen since the detector is trained with the same boosting algorithm as the CoLBP detector, which is the GentleBoost. This detector is considered for comparison for several reasons, first its implementation



FIGURE 10: Sample frames of the examined real-life sequence to give an impression about the examined environment and to illustrate the different face sizes appearing in the sequence.

is very close to the successful work of Viola and Jones detector. Furthermore, just like the CoLBP detector, Haar-like detector's weak classifier is based on decision stump. Furthermore, despite there are several extensions to the Haar-like features proposed in the literature as in [17, 19]; however, many papers compare their results with this Haar-like detector since it is available in OpenCV.

The used scanning window is similar to Viola and Jones [7] with a size of 24×24 pixels. The scanning window is shifted by $[ds \times \Delta]$, where ds is the image downsampling factor, Δ is the shifting parameters, and $[]$ is rounding operator. ds downsamples the image by a factor of 1.25 until any dimension of the image becomes smaller than the scanning window. Δ is the shifting parameter that is fixed to 1.

The Free Receiver Operator Characteristic (FROC) is plotted for CoLBP detector and Haar-like detector for all operating points. FROC is very similar to ROC, but it plots the detection rate (DR) versus number of false positive detections (nFP) instead of FP rate. The parameter β explained in Section 2.3 is used to change the operating points of the detectors.

The method of evaluation used in this paper is similar to the proposed method in [37] but only the horizontal, vertical and scale errors are considered while the rotation error is dropped. The rotation error is dropped since the CoLBP detector is to be used for surveillance purposes from a camera mounted on the ceiling; hence, only straight with no in-plane rotation is expected; thus, keeping the rotation error penalty biases the decision since it will not occur. The scaling, horizontal and vertical errors are measured between the detected eyes with respect to the manually located eye.

Furthermore, the CoLBP detector and Haar-like detector output the size and location of the face's bounding box. On the other hand, the used method of evaluation requires the position of the centers of the eyes. Therefore, the location

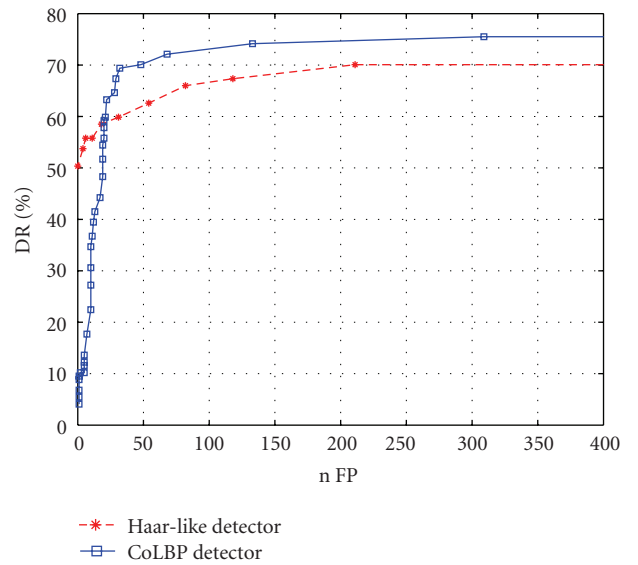


FIGURE 11: FROC for the CoLBP and Haar-like detectors using error tolerance of $\pm 2.5\%$ scaling and $\pm 5\%$ vertical and horizontal errors.

of the eyes is estimated from the bounding box output by running the detectors on many examples.

A correct face detection is considered when the detected eyes lie within *strict* face detection criterion such that the acceptable range is $\pm 2.5\%$ scaling error and $\pm 5\%$ horizontal and vertical errors [37] from the manually located eye location. Figure 11 shows the FROC for the CoLBP detector versus Haar-like detector.

It can be observed from Figure 11 that the CoLBP detector outperforms the Haar-like detector. It can also be noticed that the Haar-like detector is more consistent with its result throughout different operating points if compared to the CoLBP detector. The reason behind the consistency

issue is that the CoLBP detector is trained using two different training datasets Ole Jensen and Viola and Jones; hence, the training face images are not aligned perfectly (i.e., consistent place of the eyes and consistent cropped face area), also to mention that Viola and Jones faces dataset itself is aligned roughly as mentioned in [3]. Therefore, insensitivity to small face error occurs. While on the other hand, we have no knowledge about the training datasets used in the Haar-like detector [15], but it can be concluded from Figure 11 that the training dataset is consistent and aligned; hence, the Haar-like detector is less insensitive to small face error than our used dataset. Therefore, the following can be concluded.

- (i) CoLBP detector outperforms the Haar-like detector with 5.50% detection rate (using the operating point that achieves the highest detection rate for both detectors).
- (ii) CoLBP features have a higher discriminative power than the Haar-like features. The CoLBP detector with only 1,050 CoLBP features distributed over 9 stages could outperform the Haar-like detector which is trained with $\approx 2,122$ Haar-like features distributed over 20 stages [15].
- (iii) CoLBP detector requires less time to train the classifier than the Haar-like detector. Both detectors are trained using the GentleBoost algorithm and using the decision stump as weak classifier; hence, selecting 1,050 features from a pool of 3,164 features is less complicated than selecting $\approx 2,122$ features from a pool of 117,941 extended Haar-like features [16].
- (iv) Haar-like detector outperforms the CoLBP detector when $nFP < 20$ while CoLBP detector outperforms Haar-like detector afterwards. Therefore, the choice of the desired detector can be an application dependent; however, the training complexity difference explained in the previous point might play a crucial factor on the decision.

3.5. Detection Rate Sensitivity to Face Decision Criterion Parameters. This experiment is conducted to find the capability of the CoLBP detector to detect faces using the same evaluation dataset in the previous section however examined with wide range of error tolerance instead of the *strict* method of evaluation. The error tolerance (δ) is changed from 0% scaling, horizontal, vertical errors to 25% scaling, horizontal, vertical errors using 1% step size.

Figure 12 illustrates the result of the error tolerance range. It can be observed from Figure 12 that the detection rate can reach up to 97.28%. Therefore, having this result can further prove the capability of the CoLBP features in detecting faces as well as illustrates the effect of the training dataset, which made the system insensitive to small error.

3.6. CoLBP Detector Examined on the BioID Dataset. Due to difficulties that make reproducing identical face detectors that are implemented in the literature infeasible, and in order to have a more comprehensive comparison of the CoLBP

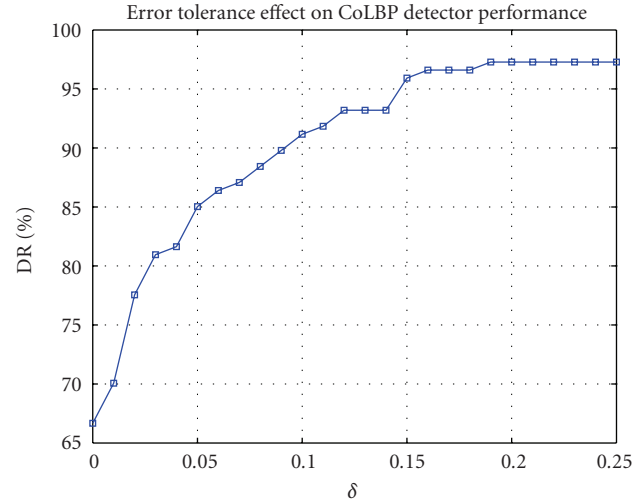


FIGURE 12: CoLBP detector performance sensitivity to different face decision criterion function parameters.

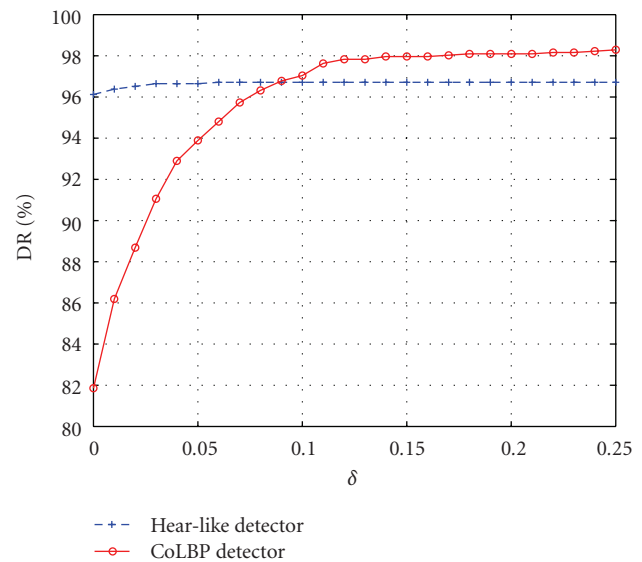


FIGURE 13: The CoLBP detector versus the Haar-like detector when examined on the BioID dataset.

detector with the state-of-the-art detectors, the CoLBP detector is applied to the real-life BioID dataset.

BioID database [38] is recorded and distributed to be used as a benchmark for face detection and recognition experiments. BioID images are recorded to illustrate *real world* scenario such as the images have variation in illumination, different background, and various face sizes. The dataset consists of 1521 gray-scale frontal face images, each image has a resolution of 384×286 pixels captured for 23 different persons.

Following the properties of the BioID dataset, it is widely examined in the literature including, but not limited to, the following works [8, 21, 39–42]. Figure 13 shows the CoLBP detector versus the Haar-like detector for the same method of evaluation range explained in Section 3.5.



FIGURE 14: Visualizing the illumination range by changing the contrast of the image from -100% to $+100\%$.

It can be concluded from Figure 13 that the CoLBP detector outperforms Haar-like detector. However, same conclusion explained in Section 3.4 can be drawn on the reasons made the Haar-like detector to be more consistent with its detection results over the entire range of different face decision criterion if compared to the CoLBP detector.

Even though several papers in the literature examined the BioID dataset; however, comparing the result still a daunting problem since different methods of evaluation are used (i.e., the method that decides whether the detected region is a face or not). However, if a comparison is conducted based on the highest detection rate achieved, then the CoLBP detector achieved 98.29% detection rate using 25% scaling and transitional error while 98.27% is reported in [8] using the Improved LBP (ILBP) features and measured using the method of evaluation explained in [38], but only 1511 images were considered in [8] instead of 1521. Furthermore, it can be observed that the CoLBP detector achieves a comparable result of $>95\%$ if compared to the state-of-the-art detectors in [21, 39–42].

It can be concluded from this experiment that the CoLBP detector is not only capable to outperform the Haar-like detector on surveillance scenarios but also on the widely examined BioID dataset. Furthermore, the CoLBP detector can be regarded as not only computationally efficient but also is capable of achieving a comparative results to several state-of-the-art face detectors that are examined on the BioID dataset.

3.7. Robustness Towards Illumination and Blurring Noise. One of the properties of the examined surveillance dataset is having a decent illumination and nonblurred frames while both types of noise are common to occur in video sequences. Therefore, the CoLBP detector performance is evaluated in various artificially added illumination and blurring scenarios. In order to have a better understanding of the tolerance to noise, then ΔDR is measured

$$\Delta DR = DR_o - DR_n, \quad (4)$$

where DR_o is the detection rate in the nonnoisy dataset, and DR_n is the detection rate when noise is applied.

The β value explained in Section 2.3 was fixed for the best PR when CoLBP detector is examined on the surveillance sequence.

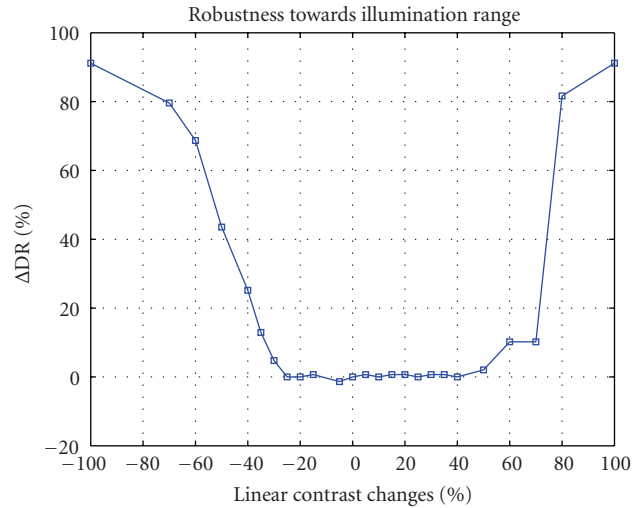


FIGURE 15: CoLBP detector robustness towards illumination changes.

3.7.1. Robustness Towards Illumination. One of the powerful facts that makes the LBP features to be superior over Haar-like features is its capability to handle illumination changes [11, 21]; therefore, this experiment is conducted to examine the robustness of the proposed CoLBP features towards illumination changes. The evaluation set was brightened and dimmed by changing the contrast of the image using linear transformation in the range from -100% to $+100\%$. Sample of the contrast range of the frames is shown in Figure 14.

The robustness of CoLBP detector towards illumination is as illustrated in Figure 15.

It can be concluded from this experiment that the CoLBP features holds the LBP features power in tolerating illumination changes. Also CoLBP features can handle a wide range of illumination changes in the range from -30% to $+70\%$.

3.7.2. Robustness Towards Blurring. The CoLBP detector is to be used in surveillance applications; therefore, camera blurring is expected. Gaussian filters of standard deviation (σ) 1, 1.4 and 2 are applied on the evaluation dataset to add blurring noise. The blurred images using these filters look as the examples in Figure 16. The robustness towards blurring noise is tabulated in Table 3.



FIGURE 16: Visualizing the Gaussian filter blur on the image.

TABLE 3: CoLBP detector robustness towards blurring.

Noise filter	ΔDR (%)
Gaussian filter with $\sigma = 1$	0
Gaussian filter with $\sigma = 1.4$	1.3605
Gaussian filter with $\sigma = 2$	2.0408

Hence it can be observed from Table 3 that the CoLBP detector has a wide range robustness towards blurring changes.

From the results presented in these experiments, several important observations can be made, and they are summarized below.

- (i) The CoLBP features are capable to extract face discriminative features.
- (ii) The co-occurrence of multiple features decreases the generalization error on the examined face dataset as well as decreases the training computational overhead in comparison to the separate features.
- (iii) The CoLBP features can achieve a faster face detector than other various examined LBP extensions.
- (iv) The CoLBP features have higher discriminative power than the Haar-like features on the examined face detection problem.
- (v) CoLBP features are not only computationally efficient and have higher discriminative power than Haar-like features but also achieve a comparative

result to the state-of-the-art face detectors when examined on the BioID dataset.

- (vi) CoLBP features hold the same properties of LBP features from an aspect of tolerating wide range of illumination changes.
- (vii) CoLBP features are capable to handle different blurring noise.

4. Conclusion

This paper introduces an idea addressing the challenging problem of face detection in surveillance sequence where the appearing faces are usually small and the video sequence is of low-resolution. The rotational LBP features which target the pixels of the image are used. The feature extraction is performed by extracting the rotational LBP features exhaustively for all possible resolutions in the examined window to target the image structure. The CoLBP features are multiple rotational LBP features occurred simultaneously, these feature are selected using SFS algorithm. The co-occurrence of features proved the capability to increase the discriminative power of the LBP features. Experiments carried out on Ole Jensen, Viola and Jones, BioID, and real-life surveillance sequence datasets show that the proposed CoLBP features are effective in boosting face detection performance and outperform state-of-the-art face detection techniques. Experiments have also shown that CoLBP features are capable to effectively handle illumination and blurring noise. While this paper concentrates on the

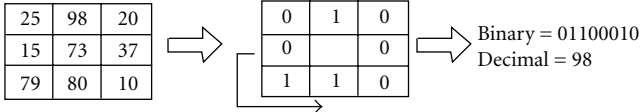
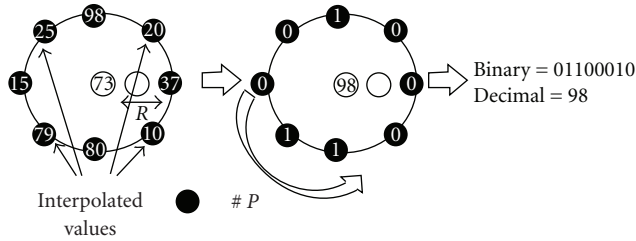


FIGURE 17: Simple LBP feature.

FIGURE 18: Rotational LBP_{8,2} feature extraction.

face detection problem, but the proven capability of the CoLBP features to extract discriminative feature with their properties to handle wide range of noise can be used for different object detection problems.

Appendix

Review of Local Binary Patterns Features

Local Binary Patterns (LBP) features were first introduced in [43]. Due to their power to detect corners, edges, spots and flat ends as well as their high tolerance to illumination [20], they are used in texture classification. Simple LBP feature extraction algorithm operates by taking the value of the center pixel in a 3×3 pixels and assuming the texture of this 3×3 matrix is the joint distribution of nine gray-scale image pixels [44]. Furthermore, it subtracts the center pixel from all surrounding pixels. The center pixel is considered as the overall luminance factor in the 3×3 matrix, and it does not provide texture information. In order to achieve scaling of gray-scale invariance and preserve the texture of the matrix, the signs of the pixels are taken. Hence,

$$\text{LBP}(x_o, y_o) = \sum_{i=0}^7 \text{sign}(g_i - g_o) 2^i, \quad (\text{A.1})$$

where $\text{LBP}(x_o, y_o)$ is the LBP value for the center pixel in the 3×3 matrix; the decimal value of $\text{LBP}(x_o, y_o)$ represents the texture for this 3×3 pixel window. g_i is the gray-scale value of the surrounding pixels, and g_o is the gray-scale value of the center pixel.

Also,

$$\text{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.2})$$

Figure 17 shows an example of simple LBP operation. Due to this procedure, 2^8 possible LBP values can be obtained from 3×3 matrices.

The LBP features are not extracted on only 1 pixel neighbor or a square window, but as in [20], they can be extracted with a circular neighbor with different radii and points. Points are considered as the number of equally spaced points that construct the LBP operator and the radius is how far the points from the central pixel lie. This LBP features type is called rotational LBP features. Rotational LBP operator is symbolized as $\text{LBP}_{P,R}$, where P, R correspond to the number of points and radius, respectively, therefore, there are 2^P possible binary words for each $\text{LBP}_{P,R}$. Figure 18 shows an example of $\text{LBP}_{8,2}$ feature extraction.

Furthermore, it was found in [20] that there is a subset of the 2^P $\text{LBP}_{P,R}$ that spans most of the texture descriptor, this subset is called Uniform $\text{LBP}_{P,R}$, $\text{LBP}_{P,R}^{u2}$. The Uniform $\text{LBP}_{P,R}$ words are the words that have only two flipping bits from 0 to 1 and 1 to 0 (e.g., 01110000).

Acknowledgments

This work is partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and ORF-RE program of the Ontario Ministry of Research and Innovation through MULTIMODAL SURVEILLANCE SYSTEM FOR SECURITY-RELATED APPLICATIONS (MUSES.SECRET) project. Also, MATLAB code available in [45] is used to implement some of the LBP features extensions.

References

- [1] E. Hjelmås and B. K. Low, "Face detection: a survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.
- [2] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.
- [3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 511–518, December 2001.
- [4] W. R. Schwartz, R. Gopalan, R. Chellappa, and L. S. Davis, "Robust human detection under occlusion by integrating face and person detectors," in *Proceedings of the 3rd International Conference on Advances in Biometrics (ICB '09)*, vol. 5558 of *Lecture Notes in Computer Science*, pp. 970–979, June 2009.
- [5] H. Moon, R. Chellappa, and A. Rosenfeld, "Optimal edge-based shape detection," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1209–1227, 2002.
- [6] K. C. Yow and R. Cipolla, "Feature-based human face detection," *Image and Vision Computing*, vol. 15, no. 9, pp. 713–735, 1997.
- [7] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [8] Y. Rodriguez, *Face detection and verification using local binary patterns*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2006.
- [9] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '96)*, pp. 203–208, June 1996.

- [10] D. Roth, M. H. Yang, and N. Ahuja, "A SNoW-based face detector," *Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861, 2000.
- [11] A. Hadid, M. Pietikäinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 797–804, Washington, DC, USA, June–July 2004.
- [12] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image and Vision Computing*, vol. 20, no. 9–10, pp. 657–664, 2002.
- [13] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved LBP under bayesian framework," in *Proceedings of the 3rd International Conference on Image and Graphics (ICIG '04)*, pp. 306–309, December 2004.
- [14] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of International Conference on Machine Learning (ICML '96)*, pp. 148–156, 1996.
- [15] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings of the International Conference on Image Processing (ICIP'02)*, vol. 1, pp. 900–903, September 2002.
- [16] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Pattern Recognition*, vol. 2781 of *Lecture Notes in Computer Science*, pp. 297–304, 2003.
- [17] T. Mita, T. Kaneko, B. Stenger, and O. Hori, "Discriminative feature co-occurrence selection for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1257–1269, 2008.
- [18] T. Mita, T. Kaneko, and O. Hori, "Joint Haar-like features for face detection," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, pp. 1619–1626, October 2005.
- [19] P. M. Tri, *Principled asymmetric boosting approaches to rapid training and classification in face detection*, Ph.D. thesis, Nanyang Technological University, 2009.
- [20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "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. 971–987, 2002.
- [21] B. Fröba and A. Ernst, "Face detection with the modified census transform," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FGR '04)*, pp. 91–96, May 2004.
- [22] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Advances in Biometrics*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 11–18, 2007.
- [23] C. Shen, S. Paisitkriangkrai, and J. Zhang, "Face detection from few training examples," in *Proceedings of IEEE International Conference on Image Processing (ICIP '08)*, pp. 2764–2767, October 2008.
- [24] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.
- [25] S. Zhao, Y. Gao, and B. Zhang, "Sobel-LBP," in *Proceedings of IEEE International Conference on Image Processing (ICIP '08)*, pp. 2144–2147, October 2008.
- [26] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," *Pattern Recognition*, vol. 42, no. 3, pp. 425–436, 2009.
- [27] A. Vezhnevets and V. Vezhnevets, "Modest AdaBoost-teaching AdaBoost to generalize better," in *Proceedings of the International Conference on the Computer Graphics and Vision (GraphiCon '05)*, pp. 322–325, Novosibirsk Akademgorodok, Russia, 2005.
- [28] T. Marill and D. Green, "On the effectiveness of receptors in recognition systems," *IEEE transactions on Information Theory*, vol. 9, no. 1, pp. 11–17, 1963.
- [29] A. W. Whitney, "Direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. 20, no. 9, pp. 1100–1103, 1971.
- [30] S. D. Stearns, "On selecting features for pattern classifiers," in *Proceedings of the International Joint Conference on Pattern Recognition*, pp. 71–75, 1976.
- [31] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Transactions on Computers*, vol. 26, no. 9, pp. 917–922, 1977.
- [32] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [33] O. H. Jensen and R. Larsen, *Implementing the Viola-Jones face detection algorithm*, M.S. thesis, Technical University of Denmark, Denmark, 2008.
- [34] P. S. Carbonetto, "Robust object detection using boosted learning," Tech. Rep., Department of Computer Science, University of British Columbia, Vancouver, Canada, 2002.
- [35] W. Louis and K. N. Plataniotis, "Weakly trained dual features extraction based detector for frontal face detection," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '10)*, pp. 814–817, Dallas, Tex, USA, March 2010.
- [36] W. Louis, K. N. Plataniotis, and Y. Man Ro, "Enhanced weakly trained frontal face detector for surveillance purposes," in *Proceedings of the 6th IEEE World Congress on Computational Intelligence (WCCI '10)*, Barcelona, Spain, July 2010.
- [37] V. Popovici, J. P. Thiran, Y. Rodriguez, and S. Marcel, "On performance evaluation of face detection and localization algorithms," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 313–317, August 2004.
- [38] O. Jesorsky, K. J. Kirchberg, R. W. Frischholz et al., "Robust face detection using the hausdorff distance," in *Audio- and Video-Based Biometric Person Authentication*, Lecture Notes in Computer Science, pp. 90–95, 2001.
- [39] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local SMQT features and split up snow classifier," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 2, pp. 589–592, April 2007.
- [40] W. K. Tsao, A. J. T. Lee, Y. H. Liu, T. W. Chang, and H. H. Lin, "A data mining approach to face detection," *Pattern Recognition*, vol. 43, no. 3, pp. 1039–1049, 2010.
- [41] K. J. Kirchberg, O. Jesorsky, and R. Frischholz, "Genetic model optimization for Hausdorff distance-based face localization," in *Proceedings of the European Conference on Computer Vision (ECCV '02)*, pp. 103–111, Springer, 2002.
- [42] P. Shih and C. Liu, "Face detection using discriminating feature analysis and support vector machine," *Pattern Recognition*, vol. 39, no. 2, pp. 260–276, 2006.
- [43] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

- [44] T. Ojala, M. Pietikainen, and T. Maenpaa, "Gray scale and rotation invariant texture classification with local binary patterns," in *Proceedings of the 6th European Conference on Computer Vision (ECCV '00)*, vol. 1842 of *Lecture Notes in Computer Science*, pp. 404–420, Dublin, Ireland, June-July 2000.
- [45] S. Paris, "Face detection toolbox," November 2009, <http://www.mathworks.com/matlabcentral/fileexchange/24092-face-detection-toolbox>.