On the Effects of Illumination Normalization with LBP-Based Watchlist Screening

Ibtihel Amara[1], Eric Granger[1], Abdenour Hadid[2]

[1] École de Technologie Supérieure, Université du Québec, Montreal, Canada
[2] Center for Machine Vision Research, University of Oulu, Finland

Abstract. Still-to-video face recognition (FR) is an important function in several video surveillance applications like watchlist screening, where faces captured over a network of video cameras are matched against reference stills belonging to target individuals. Screening of faces against a watchlist is a challenging problem due to variations in capturing conditions (e.g., pose and illumination), to camera inter-operability, and to the limited number of reference stills. In holistic approaches to FR, Local Binary Pattern (LBP) descriptors are often considered to represent facial captures and reference stills. Despite their efficiency, LBP descriptors are known as being sensitive to illumination changes. In this paper, the performance of still-to-video FR is compared when different passive illumination normalization techniques are applied prior to LBP feature extraction. This study focuses on representative retinex, self-quotient, diffusion, filtering, means de-noising, retina, wavelet and frequency-based techniques that are suitable for fast and accurate face screening. Experimental results obtained with videos from the Chokepoint dataset indicate that, although Multi-Scale Weberfaces and Tan and Triggs techniques tend to outperform others, the benefits of these techniques varies considerably according to the individual and illumination conditions. Results suggest that a combination of these techniques should be selected dynamically based on changing capture conditions.

Keywords: Illumination Normalization, Local Binary Patterns, Face Screening, Still-to-Video Face Recognition, Video Surveillance.

1 Introduction

In watchlist screening applications, systems for still-to-video FR are increasingly employed to automatically detect the presence of target individuals of interest for enhanced public security. Accurate and timely responses are required to recognize faces captured under semi-controlled or uncontrolled conditions, as found at various security checkpoint entries, inspection lanes, portals, etc. Under these conditions, face captures incorporate variations due to ambient illumination, pose, expressions, occlusion, scale, resolution and blur [3,21], and the performance of FR systems tend to deteriorate. Despite these challenges, it is generally possible to exploit spatiotemporal information extracted from video streams to improve system robustness and accuracy [11,5].

Recent developments in image analysis and recognition have shown that the Local Binary Patterns (LBP) [14] provide a simple yet powerful approach to represent faces for human computer interaction, biometric recognition, surveillance and security, etc. [2,16]. LBP is a gray-scale invariant texture operator which labels each pixel of an image by thresholding its neighborhood pixels with the intensity value of the center pixel. The resulting LBP labels can be regarded as local primitives such as curved edges, spots, flat areas, etc. The histogram of these labels over facial image can be then used as a face descriptor. Given its discriminative power, tolerance to monotonic grey-scale changes, and computational efficiently, LBP has become a well-established technique in FR[3], and has inspired many recent extensions and new research on related methods.

However, it is well known that LBP and other variants are sensitive to severe illumination changes. Variations in facial appearance caused by changes in ambient illumination conditions play an important role in the performance of any FR system applied to video surveillance. It has been shown that face images of different individuals appear more similar than images of the same individual under severe illumination variations [18].

Several techniques have been proposed in the literature for illumination invariant FR [17]. Zou et al. [25] presented a survey of techniques to manage variations in face appearance due to illumination changes using passive and active approaches. Passive approaches focus on the visible spectrum images, where face appearance has been altered by illumination variations, while active ones employ active imaging techniques to capture face images under consistent illumination conditions, or images of illumination invariant modalities.

Among passive techniques, some are specialized at either the pre-processing, the feature extraction, or the classification level [18]. At the pre-processing level, normalization techniques seek to transform facial images such that facial variations induced by illumination are removed. These approaches can be adapted for use with any FR algorithm. Techniques at the feature extraction level seek to achieve illumination invariance by using features or representations that are stable under different illumination conditions. However, some empirical studies have shown that no descriptor can ensure illumination invariant FR in the presence of severe illumination changes. Finally, classification level techniques compensate for the illumination based on the type of face model or classifier employed for FR. Assumptions regarding the effects of illumination on the face model or classifier are employed in counter measures to obtain illumination invariance.

In this study, the performance of several illumination normalization techniques is compared for representation of face captures in still-to-video FR systems using LBP descriptors, as seen in many watchlist screening applications. This empirical study focuses on passive techniques applied at the pre-processing level, and compares representative retinex, self-quotient, diffusion, filtering, means de-noising, retina, wavelet and frequency-based techniques in term of ROC and Precision Recall performance. The benefits of these approaches are assessed us-

---

[3] See LBP bibliography at `http://www.cse.oulu.fi/MVG/LBP_Bibliography`

ing faces captured in the Chokepoint video data set, with individuals walking through an array of cameras located above different portals.

The rest of this paper is organized as follows. Section 2 describes the application focus of this paper which is face screening in video surveillance. Then, Section 3 gives an introduction to the popular LBP approach to face recognition. Section 4 discusses different methods for illumination normalization. The experimental results are presented in Section 5 while a conclusion is given in Section 6.

2 Face Screening in Video Surveillance

Watchlist screening is an important application for decision support in video surveillance systems. It involves still-to-video FR according to the following steps [4]. During enrollment to a watchlist, the segmentation process isolates the regions of interest (ROIs) from reference still images (mugshots) that were previously captured under controlled conditions. Features are extracted and assembled into a discriminant and compact ROI patterns to design facial models[4]. These features are often image-based (e.g., LBP descriptors) or pattern recognition-based (e.g., PCA projections).

During operations, a video stream is captured using some video surveillance camera, and segmentation isolates the ROIs corresponding to faces captured in successive frames. A tracker is often initialized when an emergent ROI is detected far from other faces, and a track is defined to follow the movement or expression of distinct faces across consecutive frames using appearance, position and motion information. Features are extracted into ROI pattern for matching against the facial models of individuals enrolled to the watchlist. A positive prediction is produced if a matching score surpasses an individual-specific threshold. Finally, the decision function combines the tracks and classification predictions in order to recognize the most likely individuals in the scene.

Systems for still-to-video FR are typically modeled in terms of independent detection problems, each one implemented using a template matcher or classifier. These individual-specific detectors are designed with reference face samples from target and non-target individuals (from a cohort or the background model). The advantages of modular architectures with individual-specific detectors include the ease with which face models may be added, updated and removed from the systems, and the possibility of specializing pre-processing, feature extraction, matching and decision thresholds to each specific individual [6,15].

The performance of state-of-the-art FR systems applied to video surveillance is limited by the difficulty in capturing and recognizing facial regions from video streams under semi-controlled and uncontrolled capture conditions (e.g., at inspection lanes, portals and checkpoint entries, in cluttered free-flow scenes at

---

[4] A *facial model* of an individual is defined as a set of one or more reference ROI patterns (used for a template matching system), or parameters estimated from reference ROI patterns (for a classification system).

airports or casinos). In particular, performance is severely affected by the variations in ambient illumination, pose, expression, occlusion, scale, resolution, blur and ageing. Still-to-video FR is particularly challenging because very few reference samples are typically available for system design, and because of camera inter-operability – ROIs captured with still cameras (during enrollment) have different properties than those captured with video cameras (during operations). In pattern recognition literature, the situation where only one reference sample is available for system design are often referred to as a "single sample per person" (SSPP) or "one sample training" problem. Techniques specialized for SSPP in FR include multiple face representations, synthetic face generation, and enlarging the training set using auxiliary set [1]. Note that the still-to-video FR systems from the literature assume that the single face reference is consistent and representative of individuals captures in operational conditions.

3   LBP-based Face Recognition

The LBP texture analysis operator, introduced by Ojala et al. [14], is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes.

The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Fig. 1 shows an example of an LBP calculation. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor.
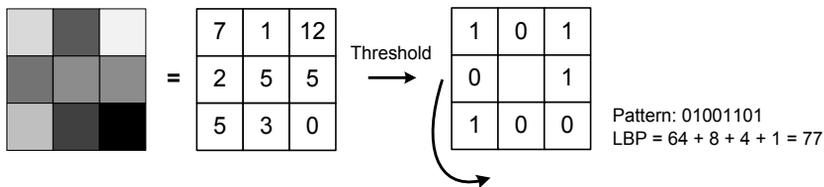

Fig. 1: The basic LBP operator.

The operator has been extended to use neighborhoods of different sizes. Using a circular neighborhood and bilinearly interpolating values at non-integer pixel coordinates allow any radius and number of pixels in the neighborhood. The notation (P, R) is generally used for pixel neighborhoods to refer to P sampling points on a circle of radius R. The calculation of the LBP codes can be easily done in a single scan through the image. The value of the LBP code of a pixel

(x_c, y_c) is given by:

$$\text{LBP}_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{1}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a circle of radius R, and s defines a thresholding function as follows:

$$s(x) = \begin{cases} 1, \text{if } x \geq 0; \\ 0, \text{otherwise.} \end{cases} \tag{2}$$

Another extension to the original operator is the definition of the so called *uniform patterns*. This extension was inspired by the fact that some binary patterns occur more commonly in texture images than others. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. This yields to the following notation for the LBP operator: $\text{LBP}_{P,R}^{u2}$. The subscript represents using the operator in a (P, R) neighborhood. Superscript $u2$ stands for using only uniform patterns and labeling all remaining patterns with a single label.

Each LBP label (or code) can be regarded as a micro-texton. Local primitives which are codified by these labels include different types of curved edges, spots, flat areas etc. The occurrences of the LBP codes in the image are collected into a histogram. The classification is then performed by computing histogram similarities. For an efficient representation, facial images are first divided into several local regions from which LBP histograms are extracted and concatenated into an enhanced feature histogram.

It is known that LBP is sensitive to severe illumination changes. As a consequence, several attempts have been made to overcome this sensitivity. For instance, Tan and Triggs [19] developed a very effective preprocessing chain for compensating illumination variations in face images. It is composed of gamma correction, difference of Gaussian (DoG) filtering, masking (optional) and equalization of variation. This approach has been very successful in LBP-based face recognition under varying illumination conditions. When using it for the original LBP, the last step (i.e. equalization of variations) can be omitted due to LBPs invariance to monotonic gray scale changes.

Aiming at reducing the sensitivity of the image descriptor to illumination changes, a Bayesian LBP (BLBP) was developed by He et al.[7]. This operator is formulated in a Filtering, Labeling and Statistic framework for texture descriptors. In the framework, the local labeling procedure, which is a part of many popular descriptors such as LBP and SIFT, can be modeled as a probability and optimization process. This enables the use of more reliable prior and likelihood information, and reduces the sensitivity to noise. The BLBP operator pursues a label image, when given the filtered vector image, by maximizing the joint probability of two images.

Table 1: Illumination normalization techniques studied in this paper.

Family	Specific Technique
Retinex	Adaptive Single-Scale Retinex (ASSR), Large and Small-Scale Features (LSSF)
Self Quotient	Multi-Scale Self Quotient (MSSQ)
Diffusion	Isotropic Diffusion (ID), Modified Anisotropic Diffusion (MAD)
Filter	Tan and Triggs (TT)
Gradient	Multi-Scale Weberfaces (MSW)
Mean Denoising	Adaptive Non Local Means (ANLM)
Retina	Retina Modeling (RM)
Wavelet	Wavelet Denoising (WD)
Frequency	Homomorphic

Liao et al.[9] noticed that adding a small offset value for comparison in LBP-like methods is not invariant under scaling of intensity values. The intensity scale invariant property of a local comparison operator is very important for example in background modeling, because illumination variations, either global or local, often cause sudden changes of gray scale intensities of neighboring pixels simultaneously, which would approximately be a scale transform with a constant factor. Therefore, a Scale Invariant Local Ternary Pattern (SILTP) operator was developed for dealing with the gray scale intensity changes in complex background. Assuming linear camera response, The SILTP feature is invariant if the illumination is suddenly changed from darker to brighter or vice versa. Besides, SILTP is robust when a soft shadow covers a background region, because the soft cast shadow reserves the background texture information but tends to be darker than the local background region with a scale factor. A downside of the methods mentioned above using one or two thresholds is that the methods are not strictly invariant to local monotonic gray level changes as the original LBP. The feature vector lengths of these operators are also longer.

In order to deal with strong illumination variations, Li et al. developed an active approach combining near-infrared (NIR) imaging with local binary pattern features and AdaBoost learning [8]. The invariance of LBP with respect to monotonic gray level changes makes the NIR images illumination invariant. The method achieved a verification rate of 90% at FAR=0.001 and 95% at FAR=0.01 on a database with 870 subjects.

4   Illumination Normalization

Changes in ambient illumination, and the resulting variations to facial appearance, are known to significantly deteriorate the performance of FR systems. Accordingly, several techniques have been proposed for illumination invariant FR [17]. Zou et al. [25] presented a survey of techniques according to passive and active approaches. *Passive approaches* focus on the visible spectrum images where face appearance has been altered by illumination variations. They include illumination variation modelling, illumination invariant features, photometric normalisation, and 3D morphable model techniques. In contrast, *active approaches* employ active imaging techniques to obtain face images captured un-

der consistent illumination condition, or images of illumination invariant modalities. Additional devices (optical filters, active illumination sources or specific sensors) are usually involved to actively obtain different modalities of face images that are insensitive to or independent of illumination change. Those modalities include 3D face information and face images in those spectra other than visible spectra, such as thermal infrared image and near-infrared hyperspatial image.

Passive approaches fall under three main types of techniques to produce illumination invariant facial images – those applied at the pre-processing, feature extraction and classification levels [18]. Pre-processing techniques seek to produce (prior to feature extraction) facial images without facial variations caused by illumination. They compensate for the illumination within any FR system, since no prior assumptions influence feature extraction or classification procedures. They may also be computationally simple, and effective at achieving illumination invariant FR. Feature extraction techniques seek to compensate for appearance variations in facial images using descriptors or representations that are stable under different illumination conditions. However, different empirical studies with LBP, Gabor wavelet-based features, and other descriptors have shown that none of these can ensure illumination invariant FR given severe illumination changes [10]. Classification-level techniques compensate for illumination changes according to the type of face model or classifier employed in the FR system. First, some assumptions regarding the effects of illumination on face models or classification procedure are made, and then based on these assumptions, counter measures are undertaken to obtain illumination invariant face models or illumination insensitive classification procedures. Managing the effects of illumination at the feature extraction level is debatable, while classification level techniques may impose difficult requirements on design data. Although they may provide the more efficient approach to illumination invariant FR, large training set must usually be acquired under a number of lighting conditions and are, furthermore, also computationally expensive.

In this paper, we focus our empirical study on passive techniques for illumination normalization at the pre-processing level. Table 1 presents the specific techniques from the literature that are considered in our study. A more detailed description of these techniques may be found in [18]. They are selected because they are the newer and more representative techniques from different families, e.g., retinex, diffusion, wavelet, frequency-based techniques.

5   Experimental Analysis

5.1   Dataset and Experimental Protocol:

To compare the performance achieved by a still-to-video FR system using different illumination normalization techniques prior to LBP, Chokepoint video dataset [24] has been employed. An array of three cameras is installed above several portals (natural choke points for pedestrian traffic) to capture 25 individuals walking through in a natural way. Videos are challenging for still-to video FR since faces are captured under semi-controlled conditions, with changes in

Normalization techniques	ID03			ID04		
	ROI from still	ROIs from videos		ROI from still	ROIs from videos	
Normalization techniques						
Tan and Triggs TT						
Large and small scale feature LSSF						
Multi-scale weberfaces MSW						
Modified Anisotropic Diffusion MAD						
Adaptive non local means ANLM						
Isotropic Diffusion ID						
Wavelet Denoising WD						
Retina model RM						
Multi Scale Quotient MSQ						
Homomorphic						
Adaptive single scale retinex ASSR						

Fig. 2: Examples of face images obtained after illumination normalization is applied to ROIs in stills and videos from individuals ID03 and ID04.

illumination, pose, scale, blur and occlusion. All 48 video sequences from the center camera, in both entering and leaving cases of Chokepoint have been con-

sidered. Cameras have a frame rate of 30 fps and the image resolution is 800 x 600 pixels.

Prior to experiments, 5 persons are randomly selected as target watchlist individuals, where just one reference still image (high-quality neutral mug-shot) is available to design each face model. These reference stills are used a priori to design templates for this FR system. The remaining people are assumed to be unknown (non-target individuals, and reflect the universal background model). Enrollment of each target individual involves isolating a ROI from the reference still image using the Viola-Jones face detection algorithm, and converting the ROI into grey scale, and then cropping it to a common size of 48x48 pixels to limit processing time. For each watchlist individual, the 11 illumination normalization techniques selected for this study (see Table 1) are used to represent the reference ROI using the INface Toolbox[5] [23,22]. At this level, 12 representations of one ROI are created in which 11 represent the normalized ROI in terms of illumination and 1 represents the original ROI (without application of illumination normalization techniques). These representations are shown for individual ID03 and ID04 of Chokepoint in Figure 2.

A division into 3x3 = 9 uniform non-overlapping patches of 16x16 pixels is performed on each ROI representations after illumination normalization. With patch-based methods, facial ROIs are divided into several overlapping or non-overlapping regions called patches, and then features are extracted locally from each patch for recognition purposes. Some specialized decision fusion techniques have been introduced in [20,13] for patch-based FR. In this paper, 59 LBP features are extracted from each patch, normalized to range between 0 and 1, and assembled into a ROI pattern of 531 features for matching. The latter are then stored as a template into a gallery. The enrollment phase produces a template gallery with 12 different templates per watchlist person (the original image plus 11 normalized images).

During the testing or operational phase, frames undergo the same processing steps as for enrollment. For each normalization technique, an ROI pattern extracted from a video frame is compared with the corresponding template of the 5 watchlist individuals. Template matching is performed with the Euclidian distance, and produces matching scores.

To assess the transaction-level performance, *receiver operating characteristic (ROC)* space is considered. A ROC curve displays the proportion of target ROIs that are correctly detected as individual of interest over the total number of target ROIs in the sequence, the true positive rate (tpr), as a function of the proportion of non-target (imposter) ROI detected as individual of interest over the total number of non-target ROIs, the false positive rate (fpr). The area under ROC curve (AUC) provides a global scalar measure that can be interpreted as the probability of classification over the range of tpr and fpr. The precision-recall (PROC) space is also considered to measure the performance in imbalanced data situation. Recall is the tpr and precision is computed as follows $pr = TP/(TP + FP)$. The AUPR measures system performance based on

---

[5] http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/

Table 2: Average pAUC(5%) performance (with standard deviation) for each watchlist individual with illumination normalization techniques.

Illumination Normalisation	ID # of Watchlist Individuals					
	ID03	ID04	ID07	ID09	ID12	Average
Entering Videos						
No Normalization	0.66±0.04	0.96±0.01	0.72±0.02	0.84±0.03	0.91±0.02	0.82±0.02
Adaptive Single Scale Retinex	0.65±0.04	0.90±0.01	0.54±0.02	0.76±0.05	0.90±0.01	0.75±0.02
Large and Small Scale Features	0.72±0.06	0.89±0.03	0.69±0.02	0.89±0.03	0.92±0.03	0.82±0.03
Multi Scale Self-Quotient	0.69±0.04	0.88±0.03	0.67±0.05	0.87±0.02	0.93±0.02	0.81±0.03
Isotropic Diffusion	0.69±0.06	0.86±0.03	0.70±0.01	0.90±0.01	0.97±0.01	0.82±0.02
Modified Anisotropic Diffusion	0.74±0.05	0.85±0.03	0.74±0.02	0.80±0.03	0.94±0.02	0.81±0.03
Tan & Triggs	0.74±0.03	0.86±0.03	0.71±0.02	0.88±0.04	0.92±0.03	0.82±0.03
Multi Scale Weberfaces	0.82±0.02	0.83±0.03	0.73±0.03	0.88±0.05	0.91±0.03	0.83±0.03
Adaptive Non-Local Means	0.71±0.02	0.89±0.02	0.66±0.03	0.69±0.04	0.84±0.03	0.76±0.02
Retina Modeling	0.73±0.05	0.85±0.03	0.69±0.02	0.90±0.03	0.91±0.05	0.82±0.03
Wavelet Denoising	0.66±0.03	0.89±0.02	0.54±0.03	0.83±0.02	0.87±0.01	0.76±0.02
Homomorphic	0.62±0.04	0.94±0.01	0.73±0.02	0.81±0.05	0.91±0.01	0.80±0.02
Leaving Videos						
No Normalization	0.67±0.08	0.91±0.03	0.79±0.02	0.91±0.02	0.94±0.02	0.84±0.03
Adaptive Single Scale Retinex	0.73±0.03	0.89±0.02	0.66±0.01	0.89±0.02	0.92±0.01	0.82±0.01
Large and Small Scale Features	0.78±0.03	0.94±0.02	0.54±0.02	0.94±0.02	0.96±0.01	0.83±0.02
Multi Scale Self - Quotient	0.74±0.03	0.82±0.07	0.74±0.02	0.92±0.01	0.93±0.02	0.83±0.03
Isotropic Diffusion	0.82±0.03	0.83±0.05	0.75±0.02	0.91±0.02	0.95±0.01	0.85±0.02
Modified Anisotropic Diffusion	0.78±0.02	0.89±0.02	0.64±0.02	0.93±0.01	0.94±0.01	0.83±0.01
Tan & Triggs	0.80±0.03	0.93±0.01	0.61±0.03	0.97±0.01	0.96±0.01	0.85±0.01
Multi-Scale Weberfaces	0.85±0.03	0.92±0.01	0.73±0.02	0.95±0.01	0.95±0.01	0.88±0.01
Adaptive Non-Local Means	0.74±0.04	0.90±0.02	0.71±0.02	0.86±0.01	0.95±0.01	0.84±0.02
Retina Modeling	0.77±0.03	0.93±0.01	0.55±0.03	0.96±0.01	0.96±0.01	0.83±0.01
Wavelet Denoising	0.71±0.02	0.91±0.02	0.66±0.02	0.87±0.01	0.92±0.01	0.81±0.01
Homomorphic	0.65±0.03	0.90±0.02	0.78±0.01	0.87±0.02	0.91±0.01	0.82±0.01

targets ROI patterns given an imbalance between target (minority) and non-targets (majority) proportions. In trajectory-level analysis, a tracking module is employed to regroup ROIs captured for a same person over successive frames. Hence, ROIs of persons in the scene are regrouped. In this paper, we show the matching scores linked to ROI patterns of each person appearing in the scene w.r.t each face model.

5.2   Results and discussion:

Results in Tables 2 and 3 present the average transaction-level performance (pAUC(5%) and AUPR) for each watchlist individual obtained by applying the 11 illumination normalization techniques over all entering and leaving videos of Chokepoint. Based on overall results, MSW and TT techniques tend to outperform the others with both entering and leaving videos.

It can however be observed that the results vary significantly according to the watchlist individual and to capturing conditions (sequence and portals). For instance, with individual ID04, applying illumination normalization decreases system performance compared to the results without any normalization (see Figure 3(c) and (d)). In contrast, with individual ID03, the pAUC(5%) and AUPR are significantly higher when normalization are applied, specially with the

Table 3: Average AUPR performance (with standard deviation) for each watchlist individual with illumination normalization techniques.

Illumination Normalisation	ID # of Watchlist Individuals					
	ID03	ID04	ID07	ID09	ID12	Average
Entering Videos						
Without Normalization	0.06±0.01	0.64±0.07	0.16±0.03	0.30±0.08	0.60±0.08	0.35±0.05
Adaptive Single Scale Retinex	0.09±0.03	0.28±0.03	0.06±0.08	0.17±0.04	0.46±0.07	0.21±0.05
Large and Small Scale features	0.18±0.06	0.40±0.04	0.14±0.03	0.51±0.08	0.63±0.01	0.37±0.04
Multi Scale Self-Quotient	0.12±0.05	0.45±0.07	0.15±0.02	0.31±0.04	0.57±0.09	0.32±0.05
Isotropic Diffusion	0.13±0.04	0.31±0.05	0.11±0.01	0.35±0.05	0.74±0.05	0.33±0.04
Modified Anisotropic Diffusion	0.13±0.04	0.34±0.07	0.17±0.02	0.28±0.06	0.68±0.09	0.32±0.05
Tan & Triggs	0.16±0.06	0.37±0.05	0.16±0.02	0.59±0.10	0.64±0.10	0.38±0.06
Multi-Scale Weberfaces	0.25±0.07	0.37±0.05	0.19±0.03	0.58±0.10	0.57±0.11	0.39±0.07
Adaptive Non-Local Means	0.11±0.04	0.51±0.06	0.14±0.02	0.07±0.01	0.53±0.07	0.27±0.04
Retina Modeling	0.22±0.07	0.32±0.05	0.16±0.02	0.63±0.09	0.66±0.10	0.40±0.06
Wavelet Denoising	0.08±0.02	0.37±0.05	0.06±0.01	0.14±0.02	0.32±0.06	0.19±0.03
Homomorphic	0.05±0.01	0.65±0.06	0.18±0.04	0.18±0.05	0.47±0.08	0.30±0.04
Leaving Videos						
Without Normalization	0.19±0.06	0.43±0.07	0.23±0.03	0.57±0.08	0.66±0.07	0.42±0.06
Adaptive Single Scale Retinex	0.14±0.02	0.26±0.06	0.11±0.01	0.41±0.04	0.58±0.03	0.30±0.03
Large and Small Scale features	0.22±0.03	0.49±0.07	0.07±0.01	0.67±0.06	0.71±0.06	0.43±0.04
Multi Scale Self-Quotient	0.11±0.01	0.31±0.09	0.19±0.04	0.59±0.05	0.66±0.04	0.40±0.04
Isotropic Diffusion	0.26±0.05	0.27±0.07	0.21±0.03	0.58±0.06	0.65±0.07	0.40±0.05
Modified Anisotropic Diffusion	0.16±0.03	0.29±0.04	0.08±0.01	0.60±0.07	0.61±0.08	0.35±0.04
Tan & Triggs	0.29±0.05	0.35±0.05	0.10±0.01	0.81±0.04	0.78±0.05	0.47±0.04
Multi-Scale Weberfaces	0.39±0.05	0.34±0.05	0.18±0.03	0.79±0.04	0.78±0.05	0.50±0.04
Adaptive Non-Local Means	0.22±0.05	0.58±0.09	0.17±0.04	0.37±0.04	0.69±0.03	0.41±0.05
Retina Modeling	0.23±0.03	0.38±0.05	0.09±0.02	0.75±0.06	0.68±0.06	0.43±0.04
Wavelet Denoising	0.09±0.01	0.37±0.08	0.09±0.01	0.30±0.03	0.46±0.06	0.26±0.03
Homomorphic	0.10±0.02	0.41±0.08	0.18±0.02	0.44±0.07	0.54±0.08	0.33±0.05

MSW technique (see Figure 3(a) and (b)). Figure 4 displays face representations of individuals ID03 and ID04.

Figures 5 and 6 present an example of trajectory-level analysis with target and non-target ROIs over time when compared to the template for ID03 and ID04 individuals, respectively. They show matching scores along with measures of brightness and sharpness [12] with MSW and TT normalization associated with each ROI captures in Chokepoint video[6]. In Figure 5, the performance of the FR system that uses MSW normalization yields the best target vs non-target discrimination, although this tends to vary along with brightness and sharpness measures. In Figure 6, the scores are already very high for individual ID04, and normalization only improves non-target scores. This reduced the target vs non-target discrimination, and the overall ROC and Precision-Recall space performance. In this last case, there is no benefit to applying a normalization technique.

In most cases, there is at least one normalization technique that provides an improvement over the case without normalization. Given the diversity of approaches, results suggest that the scores obtained from a set of normalization techniques could be combined through fusion to achieve a higher level of accu-
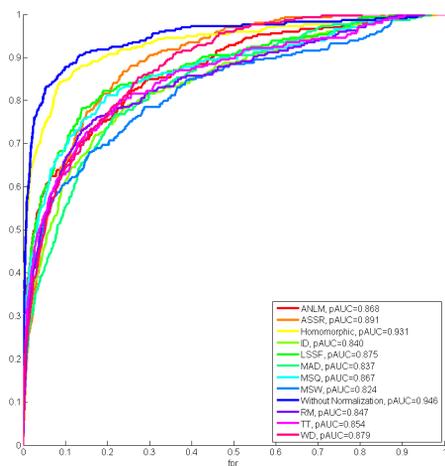
---

[6] The used Chokepoint video is: $P1E - S1 - C2$ (portal entering 1 sequence 1 and camera 2)
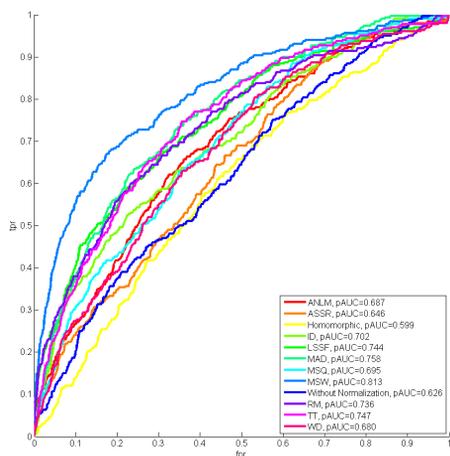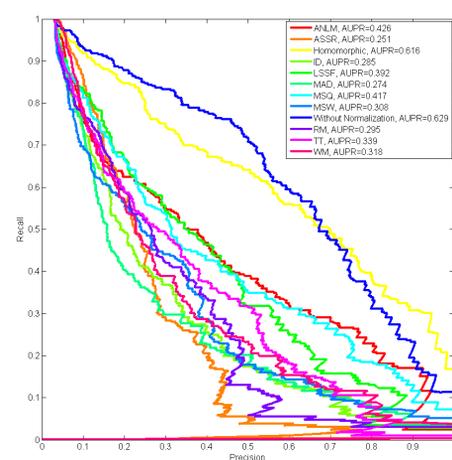
(a) ROC curves, ID03
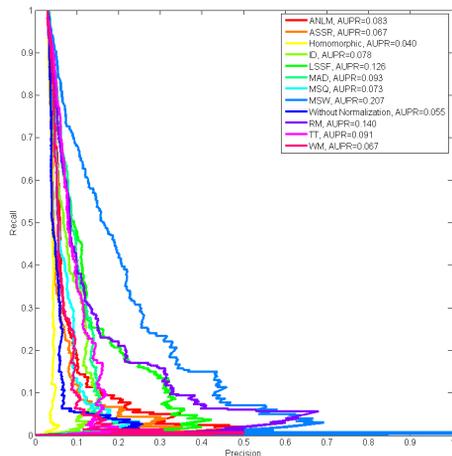
(b) Precision-Recall curves, ID03

(c) ROC curves, ID04

(d) Precision-Recall curves, ID04

Fig. 3: Transaction-level performance obtained with individuals ID03 and ID4 after using different illumination techniques.

racy and robustness. Since there is a correlation between brightness and scores achieved through normalization, a combination of these techniques should be selected dynamically based on changing capture conditions.

6   Conclusion

The popular LBP-based approach to face analysis is known to be sensitive to severe illumination changes. Based on this observation, our study investigated
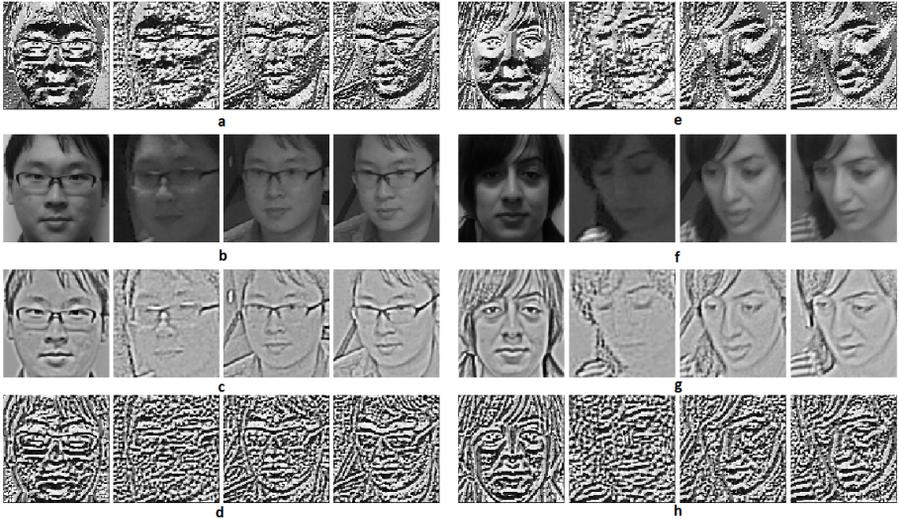
Fig. 4: Face representations of individuals ID03 and ID04. (a)LBP projection of the original ROI of ID03. (b) The original ROIs of ID03 (mug-shot and 3 from video captures). (c) Normalization of ROIs for ID03 using Multi-Scale Weberfaces. (d) LBP projection of the MSW normalized images of ID03. (e)LBP projection of the original ROI of ID04. (f) The original ROIs of ID03 (mug-shot and 3 from video captures). (g) Normalization of ROIs for ID04 using Multi-Scale Weberfaces. (h) LBP projection of the MSW normalized images of ID04.

the effect on performance of representative passive illumination normalization techniques for representation of face captures in watchlist screening with LBP. Watch-list screening is an important application for decision support in video surveillance systems.

Extensive experimental analysis on videos from the benchmark Chokepoint dataset indicated that the benefit of different techniques varies considerably according to the individual and illumination conditions. This suggests that a combination of these techniques should be selected dynamically based on changing capture conditions. Overall, the Multi-Scale Weberfaces and Tan and Triggs techniques tend to provide the most interesting results compared to other techniques.

Techniques in this study compensate for illumination changes at the preprocessing level, and may be computationally simple and effective at achieving illumination invariant FR. However, a common challenges among all theses techniques is that performance depends heavily on their implementation, and on the suitable selection of their parameters that must be set empirically. In this study, results were produced using default setting from the authors of respective techniques.
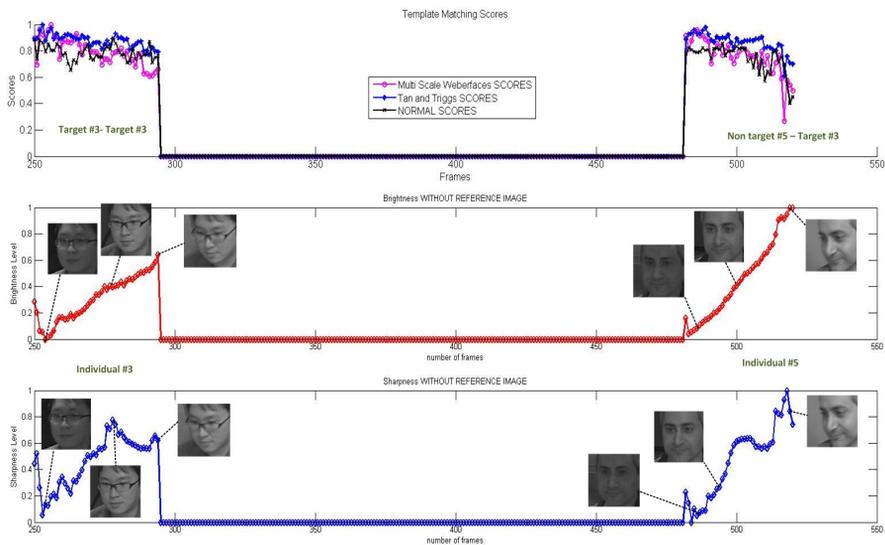
Fig. 5: Trajectory-level analysis for individual ID03 – matching scores and brightness and sharpness levels over time.
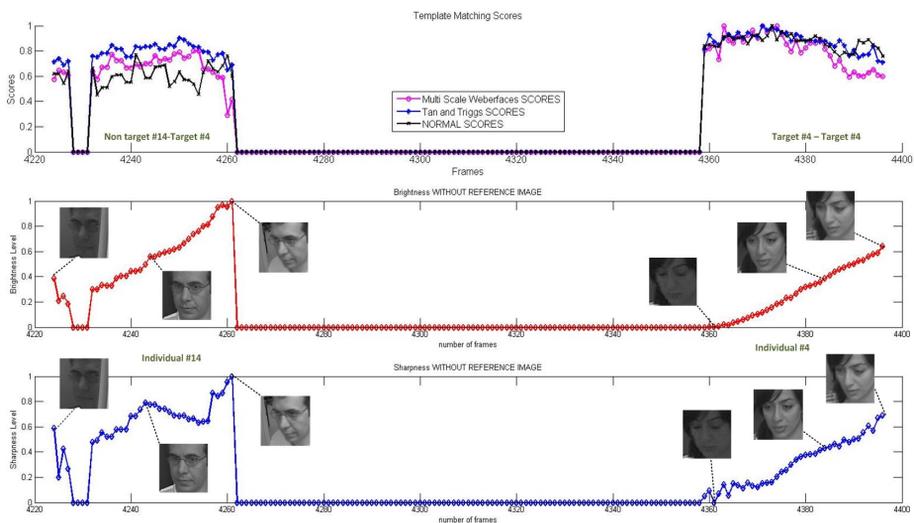

Fig. 6: Trajectory-level analysis for individual ID04 – matching scores and brightness and sharpness levels over time.

References

1. Adaptive discriminant learning for face recognition. Pattern Recognition 46(9), 2497 – 2509 (2013)
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. TPAMI 28(12), 2037–2041 (2006)
3. Barr, J.R., Bowyer, K.W., Flynn, P.J., Biswas, S.: Face recognition from video: A review. International Journal of Pattern Recognition and Artificial Intelligence 26(05) (2012)
4. Chellappa, R., Sinha, P., Phillips, P.J.: Face recognition by computers and humans. Computer 43(2), 46–55 (2010)
5. Dewan, M., Granger, E., Roli, F., Sabourin, R., Marcialis, G.L.: A comparison of adaptive appearance methods for tracking faces in video surveillance. In: The 5th International Conference on Imaging for Crime Detection and Prevention (December 16-17, 2013)
6. Ekenel, H.K., Stallkamp, J., Stiefelhagen, R.: A video-based door monitoring system using local appearance-based face models. CVIU 114(5), 596–608 (2010)
7. He, C., Ahonen, T., Pietikainen, M.: A bayesian local binary pattern texture descriptor. In: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. pp. 1–4 (Dec 2008)
8. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination Invariant Face Recognition Using Near-Infrared Images. IEEE T. PAMI 29(4), 627–639 (2007), http://dx.doi.org/10.1109/TPAMI.2007.1014
9. Liao, S., Zhao, G., Kellokumpu, V., Pietikainen, M., Li, S.: Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. pp. 1301–1306 (June 2010)
10. Marcel, S., Rodriguez, Y., Heusch, G.: On the recent use of local binary patterns for face authentication. International Journal of Image and Video Processing, Special Issue on Facial Image Processing p. 469481 (2007)
11. Matta, F., Dugelay, J.L.: Person recognition using facial video information: a state of the art. Journal of Visual Languages and Computing 20(3), 180–7 (2009)
12. Nasrollahi, K., Moeslund, T.: Face quality assessment system in video sequences. In: Schouten, B., Juul, N., Drygajlo, A., Tistarelli, M. (eds.) Biometrics and Identity Management, Lecture Notes in Computer Science, vol. 5372, pp. 10–18. Springer Berlin Heidelberg (2008), http://dx.doi.org/10.1007/978-3-540-89991-4_2
13. Nikan, S., Ahmadi, M.: Human face recognition under occlusion using lbp and entropy weighted voting. In: "ICPR". pp. 1699–1702. IEEE (2012)
14. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. TPAMI 24(7), 971–987 (2002)
15. Pagano, C., Granger, E., Sabourin, R., Gorodnichy, D.O.: Detector ensembles for face recognition in video surveillance. In: IJCNN. pp. 1–8. IEEE (2012)
16. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision Using Local Binary Patterns. Springer (2011)
17. Sharma, A., Kaushik, V.D., Gupta, P.: Illumination invariant face recognition. In: Intelligent Computing Theory. Lecture Notes in Computer Science, vol. 8588, pp. 308–319 (2014)

18. Struc, V., Pavesic, N.: Performance evaluation of photometric normalization techniques for illumination invariant face recognition. In: Zhang, Y. (ed.) Advances in Face Image Analysis: Techniques and Technologies. IGI Global, Hershey, USA (2011)

19. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Zhou, S., Zhao, W., Tang, X., Gong, S. (eds.) Analysis and Modeling of Faces and Gestures, Lecture Notes in Computer Science, vol. 4778, pp. 168–182. Springer Berlin Heidelberg (2007), `http://dx.doi.org/10.1007/978-3-540-75690-3_13`

20. Topcu, B., Erdogan, H.: Decision fusion for patch-based face recognition. In: ICPR. pp. 1348–1351. IEEE (2010)

21. De-la Torre, M., Granger, E., Sabourin, R., Gorodnichy, D.O.: Partially-supervised learning from facial trajectories for face recognition in video surveillance pp. In Press, DOI: 10.1016/j.inffus.2014.05.006 (2014)

22. Štruc, V., Pavešić, N.: Gabor-based kernel partial-least-squares discrimination features for face recognition. Informatica (Vilnius) 20(1), 115–138 (2009)

23. Štruc, V., Pavešić, N.: IGI Global (2011)

24. Wong, Y., Chen, S., Mau, S., Sanderson, C., Lovell, B.C.: Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In: IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 81–88. IEEE (June 2011)

25. Zou, X., Kittler, J., Messer, K.: Illumination invariant face recognition: A survey. In: Biometrics: Theory, Applications, and Systems, 2007. First IEEE International Conference on. pp. 1–8 (2007)