Contextual Weighting of Patches for Local Matching in Still-to-Video Face Recognition

Ibtihel Amara[1], Eric Granger[1], and Abdenour Hadid[2]
[1] Laboratory for Imagery, Vision and Artificial Intelligence
Ecole de technologie supérieure, Université de Québec, Canada
[2] Center for Machine Vision and Signal Analysis (CMVS),
University of Oulu, Finland

***Abstract*— Still-to-video face recognition (FR) systems for watchlist screening seek to recognize individuals of interest given faces captured over a network of video surveillance cameras. Screening faces against a watchlist is a challenging application because only a limited number of reference stills is available per individual during enrollment, and the appearance of face captures in videos changes from camera to camera, due to variations in illumination, pose, blur, scale, expression and occlusion. In order to improve the robustness of FR systems, several local matching techniques have been proposed that rely on static or dynamic weighting of patches. However, these approaches are not suitable for watchlist screening applications where the capturing conditions vary significantly over different camera fields of view (FoV). In this paper, a new dynamic weighting technique is proposed for weighting facial patches based on video data collected a priori from the specific operational domain (camera FoV) and on image quality assessment. Results obtained on videos from the Chokepoint dataset indicate that the proposed approach can significantly outperform the reference local matching methods because patch weights tend to grow for discriminant facial regions.**

I. Introduction

Video surveillance (VS) cameras have become omnipresent in many public places such as airports, train station, banks, etc. A common application in video surveillance is watchlist screening over multiple surveillance cameras. To enrol individuals to the watchlist, facial regions of interest (ROIs) are initially isolated from reference still images. During operations, ROIs corresponding to faces detected in the video surveillance cameras are matched against the reference ROIs of each target individual in the watchlist. In VS, a person in a scene may be tracked over several frames, and matching scores may be accumulated over a facial trajectory (a group of ROIs that correspond to the same high-quality track of an individual) for robust spatio-temporal FR. An alarm is triggered if accumulated matching scores linked to a watch-list individual surpasses an individual-specific threshold [1].

An important issue in still-to-video FR systems applied to watchlist screening is the domain shift related to differences in capture conditions of the source enrolment domain (stills images, often captured a priori under controlled conditions) with respect to the target operational domain (video frames captures under uncontrolled conditions). Indeed, the reference stills used for enrolment are usually higher quality facial images captures with a frontal pose. Probe ROIs captured from multiple surveillance cameras are often lower quality facial captures, and their appearance changes according to variations, e.g., pose, illumination, scale, blur and occlusion. Moreover, the facial models designed during enrollment for matching may not be representative of faces captured in videos because they are designed a priori with few reference stills. The resulting lack of representativeness of facial models can yield poor FR performance.

Some techniques have been proposed in FR literature to address "Single Sample Per Person" (SSPP) problems, as found in watchlist screening. In order to improve robustness to intra-class variability, FR techniques specialized for SSPP problems must often rely on multiple face representations (employing different face descriptors), synthetic generation, and auxiliary reference data [2]. For instance, techniques that exploit multiple face representations have been proposed for local matching using different patch configurations and feature extraction techniques [3] and [4]. An ensemble of exemplar SVMs based on multiple descriptors and patch configurations has also been proposed [5]. With synthetic generation techniques, different head poses have been generated from a single reference frontal facial image through 3D reconstruction [6]. In addition, target and non-target training sets have been synthetically generated to train a face classifier by applying 2D morphology on the single still frontal reference image [7]. In [8], synthetic versions of reference still facial images have been generated for various sharpness and illumination conditions by morphing stills with faces captured in the scene.

These techniques may improve the FR performance of a still-to-video FR system, although it is more costly to match faces according to several descriptors or to use several synthetic images. Complexity is an important consideration since modern cameras produce high-definition signals, and FR systems must rapidly process and manage large volumes of data from ever-growing surveillance networks. Moreover, multiple representation and synthetic generation techniques alone are only effective to the extent where reference target ROIs captured during enrolment are representative of the operational domain. In video surveillance applications, SSPP problems can also be addressed by exploiting auxiliary

The financial support from the Academy of Finland is fully acknowledged

data comprised of faces extracted from an abundance of operational videos, where non-target individuals are captured in the scene, obtained for instance during calibration of surveillance cameras. For instance, model driven techniques for domain adaptation is implemented in [9].

Local patch-based matching techniques have been shown to improve robustness of FR systems under the effects of occlusion, illumination, scale and blur [10]. Part-based or component-based techniques, like the Scale Invariant Feature Transform [11], proceed by locating discriminant or salient facial regions prior matching. These techniques typically perform well when important key facial features used for matching are easily located. In contrast, uniform patch configurations – consisting of a subdivision of ROIs into uniform patches – are more suitable for watchlist screening. Indeed, faces captured under uncontrolled capture conditions are typically distorted, misaligned and incorporate partial occlusion. To limit the complexity of face matching, this paper focuses on uniform non-overlapping patch configurations. In this paper, a new technique is proposed for dynamic weighting of patches for local face matching in still-to-video FR. This approach exploits video data collected a priori from a network of surveillance cameras. It relies on the context of the specific operational domain (camera FoV) in order to assign weights to patches. The proposed technique dynamically determines the importance of each patch based on image quality and a prior knowledge extracted from a camera FoV. Experiments using videos from the Chokepoint dataset show interesting results indicating that the proposed approach can significantly outperform the reference local matching methods.

II. Local Matching and Regional Weighting

Holistic face matching methods have been exploited in many FR systems, especially after the introduction of the Eigenfaces approach [12]. More recently, local patch-based matching techniques have been shown to improve FR performance because it is somewhat robust to facial variations due to occlusion, illumination changes, or other factors, and it also provides information on the spatial structure of a face. A comparison and a survey on local matching techniques can be found in [13] and [10].

By assigning a higher weight to patches, more importance can be placed on discriminant or salient regions in the facial image. The technique presented in [14] is performed statically with prior knowledge of the dataset and the application. In this case, weights are assigned based on distinctive facial components. Patches containing the eyes, nose and mouth were attributed high weights whereas the remaining regions were assigned lower weights. Based on their results, the system has shown an improvement when compared to local matching without weighting. A limitation of their proposed technique is that the weights are specific to their design data that only contains frontal images. When applied to video FR in a watchlist screening application, the distinctive features are not likely to occupy the same local region or sometimes can be occluded.

Another approach for local region weighting was proposed by Cheng and Chen [15]. Their concept was to divide the facial regions into uniform non-overlapping patches. Then, for each patch, multiple holistic classification algorithms are applied, and the weight values per region are defined in terms of their classification performance. The final weight is obtained through regional majority voting of the votes casted by each classifier. Although this approach has been shown to be promising, it tends to involve a complex implementation.

In this paper, the main objective is to propose a dynamic weighting scheme for local matching that exploits contextual information that can be derived during operations. Context was defined in [16] as a piece of information that one element can englobe but, once acknowledged, it may provide an estimation of a specific situation. Several context-aware techniques have recently emerged in pattern recognition literature [17]. Our hypothesis is that the weights of local patches should be adapted dynamically based the contextual information measured on probe ROIs. In watchlist applications, two source of contextual information could be exploited: (1) appearance variations due to multiple camera FoVs, and (2) appearance variations due to the non-stationary environment seen in a specific camera FoV. The first source of contextual information can be exploited using domain adaptation techniques, while the second source can be exploited through image quality assessment (IQA) measures. The rest of this section provides additional background on domain adaptation and IQA.

A. Domain Adaptation

Any distributional change (either domain shift or concept drift) can degrade performance of a still-to-video FR system. Context-aware systems could efficiently adapt to different and changing capture conditions. In the most common approach, prior expert knowledge of the expected operational domain is employed to define typical contexts and to design specialized individual detectors. Then, a suitable detector is selected dynamically among the pool for a given operational domain. In practice, however, this approach would only provide coarse adaptation because still-to-video FR systems are deployed in diverse and unknown capture conditions [17].

A source of contextual information for weighting of patches is the abundance of non-target faces captured a priori in with surveillance cameras. This can be used for domain adaptation (DA) of face models. Several transfer learning methods have recently been proposed to design accurate recognition systems [18]. Since the learning tasks and feature spaces between enrolment and operational domains are the same, but their data probability distributions are different, our transfer learning scenario is related to DA. According to the information transferred between an enrolment and operational domain, two unsupervised DA approaches from literature are relevant for still-to-video FR. Instance transfer methods attempt to exploit parts of the enrolment data for learning in the operational domain. In contrast, feature representation transfer methods exploit operational domain data to find a good common feature representation space that

reduces the difference between enrolment and operational spaces and the classification error.

B. Image Quality Assessment

Another source of contextual information for dynamic weighting of patches is the IQA of probe ROIs. In literature, IQA is categorized into two different groups: full-reference and no reference. The first category is in need of an original reference image to compare any input image and estimate its quality based on the original image. Full-reference image quality metrics have been proposed and exploited. An example is the Structural Similarity Index (SSIM) [19]. No reference IQA also called blind IQA is capable of returning a quality estimation without comparing it to a reference image. Unlike the full-reference IQA, this type is more useful in applications where reference images are not available. Many quality metrics under no reference have been suggested like head pose estimation [20] and [21], sharpness [22], contrast [23] and illumination [20].

Selecting high quality or discriminant facial ROIs at the pre-processing level (before performing feature extraction and matching) has become an interest in video FR because lower quality facial images can degrade the recognition performance of a FR system. In [24], a quality alignment module is used for selecting good quality frames in their still-to-video FR system. In [25], the authors exploited no-reference image quality for their wavelet-based FR system and using Nearest Neighbour classification. Their work demonstrated a remarkable improvement in terms of recognition rates. Similarly, in [26], the authors also used image quality for carefully selecting features/templates for better influence on the adaptive FR system's performance. The results have shown improvements compared to the non-adaptive system.

III. Our Proposed Approach: Context-Aware Weighting of Patches

A new regional weighting approach is proposed for still-to-video FR systems as needed in watchlist screening over a network of surveillance cameras. This patch weighting method accounts for local variations in capture conditions due to changes in head pose, illumination, blur, occlusions, etc. and on image quality. In addition, this proposed system relies on domain knowledge (camera view points).

The framework of a still-to-video FR system using this approach is shown in Figure 1. As any FR system, this proposed one consists in having two main phases: the design phase and the operation (or testing) phase.

Through a quick glimpse into the proposed system's framework, the first observation would be directly related to the complementary module for classification. This module is responsible for assigning local weights on non overlapping patches for each input ROI. By looking deeply into this module, a new component, aside from IQA is used which is the *specialized window*. The latter is responsible for giving more generality for the weighting technique depending on the camera view point. Indeed, one of the main characteristics of the proposed system is to provide suitable local weights for classification for every camera view point that is dependent on the quality of the input ROIs.

A. Enrollment process

The design phase of our proposed system is slightly different for traditional still-to-video FR systems. It is composed of two major parts.

Nb_{ind} individuals are randomly selected as watchlist. ROIs from the still images (mug-shots) are extracted using FD algorithm. A conversion into a gray scale and a resizing into a common size of $N_{size} \times N_{size}$ pixels is performed. The ROIs undergo *block division* into $N_p \times N_p = N_{TP}$ non overlapping patches and each patch has the size of $p_s \times p_s$. Finally, each patch of an ROI is fed to the *feature extraction* module then are saved into a set of features $\mathbf{M} = \{\mathbf{m}_1, ..., \mathbf{m}_{N_{TP}}\}$ in the gallery.

Calibration: Design of the specialized window using domain adaptation

This calibration process should be performed off-line for every camera view point. This task consists in implementing a classification process using template matching between low quality, low resolution and unlabelled probe video frames in the target domain and the enrolled high quality resolution still images of the calibration individuals in the source domain.

The *specialized window* contributes iteratively to re-weight the patches or local regions based on prior knowledge performed in advance for each camera viewpoint. This process borrows the concept of instance-weighting domain adaptation technique to assign weights on local regions. The key idea is to overcome the existing domain shift by finding a camera-specific window based on the component features and quality properties of the data distribution in the target domain.

Figure 1 shows the steps to obtain a specialized window for a certain camera view point. The calibration individuals are first enrolled into the gallery as facial models using their still reference image and undergoing the same traditional algorithm process for enrolment. Once these individuals are enrolled, the next step consists in the operational phase of this sub FR system. For one camera viewpoint, ROIs are extracted from the video sequence by performing the basic *segmentation* process (gray scale transformation and ROI resizing into $N_{size} \times N_{size}$ pixels). The segmented ROI goes through **block division** which divides the facial images into $N_p \times N_p$ having N_{TP} non overlapping blocks producing a set of blocks $\mathbf{b} = \{\mathbf{b}_1, ..., \mathbf{b}_{N_{TP}}\}$.

At this stage a two-layer process should be executed in order to calculate the *specialized window*.

1) First layer: Correlation of Local Scores and Local Image Quality: For each calibration individual, a series of tasks should be done.

1) Local matching
   This first task consists in performing direct matching. Each extracted ROI from the video frames goes

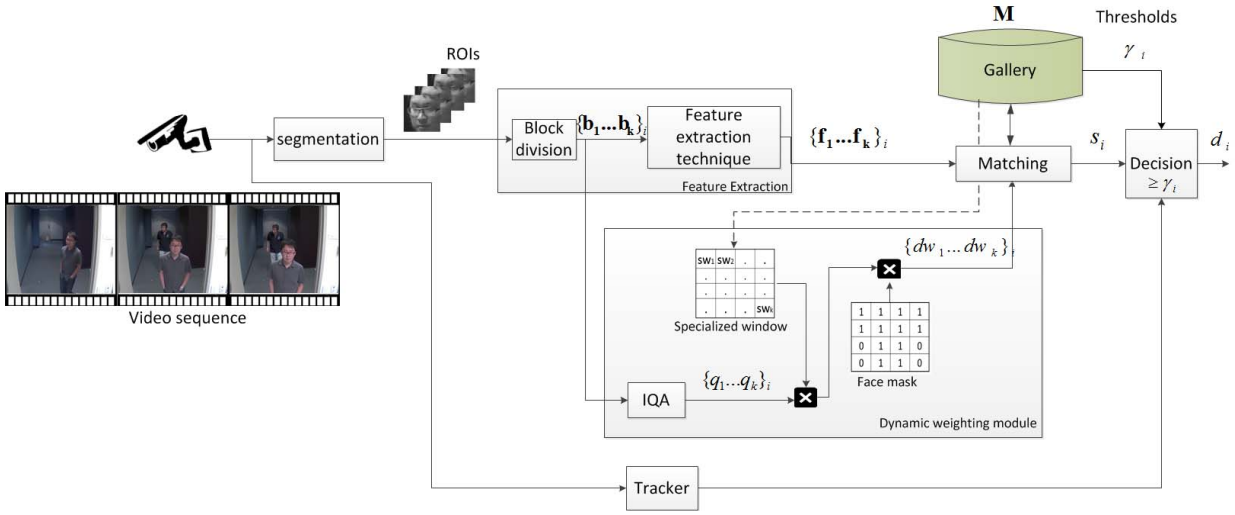

Fig. 1: Global framework of the still-to-video FR system with dynamic weighting during operations

through **segmentation** and **FE technique**. The obtained features are **locally matched** to the features from the still images of the calibration individuals saved in the gallery (one individual at a time) giving each ROI capture a set of local matching scores $\mathbf{S} = \{S_1, ..., S_{N_{TP}}\}$. For all frames from the video we would have a series of sets of local matching score.

2) Local IQA
This task consists in performing IQA locally on each patch or region of a captured ROI providing a set of local quality scores $\mathbf{q} = \{q_1, ..., q_{N_{TP}}\}$. For all frames from the video we would have a series of sets of local quality scores.

3) Patch correlation
At this stage, each individual used for calibration would have a series of sets of local matching scores and local sets of quality scores. These scores undergo **patch correlation**. This module considers one patch at a time for all frames in the video and assesses the relationship between the local matching scores and the local quality scores by providing correlation coefficient. So, for one calibration individual, a set of local correlation coefficient is obtained $\{coef f_1, ..., coef f_{N_{TP}}\}$. A primary binary window (of values 1 and 2) is created by performing **thresholding**. Each local correlation coefficient is compared to a *critical correlation coefficient* [1] r_{critic} value. If the local correlation coefficient of one patch is higher than the critical coefficient r_{critic} then the value of the corresponding patch in the primary binary window is equals to 2 if not, then it is equals to 1 providing a final product: the *specialized window* for one calibration individual and for one camera viewpoint.

[1]Pearson's critical values for significant correlation.

2) Second layer: camera specific specialized window using majority voting: A very important point that should be taken into consideration is that the design of the *specialized window* is done for each individual used for calibration per camera view point. So, the final camera specific *specialized window* that is integrated to the gallery of models (for a camera view point) is the combination of all 5 *specialized window* from all calibration people using the majority voting approach. The regions or patches containing more votes of value "2" are given that same value and correspondingly regions having more votes of value "1" are given that same value.

A basic flowchart is given in Figure 2 explaining step by step the process of designing the *specialized window*.

Through matching, correlation and majority voting a knowledge about the local regions is obtained. Indeed, for a single camera viewpoint, we were able to encourage and give importance to those regions/patches that are susceptible to contain high local matching scores and high local quality score by assigning values of "2" and to reduce the importance of certain regions by giving values of "1".

Through domain adaptation, we try to learn a better model for the source domain by: (1) Getting knowledge from the target domain; (2) Exploiting the particular information in the source domain.

B. Operation phase

The probe video frames goes through the same **segmentation** process described in the enrolment of the target individuals to obtain a resized and gray scaled facial ROI on each frame. By then, this ROI endures **block division** into N_{TP} non overlapping local patches. These patches are first taken through the **FE module** to produce a set of local features of an ROI. The most important part of this phase would be the weight calculation module in which a domain adaptive information (camera-specific specialized window) obtained from the target domain in the calibration process is

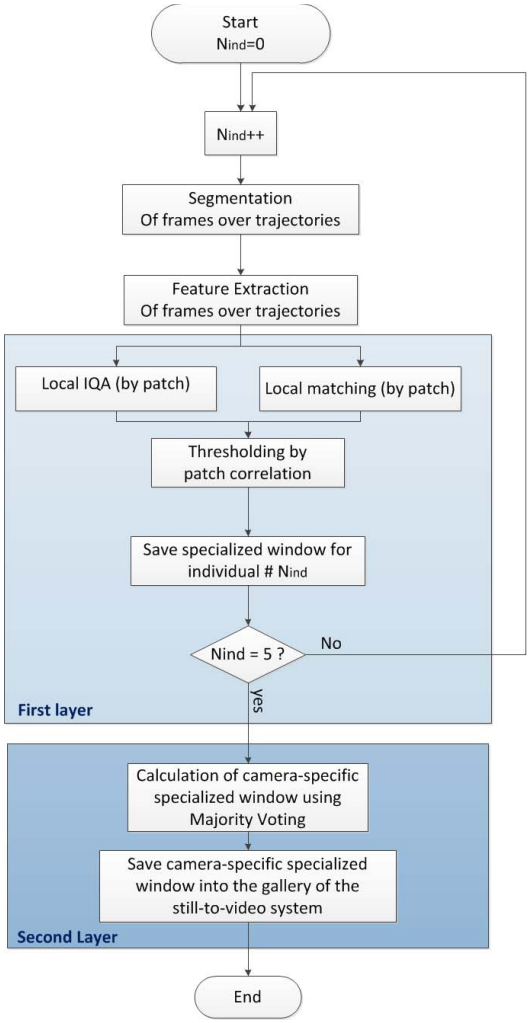

Fig. 2: Steps for calculating the *specialized window*

exploited along with contextual information based on quality to provide dynamic regional weights.

Weight Calculation using Contextual Information: For a camera viewpoint, Image Quality (IQ) is performed locally on each patch of the ROI providing a set of quality scores $\{q_1, ..., q_k\}$. The latter is multiplied to the set of values of the camera specific camera-specific *specialized window* (retrieved from the gallery) of that camera viewpoint used during the operation $\{sw_1, ..., sw_k\}$ producing a primary quality-based weights for each local area $\{q_1.sw_1, ..., q_k.sw_k\}$. Due to the presence of unwanted background a **face mask** $\mathbf{f}_{mask} = \{fm_1, ..., fm_k\}$ is applied to the primary attributed quality-based weights giving a final product of dynamic weights $\mathbf{dw} = \{dw_1, ..., dw_k\} = \{q_1.sw_1.fm_1, ..., q_k.sw_k, fm_k\}$. Dynamic weight calculation is illustrated for one single image for one specific camera recording condition in Figure 3. Finally, this window containing the set of dynamic weights is injected into the matching module for **classification** along with the corresponding set of features. The matcher produces a single value per ROI

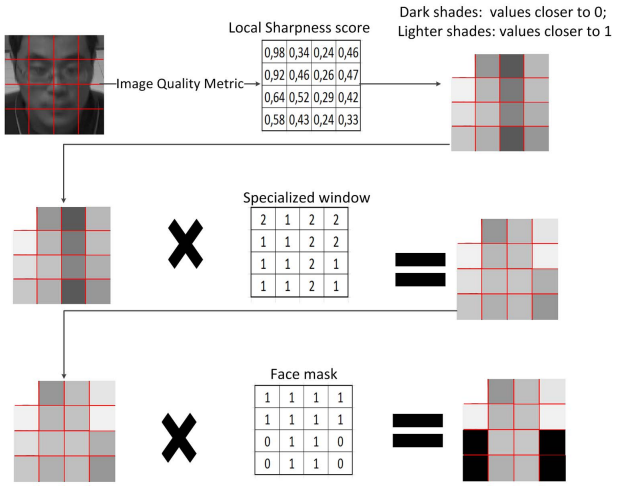

Fig. 3: Example of dynamic weight calculation for one single image of individual id1 in P2E_S1_C3 from the ChokePint dataset

comparison between 0 and 1 (having 0: likely to match and 1 unlikely to match).

IV. Experiments and Results

A. Experimental Methodology

This section presents a comparison between three techniques: the baseline (without wieghts), static weights in analogy with the work presented in [27] and the proposed technique (context-aware dynamic weighting).

This comparison considers $Nb_{calib} = 5$ randomly chosen individuals for calibration out of 29 subjects from the ChokePoint dataset. Another $Nb_{ind} = 5$ individuals are chosen as watchlist or target individuals out of the remaining 24 individuals. The remaining people are considered as non-target individuals for the testing phase. For this proposed system, all the video recordings from the Chokepoint dataset are used to assess the performance of the FR systems at different camera viewpoints. Details about each configuration and the methods used in the whole system construction is summarized in Table I.

Five independent replications are performed to validate the results of the experiments of this study. On each replication 5 different watchlist individuals are randomly chosen from the Chokepoint dataset. Then, experiments with the proposed FR systems are repeated on the same video sets for each chosen watchlist. The average results are shown over all 5 replications.

Watchlist screening corresponds to a class imbalance problem, where the number of target ROIs (positive class) processed by the FR system is usually far less than the number of non-target ROIs (negative class). Therefore, to account for this class imbalance, performance of still-to-video FR systems is evaluated using the Precision-Recall (P-R) space since both precision and recall do not consider true negatives. In particular, the global area under the P-R curve (AUPR) is presented in this paper.

TABLE I: Summary of techniques implemented for each module of baseline and proposed FR system (using dynamic weighting).

Segmentation	Face Detection: Viola-Jones Image gray-scale transformation Image resizing into 48×48 pixels
Block division	$N_{TP} = 16$ non overlapping patches
Feature Extraction	Local Binary Patterns and Local Phase Quantization
Classification during calibration	Local Chi Square $\chi^2(\mathbf{f}, \mathbf{m}) = \sum_j \sum_i \frac{(f_{ij} - m_{ij})^2}{f_{ij} + m_{ij})}$
Classification during operations	Weighted Chi Square $\chi^2(\mathbf{f}, \mathbf{m}) = \sum_j w_j \sum_i \frac{(f_{ij} - m_{ij})^2}{f_{ij} + m_{ij})}$
Image Quality Assessment	Sharpness $sharpness = \lvert lfI(x,y) - I(x,y) \rvert$ where $I(x,y)$ is the image and $lfI(x,y)$ is the filtered image
Patch correlation	Pearson's correlation rule
Thresholding	Pearson's critical value

B. Results and discussion

TABLE II: Average AUPR performance per camera for portal 1 entering after 5 replications using LBP and LPQ.

Feature Extraction	Techniques	Camera 1 (C1)	Camera 2 (C2)	Camera 3 (C3)
LBP	No weights	0.33 ± 0.041	0.37 ± 0.052	0.27 ± 0.043
	Static weights	0.39 ± 0.045	0.41 ± 0.052	0.31 ± 0.036
	Dynamic weights	$\mathbf{0.46 \pm 0.026}$	$\mathbf{0.45 \pm 0.056}$	$\mathbf{0.36 \pm 0.043}$
LPQ	No weights	0.40 ± 0.040	0.40 ± 0.051	0.36 ± 0.040
	Static weights	0.37 ± 0.047	0.41 ± 0.054	0.37 ± 0.040
	Dynamic weights	$\mathbf{0.45 \pm 0.052}$	$\mathbf{0.48 \pm 0.054}$	$\mathbf{0.40 \pm 0.039}$

TABLE III: Average AUPR performance per camera for portal 1 leaving after 5 replications using LBP and LPQ.

Feature Extraction	Techniques	Camera 1 (C1)	Camera 2 (C2)	Camera 3 (C3)
LBP	No weights	0.40 ± 0.042	0.45 ± 0.047	0.49 ± 0.061
	Static weights	0.42 ± 0.043	0.50 ± 0.052	0.48 ± 0.056
	Dynamic weights	$\mathbf{0.45 \pm 0.051}$	$\mathbf{0.51 \pm 0.058}$	0.41 ± 0.055
LPQ	No weights	0.48 ± 0.042	0.50 ± 0.043	0.52 ± 0.054
	Static weights	0.46 ± 0.041	0.50 ± 0.042	0.50 ± 0.049
	Dynamic weights	$\mathbf{0.50 \pm 0.046}$	$\mathbf{0.53 \pm 0.049}$	0.35 ± 0.046

Let us present the results per camera domain. Table II shows the average AUPR performance after 5 replications using LBP and LPQ for **portal 1 entering**. For both feature extraction techniques, the proposed system has shown an improvement overall. In fact, using LBP, an improvement of 0.13 (13%) in camera1, 8% in camera 2 and 9% for camera 3 compared to the baseline technique (no weights). As for using LPQ, an enhancement of 5% is seen in camera 1, 8% in camera 2 and 4% in camera 3.

If we look into the static weighting presented in the same Table II, this method is not stable in terms of results. For example in **portal 1 entering camera 1** using LPQ, the average AUPR decreased (0.37) compared to the baseline having an average value of AUPR equals to 0.40.

Same observations can be made to **portal 2 entering** and **leaving**. Tables IV and V present the average AUPR of **portal 2** respectively **entering** and **leaving**. The proposed

TABLE IV: Average AUPR performance per camera for portal 2 entering after 5 replications using LBP and LPQ.

Feature Extraction	Techniques	Camera 1 (C1)	Camera 2 (C2)	Camera 3 (C3)
LBP	No weights	0.24 ± 0.039	0.21 ± 0.036	0.19 ± 0.030
	Static weights	0.23 ± 0.033	0.22 ± 0.036	0.20 ± 0.036
	Dynamic weights	$\mathbf{0.27 \pm 0.041}$	$\mathbf{0.25 \pm 0.042}$	$\mathbf{0.21 \pm 0.031}$
LPQ	No weights	0.29 ± 0.036	0.23 ± 0.032	0.23 ± 0.037
	Static weights	0.29 ± 0.036	0.23 ± 0.032	0.21 ± 0.031
	Dynamic weights	$\mathbf{0.32 \pm 0.042}$	$\mathbf{0.28 \pm 0.043}$	$\mathbf{0.24 \pm 0.034}$

TABLE V: Average AUPR performance per camera for portal 2 leaving after 5 replications using LBP and LPQ.

Feature Extraction	Techniques	Camera 1 (C1)	Camera 2 (C2)	Camera 3 (C3)
LBP	No weights	0.28 ± 0.042	0.33 ± 0.041	0.31 ± 0.047
	Static weights	0.30 ± 0.045	0.35 ± 0.047	0.35 ± 0.050
	Dynamic weights	$\mathbf{0.33 \pm 0.047}$	$\mathbf{0.39 \pm 0.051}$	$\mathbf{0.41 \pm 0.055}$
LPQ	No weights	0.36 ± 0.044	0.41 ± 0.045	0.36 ± 0.040
	Static weights	0.31 ± 0.035	0.35 ± 0.035	0.34 ± 0.033
	Dynamic weights	$\mathbf{0.38 \pm 0.041}$	$\mathbf{0.43 \pm 0.044}$	$\mathbf{0.38 \pm 0.039}$

dynamic technique is outperforming compared to the baseline technique. For example, the average AUPR value has increased by 5% in camera 2 with LPQ in **portal 2 entering** and by 10% in camera 3 with LBP in **portal 2 leaving**.

Table III shows a further investigation of the average AUPR (after 5 replications) per camera for **portal 1 leaving**. The system performance at Camera 3 decreases when using the dynamic weighting for both LBP and LPQ features. The results for all three cameras (Cameras 1, 2 and 3) of **portal 1 leaving** are compared in terms of the number of captured samples, average illumination score and average sharpness score. In Figure 4, Camera 3 has the least number of samples, average illumination score and sharpness score, which explains this overall degradation of performance of the system, and specifically with the proposed dynamic weighting technique. Therefore, having a limited amount of faces captured in testing videos, and a lower ROI quality for illumination, and especially sharpness, may cause the decline in performance.

Figure 5 shows the accumulated scores over 30 frames when probe images are matches against ID#4 from the watchlist. As we can see, the proposed system provides better discrimination between the target and the non-target individuals present in the scene.

V. Conclusion

In the paper, the benefits for patch-based local face matching is shown in the case where uniform non-overlapping patches are dynamically weighted using contextual information. A new technique is proposed for dynamic weighting of patches for local matching in still-to-video FR. This approach exploits video data collected a priori from a network of surveillance cameras. The proposed technique dynamically determines the importance of each patch based on image quality and a prior knowledge extracted from a camera FoV.

The results obtained on videos from the Chokepoint dataset show that the proposed approach leads to a significant improvement in performance over baseline method and static

S= 0.42 I= 0.38	S= 0.47 I= 0.42	S= 0.40 I= 0.40	S= 0.39 I= 0.38
S= 0.45 I= 0.35	S= 0.58 I= 0.41	S= 0.45 I= 0.42	S= 0.52 I= 0.36
S= 0.46 I= 0.37	S= 0.54 I= 0.38	S= 0.40 I= 0.39	S= 0.50 I= 0.29
S= 0.39 I= 0.37	S= 0.42 I= 0.35	S= 0.32 I= 0.38	S= 0.55 I= 0.20

CAMERA 1
Average sharpness = 0.45
Average illumination = 0.36
Total number of captures = 6796

S= 0.48 I= 0.42	S= 0.49 I= 0.47	S= 0.40 I= 0.39	S= 0.40 I= 0.42
S= 0.54 I= 0.38	S= 0.59 I= 0.39	S= 0.47 I= 0.44	S= 0.48 I= 0.42
S= 0.54 I= 0.37	S= 0.66 I= 0.45	S= 0.47 I= 0.43	S= 0.58 I= 0.47
S= 0.43 I= 0.28	S= 0.44 I= 0.45	S= 0.35 I= 0.40	S= 0.52 I= 0.24

CAMERA 2
Average sharpness = 0.49
Average illumination = 0.40
Total number of captures = 6374

S= 0.45 I= 0.25	S= 0.27 I= 0.24	S= 0.28 I= 0.19	S= 0.49 I= 0.28
S= 0.31 I= 0.23	S= 0.21 I= 0.26	S= 0.29 I= 0.25	S= 0.37 I= 0.22
S= 0.38 I= 0.24	S= 0.22 I= 0.28	S= 0.26 I= 0.26	S= 0.37 I= 0.25
S= 0.49 I= 0.18	S= 0.28 I= 0.17	S= 0.30 I= 0.20	S= 0.48 I= 0.27

CAMERA 3
Average sharpness = 0.34
Average illumination = 0.24
Total number of captures = 3528

Fig. 4: Total number of facial captures, average illumination and average sharpness socre per camera in all videos in portal 1 leaving; (S) is the sharpness score and (I) is the illumination score

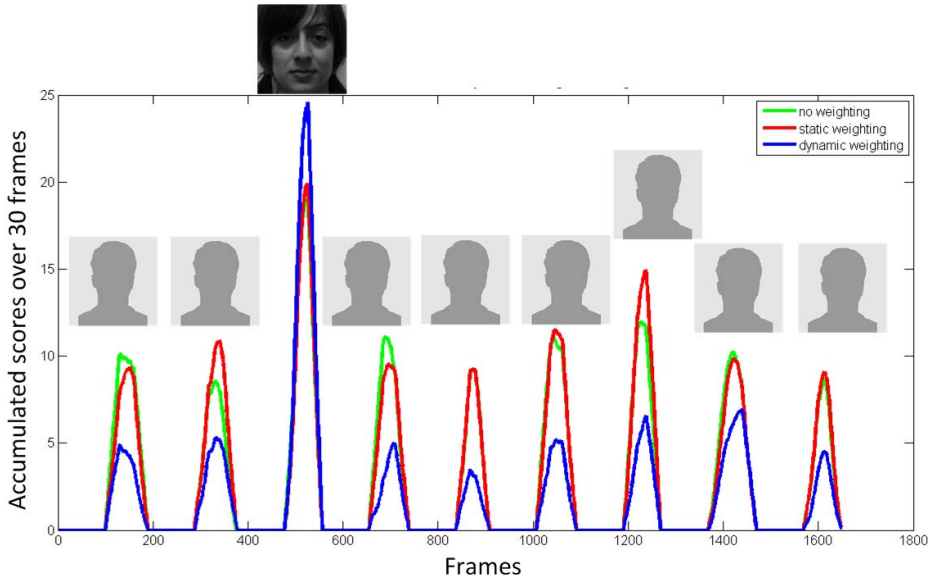

Fig. 5: Accumulated scores over 30 frames when inputs are matched against the reference for ID #4

weighting methods using the LBP and LPQ descriptors and the weighted Chi-squared similarity measure. Results suggest that quality information is correlated to the system's performance and can be leverages to improve still-to-video FR. In contrast, matching with static weighting leads to fluctuating because the weights are knowledge-based and are related to the dataset used for the estimation of the weights.

This work is by no mean complete. We plan to evaluate our proposed methodology on other databases under more challenging scenarios. Upon the publication of this work, we also plan to make the implementation code of our proposed framework publicly available for the research community.

REFERENCES

[1] M. A. A. Dewan, E. Granger, G.-L. Marcialis, R. Sabourin, and F. Roli, "Adaptive appearance model tracking for still-to-video face recognition," *Pattern Recognition*, vol. 49, pp. 129–151, 2016.

[2] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[3] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 74–81.

[4] B. Saman, G. Eric, S. Robert, and B. Guillaume-Alexandre, "Watch-list screening using ensembles based on multiple face representations," in *Pattern Recognition (ICPR), 2014 22nd Int. Con. on*. IEEE, 2014, pp. 4489–4494.

[5] S. Bashbaghi, E. Granger, R. Sabourin, and G.-A. Bilodeau, "Ensembles of exemplar-svms for video face recognition from a single sample per person," in *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE Int. Con.* IEEE, 2015, pp. 1–6.

[6] U. Park and A. K. Jain, "3d model-based face recognition in video," in *Advances in Biometrics*. Springer, 2007, pp. 1085–1094.

[7] B. Kamgar-Parsi and W. Lawson, "Toward development of a face recognition system for watchlist surveillance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 10, pp. 1925–1937, 2011.

[8] F. Mokhayeri, E. Granger, and G.-A. Bilodeau, "Synthetic face gen-

eration under various operational conditions in video surveillance," in *Int. Con. on Image Processing, Quebec, Canada*, 2015.

[9] R. Gopalan, R. Li, and R. Chellappa, "Unsupervised adaptation across domain shifts by generating intermediate data representations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 11, pp. 2288–2302, 2014.

[10] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 5, pp. 1193–1205, 2013.

[11] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B.-L. Lu, "Person-specific sift features for face recognition," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE Int. Con.*, vol. 2. IEEE, 2007, pp. II–593.

[12] J. Zhang, Y. Yan, and M. Lades, "Face recognition: eigenface, elastic matching, and neural nets," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1423–1435, 1997.

[13] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *Image Processing, IEEE Transactions on*, vol. 16, no. 10, pp. 2617–2628, 2007.

[14] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer vision-eccv 2004*. Springer, 2004, pp. 469–481.

[15] J. Cheng and L. Chen, *A Weighted Regional Voting Based Ensemble of Multiple Classifiers for Face Recognition*. Springer, 2014.

[16] A. Zimmermann, A. Lorenz, and R. Oppermann, "An operational definition of context," in *Modeling and using context*. Springer, 2007, pp. 558–571.

[17] L. Snidaro, L. Vati, J. Garcia, E. D. Marti, A.-L. Jousselme, K. Bryan, D. D. Bloisi, and D. Nardi, "A framework for dynamic context exploitation," in *Information Fusion (Fusion), 2015 18th Int. Con*. IEEE, 2015, pp. 1160–1167.

[18] J. Pan and Q. Yang, "A survey on transfer learning," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 22, no. 10, pp. 1345–1359, 2010.

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[20] K. Nasrollahi and T. B. Moeslund, "Face quality assessment system in video sequences," in *Biometrics and Identity Management*. Springer, 2008, pp. 10–18.

[21] M. Abdel-Mottaleb and M. H. Mahoor, "Application notes-algorithms for assessing the quality of facial images," *Computational Intelligence Magazine, IEEE*, vol. 2, no. 2, pp. 10–17, 2007.

[22] F. Weber, "Some quality measures for face images and their relationship to recognition performance," in *Biometric Quality Workshop. National Institute of Standards and Technology, Maryland, USA*, 2006.

[23] A. Abaza, M. A. Harrison, and T. Bourlai, "Quality metrics for practical face recognition," in *Pattern Recognition (ICPR), 2012 21st Int. Con.* IEEE, 2012, pp. 3103–3107.

[24] Z. Huang, X. Zhao, S. Shan, R. Wang, and X. Chen, "Coupling alignments with recognition for still-to-video face recognition," in *Computer Vision (ICCV), 2013 IEEE Int. Con. on*. IEEE, 2013, pp. 3296–3303.

[25] A. J. Abboud, H. Sellahewa, and S. A. Jassim, "Quality based approach for adaptive face recognition," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009, pp. 73 510N–73 510N.

[26] A. J. Abboud and S. A. Jassim, "Biometric templates selection and update using quality measures," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2012, pp. 840 609–840 609.

[27] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.