

Efficient Adaptive Face Recognition Systems Based on Capture Conditions

C. Pagano, E. Granger, R. Sabourin
Lab. d'imagerie, de vision et d'intelligence artificielle
École de technologie supérieure
Université du Québec, Montreal, Canada
cpagano@livia.etsmtl.ca
{eric.granger, robert.sabourin}@etsmtl.ca

A. Rattani, G.-L. Marcialis, F. Roli
Pattern Recognition and Applications Group
Dept. of Electrical and Electronic Engineering
University of Cagliari, Cagliari, Italy
ajita.rattani@diee.unica.it
{marcialis, roli}@diee.unica.it

Abstract—In many face recognition (FR) applications, changing capture conditions lead to divergence between facial models stored during enrollment and faces captured during operations. Moreover, it is often costly or infeasible to capture several high quality reference samples a priori to design representative facial models. Although self-updating models using high-confidence face captures appear promising, they raise several challenges when capture conditions change. In particular, face models of individuals may be corrupted by misclassified input captures, and their growth may require pruning to bound system complexity over time. This paper presents a system for self-update of facial models that exploits changes in capture conditions to assure the relevance of templates and to limit the growth of template galleries. The set of reference templates (facial model) of an individual is only updated to include new faces that are captured under significantly different conditions. In a particular implementation of this system, illumination changes are detected in order to select face captures from bio-login to be stored in a gallery. Face captures from a built-in still or video camera are taken at periodic intervals to authenticate the user having accessed a secured computer or network. Experimental results produced with the DICE dataset show that the proposed system provides a comparable level of performance to the FR system that self-updates the gallery on all high-confidence face captures, but with significantly lower complexity, i.e., number of templates per individual.

I. INTRODUCTION

Automatic FR is emerging as a key function for enhanced user authentication in computer networks, smartphones, tablets, etc. For instance, the face modality may allow for continuous user authentication in the context of controlled access to personal or confidential information on computer and network systems [11]. When a new user initially logs into a computer or network, a FR system may enroll the user (and thus design a facial model¹), according to faces captured using a computer's built-in camera. The system may then periodically validate that faces captured by the computer during operations correspond to the same user's face model.

It is well known that the performance of state-of-the-art FR technologies typically declines due to changes in capture conditions (e.g., variation in illumination, pose and scale) and human physiology [5], [13]. Moreover, only a limited number

of high-quality reference face captures are available during the enrollment process, and face models are not necessarily representative of real face to be recognized [12]. For enhanced accuracy and reliability, adaptive biometric techniques may be used to update facial models with highly confident² operational data [14]. Self-updating face models under changing semi-controlled conditions may however lead to knowledge corruption if highly-confident face captures are miss-classified. In addition, the size and complexity of a face model may required management (e.g. pruning) to bound system complexity over time.

In this paper, an adaptive FR system is proposed where self-updating is triggered by the detection of changes in capture conditions, instead of relying on high-confidence scores. Given a template matching system, the set of reference templates (facial model) of each enrolled individual is only updated to include faces that are captured during operations under significantly different conditions. If the system detects that a high-confidence face capture only incorporates a gradual change w.r.t. existing templates, this capture is not employed to update the corresponding set of user templates. In contrast, if the system detects an abrupt change, this capture is included into the set of user templates to sustain a high level of performance and restrain model size, while preserving previous-acquired knowledge.

For proof-of-concept validation, a particular implementation of this proposed system is considered. Illumination changes are monitored over time in order to select relevant and diverse faces captures to be stored in a gallery for bio-login to a secured computer network. Experiments are conducted on the DICE multi-modal dataset [15], and performance is assessed in terms of accuracy and resource requirements. Face images in the DICE dataset consists of stills captured under varying illumination conditions – gradual, recurring and abrupt patterns of change – and were explicitly collected to evaluate self-update systems.

II. SELF-UPDATING IN BIOMETRICS

Strategies for the design of adaptive biometric systems involve: (1) the *selection* of diversified and relevant reference samples to update facial models, and (2), the actual *update*

¹A facial model is defined as either a set of one or more reference face captures (used for template matching), or a statistical model estimated by training with reference captures (used for classification).

²A highly-confident capture produces a high matching score, that surpassed a self-update threshold when compared to a given facial model.

TABLE I. TYPES OF CHANGES OCCURRING IN FACE RECOGNITION.

Type of change	Examples in FR
1) random noise	– rare event and inherent noise of system (camera, matcher, etc.)
2) gradual changes	– ageing of user over time
3) abrupt changes	– new unseen capture conditions (e.g. pose angle, scale, etc.)
4) recurring contexts	– unpredictable but recurring changes in capture conditions (e.g. daily variations in artificial or natural illumination.)

of facial models (template galleries or classifier parameters). Suitable techniques for the selection of relevant samples for the self-update of a biometric system have been reviewed in [5]. The level of informativeness of a sample may be estimated using selection techniques based on the data itself, or retrieving information (predictions) from the system. Examples of selection techniques used for FR include editing algorithms such as the Condensed Nearest Neighbor method, used to manage a template gallery in a template matching systems [7].

Several techniques have been proposed in literature to self-update facial models in adaptive biometrics [14]. They are generally based on matching scores, and have been proposed for template matching systems [17]. Biometric models are first designed by storing samples from a labeled data set D_L in a template gallery \mathcal{G} . Then, during operations, similarity scores for unlabeled samples are produced through template matching. A positive prediction is output if the score surpasses the decision threshold γ^d . Predictions linked to a high degree of confidence (matching score surpasses a higher updating threshold, $\gamma^u \geq \gamma^d$), are integrated to the gallery \mathcal{G} , thereby updating the corresponding biometric models.

Although self-update methods have been shown to improve the representation of intra-class variations in facial models as well as system accuracy [14], adapting a biometric system using operational data carries an inherent risk. The trade-off between false updates and false rejections directly affects performance, and the decision threshold is a critical parameter. In addition, in template matching systems, the number of templates should be limited since it is directly related to the system's time and memory complexity. This raises the issue of selecting the most representative templates to be preserved in a gallery over time.

III. CONCEPT CHANGE IN FACE RECOGNITION

In pattern recognition literature, a *concept* is defined as the underlying data distribution of the problem at some point in time [1]. A *concept change* encompasses various types of noise, trends and substitutions in the underlying data distribution associated with a concept. A categorization of changes has been proposed by Minku et al. [10] based on severity, speed, predictability and number of re-occurrences, but noise, abrupt changes, gradual changes and recurring changes are mainly considered in the literature.

Concept change can be viewed in the context of FR, for example in continuous user authentication. In this context, changes originate mostly from variations in capture conditions

that have yet to be integrated into the facial model of the individual [13]. As shown in Table I, they may range from minor random fluctuations or noise, to sudden abrupt changes of the underlying data distribution, and are not mutually exclusive in real-world surveillance environments. In this paper, FR is performed under semi-controlled capture conditions, and concept changes are observed as illumination variations in faced captured during operations. The refinement of previously-observed concepts (e.g., faces captured for previously seen illumination) corresponds to gradual changes, and data representing newly-observed concepts (e.g., faces captured under previously unseen illumination conditions) correspond to abrupt changes. A new concept can also correspond to a recurring change as specific observation conditions may be re-encountered in the future (e.g., faces captured under natural vs. artificial lighting).

A. Measures for Detection of Concept Change

To observe changes in the underlying data distribution, several measures have been proposed in computer vision and pattern recognition literature.

Computation of quantitative and objective image quality measures have been used for quality control, to accept, reject, or reacquire biometric samples, as well as to select a biometric modality, algorithm, and/or system parameters [19]. Change detection can be performed by monitoring the values of an image-based quality over time. For example, several face image quality standards have been proposed to evaluate facial quality, such as ISO19794-5 and ICAO 9303 [18], which can be divided into: (1) image specific qualities (contrast, sharpness, compression artefact, etc.) and (2), face specific qualities (face geometry, pose angle, eye visibility, illumination, etc.). In addition, change detection may also involve comparing face captures against the gallery templates to measure image variation or distortion. Changes can also be measured after the feature extraction process of a pattern recognition system, and thus not be measured on ROI, but in their feature distribution space, and the accuracy of these measures depends heavily on the feature extraction and selection methods. Pattern recognition-based techniques fall under two categories to measure changes: those that exploit (1) classifier performance indicators, and (2) density estimation. In either case, detection of changes is then performed through thresholding.

The rest of this subsection focuses on the universal image quality index (Q) [20], which will be used in this paper to assess face image distortions. This index models the distortion of any input or probe image w.r.t. any reference image as a combination of three factors – loss of correlation, luminance distortion and contrast distortion. For instance, the quality of a face capture may be defined in terms of its luminance distortion w.r.t. a known reference face. The luminance distortion factor of Q (referred to as the luminance quality, LQ) may be used to measure global or regional illumination quality of images.

Let $\mathbf{R} = (r_1, r_2, \dots, r_n)$ and $\mathbf{X} = (x_1, x_2, \dots, x_n)$ be a reference and an input facial image signals, respectively. The universal quality index in [20] is defined as

$$Q = \frac{4\sigma_{rx}\bar{r}\bar{x}}{(\sigma_r^2 + \sigma_x^2)(\bar{r}^2 + \bar{x}^2)} \quad (1)$$

where \bar{r} and \bar{x} are the average images of the reference and input signals, σ_r^2 and σ_x^2 the variances, and σ_{rx} the covariance. Statistical features for Equ.1 may be measured locally to accommodate space-variant nature of image quality, and then combined to an overall quality measure for the entire image. A local quality index Q_k is thereby calculated with Q by sliding a window of $B \times B$ pixels from the top-left corner to the bottom-right corner of the image. For a total of K steps, the global quality index is given by:

$$GQ(\mathbf{R}, \mathbf{X}) = \frac{1}{K} \sum_{k=1}^K Q_k \quad (2)$$

The universal quality index Q of Equ.1 can be written as a product of the three factors – loss of correlation, luminance distortion and contrast distortion:

$$Q = \frac{\sigma_{rx}}{\sigma_r \sigma_x} \cdot \frac{2\bar{r} \bar{x}}{\bar{r}^2 + \bar{x}^2} \cdot \frac{2\sigma_r \sigma_x}{\sigma_r^2 + \sigma_x^2} \quad (3)$$

Luminance quality or distortion factor LQ (second component in Q) is therefore defined as:

$$LQ = \frac{2\bar{r} \bar{x}}{\bar{r}^2 + \bar{x}^2} \quad (4)$$

$LQ \in [0, 1]$ measures the proximity of the average luminance between r and x . LQ equals 1 if and only if $\bar{r} = \bar{x}$. In the same way than the GQ index, the Global LQ index GLQ can be computed, following:

$$GLQ(\mathbf{R}, \mathbf{X}) = \frac{1}{K} \sum_{k=1}^K LQ_k \quad (5)$$

B. Decision Thresholding

1) *Fixed Window and Threshold:* Explicit change detection methods as proposed by Klinkenberg et al. [8] usually monitor the evolution of performance indicators over time (such as classifier accuracy, recall or precision). Using a sliding window over the previous data batches, those measures are compared with a confidence interval of standard sample errors.

The detection of changes for one-dimensional data has been extensively studied in the area of quality control [9]. Assuming a stream of objects with a known probability p of being defective (given from product specifications) several control chart schemes have been proposed. According to the basic Shewhart control chart scheme, a window or batch of N object samples is inspected at regular intervals. The number of defective objects is counted, and an estimate \bar{p} is plotted on the chart. Using a threshold of $f\sigma$, where $\sigma = \sqrt{p(1-p)/N}$ (a typical value is $f = 3$) a change is detected if $\bar{p} > p + f\sigma$. Among the numerous control chart approaches, there are the well-known CUMulative Sum (CUSUM) charts and Sequential Probability Ratio Test (SPRT). CUSUM rely on the computation of the cumulative sum of classification errors S_t at time t , and a control limit, or a threshold, h . If $E_t > h$, a change is signalled. Reynolds and Stoumbos [16] proposed with SPRT a more successful implementation, using two thresholds h and g . If $h \leq E_t < g$, a new SPRT is started, considering the new samples to update its model.

However, in both cases, the computation of the decision

thresholds rely on the estimation of the error probability \bar{p} , as well as the probability \bar{p}^* after a change in the concept, and such information are rarely available in biometric systems. In addition, the choice of the window size N is critical to minimize the generalisation error on new examples. To address those potential issues, several authors proposed a dynamic adaptation of the size of the window encompassing the considered samples, as well as the threshold value to which the measures are compared in order to detect a change.

2) *Adaptive Window and Threshold Evaluation:* Klinkenberg and Renz [8] proposed a drift detection method to determine an optimal window size, using past batches to estimate changes in the error measure. The classifier is trained on a current data window, and the estimated error rate e_c is compared to a threshold, evaluated using the errors of the previous N batches. If e_c exceeds the threshold, a change is signalled, and a comparison with the error on the previous batch e_{c-1} determines the reduction of the training window. If the error is much larger ($e_c > 1 - \beta(1 - e_{c-1})$), an abrupt change is considered, and the training window is reduced to the current batch. Otherwise, a gradual change is supposed, and the window is reduced by a factor of γ . If no change is detected, the current batch is added to the training window. An improvement has been proposed by Baena-Garcia et al. [4], using the difference between two consecutive errors instead of a single error measure, thus improving the detection of slower gradual drifts. Similarly, Ditzler and Polikar proposed the Hellinger Drift Detection Method (HDDM) [6], to adapt the data distribution when a change is detected. When the data block D_t becomes available, its distance $\delta_H(t)$ between a reference distribution D_λ is computed. If $|\epsilon(t)| = |\delta_H(t) - \delta_H(t-1)| > \beta(t)$, $\beta(t)$ being an adaptive threshold, a change is signalled and D_λ is reset to D_t . Otherwise, D_λ is updated with samples from D_t . In the same way than Klinkenberg and Renz [8], this method considers a growing window of samples (or data blocks), which reduces itself to the current data when a change is detected, in order to update the reference distribution from which changes will be detected in the future.

To dynamically adapt the decision threshold, Alippi and Roveri proposed an adaptive evolution of the CUSUM method, dynamically updating the threshold h to the maximum difference between the cumulative sum and its minimal value [3]. This reduces the knowledge required about the hidden concept, and provides an automatic update of the change detection threshold. In addition, in HDDM [6], the adaptive threshold $\beta(t)$ is computed with previous values of $\epsilon(t)$, following $\beta(t) = \hat{\epsilon} + \gamma \cdot \sigma_{\epsilon}$, with $\hat{\epsilon}$ and σ_{ϵ} the average and standard deviation of ϵ_i for $\lambda < i < t-1$, and λ the last detected change. A t -statistic with a confidence interval of $\alpha = 0.1$ is considered to evaluate γ , then scaled by the square root of $t - \lambda - 1$. This dynamic update of the weight γ has proven more successful than the use of several standard values $\gamma = \{0.5, 1, 1.5, 2\}$ [6]. Such methods allow for dynamic adaptation of the criterion for change detection, thus basing the decision on the current detection environment and not *a priori* assumptions.

IV. SELF-UPDATE DRIVEN BY FACE ILLUMINATION

This paper only considers facial models defined as a set of reference face patterns or templates for matching. It is also

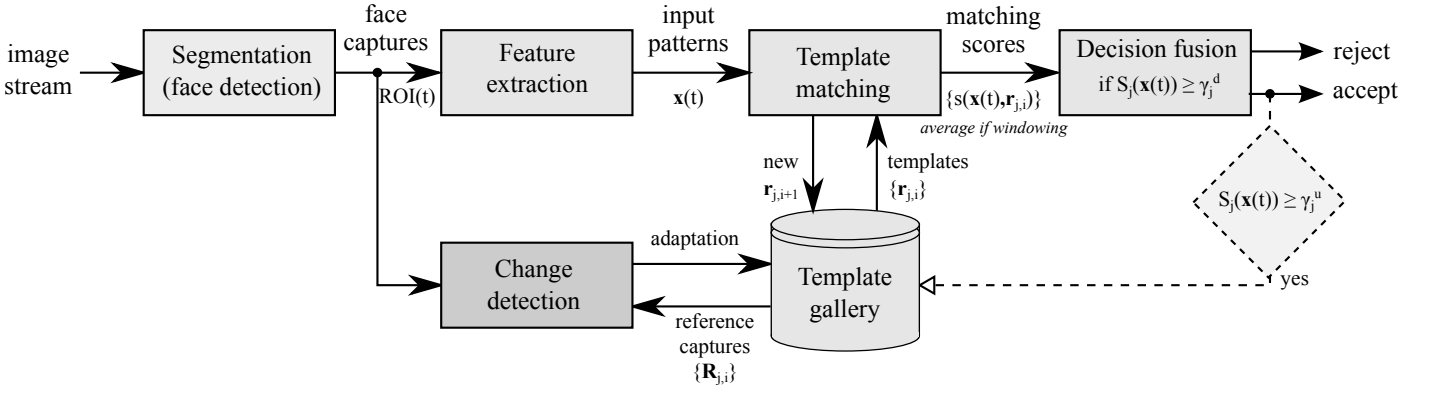


Fig. 1. A generic template matching system that allows for self-update of facial models through concept change detection.

assumed that the template matching system receives a stream of facial captures from a same individual, as found in bio-login applications. The system seeks to continuously authenticate the user logged into a computer of network.

Figure 1 presents a generic template matching system that allows to efficiently manage the self-updating of template sets based on capture conditions. Assume that an image is captured at time t using a camera. The segmentation process isolates the facial region of interest $ROI(t)$, and discriminative features are extracted, normalized and assembled into an ROI pattern $\mathbf{x}(t) \in \mathbb{R}^f$. During the enrollment of an individual j ($j = 1, \dots, J$), ROI images and patterns corresponding to a template gallery $\mathcal{G}_j = \{\{R_{j,1}, \mathbf{r}_{j,1}\}, \{R_{j,2}, \mathbf{r}_{j,2}\}, \dots, \{R_{j,I}, \mathbf{r}_{j,I}\}\}$ are stored in the system.

During operations, the classification module matches $\mathbf{x}(t)$ to the facial model (ROI patterns in set \mathcal{G}_j) of each individual j enrolled to the system. Each template match between $\mathbf{x}(t)$ and a ROI pattern $\mathbf{r}_{j,i}$ produces a similarity score $s(\mathbf{x}(t), \mathbf{r}_{j,i}) \in [0; 1]^3$. The overall score $S_j(\mathbf{x}(t))$ for individual j is combined in the decision fusion module through score-level fusion, which may be computer using several different functions, e.g., the maximum rule:

$$S_j(\mathbf{x}(t)) = \max\{s(\mathbf{x}(t), \mathbf{r}_{j,i}) : i = 1, 2, \dots, I\} \quad (6)$$

and the average rule:

$$S_j(\mathbf{x}(t)) = \frac{1}{I} \sum_{i=1}^I s(\mathbf{x}(t), \mathbf{r}_{j,i}) \quad (7)$$

Then, if $S_j(\mathbf{x}(t)) \geq \gamma^d$, the system produces a positive output prediction for individual j . The decision threshold $\gamma^d \in [0; 1]$ may be global to all users or user-specific. In traditional self-update systems [17], the overall score $S_j(\mathbf{x}(t))$ producing a positive output would also be compared to an additional threshold $\gamma_j^u \geq \gamma_j^d$, to decide whether $\mathbf{x}(t)$ and $ROI(t)$ should be stored in the set \mathcal{G}_j .

In this paper, a new self-updating system is proposed, replacing the second score thresholding γ_j^u with a change detection module (see Figure 1) based on face capture conditions. More precisely, if a change is detected in capture

conditions, the corresponding set of ROI image and pattern $\{ROI(t), \mathbf{x}(t)\}$ are selected as a new template to be added in the gallery \mathcal{G}_j . This strategy is proposed to address the issue of growing complexity and sample relevance in a self-updating system. If a face capture is added to a gallery based only on a high matching score, it may not incorporate relevant information – the system is already tuned to recognize it with a high level of confidence. Including such face capture into the gallery may lead to an unnecessary increase in the system's complexity. Selecting new templates based on their appearance under different capture conditions (pose, lighting, etc.) assures that face model will incorporate a greater diversity of representation.

Figure 2 presents a block diagram of the change detection module (embedded in Figure 1) that is specialized to detect changes in illumination conditions. In order to focus on the effect of illumination variation, the use of the Global Luminance Quality index is proposed to detect changes. At each time t , the $GLQ_{j,i}(t)$ measure (see Eq. 5) is assessed for every input ROI image $ROI(t)$ w.r.t. each reference ROI image $R_{j,i}$ in the gallery \mathcal{G}_j . Then, in order to reduce the impact of rare events, the average $GLQ_j(t)$ (over every reference ROI image in \mathcal{G}_j) is computed for change detection. A difference between two consecutive measures $\epsilon_j(t) = GLQ_j(t) - GLQ_j(t-1)$ is considered for change detection as proposed by Baena-Garcia et al. [4] and Ditzler and Polikar [6]. Finally, an adaptive threshold, as presented in [6] (see Section III-B2), is employed with a standard value of $\alpha = 0.1$. Additional details of the change detection mechanism are presented in Algorithm 1

V. SIMULATION METHODOLOGY

Proof-of-concept experiments were performed using the DIEE multi-biometric dataset [15]. This dataset was explicitly collected to evaluate the performance of self-update and co-training algorithms. It is comprised of about 4400 face and fingerprint samples collected from 49 individuals over a 1.5 year span. For each subject, 60 face still are acquired using a Canon camera over six sessions, and facial ROI are extracted and normalized to a size of 100x100 pixels. During each session, subjects are captured with different variations,

³To improve system robustness, an operating window of W inputs may be used, considering the average score $\frac{1}{W} \sum_{t'=t-W/2}^{t+W/2} s(\mathbf{x}(t'), \mathbf{r}_{j,i})$.

⁴In the same way than the scores, an operating window of W inputs may be used, considering the average GLQ $\frac{1}{W} \sum_{t'=t-W/2}^{t+W/2} GLQ_{j,i}(t)$.

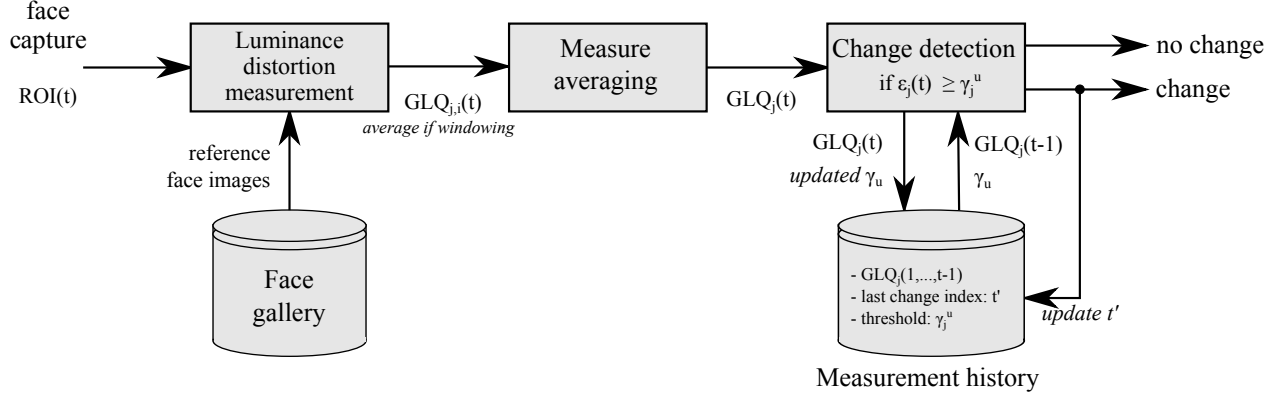


Fig. 2. Module specialized for detection of changing illumination conditions.

Algorithm 1 Self-update algorithm to adapt \mathcal{G}_j for an individual j

Inputs: - $\mathcal{G}_j = \{\{R_{j,1}, \mathbf{r}_{j,1}\}, \dots, \{R_{j,I}, \mathbf{r}_{j,I}\}\}$ // template set \mathcal{G}_j
 - $\epsilon(t^*), t^* = 1, \dots, t-1$ // history for change detection
 - t' // index of the last detected change
 - $\{ROI(t), \mathbf{x}(t)\}$ // input ROI image and pattern
Outputs: - $\mathcal{G}'_j = \{\{R_{j,1}, \mathbf{r}_{j,1}\}, \dots, \{R_{j,I}, \mathbf{r}_{j,I'}\}\}$, $I' \geq I$ // updated set
 - $\epsilon_j(t^*), t^* = 1, \dots, t$ // updated history
 - t' // updated index of the last change
 - $d_j(\mathbf{x}(t))$ // decision over $\mathbf{x}(t)$

```

1: Initialize updated gallery  $\mathcal{G}'_j \leftarrow \mathcal{G}_j$ 
2: for  $i = 1, \dots, I$  do // Score estimation for all templates  $\mathbf{r}_{j,i} \in \mathcal{G}'_j$ 
3:    $s(\mathbf{x}(t), \mathbf{r}_{j,i}) \leftarrow \frac{\sqrt{f} - d_{Euc}(\mathbf{x}(t), \mathbf{r}_{j,i})}{\sqrt{f}}$ 
4: end for
5: Compute the fused score  $S_j(\mathbf{x}(t))$  according to Eq. 6 or Eq. 7
6: if  $S_j(\mathbf{x}(t)) \geq \gamma_j^d$  then
7:    $d_j(\mathbf{x}(t)) \leftarrow \text{Accept}$ 
8:   for  $i = 1, \dots, I$  do // GLQ estimation for all templates  $\mathbf{r}_{j,i} \in \mathcal{G}'_j$ 
9:      $GLQ_{j,i}(t) \leftarrow$  GLQ measure between  $ROI(t)$  and  $\mathbf{r}_{j,i}$  (Eq. 5)
10:  end for
11:   $GLQ_j(t) \leftarrow \frac{1}{I} \sum_{i=1}^I GLQ_{j,i}(t)$  // Average of the GLQ measures
12:   $\epsilon_j(t) = GLQ_j(t) - GLQ_j(t-1)$  // Store the GLQ difference into  $\epsilon$ 
13:  if  $\epsilon_j(t) \geq \gamma_j^u$  then
14:     $\mathcal{G}'_j \leftarrow \mathcal{G}'_j \cup \{ROI(t), \mathbf{x}(t)\}$  // Include image and pattern
15:    // surpassing  $\gamma_j^u$  in new gallery
16:    Update  $t' \leftarrow t$ 
17:  end if
18:  Update threshold  $\gamma_j^u = \epsilon_j + t_{\alpha/2} \cdot \frac{\sigma_{\epsilon_j}}{\sqrt{t-t'-1}}$ 
19: else
20:    $d_j(\mathbf{x}(t)) \leftarrow \text{Reject}$ 
21: end if

```

most notably illumination, pose and expression, ranging from gradual to abrupt changes.

Prior for experiments, the facial images of each person have been ranked and organized by average brightness to exhibit a gradually greater distortion of the illumination conditions with respect to image 1 of the dataset, in order to simulate an environment with gradual changes in the observation conditions. As an example, the stream of faces captured for individual 8 is shown in Figure 3. In addition, Features have been extracted from all facial ROIs using the Multi-Bloc LBP algorithm [2] with 3x3, 5x5 and 9x9 pixel windows, and concatenated with the gray-scale pixel intensity values. ROI patterns are then reduced to 32 features using Principal Component Analysis,

and each feature has been normalized between 0 and 1.

Three different strategies have been considered for validation:

- 1) Update \mathcal{G}_j with only the first ROI pattern of each individual j (no self-update).
- 2) Update \mathcal{G}_j with all ROI patterns that were previously processed for an individual j (always self-update).
- 3) Update \mathcal{G}_j only with ROI images and patterns that triggered a significant change in illumination conditions for individual j (proposed approach).

The experimental protocol is as follows. For each individual j , two thirds of the ROI patterns are presented as the input stream $\{\mathbf{x}(t)\}$, for $t = 1, 2, \dots, 40$, reserving every third image to test the final system. Then, for each individual, the fully trained systems are presented with the remaining images to produce genuine scores, as well as images from the other individuals for impostor scores. Experiments were replicated for 10 individuals selected at random from the DIFE dataset. Only image streams from these individual have been considered in this paper as we seek to study the effect of self-update over scores and capture conditions.

Initially, the user gallery contains the first ROI image and pattern, thus $\mathcal{G}_j = \{\mathbf{R}_{j,1}, \mathbf{r}_{j,1}\}$, and t' (the last time where a change has been detected) is set to 1. For subsequent face



Fig. 3. Stream of faced captured for individual 8 of the DIFE dataset.

captures, the matching score between $\mathbf{x}(t)$ and the reference pattern $\mathbf{r}_{j,i}$ is computed using:

$$s(\mathbf{x}(t), \mathbf{r}_{j,i}) = \frac{1}{\sqrt{f}} \left[\sqrt{f} - d_{Euc}(\mathbf{x}(t), \mathbf{r}_{j,i}) \right] \quad (8)$$

where $f=32$ is the dimension of the feature space, and $d_{Euc}(\mathbf{x}(t), \mathbf{r}_{j,i})$ the Euclidean distance. The average score $S_j(\mathbf{x}(t))$ between ROI pattern $\mathbf{x}(t)$ and all templates in \mathcal{G}_j are computed using the average score-level fusion rule (see Eq. 7).

The GLQ index for $ROI(t)$ is computed with a patch size of 8×8 pixels, for a total of 16 steps. The $GLQ_j(t)$ value is computed between $ROI(t)$ and all reference images $\{\mathbf{R}_{j,i}\}$ in \mathcal{G}_j . If the difference, $\epsilon_j(t) = GLQ_j(t) - GLQ_j(t-1) > \gamma_j^u$, then $\{ROI(t), \mathbf{x}(t)\}$ are added to \mathcal{G}_j . Then the update threshold is updated: $\gamma_j^u = \bar{\epsilon}_j + C \cdot \sigma_j$, where $\bar{\epsilon}_j$ and σ_j are the average and standard deviation values of $\epsilon_j(k)$ for $k = t', \dots, t$, and $C = t_{\alpha/2} / (\sqrt{t - t' - 1})$, the t -statistic constant for α ($\alpha = 0.1$ in experiments) $t_{\alpha/2}$ scaled by the square root of the length of the sequence since the last change.

VI. SIMULATION RESULTS

Figure 4 shows the average genuine scores $S_j(\mathbf{x}(t))$ (left column) and luminance quality measure $GLQ_j(t)$ (right column) of the proposed and reference systems, for the 10 individuals selected from the DIFE dataset. The scores and quality are shown for window sizes W of 1, 2 and 5 captures.

First of all, it can be observed that the use of a sliding window ($W = 3$ or 5) enables fewer fluctuations in scores and the quality measures, without affecting their overall behaviour. For each window size, it can be observed that the genuine scores produced by each system tend to gradually decline over the time, because of the gradual illumination changes the input stream. It is also a normal by-product of the average score-level fusion rule over a growing number of template in galleries. However, it can be noted that, from image 5 until the end of the stream, both systems that perform self-updating produce significantly higher genuine scores than system 1. The scores produced by system 2 and 3 are however statistically similar.

The quality index plotted in Figure 4 shows the overall illumination distortion of input face images over time w.r.t. the reference face images in \mathcal{G}_j . Updating \mathcal{G}_j based on illumination changes provides a significantly smaller decline in illumination quality over the image stream. Result suggest that the proposed approach allow to select a robust set of templates with a diverse representation of an individual under a variety of capture conditions.

For a more detailed analysis, Figures 5(a) and 5(b) show the genuine scores $S_j(\mathbf{x}(t))$ obtained by the proposed and reference systems for individuals 8 and 19 respectively. The scores are shown for systems with a window size of $W = 5$ samples. The black vertical bars in each graph correspond to changes detected by the proposed system in response to varying lighting conditions. While the same observations as with the average values can be made with the individual 8 (the system 1 produces lower scores than the others), the use of self-updating doesn't generate any improvements for the individual 19 before the 20th frame. This behaviour can be explained by the different evolution of the illumination conditions for both

TABLE II. AVERAGE CLASSIFICATION PERFORMANCE AND SYSTEM COMPLEXITY FOR THE 10 INDIVIDUALS.

Window size	Gallery with first sample		Gallery with all previous samples		Gallery with selected samples	
	AUC	Comp.	AUC	Comp.	AUC	Comp.
W = 1	0.70 ± 0.10	1 ± 0	0.91 ± 0.03	40 ± 0	0.84 ± 0.05	4.4 ± 0.6
W = 3	0.74 ± 0.12	1 ± 0	0.95 ± 0.02	40 ± 0	0.95 ± 0.03	5.0 ± 0.7
W = 5	0.75 ± 0.12	1 ± 0	0.96 ± 0.02	40 ± 0	0.93 ± 0.04	4.8 ± 0.5

individuals. Although the illumination changes dramatically and rapidly in the input stream of individual 8 the illumination of captured faces of individual 19 remain comparable for the first 20 captures. Individual 8 thus represents a case in which a template-based matching system significantly benefits from a change-detection driven update strategy, as opposed to the data stream of the individual 19, where the capture conditions are similar enough not to always require a self-update strategy.

Finally, classification performance is presented in Table II, as the average area under the ROC curve (AUC) for the proposed approach versus the two reference methods, and for the 10 individuals. It can be observed that using only one template for \mathcal{G}_j produces lower recognition performance for every window size. For a window of $W = 1$ ROI pattern, adding reference patterns and images based on illumination allows to increase the AUC from 0.70 ± 0.10 to 0.84 ± 0.05 , but using all reference pattern and images increases the AUC even further up to 0.91 ± 0.03 , at the expense of a higher memory complexity. However, for window sizes of 3 and 5 patterns, the difference between the 2 best systems becomes statistically insignificant, as they respectively produce AUCs of 0.95 ± 0.03 and 0.95 ± 0.02 , and 0.93 ± 0.04 and 0.96 ± 0.02 . Comparing the systems complexity, measured as the amount of reference templates stored in the gallery \mathcal{G}_j , reveals that storing those relevant templates not only produces a comparable performance improvement than the adaptive system storing all 40 reference templates, but also with a considerably lower complex computational complexity in terms of CPU cycles and storage, with an average of 4.4 ± 0.6 , 5.0 ± 0.7 and 4.8 ± 0.5 templates for the three window sizes.

VII. CONCLUSION

The reliable authentication of individuals logged in to a computer network is challenging when the appearance of facial captures changes with capture conditions (e.g., variation in illumination, pose and scale). As a preliminary investigation, this paper focuses on the impact of illumination changes on a template matching systems for FR. In traditional self-updating systems, face captures are added to the template gallery based on matching scores, which may lead to growing galleries with many templates that do not incorporate relevant information. In this paper, an adaptive FR system is proposed where self-updating is driven by changes over time in capture conditions. Assuming that the system receives a steam of facial captures

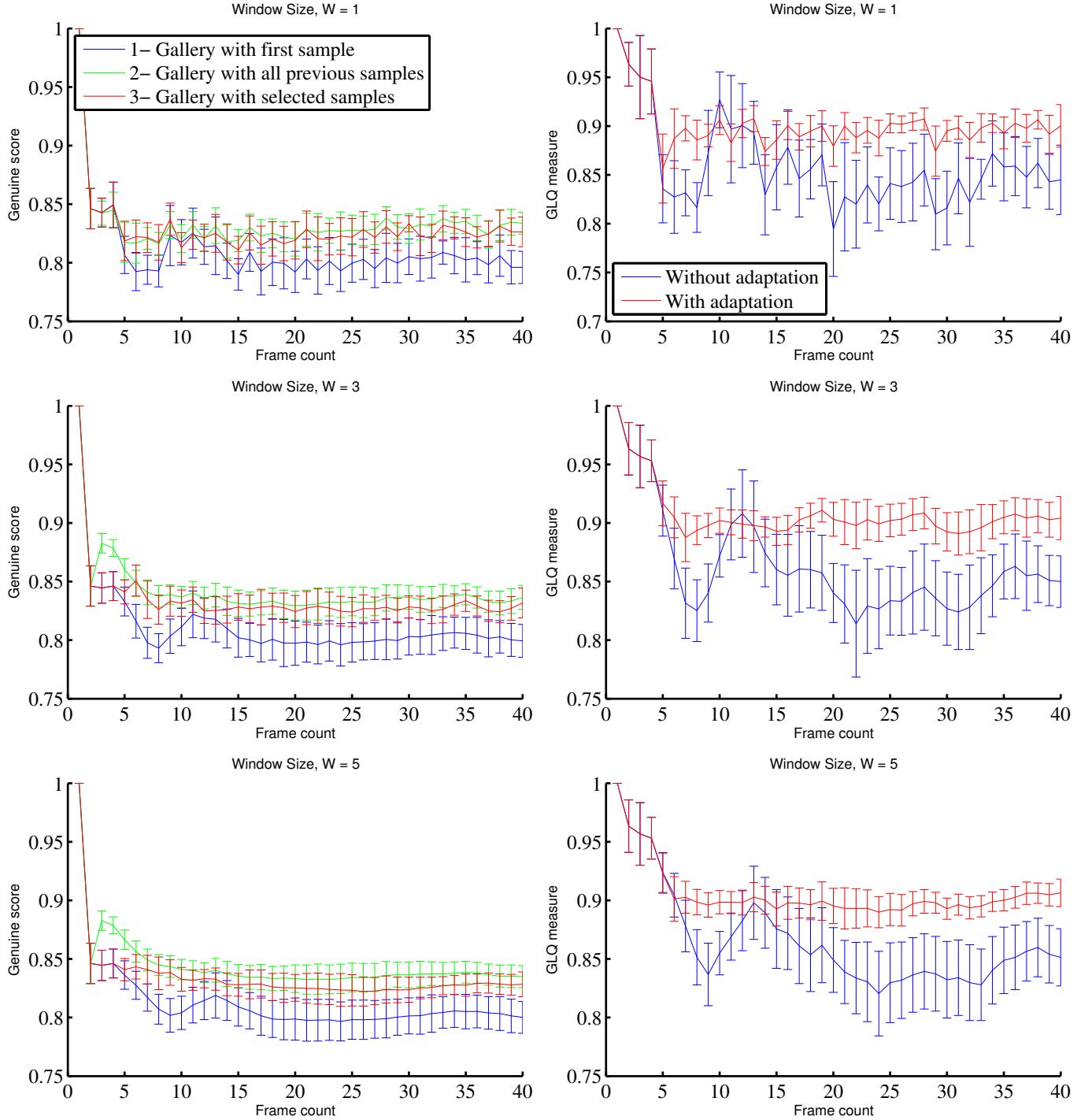
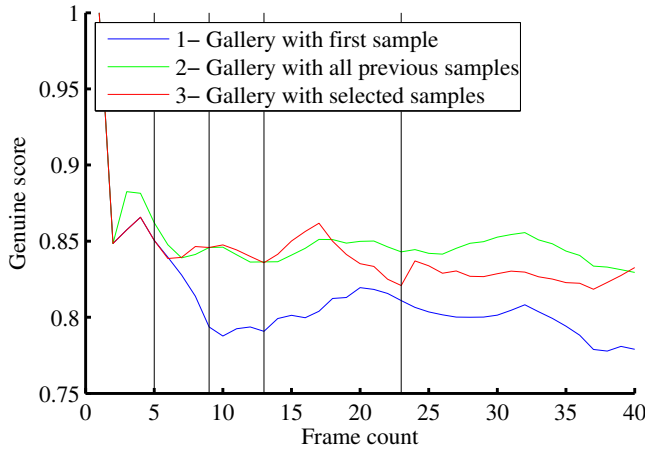


Fig. 4. Average genuine score and illumination quality measure for window sizes of 1, 3 and 5 samples

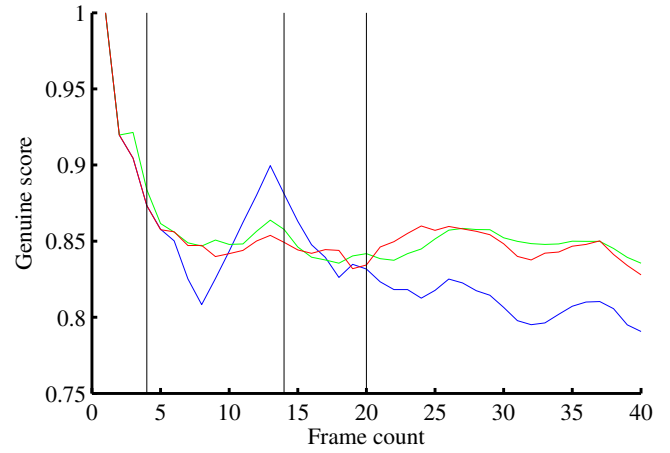
at a regular interval from the built-in computer camera, it monitors the image illumination quality in order to detect possible changes in the illumination conditions. Only when a high confidence face capture exhibits an abrupt pattern of change will it be automatically included in the set of user templates. This enables to sustain a high level of performance while restraining the complexity of the user model.

Experiments have been conducted with face images from

the DLEE multi-modal dataset exhibiting recurring, gradual and abrupt changes. Simulation results indicate that using the proposed strategy to perform self-update based on changes in illumination yields higher matching scores than traditional template matching systems with only one reference sample, yet exhibit the same performance than a more complex system storing every high-confidence sample of the data stream. The added cost of the proposed approach in the evaluation of distortion on input face captures



(a) Individual 8



(b) Individual 19

Fig. 5. Average genuine score for the use template set \mathcal{G}_j as a function of the number of input face captures.

Although encouraging, results should be extended to different, and larger datasets that feature different capture condition. Changes in other capture conditions should be explored in addition to illumination distortion. In addition, robustness of the proposed strategy against impostors should be evaluated in order to assess the discriminative capability of the adaptive system. Finally, the system's complexity could be further improved, using a pruning strategy that removes possible redundant faces images accumulated over time.

ACKNOWLEDGMENT

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada, and the Ministère du Développement Économique, de l'Innovation et de l'Exportation du Québec.

REFERENCES

- [1] A. A. Narasimhamurthy and L. I. Kuncheva. A framework for generating data to simulate changing environments. In *Proc. of the 25th IASTED International Multi-Conf.: artificial intelligence and applications*, pages 84–389, 2007.
- [2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–41, December 2006.
- [3] C. Alippi and M. Roveri. An adaptive cusum-based test for signal change detection. In *ISCAS 2006: 2006 IEEE International Symposium on Circuits and Systems, May 21, 2006 - May 24, 2006*, pages 5752–5755, Politecnico di Milano, DEI, Milano, Italy, 2006. Institute of Electrical and Electronics Engineers Inc.
- [4] M. Baena-Garcia, J. Del Campo-Ávila, R. Fidalgo, A. Bifet, R. Gavalda, and R. Morales-Bueno. Early drift detection method. In *Fourth International Workshop on Knowledge Discovery from Data Streams*, 2006.
- [5] M. De-la Torre, E. Granger, R. Sabourin, and D. O. Gorodnichy. Partially-supervised learning from facial trajectories for face recognition in video surveillance. *Information Fusion*, (0):–, 2014.
- [6] G. Ditzler and R. Polikar. Hellinger distance based drift detection for nonstationary environments. In *Symposium Series on Computational Intelligence, IEEE SSCI 2011 - 2011 IEEE Symposium on Computational Intelligence in Dynamic and Uncertain Environments, CIDUE 2011, April 11, 2011 - April 15, 2011*, pages 41–48, Dept. of Electrical and Computer Engineering, Rowan University, Glassboro, NJ, United States, 2011. IEEE Computer Society.
- [7] B. Freni, G.-L. Marcialis, and F. Roli. Template selection by editing algorithms: A case study in face recognition. In *Proc. Joint Int. Association of Pattern Recognition International*, volume 5342 of *LNCS*, pages 745–754, Orlando, USA, 2008.
- [8] R. Klinkenberg and I. Renz. Adaptive information filtering: Learning in the presence of concept drifts. In *Workshop Notes of the ICML/AAAI-98 Workshop Learning for Text Categorization*, pages 33–40. AAAI Press, 1998.
- [9] L. I. Kuncheva. Using control charts for detecting concept change in streaming data. Technical report, School of Computer Science, Bangor University, 2009.
- [10] L. Minku, A. White, and X. Yao. The impact of diversity on on-line ensemble learning in the presence of concept drift. *IEEE Trans. on Knowledge and Data Engineering*, 22(5):730 – 742, 2010.
- [11] N. Niinuma, U. Park, and A. Jain. Soft biometric traits for continuous user authentication. *IEEE Trans. on Information Forensics and Security*, 5(4):771–780, 2010.
- [12] C. Pagano, E. Granger, R. Sabourin, and D. O. Gorodnichy. Detector ensembles for face recognition in video surveillance. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–8. IEEE, 2012.
- [13] C. Pagano, E. Granger, R. Sabourin, F. Roli, and L. Marcialis, G. Adaptive ensembles for face recognition in changing video surveillance environments. *Information Sciences*, 2014.
- [14] A. Rattani, B. Freni, G. Luca Marcialis, and F. Roli. Template update methods in adaptive biometric systems: A critical review. In *Lecture Notes in Computer Science (included Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5558, pages 847 – 856, Alghero, Italy, 2009.
- [15] A. Rattani, G.-L. Marcialis, and F. Roli. A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study. *International Journal of Biometrics*, 5(4):266 – 287, 2013.
- [16] M. R. Reynolds and Z. G. Stoumbos. General approach to modeling cusum charts for a proportion. *IIE Transactions (Institute of Industrial Engineers)*, 32(6):515–535, 2000.
- [17] F. Roli, L. Didaci, and G.L. Marcialis. Template co-update in multi-modal biometric systems. In *International Conference on Biometrics*, volume 4642, pages 1194 – 1202, Seoul, Korea, August 2007.
- [18] J. Sang, Z. Lei, and S. Z. Li. Face image quality evaluation for iso/iec standards 19794-5 and 29794-5. In *Proc. Proc. Int'l Conference on Biometrics*, volume 5558 of *LNCS*, page 229–238, 2009.
- [19] H. Sellahewa and S. A. Jassim. Image-quality-based adaptive face recognition. *IEEE Transactions on Instrumentation and Measurement*, 59(4):805–13, 04 2010.
- [20] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–4, 03 2002.