

# SYNTHETIC FACE GENERATION UNDER VARIOUS OPERATIONAL CONDITIONS IN VIDEO SURVEILLANCE

Fania Mokhayeri<sup>†</sup>, Eric Granger<sup>†</sup>, Guillaume-Alexandre Bilodeau<sup>\*</sup>

<sup>†</sup>LIVIA, École de technologie supérieure, Université du Québec, Montréal, Canada

<sup>\*</sup>LITIV, Polytechnique Montréal, Montréal, Canada

fmothayeri@livia.etsmtl.ca, eric.granger@etsmtl.ca, gabilodeau@polymtl.ca

## ABSTRACT

In still-to-video face recognition (FR), the faces captured with surveillance cameras are matched against reference stills of target individuals enrolled to the system. FR is a challenging problem in video surveillance due to uncontrolled capture conditions (variations in pose, expression, illumination, blur, scale, etc.), and the limited number of reference stills to model target individuals. This paper introduces a new approach to generate multiple synthetic face images per reference still based on camera-specific capture conditions to deal with illumination variations. For each reference still, a diverse set of faces from non-target individuals appearing in the camera viewpoint are selected based on luminance and contrast distortion. These face images are then decomposed into detail layer and large scale layer using an edge-preserving image decomposition to obtain their illumination dependent component. Finally, the large scale layers of these images are morphed with each reference still image to generate multiple synthetic reference stills that incorporate illumination and contrast conditions. Experimental results obtained with the ChokePoint dataset reveal that these synthetic faces produce an enhanced face model. As the number of synthetic faces grows, the proposed approach provides a higher level of accuracy and robustness across a range of capture conditions.

*Index Terms*— Video Surveillance, Face Recognition, Image Morphing, Synthetic face generation

## 1. INTRODUCTION

Face recognition (FR) in video surveillance has received significant interest due to covert capture of faces using surveillance cameras, flexible control, high performance to cost ratio as well as the possibility of analysis of live feeds [1]. Watchlist screening is a common application of still-to-video FR systems, where facial models<sup>1</sup> used for matching are designed using regions of interest (ROIs) isolated from reference still images of target individual enrolled to the system. Then, FR system seeks to determine if faces captured using video surveillance cameras correspond to facial models of target individuals [2].

In recent years, public security organizations have deployed several video surveillance cameras. Despite recent advances that occurred in FR systems, [3] [4] [5] [6] [7], designing a robust system for still-to-video FR in video surveillance under semi-controlled and uncontrolled capture conditions remains a challenging problem. This is due in part to the limited number of representative reference stills per target individuals. In addition, ROIs isolated from reference still images may differ significantly

from those captured in videos, due to camera inter-operability. Finally, faces captured in operational videos vary due to pose angle, expression, illumination, blur, resolution, scale and occlusion. This paper focuses on the variations in facial appearance related to changes in illumination and contrast. These are among the most significant factors affecting FR performance. Variations in illumination and contrast are typically caused by operational conditions that are specific to a camera viewpoint, such as lighting condition, shadow, scale and proximity of faces to the camera [8].

Over the past few decades, a wide variety of approaches have emerged to compensate lighting variations that affect FR performance. These methods can be grouped into techniques for (1) illumination normalization, (2) illumination invariant feature extraction, (3) modeling the illumination conditions, and (4) face image synthesizing. Normalization approaches try to adjust the images to a normal lighting condition. Histogram equalization, gamma correction, histogram specification/matching, and oriented local histogram equalization are basic methods in this area [9] [10]. These techniques are effective, but their ability to handle uneven illumination variations remains limited [11] [12]. Illumination-invariant representation strategies extract facial descriptors that are robust to variant lighting such as edge maps, geometrical features, derivatives of the gray-level e.g., the local binary patterns (LBP) and Gabor features [13] [14]. Due to the change in the illumination direction, these methods are insufficient to overcome lighting variations in practical applications. Another strategy involves modeling the illumination conditions that allow restoring illumination invariant features, and then deploying the features for FR. Authors in [15] compensated illumination variations by exploiting the fact that a bi-convex set is formed by images under non-uniform blurring and varying illumination. However, it is difficult with this technique to estimate the lighting and reflectivity source from a single face image. Finally, face image synthesizing approaches attempt to generate virtual images under different appearance. Adding extra samples to the gallery produces diverse face representations and accordingly improves (intra-class) robustness of face models to various capture conditions [16]. The benefit of the synthesizing approach compared to other illumination compensation methods is that the sample generated is still an image, instead of abstract features. Moreover, this is useful for the single sample per person (SSPP) scenario in which there is only one sample for each enrolled individual into the gallery. Authors in [17] proposed to use illumination transferring technique to compensate the lack of the illumination information in the training set. A common approach for synthetic face generation is to reconstruct the 3D model of the face from its 2D face image. The major drawback of 3D-based techniques [18], is their high computational cost. In contrast to face synthesizing approaches based on 3D reconstruction, the proposed approach in this paper allows generating new samples without any knowledge of 3D shape.

This work was supported by FRQ-NT REPARTI strategic cluster.

<sup>1</sup>A facial model is defined as either a set of reference face samples (used in template matching), or a statistical model estimated through training with reference captures (used in neural statistical classification) corresponding to a target individual.

This research aims to address SSPP problem of still-to-video FR by generating multiple synthetic face images per reference still in order to improve the representativeness of face models. To this end, a novel approach is introduced to generate virtual face images under different lighting and contrast based on the abundance of unconstrained face captures in videos of non-target individuals. With the proposed approach, a set of non-target face ROIs captured in video sequences with diverse illumination and contrast are selected and their large scale layers are extracted to obtain their illumination dependent component. Each reference image of the gallery is morphed with the decomposed large scale layers to transfer the operational environment conditions (illumination, contrast, noise) of non-target samples to virtual samples that are then added to gallery of an individual.

The contributions of this paper include: (1) a method for selecting a set of facial ROIs from videos with non-target individuals that incorporates a diversity in illumination and contrast, as found in the operational environment; (2) a 2D image morphing technique to generate a set of synthetic face images from a reference still ROI and diversified video ROI; (3) experimental results that show adding synthetic reference faces to a watchlist can improve still-to-video FR performance.

## 2. A SYNTHETIC FACE GENERATION METHOD

This study introduces a new face synthesizing approach that generates multiple synthetic face images per reference still, each one under different illumination and contrast conditions that are specific to an operational condition. The overall process of the proposed approach is depicted in Figure 1.

According to this approach, several ROIs are isolated from video sequences where non-target individuals appear in a specific camera viewpoint. Video ROIs that incorporate the most diverse variations in illumination and contrast w.r.t. a reference still ROI, with almost frontal face poses, are selected as follows. Video ROIs are extracted  $\mathbf{ROI}_N^i(x, y)$ , and the head pose of each ROI in each frame is estimated. The ROI with frontal viewpoint are selected. Then, the illumination distortion between each ROI still image in the watchlist  $\mathbf{ROI}_W^i(x, y)$  and each video  $\mathbf{ROI}_N^i(x, y)$  are measured using global luminance quality (GLQ) factor [19].  $GLQ(x, y)$  measures the proximity of the average luminance between a still and probe video ROIs locally by utilizing sliding window:

$$GLQ(x, y) = \frac{1}{n} \sum_{m=1}^n \frac{2 \cdot \mu_m(x) \cdot \mu_m(y) + C_1}{\mu_m(x)^2 + \mu_m(y)^2 + C_1} \quad (1)$$

where  $\mu(x)$  and  $\mu(y)$  denotes mean values of image block of reference and input images respectively, and  $m$  is sliding step corresponding to block size.  $C_1$  is constant value defined as  $C_1 = (K_1 L)^2$  that  $L$  is the dynamic range of the pixel values and  $K_1$  is experimentally determined [20].

Next, the contrast distortion between each  $\mathbf{ROI}_W^i$  and each  $\mathbf{ROI}_N^i(x, y)$  are measured using global contrast quality (GCQ) factor [19] defined as:

$$GCQ(x, y) = \frac{1}{n} \sum_{p=1}^n \frac{2 \cdot \sigma_p(x) \cdot \sigma_p(y) + C_2}{\sigma_p(x)^2 + \sigma_p(y)^2 + C_2} \quad (2)$$

where  $\sigma(x)$  and  $\sigma(y)$  denotes standard deviation of image block of reference and input images respectively, and  $p$  is sliding step corresponding to block size.  $C_2$  is constant value defined as  $C_2 = (K_2 L)^2$  where  $L$  is the dynamic range of the pixel values and  $K_1 \ll 1$ .

Afterwards,  $GLQ(\mathbf{ROI}_W, \mathbf{ROI}_N)$  and  $GCQ(\mathbf{ROI}_W, \mathbf{ROI}_N)$  are normalized. Then, clustering of the non-target ROIs is per-

formed on the normalized GLQ and GCQ in 2D space using K-means to allow selection of diverse illuminations and contrasts. In order to determine an optimal number of clusters, Dunn index is exploited. The aim of this index is to discover clusters that are compact, and well separated, where the means of different clusters are sufficiently far apart, as compared to the internal cluster variance [21]. Then, the representative ROIs of each cluster (the ROI closest to the cluster's centroid) are determined. This process is repeated  $T$  times for non-target videos employed for design process.

Next, their illumination dependent component is extracted. To do so, ROIs are decomposed into two layers including large-scale layers  $\mathbf{L}_i^i(x, y)$  (similar to shading) and detail layers  $\mathbf{L}_d^i$  (containing reflectance) using weighted least square (WLS) filter [22]. The  $\mathbf{L}_d^i(x, y)$  can be considered as illumination independent, and  $\mathbf{L}_i^i(x, y)$  as illumination dependent. Edge-preserving filter is further applied to smooth the  $\mathbf{L}_i^i(x, y)$ . Edge-preserving filters preserve the edges of the image [23].

After extracting large scale layers, each  $\mathbf{ROI}_W^i$  is then morphed with the  $\mathbf{L}_i^i(x, y)$  to generate multiple virtual samples under new illumination conditions. The morphing process is performed according to the following steps:

1. detect the landmark points using active shape model to locate corresponding feature points in ROIs [24];
2. define a triangular mesh over feature points via the Delaunay triangulation technique;
3. coordinate transformations between triangles, with affine projections on points;
4. warp each triangle separately from the source to destination using mesh warping technique. This method moves triangular patches to the newly established location to align two ROIs before cross-dissolving;
5. cross-dissolve the triangulated ROIs considering warped pixel locations:

$$f_{morph}(x, y) = \alpha f_1(x, y) + (1 - \alpha) f_2(x, y) \quad (3)$$

where  $\mathbf{f}_1(x, y)$  and  $\mathbf{f}_2(x, y)$  are the pixels of reference and destination images of the still and video ROIs respectively and  $\alpha$  is morphing weight ( $0 < \alpha \leq 1$ ). This equation demonstrates moving the feature points at new location  $\mathbf{f}_{morph}(x, y)$  takes  $\alpha$  percentage from  $\mathbf{f}_1(x, y)$  and  $1 - \alpha$  from  $\mathbf{f}_2(x, y)$  [25]. The output is a set of synthetic ROI  $I_S$  with smooth variations of illumination and contrast. The face generation process is formalized in Algorithm 1.

## 3. EXPERIMENTAL ANALYSIS

### 3.1. Methodology and Dataset

In this work, a particular implementation of the proposed system has been proposed to assess the performance of the synthesizing method and assess our hypothesis that generating synthetic reference faces can improve recognition performance. The presented system for FR is composed of a group of matchers dedicated to each individual enrolled to the system. (original image and synthetic ones). During the design phase, few synthetic samples are generated from each single high-quality reference still. The generated samples are then added to the gallery to enlarge the training set and consequently improve the representative ability of face models. The facial models are designed using ROIs extracted from reference still and corresponding virtual ones. During operations, ROIs captured from faces of probe individuals are matched against the enrolled facial models during enrollment to detect the presence of watchlist individuals.

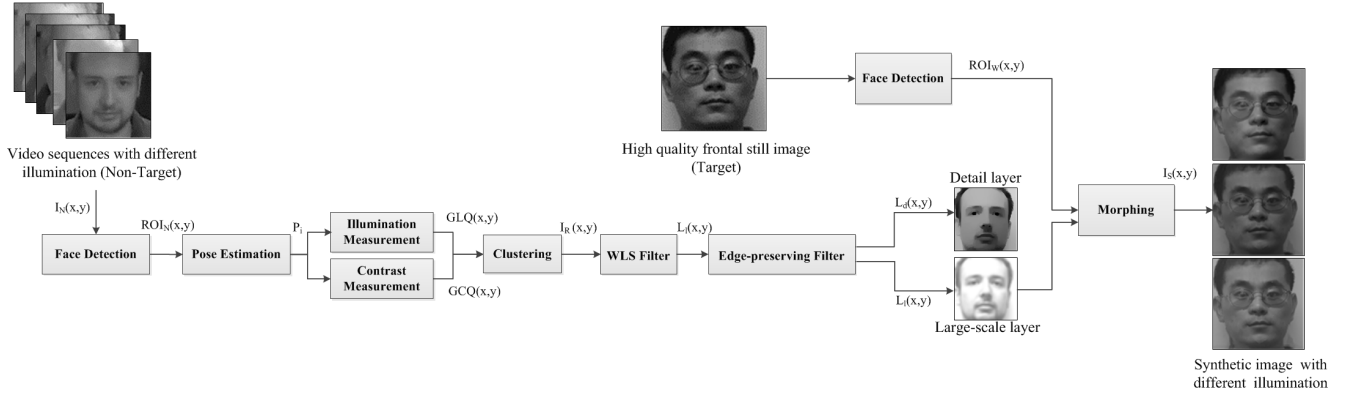


Fig. 1. Block diagram of the proposed system applied to illumination synthesis.

### Algorithm 1: Algorithm for synthetic face generation

**Input:** Gallery frames  $\{\mathbf{I}_W^i(x, y) : i = 1, \dots, n\}$ ,  
Non-target videos with different illumination conditions  
 $\{\mathbf{I}_N^i(x, y) : i = 1, \dots, N\}$   
**Output:** Set of synthetic face images under different illumination  
 $\{\mathbf{I}_S^i(x, y) : i = 1, \dots, m\}$

- 1 **for each**  $\{\mathbf{I}_N^i(x, y)\}$  **frames do**
- 2     Apply *segmentation* to  $\{\mathbf{I}_N^i(x, y)\}$  to detect ROIs  
 $\{\mathbf{ROI}_N^i(x, y) : i = 1, \dots, m\}$  corresponding to faces in the frames  
*count* = 0
- 3     **while** *count*  $\leq T$  **do**
- 4         Estimate head pose of the ROIs  $\{\mathbf{P}_i : i = 1, \dots, m\}$   
**while**  $\{\mathbf{P}_i\} \leq T^\circ$  **do**
- 5             Measure illumination of the selected non-targets  
 $\{GLQ_i(x, y) : i = 1, \dots, m\}$
- 6             Measure contrast of the selected non-targets  
 $\{GCQ_i(x, y) : i = 1, \dots, m\}$
- 7             Normalize  $GLQ_i(x, y)$  and  $GCQ_i(x, y)$
- 8             Find optimal number of clusters  $n_k$  based on Dunn index
- 9             Perform clustering ( $k=n_k$ ) on the normalized  
 $\{GLQ_i(x, y)\}$  and  $\{GCQ_i(x, y)\}$
- 10             Determined the representative images of each cluster  
 $\{\mathbf{I}_R^i(x, y)\}$
- 11             Extract large-scale layer  $\{\mathbf{L}_l^i(x, y)\}$  of the  $\{\mathbf{I}_R^i(x, y)\}$
- 12             Perform edgepreserving filter on  $\{\mathbf{L}_l^i(x, y)\}$
- 13         **for each**  $\{\mathbf{I}_W^i(x, y)\}$  **frames do**
- 14             Apply *segmentation* to  $\{\mathbf{I}_W^i(x, y)\}$  to detect ROIs  $\{\mathbf{ROI}_W^i(x, y)\}$
- 15             **for each**  $\{\mathbf{L}_l^i(x, y)\}$  **do**
- 16                 Detect landmark points of  $\{\mathbf{ROI}_W^i(x, y)\}$  and  $\mathbf{L}_l^i(x, y)$
- 17                  $\mathbf{V}_1 = \text{ASMLandmarks}(\{\mathbf{ROI}_W^i(x, y)\})$
- 18                  $\mathbf{V}_2 = \text{ASMLandmarks}(\mathbf{L}_l^i(x, y))$
- 19                 Define a triangular mesh over the points  $\mathbf{V}_1$  and  $\mathbf{V}_2$
- 20                  $\mathbf{D}_1 = \text{Delaunay}(\mathbf{V}_1)$ ,  $\mathbf{D}_2 = \text{Delaunay}(\mathbf{V}_2)$
- 21                 Coordinate transformations between triangles
- 22                  $\mathbf{D}_{new1} = \text{Affine}(\mathbf{V}_1)$ ,  $\mathbf{D}_{new2} = \text{Affine}(\mathbf{V}_2)$
- 23                 Warp each triangle from source to destination
- 24                  $\mathbf{W}_1 = \text{Meshwarp}(\mathbf{V}_1, \mathbf{D}_{new1})$
- 25                  $\mathbf{W}_2 = \text{Meshwarp}(\mathbf{V}_2, \mathbf{D}_{new2})$
- 26                 Cross-dissolve  $\mathbf{W}_1$  and  $\mathbf{W}_2$
- 27                  $\{\mathbf{I}_S^i(x, y)\} = \text{CrossDissolve}(\mathbf{W}_1, \mathbf{W}_2, \alpha)$

The overall block diagram of the proposed FR system is shown in Figure 2. During the enrollment, a set of non-target videos are collected and their ROIs are extracted and the head pose of each ROI in each frame is estimated. The ROI with the face pose angles less than  $3^\circ$  are selected. Then,  $GLQ_i(x, y)$  and  $GCQ_i(x, y)$  between each ROI isolated from still, and the video ROIs of various non-target individuals are measured. Next, clustering is performed on the normalized  $GLQ_i(x, y)$  and  $GCQ_i(x, y)$  in 2D space using K-means and the representative image of each cluster is determined.

The optimal number of clusters obtained using Dunn index is typically around  $k = 4$ . This process is repeated 3 times for 10 different sets of non-target videos. So, for each watchlist individual, 12 non-target images are selected. Each watchlist individual is then morphed with the decomposed large scale layers. A total of 12 synthetic face images with diverse illumination and contrast for each still individual are generated and added to the watchlist gallery to create a new gallery. During design phase, for each face of gallery, segmentation is performed and the ROI is then scaled to a common size  $48 \times 48$  to limit processing time. Next, a division into  $3 \times 3 = 9$  uniform non-overlapping patches is performed on each ROI representations. Next, uniform pattern of 59 local binary pattern features of face patches using in the single reference ROI and corresponding synthetic ones are extracted to generate diverse face representations. The extracted features are normalized to range between 0 and 1, and assembled into a ROI pattern of features for matching. The latter are then stored as a template into a gallery. The enrollment phase produces a template gallery with 13 different templates per watchlist person (the original image plus 12 synthetic images). During the operational phase, frames undergo the same processing steps as for the enrollment and following that, template matching is applied that matches the facial models of probes against those models stored in the gallery during enrollment. Each matcher provides a similarity score between every patch of the input vector and the corresponding patch template in the gallery via Euclidian distance. Output scores from matchers are fed into the fusion module after score normalization. A face tracker also regroups faces from each different person and accumulates positive predictions over time for robust spatio-temporal recognition. A positive prediction is produced if a matching score surpasses an individual-specific threshold. Finally, the decision function combines the tracks and matching predictions in order to recognize the most likely individuals in the scene.

To assess the transaction-level performance of the proposed FR system, partial ROC curve (pAUC), area under precision-recall space (AUPR) and F1-measure for a desired false positive rate of 1%, are considered. Prior to each replication, 5 persons are randomly selected as target watchlist individuals. The remaining individuals are used in the operational phase as non-target subjects. This process is repeated 5 times.

In order to validate the performance achieved by developed FR system for watchlist screening applications, ChokePoint video dataset has been employed. The dataset includes an array of three cameras placed above two portals to capture subjects walking through them. It contains 54 video sequences in portal 1 and 29 subjects in portal 2. Captured face frames have variations in terms of illumination conditions, pose, sharpness, as well as misalignment due to automatic face localization/detection [26].

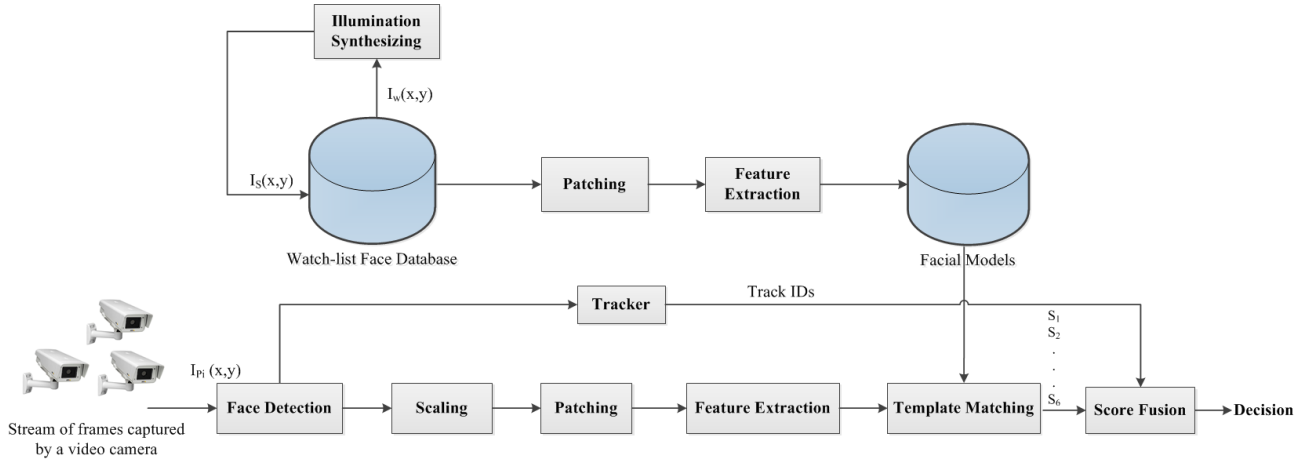


Fig. 2. Synthetic-based still-to-video face recognition system

Table 1. Average transaction-level results of the proposed system with 12 synthetic images on 5 replications with ChokePoint dataset

Recognition system	pAUC(20%)	AUPR	F1-measure
BFR	0.298 ± 0.105	0.160 ± 0.054	0.187 ± 0.068
synthetic FR	0.401 ± 0.096	0.218 ± 0.029	0.278 ± 0.046
patch-based BFR	0.461 ± 0.462	0.293 ± 0.283	0.331 ± 0.330
patch-based synthetic FR	0.547 ± 0.169	0.367 ± 0.145	0.363 ± 0.139

### 3.2. Results and Discussion

Figure 3 represents examples of the images generated with assistance of the proposed synthesizing algorithm.

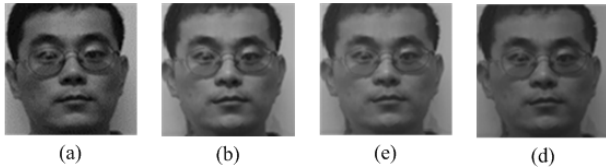


Fig. 3. Synthetic face images with different illumination; (a) original image, (b-d) synthetic relighted images (ID#1 ChokePoint dataset)

Results in Table 1 present the average transaction-level performance (pAUC(20%), AUPR and F1-measure) in the baseline FR system (BFR) with only one sample per person and the FR system with extra images under various illumination conditions. Moreover, it compares the results obtained with and without patching. As shown in Table 1, the recognition system with extra samples under varying illumination outperform the baseline system because of its robustness to the illumination variations. Furthermore, the results achieved via the patch-based technique provide a higher level of performance compared to the baseline system. Since, the extracting features from each patch allow exploiting more discriminant information and consequently yield better matching performance.

Table 2 compares the average transaction-level performance in the synthetic FR system based on the number of synthetic samples. It can be concluded that the number of images added to the gallery has a direct impact on the recognition rate and time complexity. Increasing synthetic images enhances the system performance; how-

ever, reduces time efficiency. Therefore, there should be a trade-off between performance and computational cost associated with increased number of samples.

Table 2. Average transaction-level results of proposed FR system based on number of synthetic images with ChokePoint dataset

Number of images	pAUC(20%)	AUPR	F1-measure
1 (baseline)	0.298 ± 0.105	0.160 ± 0.054	0.187 ± 0.068
4	0.315 ± 0.042	0.187 ± 0.032	0.228 ± 0.083
8	0.378 ± 0.071	0.201 ± 0.103	0.255 ± 0.037
12	0.401 ± 0.096	0.218 ± 0.029	0.278 ± 0.046

It can however be observed that the results vary according to the watchlist individuals. This is demonstrated by the results, for instance, with individual ID#01, adding extra images improve the performance pAUC=0.118 to pAUC=0.378, but in individual ID#04 from pAUC=0.319 to pAUC=0.339.

### 4. CONCLUSION

Given the limited number of reference face stills that are available to design a system for still-to-video FR, a new approach is proposed in this paper to generate multiple synthetic face images per reference still based on camera-specific capture conditions. This approach exploits the abundance of diverse facial ROIs from non-target individuals that appears in a specific camera viewpoint. An algorithm for image morphing allows to generate a set of diverse images with a smooth transition of illumination. It is able to accurately convey a range of synthetic face images that integrate a diversity of illumination and contrast. Experimental results with videos from the ChokePoint dataset show that the proposed approach can improve the representativeness of facial models to variations in illumination and contrast that are specific to faces captured by a camera viewpoint. As the number of synthetic faces are grows, the performance of a baseline still-to-video FR systems using template matching significantly improves. It is worth mentioning that this method can be generalized to transfer other appearance variations such as shadow and blur to various object in a wide range of applications. In order to further improve robustness of still-to-video FR systems, future research should include methods like 3D reconstruction to generate synthetic faces based on variations in pose and expression of a target individual.

## 5. REFERENCES

- [1] P. J. Phillips, P. Grother, and R. Micheals, *Evaluation methods in face recognition*, Springer, 2011.
- [2] C. Pagano, E. Granger, R. Sabourin, G. L. Marcialis, and F. Roli, "Adaptive ensembles for face recognition in changing video surveillance environments," *Information Sciences*, vol. 286, pp. 75–101, 2014.
- [3] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [4] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 3, pp. 328–340, 2005.
- [5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.
- [6] 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.
- [7] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*. IEEE, 2014, pp. 1701–1708.
- [8] H. Song, X. Lin, and Z. Liu, "A face recognition under varying illumination," in *Advances in Image and Graphics Technologies*, pp. 120–128, 2014.
- [9] H. D. Liu, M. Yang, Y. Gao, and C. Cui, "Local histogram specification for face recognition under varying lighting conditions," *Image and Vision Computing*, vol. 32, no. 5, pp. 335–347, 2014.
- [10] P. H. Lee, S. W. Wu, and Y. P. Hung, "Illumination compensation using oriented local histogram equalization and its application to face recognition," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4280–4289, 2012.
- [11] H. Shahamat and A. A. Pouyan, "Face recognition under large illumination variations using homomorphic filtering in spatial domain," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 970–977, 2014.
- [12] I. Amara, E. Granger, and A. Hadid, "On the effects of illumination normalization with lbp-based watchlist screening," in *Lecture Notes in Computer Science*. Springer, 2015, pp. 173–183.
- [13] X. Xie, J. Lai, and W. Zheng, "Extraction of illumination invariant facial features from a single image using nonsubsampling contourlet transform," *Pattern Recognition*, vol. 43, no. 12, pp. 4177–4189, 2010.
- [14] S. Bashbaghi, E. Granger, R. Sabourin, and G. A. Bilodeau, "Watch-list screening using ensembles based on multiple face representations," in *Pattern Recognition (ICPR), 22nd International Conference on*. IEEE, 2014, pp. 4489–4494.
- [15] A. Punnappurath, A. N. Rajagopalan, S. Taheri, R. Chellappa, and G. Seetharaman, "Face recognition across non-uniform motion blur, illumination, and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 2067–2082, 2015.
- [16] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 643–660, 2001.
- [17] L. Zhuang, A. Y. Yang, Z. Zhou, S. S. Sastry, and Y. Ma, "Single-sample face recognition with image corruption and misalignment via sparse illumination transfer," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2013, pp. 3546–3553.
- [18] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [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] C. Pagano, E. Granger, R. Sabourin, A. Rattani, G. L. Marcialis, and F. Roli, "Efficient adaptive face recognition systems based on capture conditions," in *IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)*, Orlando, FL, United State, 2014.
- [21] S. Günter and H. Bunke, "Validation indices for graph clustering," *Pattern Recognition Letters*, vol. 24, no. 8, pp. 1107–1113, 2003.
- [22] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," in *ACM Transactions on Graphics (TOG)*, 2008, vol. 27, p. 67.
- [23] B. Gu, W. Li, M. Zhu, and M. Wang, "Local edge-preserving multiscale decomposition for high dynamic range image tone mapping," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 70–79, 2013.
- [24] W. Hu, X. Zhou, W. Li, W. Luo, X. Zhang, and S. Maybank, "Active contour-based visual tracking by integrating colors, shapes, and motions," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1778–1792, 2013.
- [25] 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.
- [26] 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), IEEE Computer Society Conference on*. IEEE, 2011, pp. 74–81.