Face Re-Identification for Digital Signage Applications

G. M. Farinella[1], G. Farioli[1], S. Battiato[1], S. Leonardi[2], and G. Gallo[1]

gfarinella@dmi.unict.it, gfarioli88@gmail.com, battiato@dmi.unict.it,
salvo.leonardi@adviceweb.it, gallo@dmi.unict.it
[1]Image Processing Laboratory,
Department of Mathematics and Computer Science
University of Catania, Italy
http://iplab.dmi.unict.it
[2]Advice S. C., Catania, Italy

Abstract. The estimation of soft biometric features related to a person standing in front an advertising screen plays a key role in digital signage applications. Information such as gender, age, and emotions of the user can help to trigger dedicated advertising campaigns to the target user as well as it can be useful to measure the type of audience attending a store. Among the technologies useful to monitor the customers in this context, there are the ones that aim to answer the following question: is a specific subject back to the advertising screen within a time slot? This information can have an high impact on the automatic selection of the advertising campaigns to be shown when a new user or a re-identified one appears in front the smart screen. This paper points out, through a set of experiments, that the re-identification of users appearing in front a screen is possible with a good accuracy. Specifically, we describe a framework employing frontal face detection technology and re-identification mechanism, based on similarity between sets of faces learned within a time slot (i.e., the models to be re-identified) and the set of face patches collected when a user appears in front a screen. Faces are pre-processed to remove geometric and photometric variability and are represented as spatial histograms of Locally Ternary Pattern for re-identification purpose. A dataset composed by different presentation sessions of customers to the proposed system has been acquired for testing purpose. Data have been collected to guarantee realistic variabilities. The experiments have been conducted with a leave-one-out validation method to estimate the performances of the system in three different working scenarios: one sample per presentation session for both testing and training (one-to-one), one sample per presentation session for testing and many for training (one-to-many), as well as considering many samples per presentation sessions for both testing and training (many-to-many). Experimental results on the considered dataset show that an accuracy of 88.73% with 5% of false positive can be achieved by using a many-to-many re-identification approach which considers few faces samples in both training and testing.

1 Introduction and Motivation

Digital signage is considered a revolutionary research area which aims to build advanced technologies for the out-of-home advertising. More specifically, with the term "digital signage" are referred the smart screens employed to show advertising content to a broad audience in a public/private area (e.g., store, airport, info office, taxi, etc.). The advertising screens are usually connected to the internet and are able to perform a series of "measurements" on the audience in front of the screen which are then exploited for marketing purposes (e.g., the screen reacts differently depending on the measurements). Taking into account the survey of the Aberdeen Research [1], the global market around the digital signage will expand from $1.3 billion in 2010 to almost $4.5 billion in 2016. The global market includes the displays, media players, as well as advanced software technologies for audience measurements. The only market of signage displays in 2014 is projected to reach more than 20 million units, with a growing in the years to come, and a total shipments hitting 25.8 million units by 2016 [2]. This rapid growth is due to the fact that digital signage gives to the organizations the possibility to advertise targeted and personalised messages to the audience in front of a smart screen. Thousands of organizations (e.g., retailers, government institutions, etc.) have already realized the benefits of the digital signage increasing the revenue related to their products or offering a better service in terms of given information to the audience.

In the context of digital signage, soft biometrics data inferred from the face of the user in front to an advertising screen [3] (such as gender identification and age estimation) are used to collect information to be exploited for users profiling. Adhoc advertising campaigns are then showed, taking into account of the collected information. Recent works demonstrate that computer vision techniques for face detection, age and gender recognition, classification and recognition of people's behavior can provide objective measurements (e.g., time of attention) about the people in front of a smart display [4,5,6,7]. Systems able to learn audience preferences for certain content can be exploited to compute the expected view time for a user, in order to organize a better schedule of the advertising content to be shown [5]. The audience emotional reaction can be also captured and analysed to automatically understand the feeling of the people to a campaign (e.g., to understand the attractiveness of a campaign with respect to another) [6]. Recent studies demonstrate that through computer vision methods it is possible quantify the percentage of the people who looked-at the display, the average attention time (differentiating by gender), the age groups who are more most responsive to the dynamic or static content [7].

Although the explosion of the field, in both academia and industry, it seems that measurements about the re-identification of a person in front a smart screen has been not taken into account in the context of digital signage. The information collected with a re-identification engine could be useful to answer the following question: is a specific person back to the advertising screen within a time slot? An automatic answer to this question obtained with a computer vision algorithm for the re-identification of person can be extremely useful to drive the

behaviour of the smart advertising screen which can automatically understand if the person is a new user or it has been seen already within a time slot. Looking at the humans' behaviour, the re-identification of a person is one of the most important feature used by the owner of stores to modify their behaviour in the presentation of the products or to propose special and personalised discounts. A huge number of digital signage scenarios can exploit the information about the re-identification. Among the others, the re-identification can be useful in giving personalised information at an ATM Bank and contextually can be exploited to improve the security during the withdrawal at the ATM (e.g., in a ATM session, the person who is taking the money should be the same of the re-identified one who has inserting the pin number).

In computer vision literature different methods for person re-identification have been proposed [9]. However, differently of the application contexts belonging to digital signage where in most of the cases only the person face is acquired to measure the information, the classic re-identification (e.g., in surveillance) is based on the exploitation of features extracted considering global appearance of an individual (e.g., clothing). Few works consider the re-identification based only on the person face. In [8] a re-identification method which include information of the face is proposed. The authors exploit face features jointly with other information (like hair, skin and clothes color) to re-identify a person.

In this paper, building on face recognition technologies, we propose and evaluate a re-identification system which works by considering only the face of the user in front an advertising screen. To this purpose we consider a dataset composed by $s = 100$ different presentation sessions (10 different customers per 10 different presentation to the system). Each session has been coupled with the remaining ones to produce a final set composed by $100 \times 99 = 9900$ different session pairs (training-session, testing-session) to be used for testing purposes. Data have been collected to guarantee variability during acquisition time (different acquisition period, geometric and photometric variability, faces appearance variability). Experimental results show that an accuracy of 88.73% with 5% of false positive can be achieved by using a many-to-many re-identification approach which considers few faces samples in both training and testing.

The reminder of this paper is organized as following. In Section 2 the proposed digital signage scenario is defined. In Section 3 the proposed approach is described, whereas experimental settings and results are reported in Section 4. Finally, Section 5 concludes the paper with hints for future works.

2 Proposed Digital Signage Scenarios

One of the key features to be a successful salesperson or a good front desk person in a info-point is the ability of identifying customers. In particular, it is well known that when a customer is re-identified it is more simple to offer the most appropriate products or information to the customer. A possible re-identification scenario is the one in which a person has asked some information about a cultural heritage place in a info point and comes back after few minutes to the desk. Depending on the information offered in the first discussion, the person at the front office can predict that some extra information are needed

Fig. 1. A possible digital signage working scenario of a face re-identification engine. The system has to be able in re-identifying a customer within a monitoring session defined by a time slot.

by the customer (e.g., the opening time, ticket price, etc.). This prior can help the front desk person being reactive to reply on further answers as well as in providing extra information to offer a better service. Similarly, if a customer has asked for a particular product (e.g., a perfume of a particular brand) and then comes back within a short time slot, there could be probability that he wishes to compare prices with respect to other products of the same type (e.g., in case he/she has not yet brought the previous seen products) or need to buy (or ask information about) other related products. A lot of other similar scenarios can be imagined. All of them, share the same customer's behaviour: he/she is back to the salesperson or to the front desk person within a time slot. In these cases the re-identification of the customer helps to be more effective in the service. Hence, independently from other collected information (name, age, gender, etc.) the re-identification information is useful.

The above scenarios can be straightforward translated in the context of Digital Signage: the smart screen has to be able to re-identify the customer in front to the display who has been seen within a time slot to trigger the right advertising campaign or to offer more information with respect to the ones searched previously by customer. Differently from the aforementioned scenarios in stores, in the case of Digital Signage the only information useful to perform the re-identification is usually the face of the customer which is visible by the camera of the smart screen.

In the considered context the Digital Signage system has to be designed to work as summarized in Fig. 1. During a monitoring session defined by a specific time slot, the system detects and classifies customers appearing in front of the screen as new or re-identified users. The re-identification engine should be robust to deal with both geometric (e.g., scale, orientation, point of view) and photometric (e.g., luminance) variabilities.

In the following sections we will detail all the key "ingredients" useful to build a Face Re-Identification engine. We will also report the experiments done to assess the performances of the proposed system. We have tested the case of classifying the user as new or re-identified by setting a dynamic time slot equal

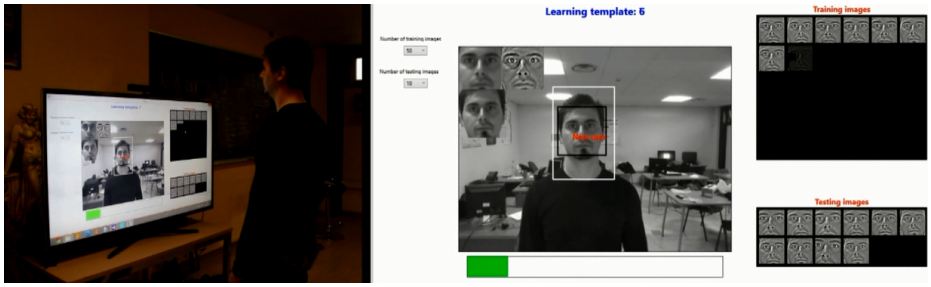

Fig. 2. The developed face re-identification engine in action.

to the time occurred from the previously seen customer. So, when a customer appears at the digital signage screen, it should be recognised if he is the same of the previous one or not. The results show that a re-identification accuracy of 88.73% with 5% of false positive can be obtained exploiting the current state of the art computer vision technologies.

3 Proposed Framework

In this section we detail all the component involved into the pipeline proposed for the face re-identification framework. As first, the face of the customer is detected (Fig. 3). Then the subimage containing the face is pre-processed to remove variabilities due to scale, rotation and lights condition changes (Fig. 4 and Fig. 5). From the obtained image the Local Ternary Patterns (LTP) [10] are extracted and the spatial-based distributions of LTP are considered as final representation of the detected face. For the re-identification purpose (Fig. 6), we employ the χ^2 distance between a set of N representations obtained considering N frames in which the face of the customer is detected, and a set of M representations related the face of the previously seen customer. We use the Kinect [11] as acquisition system by exploiting the skeleton tracking library to detect and track the customer. The aforementioned face re-identification pipeline is performed exploiting the Kinect's RGB channel in a region surrounding the skeleton's head position.

3.1 Face Detection

As first stage of the Face Re-Identification pipeline the face of the user has to be detected. To this purpose we exploit the well-known Viola and Jones object detection framework [12]. It is able of processing images in real time achieving high face detection rates. The solution exploits the "Integral Image" representation so that Haar-like features can be computed at any scale or location in constant time. A learning algorithm based on AdaBoost [13] is used to select the most discriminative features bases for classification purposes. Combining different classifiers in a "cascade" the background regions of the image are discarded while faces are detected. In our experiments we constrain the Face Re-Identification to frontal-faces so that both eyes are visible into the detected face. This is useful to

have references points (i.e., the eyes) to align the detected faces hence removing variabilities due to scale and rotation. As for the face detection, the eyes are localized through the Viola and Jones framework (Fig. 3).

Fig. 3. Customer's face detected by the face re-identification framework. The face is considered for further processing only if both eyes are detected. Eyes position are used to remove scale and rotation variabilities before feature extraction.

Fig. 4. Top row: Customer's face detected in different frames of a session by the Face Re-Identification framework. Bottom Row: Corresponding faces after rotation and scale alignment based on eyes positions.

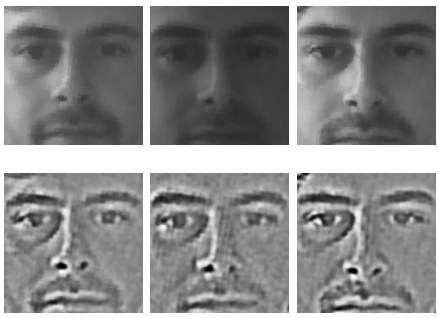

Fig. 5. Top row: Customer's face detected in different frames of a session and aligned by the Face Re-Identification framework. Bottom Row: Corresponding faces after photometric normalization [10].

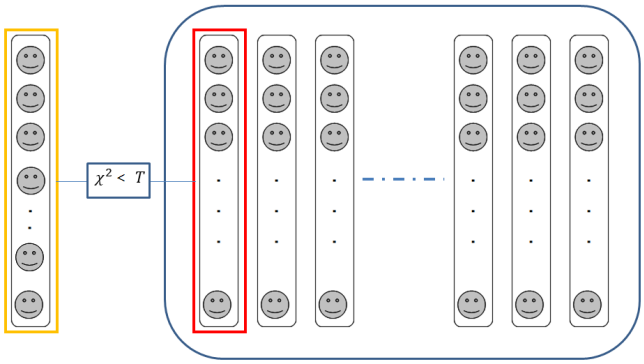

Fig. 6. For the re-identification purpose we compute the χ^2 distance between a set of N representations obtained considering N frames in which the face of the customer is detected (in orange), and a set of M representations related the face of the last seen customer (in red).

3.2 Face Pre-Processing

When a frontal face containing both eyes is detected, a pre-processing step is performed to prepare image containing the face for the feature extraction. As first, geometric variability due rotation and scale changes are removed by mean of the parameters obtained trough an affine transformation of the detected eyes coordinates with respect to fixed eyes positions (Fig. 4). Then to counter the photometric variabilities of illumination, local shadowing and highlights, the normalization pipeline suggested in [10] is employed (Fig. 5). The pipeline is composed by the following three ordered steps: gamma correction, difference of gaussian (DoG) filtering and contrast equalization. All the details can be found in [10].

3.3 Feature Extraction and Face Representation

At this stage the faces have been detected and aligned. As in the problem of Face Recognition [14], for the Face Re-Identification the faces have to be represented with discriminative descriptors to deal with facial expression, partial occlusions and other changes. Moreover, the signature of a face should be computed in real-time. Different papers addressing the problems related to the measurement of soft biometrics (e.g., gender, ages, etc.) for digital signal applications exploit features which are variants of the so called Local Binary Patterns (LBP) [15,16,17]. LBP are robust to lighting effects because they are invariant to monotonic gray-level transformations, and are powerful in discriminating textures. The LBP is an operator useful to summarise local gray-level structures. It takes a local neighborhood around a pixel p, performs a thresholding of the pixels of the neighborhood by considering the value of the pixel p and uses the resulting binary-valued patch as a local image descriptor. By considering a neighborhoods of 3×3 around p, LBP gives a 8 bit code (i.e., only 256 possible codes). The codes extracted for each pixel of the patch containing the face can

be then summarized in a normalised histogram and used as final representation for soft biometric measurements. As a drawback LBP tends to be sensitive to noise, especially in near-uniform image (as in the case of facial regions). For this reason in our system we employ a generalization of the LBP, the so called Local Ternary Patterns (LTP) [10]. Differently than LBP where the central pixel p is considered as threshold to obtain a binary code, the LTP operator gives three possible values as output for each neighbor of p. Let p' a neighbor of p and t a user-specified threshold[1]. The LTP operator is defined as follows:

$$LTP(p,p',t) = \begin{cases} 1 & \text{if } p' \geq p+t \\ 0 & \text{if } |p'-p| < t \\ -1 & \text{if } p' \leq p-t \end{cases} \quad (1)$$

For the encoding procedure the LTP code is splitted into positive and negative patterns which are considered unsigned to compute two different binary codes (i.e., Upper Pattern and Lower Pattern). For representation purposes, the two distributions of Upper and Lower Patterns are computed and concatenated.

As suggested in other works related the field of soft biometrics estimation [18], we divide the input image with a regular grid (7×7 in our experiments). The LTP distributions are computed on each cell of the grid to encode spatial information and hence improve the face re-identication results.

3.4 Face Re-Identification

As last step the framework re-identifies a customer within a monitoring session defined by a given time slot (Fig. 1). We consider the case in which the time slot is equal to the amount of time in which the last customer has been detected by the digital signage system. This means that the re-identification engine has to be able to answer the following question: is the current customer and the previous seen customer the same person? The proposed face re-identification mechanism works as described in the following. During a customer session (i.e., the customer is interacting with the smart display) the face re-identification engine collects M faces templates as described in Sections 3.1 and 3.2, and represents them as described in the previous section. This set of templates is considered the training dataset for the re-identification of the next customer. When the next customer shows up in front the smart monitor, the proposed system analyses its face, collects and represents N faces templates (testing images) and finally compares them with the M training images learned during the previous monitoring session (Fig. 6). To this purpose the χ^2 distance is employed. For every represented template of the current customer, the smallest χ^2 distance to any represented template of the training is computed. Then the smallest distance is considered to decide if the current customer and the previous one are the same person (re-identification) or if there is a new user. The decision threshold is fixed to have a low number of false positives (i.e., low number of cases in which a customer B is wrongly re-identified as previous seen customer A rather than as new user).

[1] In our experiments the LTP threshold t has been fixed as suggested in [10].

The number of training templates could be greater than the number of testing template (i.e., $M > N$). Indeed, although the collection of training images of a new user does not have impact to the system behaviour and it is transparent for the customer in front the digital screen and can have duration equal to the time spent by the customer in front the screen, during testing the system have to react in real time (or at least as soon as he get the N faces templates) so that personalised information can be offered to the final user. Our experiments demonstrates that few templates can be used for re-identification purpose in both training ($M = 50$) and testing ($N = 6$).

3.5 Customer Tracking

In real scenarios of digital signage it is reasonable to imagine that the customer is surrounded by other people (which look or not at the smart screen). To properly track the customer we use the Kinect and its standard skeleton tracking. This allows to recognize the closest person to the screen which we assume to be the customer interacting with the advertising screen. All the processing needed for the face re-identification (see previous section) is performed exploiting the Kinect's RGB channel. Since we track the skeleton, the customer's head position is known and the face re-identification pipeline can be performed only in the region surrounding the head of the customer (i.e., white bounding box in Fig. 2).

4 Experimental Settings and Results

To asses the performances of the proposed framework we have collected a dataset composed by 5000 faces. Specifically, we have acquired 100 different presentation sessions (each composed by 50 face patches) related to 5 male and 5 female different users who have been presented at the system 10 different times. The data have been acquired at different time (in a range of a couple of months), in different environments to guarantee variability in the subjects' appearance (beard, makeup, tan), as well as by considering different photometric and geometric conditions (lights, point of view, clutter background). Each presentation session has been coupled with the remaining ones to produce a final set of $100 \times 99 = 9900$ users' pairs (training-session, testing-session) to be used in the experiments. Each pair of users has been labeled as a session containing the same person or not.

In each session, a customer showed up in front the smart screen and a RGB-D video of a couple of minutes has been acquired with a Kinect device [11]. From each session the first $M = 50$ face templates have been detected, processed and represented as described in previous section. For testing purpose we have used a leave one out approach [19]. At each run, the templates related to one customer presentation session are considered as training model, whereas the remaining sessions templates are considered as testing set. Different tests have been done to assess the influence of the size M of the set of faces template to be used as training in a session, and the size N of the set of faces to be used as testing in a session. The final results are obtained by averaging over all runs. Since the face re-identification results depends from the threshold used to make the re-identification decision (see Section 3.4) we evaluate the proposed method making use of ROC curves [19].

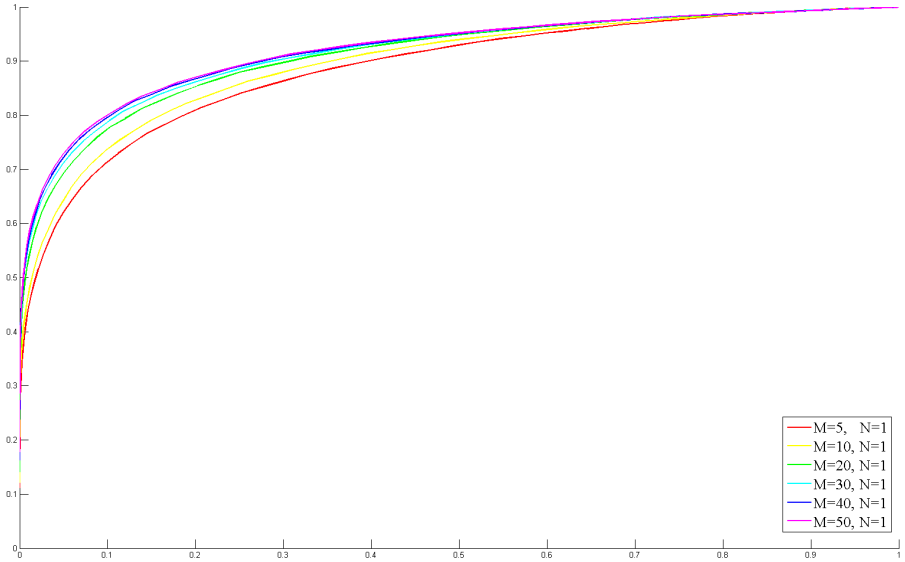

Fig. 7. Face re-identification results at varying of the number of training templates. The vertical axis corresponds to the True Positive rate, whereas the horizontal axis is related to the False Positive rate.

In Fig. 7 are reported the ROC curves results obtained at varying of the parameter M related to the faces templates to be considered as model in the training set. In this test the number of faces template to be used for testing is fixed to $N = 1$. In Fig. 8 are reported the ROC curves results obtained at varying of the parameter N related to the number of faces template to be considered as testing set. The number of templates to be used for training is instead fixed to $M = 50$. The related results are compared in Fig. 9. The tests showed that by increasing the number of patterns to be considered for both training and test sets the face re-identification performances improve. It should be noted that for the testing case (i.e., when a user shows up to the screen and should be recognized as new user of same person of the previous customer) by increasing the number of patterns more than 6 per session does not gives much improvements in the results (Fig. 8). Considering few face templates in both training ($M = 50$) and testing ($N = 6$) the re-identification accuracy stated at 88.73% with 5% of false positive. The proposed face re-identification framework (Fig. 2) works in real time as demonstrated by the video at the following link: `http://iplab.dmi.unict.it/download/VAAM2014`.

5 Conclusions and Future Works

This paper has addressed the problem of face re-identification in the context of digital signage applications. Motivations, scenarios and the main ingredients to build a face re-identification framework are described. Quantitative evaluation is reported to assess the proposed framework. Further works will be devoted to

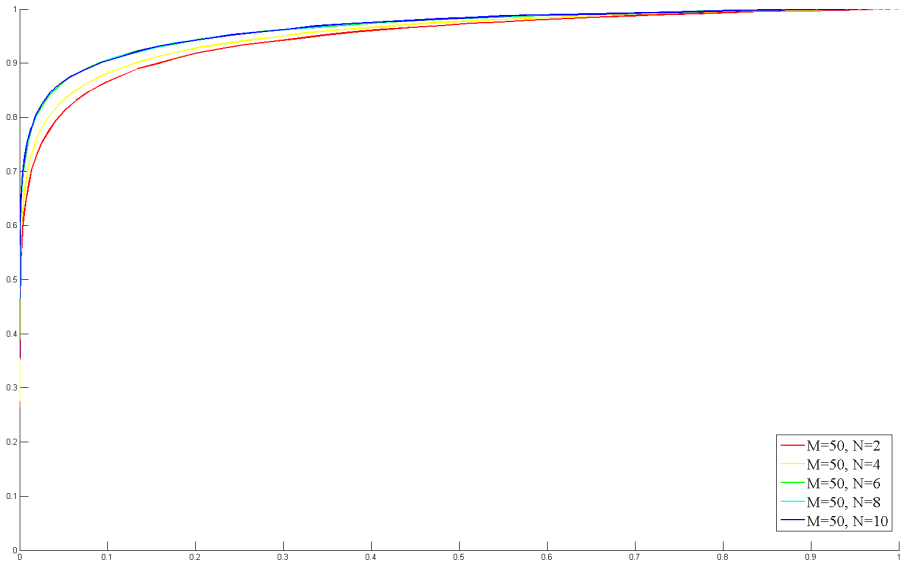

Fig. 8. Face re-identification results at varying of the number of testing templates. The vertical axis corresponds to the True Positive rate, whereas the horizontal axis is related to the False Positive rate.

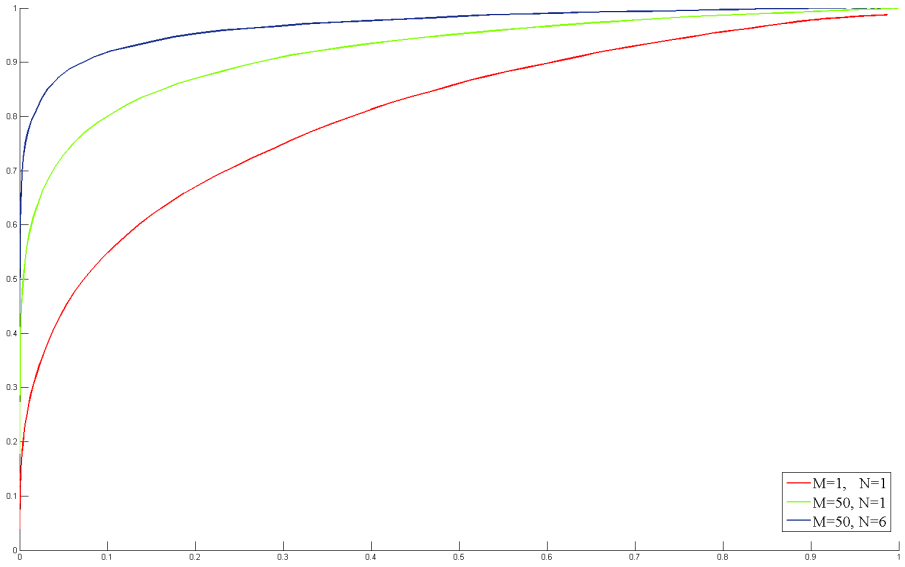

Fig. 9. Comparison of face re-identification results at varying of the number of training and testing templates. The vertical axis corresponds to the True Positive rate, whereas the horizontal axis is related to the False Positive rate.

improve the performances of the proposed baseline face re-identification method by reviewing and refining all the steps involved into the face re-identification pipeline and augmenting the face representation to consider depth information (e.g., for faces alignment purpose). Moreover, an in-depth study by considering the re-identification of the last $K > 1$ customers on a larger dataset collected in real scenarios (e.g. stores, info points) should be done.

References

1. Aberdeen Research: Digital Signage: A Path to Customer Loyalty, Brand Awareness and Marketing Performance, 2010
2. Khatri, S.: Digital Signage and Professional Display Market Set for Solid Growth in 2012. Signage & Professional Displays Market Tracker Report, an April 2012
3. Ricanek, K., Barbour, B.: What Are Soft Biometrics and How Can They Be Used?. Computer , 44(9), 106-108, 2011
4. Batagelj, B., Ravnik, R., Solina, F.: Computer vision and digital signage. Tenth International Conference on Multimodal Interfaces, 2008
5. Müller, J., Exeler, J., Buzeck, M., Krüger, A. : ReflectiveSigns: Digital Signs that Adapt to Audience Attention. Proceedings of Pervasive 2009
6. Exeler, J., Buzeck, M., Müller, J.: eMir: Digital Signs that react to Audience Emotion. 2nd Workshop on Pervasive Advertising, 38-44, 2009
7. Ravnik, R., Solina, F.: Audience Measurement of Digital Signage: Quantitative Study in Real-World Environment Using Computer Vision. Interacting with Computers, 25(3), 218-228, 2013
8. Dantcheva, A., Dugelay, J.-L.: Frontal-to-side face re-identification based on hair, skin and clothes patches. IEEE International Conference on Advanced Video and Signal-Based Surveillance, 309-313, 2011
9. Vezzani, R., Baltieri, D., Cucchiara, R.: People Re-identification in Surveillance and Forensics: a Survey. ACM Computing Surveys, 46(2), 2013.
10. Tan, X., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. IEEE Transactions on Image Processing, 19(6), 1635-1650, 2010
11. Microsoft Kinect: http://www.microsoft.com/en-us/kinectforwindows/
12. Viola, P., Jones, M. J.: Robust Real-Time Face Detection. International Journal of Computer Vision 57(2), 137-154, 2004
13. Schapire, R. E.:A brief introduction to boosting. International Joint Conference on Artificial intelligence, 1999
14. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Survey 34(4), 399485, 2003
15. Ahonen, T., Hadid, A., Pietikinen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 28(12), 2037-2041, 2006
16. Wang, J.-G., Yau, W.-Y., Wang, H. L.: Age Categorization via ECOC with Fused Gabor and LBP Features. Workshop on Applications of Computer Vision, 2009
17. Hadid, A., Pietikinen, M.: Combining Appearance and Motion for Face and Gender Recognition from Videos. Pattern Recognition 42(11), 2818-2827, 2009
18. Ylioinas, J., Hadid, A., Pietikainen, M.: Age Classification in Unconstrained Conditions Using LBP Variants. International Conference on Pattern Recognition, 2012
19. Webb, A. R.: Statistical Pattern Recognition (2nd Etition). John Wiley & Sons, LTD., 2002