A Talking Profile to Distinguish Identical Twins

Li Zhang[1], KengTeck Ma[1], Hossein Nejati[1], Lewis Foo[1],
Dong Guo[2], Terence Sim[1]
{lizhang,ktma,lewis,tsim}@comp.nus.edu.sg,
nejati@nus.edu.sg, dnguo@fb.com

[1] *School of Computing, National University of Singapore*
[2] *Facebook Inc., Palo Alto, CA*

Abstract

Identical twins pose a great challenge to face recognition due to high similarities in their appearances. Motivated by the psychological findings that facial motion contains identity signatures and the observation that twins may look alike but behave differently, we develop a talking profile to use the identity signatures in the facial motion to distinguish between identical twins. The talking profile for a subject is defined as a collection of multiple types of usual face motions from the video. Given two talking profiles, we compute the similarities of the same type of face motion in both profiles and then perform the classification based on those similarities. To compute the similarity of each type of face motion, we give higher weights to more abnormal motions which are assumed to carry more identity signature information.

Our approach, named Exceptional Motion Reporting Model (EMRM), is unrelated with appearance, and can handle realistic facial motion in human subjects, with no restrictions of speed of motion or video frame rate. We first conduct our experiments on a video database containing 39 pairs of twins. The experimental results demonstrate that identical twins can be distinguished better by the talking profiles over the traditional appearance based approach. Moreover, we collected a non-twin youtube dataset with 99 subjects. The results on this dataset verified that the talking profile can be the potential biometric. We further conducted an experiment to test the robustness of talking profile to the time. Videos from 10 subjects which spans across years or even decades in their lives are collected. The results indicated the robustness of talking profile to the aging process.

Keywords:
talking profile, identical twins, abnormal motions, EMRM

1. Revision Statement

Thanks for the valuable comments. In the revised version, we have conducted more experiments for comparison. Three more baseline algorithms suggested by the reviewers are tested: face recognition via independent component analysis, face recognition via locality projection preserving and a commercial face matcher Luxand faceSDK.

We would like to clarify the issues pointed out as follow:

1). The size of the dataset is a little bit small. A large size would be better.

We would like to test our algorithm on a larger database, but the database for twins is really hard to collect. To the best of our knowledge, our database is already the largest video twin databases in the entire research community.

2). There are two "??" in Section 4.2. Please revise them with the right number.

We have revised it accordingly.

3). Performance comparison with other methods.

We have conducted some more experiments to compare our algorithms with suitable state-of-the-art algorithms. Three more recent algorithms are used: face recognition with Independent component analysis, face recognition with Local preserving projection and a commercial face matcher Luxand faceSDK. The final results further demonstrate the superiority of using facial motion to distinguish identical twins instead of static appearance.

4). What is the suitable length of the video?

Currently the videos are 45 to 60 seconds. We choose this length for our experiment mainly through empirical observation. We find that the subjects have performed enough motions within this time interval. This paper is mainly to verify the possibility of using abnormal motion as a biometric to distinguish identical twins. This has been verified via the chosen video length, even though it may be the best length.

5). Some variations in the videos including face motions might be native but some variations might be accidentals. How to distinguish these two kinds of variations?

We acknowledge this question is very important in real application. The main idea in this paper is to evaluate the effectiveness of using pure natural face/head motion as a biometric. However, in real application, natural motion is always mixed with some accidental motions (such as slapping one's face during talking).

Detection of natural (native) face motion and accidental motion is actually very hard in general setting. We overcome this problem in this paper through data collection setting. We encourage the subjects to talk in a more natural and free manner, while discourage the talking with some tasks. In this way, the final experimental videos rarely contain the significant accidental face motion. We are looking for solutions to separate the natural motion from accidental motion.

2. Introduction

The occurance of twins has progressively increased in the past decades as twins birth rate has risen to 32.2 per 1000 birth with an average 3% growth per year since 1990 [1]. With the increase of twins, identical twins are becoming more common as well. This, in turn, is urging biometric identification systems to accurately distinguish between twin siblings. Although identical twins represent a minority (0.2% of the world's population), it is worth noting that they equal the whole population of countries like Portugal or Greece. Therefore failing to identify them is a significant hindrance for the success of biometric systems. Identical twins share the same DNA code and therefore they look extremely alike. Nevertheless, some biometrics depend not only on the genetic signature but also on the individual development in the womb. As a result, identical twins have some different biometrics such as fingerprint and retina. Several researchers have taken advantage of this fact and have shown promising results in automatic recognition systems that use these discriminating traits: fingerprint [2], palmprint [3], iris [4] and combinations of some of the above biometrics [5]. However, these biometrics require the cooperation of the subject. Thus, it is still desirable to identify twins by pure facial features, since they are non-intrusive, they do not require explicit cooperation of the subject and are widely available from photos or videos captured by ordinary camcorders. Unfortunately, the high similarity between identical twins' appearance is known to be a great challenge for face recognition systems. The performance of various face appearance based approaches on recognizing identical twins has recently been questioned [5, 6]. They both confirmed the difficulties encountered by appearance based face recognition systems on twin databases, and strongly suggested the need for new ways to improve performance of recognizing identical twins.

In psychology, it has been demonstrated that the human visual system utilizes both appearance and facial motion for face recognition [7, 8]. Appearance information provides the first route for face recognition, while the dynamic signatures information embedded in facial motion are processed in the superior temporal sul-

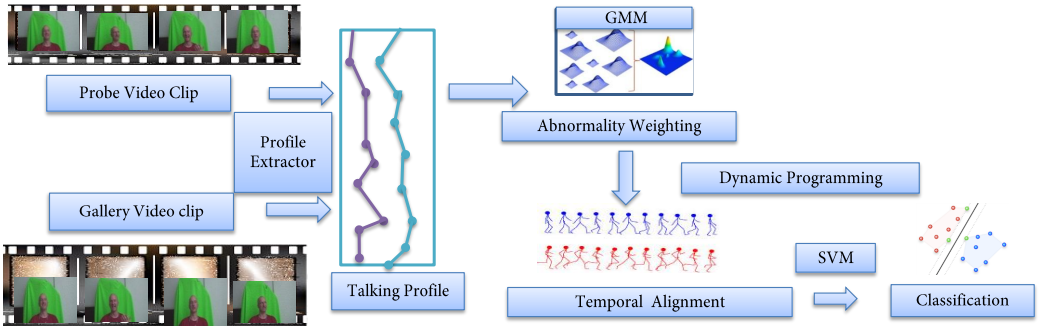

Figure 1: For each video, we extract its talking profile consisting of various types of face motions during free talking. In our experiments, six types of usual motions are included: 2D in-plane head translation, pose change, gaze change, pupil movement and eye/mouth opening-closing change. Each type of face motion in talking profile is a sequence of local motions between two adjacent frames. For two corresponding face motions who are same type in two talking profiles, we assign the weights to every local action based on its abnormality, and then perform a temporal alignment by minimizing the abnormality object functions. The similarity of these two corresponding face motions is then computed as the weighted summation of local motion similarity. Finally, a classification using support vector machine is performed on the similarities of all corresponding face motions in the talking profiles.

cus and provide a secondary route for face recognition. This secondary route for face recognition, also known as *supplemental information hypothesis*, is supported by many works both in psychology and computer vision [7, 9].

The traditional appearance based face recognition simulates the first route to recognize faces, but fails to effectively distinguish between identical twins. Considering both studies support that facial motion contains identity information, and the observation that twins may look alike but behave differently, we propose to use talking profiles which consists of multiple type of usual face motions to recognize twins. Our intention is to simulate the secondary route for face recognition. The flowchart of our approach, named Exceptional Motion Report Model (EMRM), is illustrated in Figure 1.

1) Given a video, the talking profile is extracted. The profile consists of 6 types of usual face motions, such as pose change and 2D in-plane movement. Each type of motion in the profile is represented as a sequence of local motions between two adjacent frames.

2) From the computed two talking profiles, we compute the similarity of each pair of corresponding face motions of the same type from both talking profiles. Given a pair of corresponding face motions, since they may be *unsynchronized*,

i.e. different frame rate or speed, we perform a motion alignment in advance. The alignment is achieved by minimizing the abnormality function. A gaussian mixture model is employed to estimate the abnormality of each local motion. This abnormality scheme is inspired from [10] in psychology which proves that humans use exceptional motions to identify faces. This is also observed in real life that humans always use a person's peculiar head motion (*e.g.* tilt) rather than common head movement to aid recognition. After alignment, the similarity of this pair of corresponding face motions is computed as the weighted sum of local motion similarity.

3) From the similarities of every pair of corresponding face motion sequences in the talking profiles, we perform a SVM classification on those similarities. Also, our model is set in verification mode, that is, to claim the faces in two videos are genuine or imposters.

We test our algorithm by conducting several experiments on a free talking video database with 39 pairs of identical twins[1]. Our experimental results shows that compared with traditional appearance based approaches, the talking profile can be used to accurately distinguish between identical twins. We further apply talking profile on a free talking video database of 99 subjects from YouTube. Results from our second set of experiments are in agreement with the psychological findings that facial motion contains identity signatures and demonstrate that the talking profile has potential to be used in biometrics. To test the robustness of talking profile against time, we further collected an cross-age video database with 10 subject whose videos set are across years and conducted the experiments on this database. The results indicated that the talking profile is relatively robust to the time change.

The contributions of our work are four fold: 1) we show that talking profiles can be used to distinguish between identical twins. 2) we propose a novel EMRM to analyze facial motion in video, which also provides a general framework of using abnormality for recognition. 3) our experiments on YouTube dataset supports psychological findings that facial motion does provide identity signatures. 4) our experiments on cross-age video database demonstrate the robustness of talking profile against the aging process. We continue by introducing existing twins recognition and motion based face recognition works in Section 3. In Section 4, we describe the details of our model. We present our dataset and experiments in Section 5, concluding in Section 6.

[1]This is the largest video database of identical twins, known to authors

3. Related Work

3.1. Face Recognition for Identical Twins

To our best knowledge, there are limited works on identical twin recognition using 2D face biometric [5, 6, 11]. Sun et al. [5] were the first to evaluate the performance of appearance based face recognition to distinguish between twins. They compared with performances of iris, fingerprint and a fusion of them. Their database was collected in 2007 at the fourth Annual Festival of Beijing Twins Day. The face subset used in the experiments contained 134 subjects, each having around 20 images. All images were collected during a single session over a short interval. Experiments were conducted using the FaceVACS commercial matcher and showed that identical twins are a challenge to current face recognition systems. Phillips *et.al* [6] thoroughly extended the analysis of the performance of face recognition systems in distinguishing between identical twins on another database collected at the Twins Days festival in Ohio in 2009 and 2010. It consisted of images of 126 pairs of identical twins collected on the same day and 24 pairs with images collected one year apart. Facial recognition performance was tested using three of the top submissions to the Still Face Track at Multiple Biometric Evaluation 2010. Based on their experimental results, the best performance was observed under ideal conditions (same day, studio lighting and neutral expression). But under more realistic conditions, distinguishing between identical twins was very challenging. Klare *et.al* [11] analyzed the features of each facial component to distinguish identical twins from the same database in [6]. They also analyzed the possibility of using facial marks to distinguish identical twins. They also confirmed the challenge of recognizing identical twins merely based on appearance. All these works showed the need for new approaches to help improve performance when recognizing identical twins.

3.2. Motion Based Face Recognition

Psychological studies have shown that humans better recognize faces with expressive motion. Hill and Johnston [7] showed in their experiments that humans utilize rigid head movements in face recognition. Thornton and Kourtzi [12] observed that showing moving face images rather than static face images in training sessions improved human subjects' performances in face recognition. Pilz *et al.* [9] claimed further that moving images not only increased recognition rate, but also reduced reaction time. These psychological findings imply that face motions contain considerable identity information which is fairly reliable for face recognition. There are some recent attempts [8, 13] in computer vision to use face

motions for face recognition. They computed either a dense or sparse displacement on tracked points and used it to identify general human subjects. Tulyakov *et.al* [8] manually marked some landmark points and used their displacement as features for face recognition. Ye *et.al* [13] proposed to compute the dense optical flow from the neutral to the apex of smile, then used the optical flow field as feature for recognition. Later, they designed a local deformation pattern as a feature and tried to find identity signature between different expressions. All these works achieved some breakthrough in motion based face recognition, but all of them only considered general population and required the cooperation of subjects. Moreover, among all possible face motions, only facial expressions are considered in these works, while many other types of face motions remain unexplored.

One recent work tried to apply motion based face recognition on the identical twins problem [14]. Their proposal only focused on expression and required fixed heads position without any movement. Our work is different from [14] in three aspects. First, in our model, we do not require any face alignment while they required very accurate face alignment. Second, our work uses 6 different types of usual face motions while they only used facial expressions. Third, our algorithm can handle occurrence of multiple and unknown motions in a single video, while they can only handle the existence of one expression per video.

4. Proposed Algorithm

In this section we describe the details of our model, Exception Motion Reporting Model (EMRM), to use talking profiles for face recognition. We follow the setting in Labeled Face in the Wild [15] to set EMRM in verification model, that is, to claim the faces in two videos are imposters or genuines. There are in total five steps in the EMRM. The first step is to extract the talking profile from each video. The second and third steps are to assign the abnormality weight and compute local motion similarity. The fourth step is to align two face motion sequence of the same type from two talking profiles and compute their similarity. The final step is to use the similarities from all the corresponding pairs in talking profiles for classification using support vector machine.

4.1. Extracting Talking Profile

For each video, the talking profile is comprised of multiple types of usual face motions. In our work, it includes 6 types: 2D in-plane translation, pose change, gaze change, pupil movement and eye/mouth open-closing magnitude

(i.e. the extent that the eyes/mouths are open). Various tools have been released to extract these information from the video, such as Luxand, Pitpatts and Omron SDK. In our implementation, we use Omron SDK to extract the talking profile. Note that each type of face motion in talking profile is a sequence of local motions between two adjacent frames. We perform a sampling before processing to test the robustness when the probe and the gallery have different frame rates and also save computation workload. Assume the sample rate is FPS (*i.e.* the local motion is computed between each FPS frames). We define talking profiles as $TP = \{\phi_{head}, \phi_{gaze}, \phi_{pose}, \phi_{pupil}, \phi_{eye}, \phi_{mouth}\}$. Assume $\phi_i \in TP$ is one type of face motion, then it can be expressed as follows in the temporal order, where ς_i is a local motion:

$$\phi_i \quad = \quad \{\varsigma_1, \varsigma_2, \cdots, \varsigma_t\} \tag{1}$$

4.2. Encoding Local Motion Abnormality

For a talking profile, there are 6 different types of face motions, and each type of face motion is a sequence of local motions. In this section, we address how to encode the abnormality weight to each local motion. The motivation is from psychological studies [10] which prove that the human visual system uses visual abnormality for recognizing faces.

Considering human variations (e.g. ethnicity, gender, etc), we employ a Gaussian Mixture Model, $G = \{g_1, g_2, g_3, ..., g_\tau\}\ \forall i,\ g_i \sim N(\mu_i, \sigma_i)$, to estimate the local motion distribution in each type of face motion space. μ_i and σ_i are estimated by using expectation maximization [16]. In our implementation, we will run the gaussian mixture model six times as we have 6 types of face motions. Then given a local motion $\varsigma \in \phi_i$, we use maximum likelihood to find its intrinsic gaussian distribution as κ, where $\varsigma \in g_\kappa \iff \kappa = \underset{i}{argmax} P(\varsigma|g_i)$. After knowing its gaussian distribution, the probability of this local motion can be computed as $P(\varsigma|g_\kappa)$. Then, we approximate its abnormality, ω, as $\omega(\varsigma) = 1 - P(\varsigma|g)$.

4.3. Computing Local Motion Similarity

Given two talking profiles, each with 6 type of face motions, we need to compute the similarity of corresponding face motions from both talking profiles. In our work, six similarities in total will be computed, each for one type of face motion. Here, we choose pose change as an example. Assume there are two pose change sequences from the two talking profiles, $\phi = \{\varsigma_1, \varsigma_2, \cdots, \varsigma_n\}$ and

$\varphi = \{\varsigma'_1, \varsigma'_2, \cdots, \varsigma'_m\}$, we first define how to compute the local motion similarity between two local motions $\varsigma_i \in \phi, \varsigma'_j \in \varphi$ and then perform motion alignment.

We define the local motion similarity W in Eq 2.

$$
\begin{aligned}
W(\varsigma_i, \varsigma'_j) &= \frac{\omega(\varsigma_i) * \omega(\varsigma'_j) * (Sim(\varsigma_i, \varsigma'_j))}{(D_{RAD}(g_s, g'_t)^2 + C_1)} && (2)\\
\omega(\varsigma_i) &= (1 - P(\varsigma_i|g)); \ \varsigma_i \in g && (3)\\
s &= \underset{k}{argmax} P(\varsigma_i|g_k) && (4)\\
t &= \underset{k}{argmax} P(\varsigma'_j|g_k) && (5)
\end{aligned}
$$

where $\omega(\varsigma_i)$ and $\omega(\varsigma'_j)$ are the abnormality for each local motion, ς_i and ς'_j, respectively. $Sim(\varsigma_i, \varsigma'_j)$ is the similarity of the motion in Euclidean space. $D_{RAD}(g_s, g'_t)$ is the difference between two gaussian distributions, $g_s$and g'_t. g_s is the intrinsic gaussian distribution of ς_i, and g'_t is the intrinsic gaussian distribution of ς'_j. Note $g_s$and g'_t can be same or different. To estimate the difference between two gaussian distributions, the common way is the Kullback-Leibler Divergence [17], defined as Eq 6.

$$D_{KL}(p \parallel q) \doteq \int p(x) log_2(\frac{p(x)}{q(x)}) dx \quad (6)$$

where $p(x)$ and $q(x)$ are distributions. KL distance is non-negative and equal to zero iff $p(x) \equiv q(x)$, however, it is asymmetric. Thus, we use its symmetrical extension, Resistor-Average Distance (RAD), defined as Eq 7.

$$D_{RAD}(p, q) = [D_{KL}(p \parallel q)^{-1} + D_{KL}(p \parallel q)^{-1}]^{-1} \quad (7)$$

Similar to KL, RAD is non-negative and equal to zero iff $p(x) \equiv q(x)$. Moreover, it is symmetric.

From Eq. 2, several points can be concluded. First, we reward the more abnormal local motions, seen ς_i and ς'_j. The larger ς_i and ς'_j are, the more abnormal $W(\varsigma_i, \varsigma'_j)$ will be. Secondly, the higher the similarity of the local motions in Euclidean space is, the higher the final local motion similarity is, seen $Sim(\varsigma_i, \varsigma'_j))$. In our implementation, we define the similarity in Euclidean space as the inverse of the Euclidean distance between ς_i and ς'_j. Note ς_i and ς'_j are local motions from the same type of face motion, thus they are vectors of same length in Euclidean space. Thirdly, if the intrinsic distribution of ς_i and ς'_j is same, then $D_{RAD}(g_s, g'_t$ is equal to zero, otherwise it will be larger than 0. Therefore, $W(\varsigma_i, \varsigma'_j)$ penalizes the situation when ς_i and ς'_j are from different gaussian distribution. C_1 is a constant to avoid zero division. We set it to $C_1 = 1$ in our implementation.

4.4. Aligning Motion Sequences

In previous section, we describe how to compute the similarity between two local motions of the same type. Next we need to align these two motion sequences in temporal order (*i.e.*find the best matches between two sequences). Mathematically, we maximize the total local motion similarity score $\Upsilon(\phi, \varphi)$ between ϕ and φ as follows:

$$\max \Upsilon(\phi, \varphi) = \{\varsigma_{i1} : \varsigma'_{j1}, \varsigma_{i2} : \varsigma'_{j2}, ...\varsigma_{ik} : \varsigma'_{jk}\}$$
$$s.t. \forall i_s, i_t, j_s, j_t, i_s > i_t \Rightarrow j_s > j_t$$

where i_s, i_t, j_s, j_t represent the frame number in temporal order. Temporal consistency is imposed in the constraint. This maximization problem is similar to finding the longest common sub-sequence, with addition of element continuous match scoring, instead of a binary $1/0$ match scoring. Based on the local motion similarity described above, we propose a dynamic programming script to calculate the maximum matching score given two feature sequences. The rules of this dynamic programming are described in Algorithm 1. The general idea of Algorithm 1 is to continuously update the current best match up to action $\varsigma_i \in \phi$ in Table $\gamma(\phi, \varphi)$. As convention, we use $|V|$ to denote the length of vector V. Rule 1 is the initialization. Based on Rule 2, if there is only one local motion to match (i.e. either $|\phi| = 1$ or $|\varphi| = 1$), then the matching weight would be $W(\varsigma_i, \varsigma'_j)$. Based on Rule 3 (when there are more local motions to match), at each step we either abandon previous matches to local motion ς'_j and decide to match local motion ς_i with ς'_j (i.e. $W(\varsigma_i, \varsigma'_j)$) or prefer to keep the previous match to local motion ς'_j (i.e. $\gamma_{i-1,j}(\phi, \varphi)$) and match the ς_i to one of the next local motions in the temporal order (i.e. $W(\varsigma_i, \varsigma'_{j'})\ s.t.\ j' = j + 1, ..., m$). Finally, based on the outcome of Rule 3, all of the next rows of the γ Table will be updated.

4.5. Performing Classification

Through the above steps, we align two face motion sequences of the same type from both talking profiles in $\Upsilon(\phi, \varphi)$. Then we can compute the final similarity for pose change in talking profile scores as the summation of corresponding local motion similarity in $\Upsilon(\phi, \varphi)$. We repeat those steps to compute the similarity for gaze change, 2D in-plane change, pupil movement, eye/mouth opening-closing, respectively. Finally we employ support vector machine to perform classification [18] by using those similarities for verification.

Algorithm 1 Dynamic programming to align two sequences by maximizing total local motion similarities.

$\Upsilon(\phi, \varphi)$: Table $\gamma(\phi, \varphi)$ is an $n \times m$ table where $|\phi| = n$ and $|\varphi| = m$.
i and j are row index and column index of table γ
Initialize $i = 0$
Start Loop i
    Rule 1: if $i <= 0$ then, $\gamma_{i,j}(\phi, \varphi) = 0$
    Rule 2: else if $i = 1$ then, $\gamma_{i,j}(\phi, \varphi) = W(\varsigma_i, \varsigma'_j)$
    Rule 3: else if $i <= n$
        Start Loop j
        $\gamma_{i,j}(\phi, \varphi) = \max(W(\varsigma_i, \varsigma'_j), \max(\gamma_{i-1,j}(\phi, \varphi) + W(\varsigma_i, \varsigma'_{j'})))$, j' varies from $j + 1, ..., m$
        $\gamma_{i+1,j}(\phi, \varphi) = \gamma_{i+1,j}(\phi, \varphi) + \gamma_{i,j}(\phi, \varphi)$
        End Loop j
End Loop i
$\Upsilon(\phi, \varphi) = \underset{j}{argmax}(\gamma_{n,j}(\phi, \varphi))$

Figure 2: Some examples from identical twins database.

5. Experiments

5.1. The Identical Twins Dataset

In experiments, we collected an identical twin free talking database at the Sixth Mojiang International Twins Festival held on 1st May 2010 in China. This database includes Chinese, Canadians and Russians. There are 39 pairs of twins (78 subjects) and each subject has at least 2 video clips at approximately 45 seconds. A Sony HD color video camera is used to capture the video clips. They do not constrain the face position while speaking so the expression, head and neck movement of each participant is realistic in each clip. Figure 2 shows 6 subjects (3 pairs of identical twins) from this database. To the best of our knowledge, this is the largest twin video database in the entire research community.

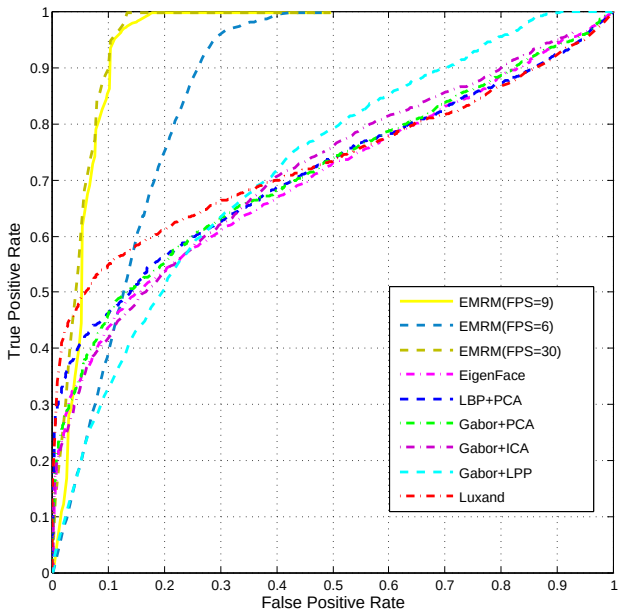

Figure 3: Comparison between traditional appearance approaches and our proposed EMRM at different sampling rate on identical twins database

5.2. Experiment 1: Traditional Appearance Based Approach on Twins

We chose six facial appearance approaches, Eigenface [19], Local Binary Pattern [20], Gabor [21], Independent Component Analysis [22, 23], Locality Preserving Project [24] and a commercial face matcher "Luxand faceSDK", as baseline to compare our approach with the performance of using appearance to distinguish between identical twins. For each twin subject, we randomly select 8 images from the talking videos. The images are then registered by eye positions detected by STASM [25] and resized to to 160 by 128. For Eigenface, we vectorized gray intensity in each pixel as feature and performed principle component analysis to reduce the dimension. For LBP, we divided the image into 80 blocks. For each block, we extract the 59-bins histogram. For Gabor, we used 40 Gabor (5 scales, 8 orientation) filters and set the kernel size for each Gabor filter to 17 by 17. Principle component analysis is performed to reduce the feature dimension for LBP and Gabor. For Independent Component Analysis (ICA), we use Gabor as representation of the image. Suggested by [22], we employ the architecture I which we find a set of statistically independent basis image. For LPP, we also use Gabor as representation of the image. LPP is a dimension reduction technique that preserve the locality after projection. For the commercial face matcher, we

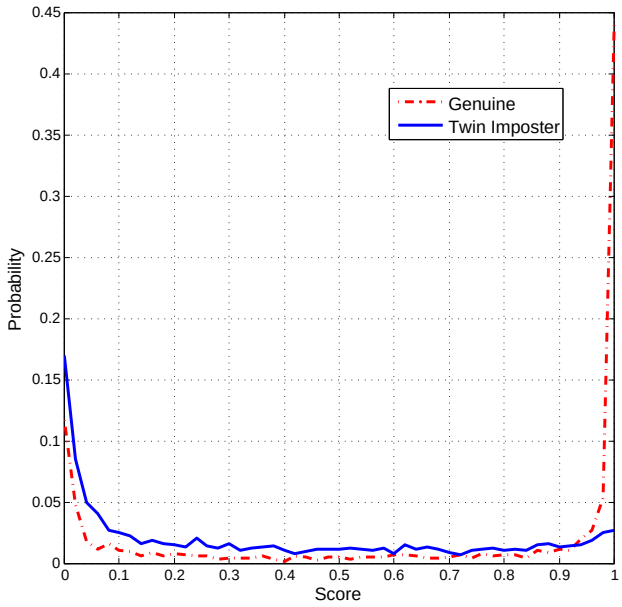

Figure 4: genuine and twin imposter score distribution of Luxand faceSDK in twin database

use Luxand faceSDK. Luxand, Inc. is a private hi-tech company formed in 2005. Luxand faceSDK can output a similarity score range from 0 to 1 given two images. Here, 0 represents the most dis-similar, while 1 represents the most similar.

The experimental result is shown in Figure 3. From this figure, we can see that identical twins indeed pose a great challenge to appearance based approaches. If the threshold is set when false accept rate is equal to false reject rate, the accuracy is 0.644 for Eigenface , 0.654 for LBP, 0.658 for Gabor, 0.656 for ICA, 0.666 for LPP and 0.674 for Luxand faceSDK separately. We can also clearly see that there is no huge difference between Intensity, LBP, Gabor, ICA, LPP and Luxand faceSDK for twin verification. This result verifies the twin challenge. To better illustrate the result, we also demonstrate the score distribution of twin imposter and genuine of Luxand faceSDK in Figure 4. From this figure, we can clearly see that twin imposter can have very high similarity score even they are not same subject.

5.3. Experiment 2: Performance with Same Sample Rate

In this experiment, we evaluate the performance of using talking profile to distinguish between identical twins. The sample rate for probe and gallery are set to be same. For classification, we use 60% of the videos for training and

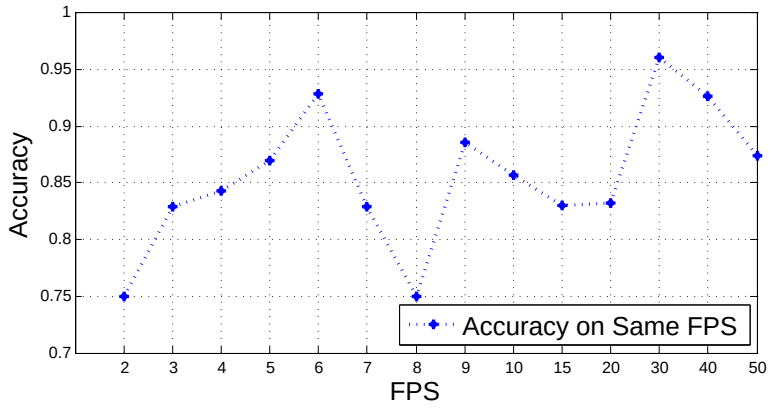

Figure 5: Experiment 2: Accuracy when gallery and probe have same FPS

the remaining 40% for testing. The training data and testing data are mutually exclusive and the training and testing videos are from different recordings. Our experiment test the performance when FPS is equal to 2, 3, 4, 5, 7, 10, 15, 20, 30, 40, 50, and 60, respectively. For example, if the FPS is 25 and the video frame rate is 50, then only 2 frames are sampled per second. The verification accuracy is shown in Figure 5.

From Figure 5, several points can be observed. First, the verification accuracies on identical twins are above 0.90. Compared with the best accuracy in [14] where it is 0.82 using facial expression and the best accuracy using appearance in aforementioned section, our proposal shows a great improvement. This experimental results again verify the hypothesis that twins look similar but behave differently. Secondly, we can see three local maximum along the entire ranges of FPS rate: the first local maximum is at $FPS = 6$, then the second maximum is at $FPS = 9$, and the last one is around 30. These three peaks suggest that identical twins can be better recognized by talking profile with different different speeds (fast, medium, and slow). Three ROCs are for these three different sample rates in Figure 3 clearly demonstrates the superiority of our proposal against traditional appearance based approaches on identical twins.

5.4. Experiment 3: Performance with Different Sample Rate

In this experiments, we consider the scenario when the probe and gallery have different FPS. Among the gallery, we assume the FPS for all videos is the same. It is practical in real applications because gallery is usually pre-collected. Our motivation to use different FPS rates between probe and gallery is to test the

Performance \ FPS in Gallery	2	3	4	5	6	7	8	9	10
Average accuracy	0.321	0.407	0.482	0.756	0.738	0.753	0.619	0.735	0.550
Variance	0.096	0.0951	0.069	0.031	0.073	0.022	0.081	0.071	0.069

Table 1: Experiment 3: performance with different sample rate between probe and gallery

robustness of our algorithm when the video frame rate is not fix. For example, the gallery video may be captured by a 50fps camera, while the probe video may be captured by 100fps camera. Given a FPS in gallery, we compute the average accuracy of various FPS in probe to evaluate the performance. For example, if the FPS in the gallery video is 2, then we test the accuracy for each FPS except 2 in the probe video, and use the average accuracy to represent the performance. The performance in such setting is presented in Table 1. The variance distance between the largest and the smallest is also listed. In terms of accuracy, for the identical twins database, the best performance can be achieved when the FPS in gallery is 5 or 7. When the FPS in gallery is either too large or small, the performance degrades significantly. We think the reason is because if the FPS is in the middle, the local motions in the gallery video at least have some overlap with the local motions in the probe video, otherwise the local motions in the gallery would jump too much or too little.

5.5. Experiment 4: Performance of Single Type of Motion

It would be interesting to see the individual discriminating ability of each type of face motion in talking profile to distinguish between identical twins. Hence, we conducted experiments with the same FPS in gallery and probe videos. Various FPSs, such as $FPS = 2, 3, 4, 5, ..., 10$ are tested and we compute the average performance to evaluate the discriminating ability of each type of facial motions on same FPS settings. We also use the average accuracy of each single type of face motion for evaluation. The final result is shown in Tab 2. We can see that the best performance of single type of face motion *i.e.*, 2D in-plane face translation, is less than 0.50. This result shows that even though the individual discriminating ability of each type of facial motion is low, together they convey enough identity specific information for recognition, as shown in Experiment 1. The reason may be because identical twins may have different face motions but not restrict to one single type. For example, for pair A, their pose changes are different and gaze changes are same, while for pair B, it is reversed.

Performance \ Type of Motion	Face	Gaze	Eye	Mouth	Pose	Pupil
Average accuracy	0.447	0.331	0.377	0.324	0.324	0.324

Table 2: Experiment 4: Accuracy on twins database for each motion

Figure 6: Some examples from Youtube non-twin database

5.6. Discussion on Twins

Several points can be concluded from the experiments: 1) Twins can be distinguished by talking profiles. As shown in our first experiment, our algorithm obtains the best accuracy to recognize identical twins up to now. This conclusion verifies the observation that parents of twins prefer to use the motion of their children for recognition. The proposal in this work presents a new way to recognize identical twins, as suggested in previous research [5, 6, 11].

2)Though the proposed talking profile can provide enough identity information for twin recognition, each type of face motion in talking profile lacks such discriminating power. This proves that there exist some differences of face motions between identical twins, but those differences only occurs on some types of face motions and such difference may be subject dependent.

3)Synchronization of motion sequence is also an important factor affecting recognition performance. Different FPS can significantly degrade the accuracy, as seen in experiment 1 and 2. With same FPS, the accuracy can be as high as 0.90, while for different FPS, it reduces to at best, around 0.70, when FPS in probe is 5 or 7.

5.7. Youtube Non-twin Database

Besides the identical twins database, we further investigate the possibility of using talking profiles for non-twin population. To verify it, we collected a moderate database from Youtube. It contains 99 subjects in 228 clips of 45 seconds each. These videos only have a single person talking. The quality of the video ranges from medium to high due to the variation of webcams. The person is either sitting or standing still and the environment can be either indoor or outdoor

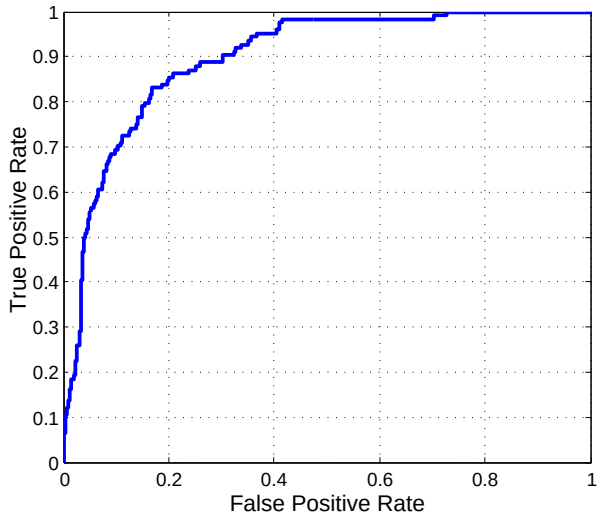

Figure 7: ROC of EMRM on Youtube Non-twin database

without controlled lightings. These type of videos range from speeches, technical talks to interviews. The gender and ethnicity of the speakers are diverse. Some examples from this database are shown in Figure 6.

To conduct experiment, we use 60% of the videos for training, and 40% for testing. The training and testing are mutually exclusive. Based on our preliminary works on 20 subjects, the sample rate equal to 5 is a good balance between efficiency and accuracy. Therefore, we use sample rating equal to 5 in the following experiments. The verification accuracy for Youtube Non-twin database is around 0.87, the corresponding ROC curve is shown in Figure 7.

Although our focus in this work is on distinguishing between identical twins, this experiment on YouTube Non-twins dataset indicates the potential of talking profile to be used as a biometric for generic (non-twin) populations. We also conduct the experiment to investigate the performance of each type of face motion on Youtube dataset, similar to the experiment 4. The accuracies for face 2D in-plane translation, gaze change, pose change, pupil movement, eye open-close and mouth open-close magnitude are 0.45, 0.34,0.32,0.32,0.45 and 0.40. The result is also consistent with identical twins database. From this experiment, we can see that even though single type of facial motion cannot provide enough signature information for recognition, their combination can be used for recognition.

Figure 8: Examples of Youtube Cross Age database

5.8. Youtube Cross-Age Database

To test the robustness of talking profile against time, we further collected a moderate database from Youtube. The videos of the subjects are all cross many years. An example is shown in Figure 8. There are in total 10 subjects in the database due to the difficulty of collecting.

In our experiments, we also set 60% of the data as training and remaining 40% as testing. The subjects are mutually exclusively. The accuracy is 0.715 when the sampling rate is set to 5. The corresponding ROC curve is shown in Figure 9. From this experiment, we can see that even though the time degrades the performance of our talking profile from 0.87 to 0.715, our talking profile can still provide some signature identity information through years. This further verified that our proposed model is invariant to the aging process. We will conduct experiments on larger dataset in future to further validate the robustness of our approach for unconstrained verification in the general population

6. Conclusion

Distinguishing between identical twins is a challenging problem in face recognition. In this paper, we verify that the talking profile can be used to distinguish between identical twins. To use talking profiles, we proposed a framework, EMRM, to effectively use identity-related abnormalities in face motions, with explicit focus on temporal information in the motion sequences. The experimental results on 2 databases, collected under free talking scenario, verified the robustness of our algorithm with both fixed and variable frame rates. We also suggest the most discriminating face motion type and best gallery video sampling rates for archival to achieve best performance for twin and non-twin subjects. Finally,

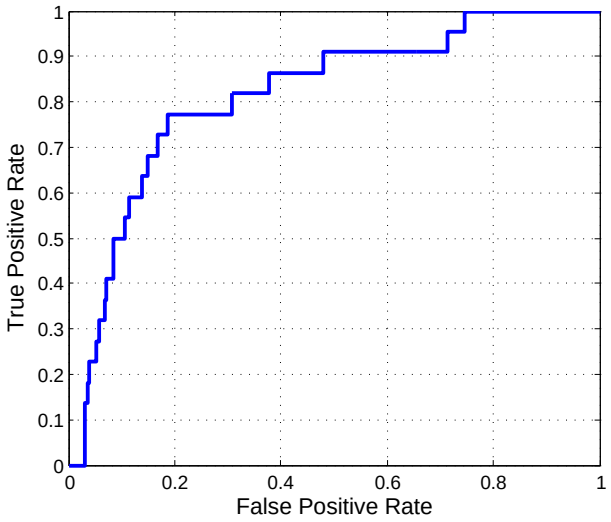

Figure 9: ROC of EREM for Youtube Cross-Age database

our results on the YouTube non-twins database shows potential of talking profile to be used for general subject recognition.

For future works, several points need more efforts. Firstly, at current stage, during data collection setting, we are encouraging the subjects to perform face and head motion in a natural free talking, while discouraging face and head motion with some specific task, such as slapping one's face. In such way, the data mainly contains the natural motions instead of accidental motions. In practical scenario, it is very important yet very challenging to separate the natural motion with accidental motion. Secondly, the video length in our current database is around 45 to 60 seconds. We set such length via the empirical observation that enough face and head motion have been captured. It is still desired to analysis the optimal video length for our algorithm. Thirdly, though the current experimental results are promising, we would investigate the scalability of our model on even larger dataset, to explore the scalability and stability of face motion features. Fourth, we will test the stability of talking profiles with increased time intervals. Finally, we will explore if the accuracy can be boosted by cascading face motions at different sample rates.

[1] J. Martin, H. Kung, T. Mathews, D. Hoyert, D. Strobino, B. Guyer, S. Sutton, Annual summary of vital statistics: 2006, Pediatrics.

[2] A. Jain, S. Prabhakar, S. Pankanti, On the similarity of identical twin fingerprints, Pattern Recognition 35 (11) (2002) 2653–2663.

[3] A. Kong, D. Zhang, G. Lu, A study of identical twins' palmprints for personal verification, Pattern Recognition 39 (11) (2006) 2149–2156.

[4] J. Daugman, C. Downing, Epigenetic randomness, complexity and singularity of human iris patterns, Proceedings of the Royal Society of London. Series B: Biological Sciences 268 (1477) (2001) 1737.

[5] Z. Sun, A. Paulino, J. Feng, Z. Chai, T. Tan, A. Jain, A study of multibiometric traits of identical twins, SPIE.

[6] P. Phillips, P. Flynn, K. Bowyer, R. Bruegge, P. Grother, G. Quinn, M. Pruitt, Distinguishing identical twins by face recognition, in: FG 2011.

[7] H. Hill, A. Johnston, Categorizing sex and identity from the biological motion of faces, Current Biology 11 (11) (2001) 880–885.

[8] S. Tulyakov, T. Slowe, Z. Zhang, V. Govindaraju, Facial expression biometrics using tracker displacement features, in: Proc. CVPR, 2007. doi:10.1109/CVPR.2007.383394.

[9] K. S. Pilz, I. M. Thornton1, H. H. Bülthoff, A search advantage for faces learned in motion, Experimental Brain Research 171 (2006) 436–447.

[10] M. Unnikrishnan, How is the individuality of a face recognized?, Journal of theoretical biology 261 (3) (2009) 469–474.

[11] B. Klare, A. Paulino, A. Jain, Analysis of facial features in identical twins, in: Biometrics (IJCB), 2011 International Joint Conference on, IEEE, 2011, pp. 1–8.

[12] I. M. Thornton, Z. Kourtzi, A matching advantage for dynamic human faces, Perception 31 (2002) 113–32.

[13] Y. Ning, T. Sim, Smile, youre on identity camera, in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, 2008, pp. 1–4.

[14] L. Zhang, N. Ye, E. Marroquin, D. Guo, T. Sim, New hope for recognizing twins by using facial motion.

[15] G. Huang, M. Mattar, T. Berg, E. Learned-Miller, et al., Labeled faces in the wild: A database forstudying face recognition in unconstrained environments.

[16] T. Moon, The expectation-maximization algorithm, Signal Processing Magazine, IEEE 13 (6) (1996) 47–60.

[17] T. Cover, J. Thomas, Elements of information theory, Wiley, New York, 1991.

[18] C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (2011) 27:1–27:27.

[19] M. Turk, A. Pentland, Face recognition using eigenfaces, in: Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on, IEEE, 1991, pp. 586–591.

[20] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, Computer Vision-ECCV 2004 (2004) 469–481.

[21] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, Image processing, IEEE Transactions on 11 (4) (2002) 467–476.

[22] M. S. Bartlett, J. R. Movellan, T. J. Sejnowski, Face recognition by independent component analysis, Neural Networks, IEEE Transactions on 13 (6) (2002) 1450–1464.

[23] W. Deng, Y. Liu, J. Hu, J. Guo, The small sample size problem of ica: A comparative study and analysis, Pattern Recognition.

[24] X. Niyogi, Locality preserving projections, in: Advances in neural information processing systems 16: proceedings of the 2003 conference, Vol. 16, The MIT Press, 2004, p. 153.

[25] S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, ECCV (2008) 504–513.