

New Hope for Recognizing Twins by Using Facial Motion

Li Zhang¹, Ning Ye², Elisa Martinez Marroquin³, Dong Guo⁴, Terence Sim¹

¹ School of Computing, National University of Singapore, ² Bioinformatics Institute, Singapore

³ La Salle Universitat Ramon Llull, Spain, ⁴ Facebook

{lizhang,tsim}@comp.nus.edu.sg, yening@bii.a-star.edu.sg, elisa@salleurl.edu, dnguo@fb.com

Abstract

Distinguishing between identical twins is the Holy Grail in face recognition because of the great similarity between the faces of a pair of twins. Most existing face recognition systems choose to simply ignore it. However, as the population of twins increases quickly, such an "ostrich strategy" is no longer acceptable. The biometric systems that overlook the twins problem are presenting a serious security hole.

Inspired by recent advances in motion-based face recognition techniques, we propose to use facial motion to address the twins problem. We collect a twins facial expression database and conduct a series of experiments in two assumed scenarios: the Social Party Scenario and the Access Control Scenario. The experimental results show that facial motion outperforms facial appearance in distinguishing between twins. Based on this finding, we propose a two-stage cascaded General Access Control System, which combines facial appearance with facial motion. The experimental results show that, compared with an appearance-based face recognition system, this cascaded system is much more secure against an "evil-twin" imposter attack, while performing as good for normal population.

1. Introduction

The incidence of twins has progressively increased in the past decades. Twins birth rate has risen to 32.2 per 1000 birth with an average 3% growth per year since 1990 [9]. Since the increase of twins, identical (monozygotic) twins are becoming more common. This, in turn, is urging biometric identification systems to accurately distinguish between twin siblings. Failing to identify them is a significant hindrance for the success of biometric systems.

Identical twins share virtually 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 ad-

vantage of this fact and have shown encouraging results in automatic recognition systems that use these discriminating traits: fingerprint [5, 13], palmprint [6], iris [2], speech [1] and combinations of some of the above biometrics [14]. 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 is known to be a great challenge for face recognition systems. Different approaches have been successfully proposed for general 2D face recognition and just recently their performance on twins has been questioned. Sun *et al.* [14] pioneered the analysis of face recognition on a database of identical twins. They tested the performance of an appearance-based method and showed that the match score distributions from twin imposters and genuine subjects had a large overlap, which made it difficult to distinguish between them. Phillips *et al.* [10] completed the analysis and presented the most extensive investigation of still-image face recognition performance on twins to date. They confirmed the difficulties encountered by state-of-the-art still-image face recognition systems to distinguish between twins and suggested looking for new research ideas to help improve the performance.

Knowing the difficulties that still-image based face recognition methods encounter in distinguishing between twins, we propose to use motion-based facial features to address the identical twins problem. Our proposal is leveraged on three facts: 1) Humans use facial motion as an additional cue for identification [12]. 2) Lifestyle affects the idiosyncratic facial expressions [3]. 3) Dynamic facial differences are empirically observed between twins performing facial expressions.

Psychologists have shown that faces with changing facial expressions are significantly easier to recognize by humans than static images [7, 12, 15]. Furthermore, research in psychology conducted by Fraga *et al.* reports that identi-

cal twins may exhibit differences associated with different environments and lifestyles [3]. In addition, during the collection of our database, we observed that parents used this principle to distinguish their twin children.

The goal of our research is to measure the capability of motion-based face recognition methods to distinguish between identical twins and to propose a solution to improve the performance of existing face verification systems against twin impostors. Many face verification systems have been successfully implemented and deployed in recent years. It is interesting to protect them against twin impostors while preserving their core structure. In this paper we propose to couple existing appearance-based verification systems with a motion-based module.

We review the state-of-the-art in motion-based face recognition and evaluate the two methods that can perform on various types of expressions: face recognition by tracked displacement features [16] and face recognition by using local deformation feature [17]. To the best of our knowledge, we are **the first** to use facial dynamics from expressions to distinguish identical twins. Our results demonstrate that motion-based face recognition can distinguish between twin siblings in verification mode. This motivates us to propose the cascade of appearance-based face recognition and motion-based verification for detection of twin impostors. This cascade could be considered by commercial vendors to gain robustness to twin impostors in general face verification systems. Moreover, we present the first twins expressive database, with videos of all six basic facial expressions, which we plan to release for public use in the near future.

2. Related work

As face recognition systems continues to improve, they are required to perform in more extreme conditions or challenging situations. However, until recently, the ability to distinguish between identical twins has been overlooked and the evaluation of still-image face recognition systems has traditionally skipped the twins test [4]. To the best of our knowledge only two recent papers have completely analyzed the ability of 2D still-image face recognition algorithms to recognize identical twins [14, 10].

Sun et al. [14] evaluated the performance of using face to distinguish twins but also iris, fingerprint and a fusion of them. The tests were conducted on a database 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. Despite the large number of facial images for each person, only two (template and query) were used in the experiments, due to the high similarity of images taken over a very short time interval. Face recognition experiments were conducted using the FaceVACS commercial matcher. The result showed

that identical twin impostor distribution had a greater overlap with genuine distribution than general impostor distributions. Their literal conclusion was that identical twins are a real challenge to face recognition systems.

Phillips et al [10] have thoroughly extended the analysis of the performance of face recognition systems in distinguishing identical twins, with the aim of drawing a conclusive assessment of current possibilities and challenges. The experiments were conducted on a 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. Although video of the twins' faces was simultaneously recorded, the paper only reports experiments based on 2D face still photographs. Facial recognition performance was tested using three of the top submissions to the Still Face Track at Multiple Biometric Evaluation 2010 [4]. Their main contribution is a detailed covariant analysis of static face recognition methods for identical twins. The best performance was observed under ideal conditions (same day, studio lighting and neutral expression). However, under more realistic conditions, they conclude that the problem is very challenging and requires new research ideas to help improve the performance.

Still-image face recognition systems rely only on the appearance in a static image and thus find tremendous difficulties in distinguishing identical twins, which look very much alike. Psychological studies have shown that humans better recognize faces with expressive motion [7]. It has been observed that when moving-expressive faces are used for training, not only the recognition rates increase [15, 12] but also the reaction time is reduced [11]. Importantly, it is not simply an effect of additional samples [12].

Inspired by these findings, computer vision researchers have raised different proposals for motion-based face recognition. They compute either a dense optical flow or sparse displacement on tracked points and use these motion estimations to identify the human subject. Most of them require the probe to perform a particular expression. To the best of our knowledge only the works by Tulyakov et al [16] and Ye and Sim [17] perform on different types of expressions. We will focus on them in this paper, since we want to evaluate different expressions to distinguish twins.

3. Data and Method

3.1. Data

We collected a twins facial expression database at the Sixth Mojiang International Twins Festival during Mayday 2010 in China. It includes Chinese, Canadian and Russian subjects summing a total of 27 pairs of twins. An example can be seen in Figure 1. For each subject, we recorded three still images and twelve video clips, two for each of the six



(a) Twins Pair 20, elder (b) Twins Pair 20, younger

Figure 1. An example of identical twins

basic facial motions (i.e. expressions): joy, anger, surprise, sad, fear and disgust. A Sony HD color video camera was used. We did not constrain the face position so that expression of each participant was realistic, thus there was some head motion and pose change in each video clip. We are preparing the database to become a public testbed.

3.2. Experimental Scenarios

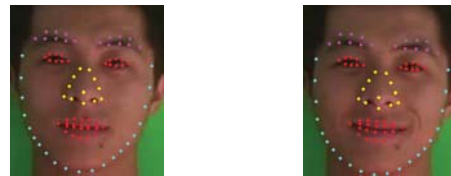
The tests are conducted in experiments that resemble two real life scenarios: Social Party Scenario and the Access Control Scenario. In the Social Party Scenario, imagine you attend a party where you are introduced to a pair of identical twins (who look and dress alike). After dinner, one of the twins strikes up a conversation with you, leaving you guessing who you are speaking to. In the Access Control Scenario, one twin sibling is an authorized user while the other is not. The security challenge is to grant access to the right twin and deny access to the twin impostor, without prior knowledge of the existence of a twin sibling. In the Social Party Scenario, the system has to distinguish between two twin siblings (knowing that they are twins) hence it only needs an intra-class comparison to get the result with the highest score (i.e. without any pre-set threshold); in Access Control Scenario, the system has to compare the score with a pre-set threshold and decide whether the probe is genuine or not.

3.3. Algorithm

We conduct several experiments on the twins facial expression database so as to evaluate motion-based face recognition in distinguishing identical twins. We test two motion-based algorithms on twins: simple sparse displacement algorithm which uses sparse displacement as a feature [16] and dense displacement algorithm which uses dense displacement and deformation on the entire face as a feature [17]. We choose these algorithms because they are the only two that can perform motion-based face recognition on different facial expressions, as far as we know. Besides, we use a state-of-the-art appearance-based face recognition software Luxand [8] for comparison purpose. Note all algorithms are run in pairwise verification mode. In the following sections, we give an overview of two motion-based face recognition algorithms.

3.3.1 Sparse Displacement Algorithm (SDA)

We track several key points at the neutral and apex of different expression face. Then we calculate the displacement of these points and regard this displacement vector as feature. Different from original method, in our implementation, since there is slight global head motion and pose change, we first use the positions of two eye-center¹ to align the face and track 87 landmark points automatically in neutral and apex of expressions as in Figure 2. Then we put each point's displacement together as the feature vector and normalize this vector by its L_2 norm. The pairwise score will be the Euclidean distance of two displacement vectors. We conduct the experiment on our twins database for all six basic expressions. This method is very sensitive to the repeatability of human expressions and can be applied in fix expression (i.e. gallery and probe must share the same expression). We call this method SDA (Sparse Displacement Algorithm) for abbreviation.



(a) Landmarks in neutral face (b) Landmark in apex of smile
Figure 2. Landmark Points Tracking

3.3.2 Dense Displacement Algorithm (DDA)

We track the face along the video and use eye-center position for alignment. Then we warp the face to meanface and construct the deformation feature as in [17]). Then we perform pairwise verification and compute the verification score as the weight summation of deformation feature similarity. We name this method DDA (Dense Displacement Algorithm) for abbreviation.

3.4. Performance Evaluation

In Social Party Scenario, the performance is evaluated by accuracy, which is the ratio of making correct guesses about who-is-who between a pair of twins. In Access Control Scenario, we evaluate the performance in terms of Twins-EER (Twins Equal Error Rate). Twins-EER is where FRR (False Reject Rate) meets Twins-FAR (Twins False Accept Rate). Twins-FAR is the ratio of misrecognizing one of the twins as the other.

4. Experimental results

We conduct the experiments in two scenarios and compare the performance of Luxand, SDA and DDA in each

¹These points are tracked by commercial software from Omron Corporation

scenario, separately. In the experiments, the tests on SDA and DDA generally use more samples than the tests on Luxand do. The reason is that Luxand uses static neutral face images, usually one or two per subject, while SDA and DDA use various kinds of facial expressions, up to six per subject in our dataset. The numbers of testing samples for SDA and DDA are also slightly different, due to their different input requirements.

4.1. Social Party Scenario

For Luxand, for each pair of twins we enroll two neutral images in gallery, one from the elder and the other from the younger. The similarity score of Luxand varies from 0 to 1.0. If the score for genuine is larger than for twin-imposter, it is a hit guess, otherwise it is a miss. We test 150 guesses in total and we hit 120 times. Thus, the accuracy of Luxand is 0.8. For SDA, we get 324 times guess in total and we hit 275 times, the overall accuracy is 0.85. For DDA, we get 360 times guess and hit 311 times, the overall accuracy is 0.864. Moreover, we list the accuracy of each expression both in SDA and DDA in Table 1.

	Smile	Anger	Surprise	Sad	Fear	Disgust
SDA	0.833	0.889	0.907	0.778	0.889	0.800
DDA	0.933	0.917	0.857	0.917	0.839	0.768

Table 1. SDA and DDA performance in each expression

4.2. Discussion of Social Party Scenario

Several points can be concluded from above experiment:

1) Motion performs better than pure appearance methods in Social Party Scenario. The accuracy of Luxand is 0.8, while the accuracy of SDA and DDA are 0.85 and 0.864, respectively. We can see that for nearly all expressions SDA and DDA perform better than Luxand, except sad in SDA and disgust in DDA. Note that Luxand is run under ideal conditions and only frontal neutral views.

2) In some particular expressions, for example surprise in SDA and smile in DDA, the accuracies are 0.907 and 0.933, which are both more than 10% higher than 0.8 in Luxand. This suggests us that we can find a very good feature in particular expressions to distinguish twins.

3) On the whole, DDA perform better than SDA, since the overall accuracy is 0.864 in DDA and 0.85 in SDA. And in some particular expressions, smile, anger and sad, DDA performs dramatically better than SDA, because the average accuracy of SDA in these three expressions is 0.833, while the average accuracy of DDA in these three expressions is 0.922. However, DDA needs more computation and requires more stable displacement tracking, because DDA extracts dense displacement from each pixel rather than sparse displacement.

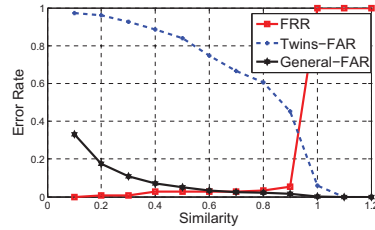
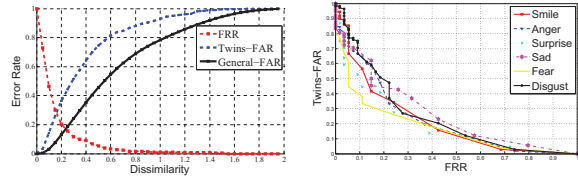
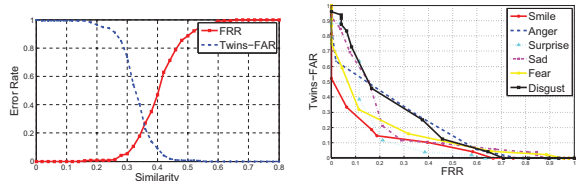


Figure 3. Performance of Luxand



(a) FRR-TwinsFAR of SDA (b) FRR-TwinsFAR of each expression in SDA

Figure 4. Performance of SDA in Access Control Scenario



(a) FRR-TwinsFAR of DDA (b) FRR-TwinsFAR of each expression in DDA

Figure 5. Performance of DDA in Access Control Scenario

4.3. Access Control Scenario

In Luxand, there are in total 150 genuine and 150 twins-imposter scores. Figure 3 shows the FRR-TwinsFAR curve. The Twins-EER is around 0.35 (where the two curves meet). The big jump in Figure 3 is due to the high similarity between images of the same subject. There are almost no lighting or pose changes between the images, because the images were taken in a short interval. Hence all the similarity scores for genuine are very close to one in the experiments. In SDA, there are 324 genuine distances, 648 twins-imposter distances in total. Figure 4(a) shows FRR-TwinsFAR curve. The overall Twins-EER is 0.28, which is 20% lower than the Twins-EER of Luxand (0.35). Figure 4(b) shows the FRR-TwinsFAR curves in SDA for individual facial expressions. The Twins-EERs from different facial expressions vary between 0.259 and 0.31 with facial expression of surprise giving the lowest. In DDA, we test 360 twins-imposter and 360 genuine in total. Figure 5(a) shows FRR-TwinsFAR curve. The overall Twins-EER is 0.25, which is significantly better than SDA and Luxand. Figure 5(b) shows the FRR-TwinsFAR in DDA for individual expressions. The facial expression of smile gives the

lowest Twins-EER (0.18).

4.4. Discussion of Access Control Scenario

Firstly, motion performs better than Luxand to distinguish identical twins in terms of Twins-EER in Access Control Scenario. Secondly, between two different motion-based face recognition methods, DDA can achieve better result than SDA to distinguish between twins in access control scenario, not only in the overall Twins-EER but also the best Twins-EER of individual expression (smile in DDA, surprise in SDA). However, DDA is more computation expensive and sensitive to tracking algorithm, which may be the reason why the worst Twins-EER of individual expression to DDA is worse than the worst expression to SDA (sad in SDA, anger in DDA).

5. General Access Control System

We consider the General Access Control System (ACS) taking into account genuine, twins-imposter and general-imposter. This is the most general case, where the system does not require the probe twin’s feature in the gallery or prior knowledge of the existence of twins.

5.1. Performance of Appearance or Motion Only

We investigate the performance of Luxand and SDA in General Access Control System. Besides Twins-EER, we also evaluate the performance in terms of General-EER (general equal error rate). General-EER is where General-FAR and FRR meet. General-EER is the ratio of misrecognizing the person as others who are not his/her twin sibling. The overall EER is the combination of Twins-EER and General-EER. In Luxand, we compute 4200 general-imposters scores. In SDA, we compute 33696 general-imposters scores in six expressions. We choose SDA instead of DDA due to its high efficiency of computation.

Figure 3 shows that we cannot choose a threshold that can distinguish both general-imposter and twins-imposter well at the same time in Luxand. If the threshold is set to around 0.5, General-FAR is less than 0.03 while Twins-FAR is as high as 0.7, which means that with this threshold Luxand is secure to general imposter attacks but fails in "evil-twin" imposter attacks. If Luxand is tuned to distinguish between twins, then the Twins-EER is around 0.35 with a threshold set at 0.9, Luxand is secure to "evil-twin" imposter attacks, but fails in general imposter attacks. On the other hand, Figure 4(a) shows that the threshold for Twins-EER is 0.154, which is not that far from the threshold (0.236) for General-EER. We can best avoid general-imposter attacks and "evil-twin" attacks at the same time. However, we wanted to further explore the possibility of improving the performance, trying to get closer to the General-EER of 0.03 that Luxand shows when only general population (non-twins) is considered. Please note in Figure 3 the

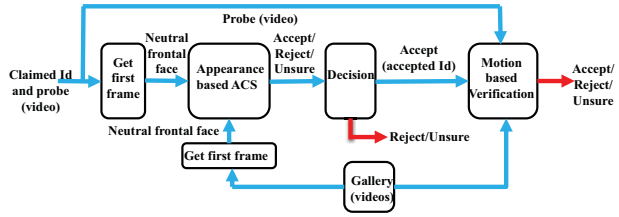


Figure 6. Cascade for General Access Control System

horizontal axis is similarity score while in Figure 4(a) it is dissimilarity score (the Euclidean Distance of two displacement).

5.2. Proposal of a Cascade Approach

We propose to use a cascade system that combines a motion-based verification module after an appearance-based module, as shown in Figure 6. It retains the best of both approaches; appearance-based can better avoid general imposter attacks, while motion-based verification better avoids "evil-twin" attacks.

This General ACS works as following: 1) Probe input is now a video of facial expression. 2) A pre-processor extracts first frame of neutral face from video and gives it (together with the claimed identity) to the first stage appearance-based verifier. 3) If first stage appearance-based verifier outputs "reject", then final decision is "reject". There is no need to go to second stage motion-based verifier. 4) If first stage appearance-based verifier outputs "accept", then claimed id and video will be given to second stage motion-based verifier. Final decision is the decision of the second stage motion-based verifier. 5) If first stage appearance-based verifier outputs "unsure" then final decision is "unsure". There is also no need for second stage.

Several points shall be addressed:

1) Our cascade scheme works without requiring the twin’s features in the gallery and without knowing the existence of twins as a prior.

2) Our cascade scheme cannot worsen overall FAR, since at most the second stage can say "accept" when the first stage says "accept". Moreover, our scheme improves FAR, this occurs when first stage accepts imposter no matter it is general imposter or twin imposter and second stage rejects these imposter correctly.

3) Our cascade scheme could worsen FRR, this occurs when first stage says "accept" correctly but second stage says "reject" incorrectly. However, since Access Control Systems usually require low FAR and can tolerate high FRR, our scheme can’t hurt, but can in fact help against attacks from twin imposters.

4) Our cascade scheme is based on current appearance-based Access Control System, thus existing Access Control Systems can be easily augmented to use our cascade scheme without much modification.

To confirm the potential of our proposal, we conduct a simple experiment. In this experiment, we store smile videos from 15 pairs of twins in the gallery. When an unknown human subject comes as probe, we record his/her smile video clip. The first neutral face frame is extracted as input to the appearance-based module. In the first stage appearance-based verifier, Luxand is used. In the second stage, motion-based verifier, DDA is used. We set the threshold of Luxand to 0.8 such that we can strictly exclude general imposters. We set the threshold of DDA to 0.36 corresponding to Twins-EER of DDA such that we can exclude twin-imposter. We verify ten human subjects in total and each for 5 times (one time for claiming genuine, one time for claiming twin-imposter and the remaining three times claiming for general-imposter). We compare our cascade approach with Luxand. There are nine "evil-twins" imposter attack recognized as genuine in Luxand, while there is only one "evil-twin" imposter attack recognized as genuine in our cascade scheme. This result proves that our cascade scheme can be more secure to "evil-twin" imposter attack. However, as we explain above, our cascade scheme may worsen FRR, in our experiment there is one genuine recognized as imposter attack. In a real system, if such situation happens, we just need to repeat the genuine probe until it is recognized correctly.

In summary, our cascade 1) keeps the General-FRR and General-FAR (for general population, non-twins) as good as it is in the appearance-based method, since the threshold is tuned for general population; 2) improves the overall FAR by improving the Twins-FAR at the motion-based verifier and 3) provides protection to twin imposter attacks even if the system does not know that any subject in the gallery has a twin. This is because our method learns the distribution of twin and non-twin feature similarities (distances) separately, rather than the distribution of the features themselves.

6. Conclusions and Future Work

We have proposed facial motion to distinguish identical twins. Our goal is to analyze the discriminative power of different facial expressions, thus we have analyzed the performance of the two motion-based face recognition algorithms that perform with different expressions. We have appraised the comparison with state-of-the-art still-image face recognition methods. Although two still-image face databases for twins have been reported before, to our best knowledge we have collected the first expressive twins database.

On this moderate sized database, the experiments show the capability to identify identical twins in verification mode with best Twins-EER being 0.18. Based on the results, we have proposed a cascading of appearance-based verifier and motion-based verifier that reaches an accu-

racy of 0.96. The experiments have measured the effect of different facial expressions and have pointed out smile as the most discriminative expression for twins identification. Since motion-based feature is extracted from expression, the method is robust to different expressions. Moreover, as facial motion is becoming much easier to obtain, since video cameras have become ubiquitous, we believe that the proposed cascading could be considered by commercial parties to gain robustness to twin imposters in general face recognition systems. However, due to the limited size of the database, this research only provides a baseline evaluation. Future work will expand the size of the dataset.

7. Acknowledgement

Elisa Martinez acknowledges the support of the Mobility Program from the Spanish National R&D Plan 2008-2011.

References

- [1] A. Ariyaeeinia, C. Morrison, A. Malegaonkar, and S. Black. A test of the effectiveness of speaker verification for differentiating between identical twins. *Science & Justice*, 2008.
- [2] J. Daugman and C. Downing. Epigenetic randomness, complexity and singularity of human iris patterns. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 2001.
- [3] M. Fraga, E. Ballestar, M. Paz, S. Roperro, F. Setien, M. Ballestar, D. Heine-Suñer, J. Cigudosa, M. Urioste, J. Benitez, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 2005.
- [4] P. Grother, G. Quinn, and P. Phillips. Report on the evaluation of 2d still-image face recognition algorithms. *NIST*, 2010.
- [5] A. Jain, S. Prabhakar, and S. Pankanti. On the similarity of identical twin fingerprints. *Pattern Recognition*, 2002.
- [6] A. Kong, D. Zhang, and G. Lu. A study of identical twins' palmprints for personal verification. *Pattern Recognition*, 2006.
- [7] K. Lander, F. Christie, and V. Bruce. The role of movement in the recognition of famous faces. *Memory & Cognition*, 1999.
- [8] LuxandSDK. <http://www.luxand.com/facesdk>. Detect and Recognize Faces with Luxand FaceSDK.
- [9] J. Martin, H. Kung, T. Mathews, D. Hoyert, D. Strobino, B. Guyer, and S. Sutton. Annual summary of vital statistics: 2006. *Pediatrics*, 2008.
- [10] P. Phillips, P. Flynn, K. Bowyer, R. Bruegge, P. Grother, G. Quinn, and M. Pruitt. Distinguishing identical twins by face recognition. In *FG 2011*.
- [11] K. S. Pilz, I. M. Thornton1, and H. H. Bülthoff. A search advantage for faces learned in motion. *Experimental Brain Research*, 2006.
- [12] D. Roark, S. Barrett, M. Spence, H. Abdi, and A. O'Toole. Psychological and neural perspectives on the role of motion in face recognition. *Behavioral and cognitive neuroscience reviews*, 2(1):15, 2003.
- [13] S. Srihari, H. Srinivasan, and G. Fang. Discriminability of fingerprints of twins. *Journal of Forensic Identification*, 2008.
- [14] Z. Sun, A. Paulino, J. Feng, Z. Chai, T. Tan, and A. Jain. A study of multibiometric traits of identical twins. *SPIE*, 2010.
- [15] I. M. Thornton and Z. Kourtzi. A matching advantage for dynamic human faces. *Perception*, 31:113–32, 2002.
- [16] S. Tulyakov, T. Slowe, Z. Zhang, and V. Govindaraju. Facial expression biometrics using tracker displacement features. In *CVPR*, 2007.
- [17] N. Ye and T. Sim. Towards general motion-based face recognition. In *CVPR*, 2010.