Database Cross Matching: A Novel Source of Fictitious Forensic Cases

Abhishek Dutta Raymond Veldhuis Luuk Spreeuwers
Signals and Systems Group, EEMCS,
University of Twente, Netherlands.
{a.dutta,r.n.j.veldhuis,l.j.spreeuwers}@utwente.nl

Abstract

Due to privacy concern and data protection laws, it is very difficult to obtain real forensic data for forensic face recognition research. In this paper, we introduce the concept of Database Cross Matching (DCM) as a novel source of fictitious but challenging forensic cases. DCM refers to the task of finding the subjects that are common in two different data sets. For most pairs of independent data sets, there will be no common subjects. However, for some data sets captured at the same institution, but independently and at different times, there is a high probability of finding some common subjects. We demonstrate the feasibility of DCM using the PIE and MultiPIE data set that were captured at the same institution in 2000 and 2004 respectively. We denote the task of finding the subjects that are common in PIE and MultiPIE data as PIE $\cap$ MultiPIE problem. Evaluation of the five face recognition systems applied to the PIE $\cap$ MultiPIE problem show that DCM can indeed create very challenging forensic problems.

1 Introduction

Automatic face recognition systems have a great potential to become a reliable and robust forensic tool. Sufficiently large data sets simulating forensic setting are required to achieve this goal. Data collected from real forensic cases usually have limited number of samples per subject and therefore is not sufficient for research purposes which require a large number of images per subject under different setting (e.g. pose, illumination, age, etc) including a high quality sample as the ground truth. Moreover, acquiring real forensic data is very difficult due to privacy concerns and data protection laws.

In this paper, we introduce the concept of Database Cross Matching (DCM) as a novel source of fictitious forensic data. DCM starts by first determining the subjects (or participants) that are common in two different data sets: the ground truth. With the ground truth to hand, we can create fictitious but challenging forensic cases in which a poor quality trace is taken from one of the two data sets and a relatively better quality suspect reference set is created from another. The results reported by different face comparison schemes and algorithms can be compared against the ground truth to evaluate their performance in a forensic setting.

There are two ways to establish the ground truth for a DCM problem. First is to request the authors of the two data sets to compare their meta data (name, date of birth, etc) collected during data set capture and publish the ground truth without revealing the identity. Second, is to compare good quality facial images of subjects present in the two data sets using a state-of-the-art face comparison system.

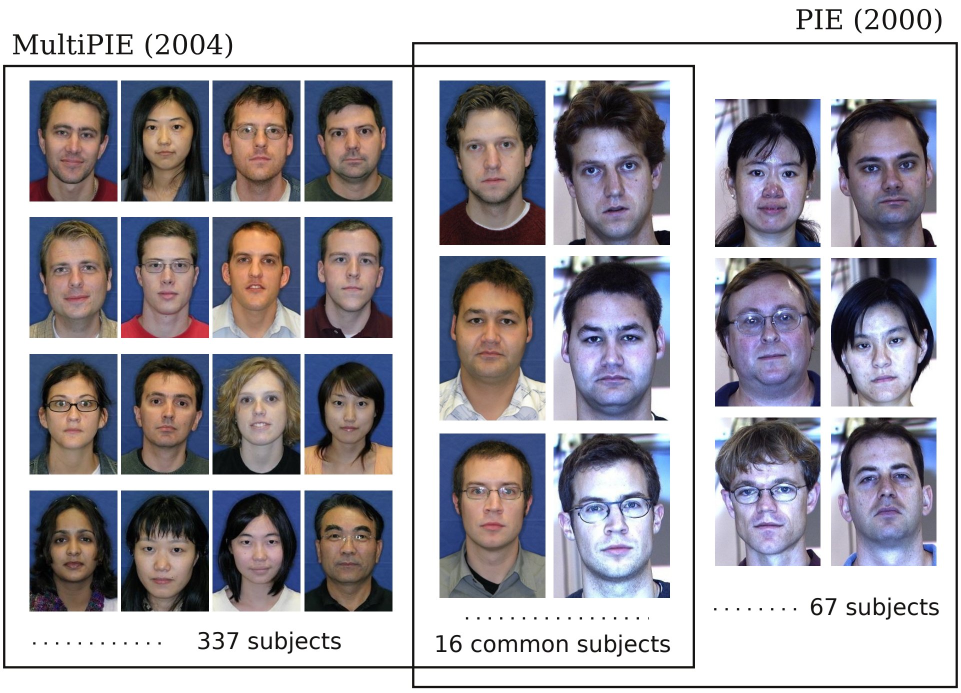


Figure 1: Illustration of Database Cross Matching on PIE and MultiPIE data sets captured in 2000 and 2004 respectively.

For most pairs of independent data sets, there will be no common subjects. However, for data sets captured at the same institution, but independently and at different times, there is a high probability of finding some common subjects. We demonstrate the feasibility of DCM using two data sets (PIE [6] and MultiPIE [3]) captured at the same institution in 2000 and 2004 respectively as shown in Figure1. We denote the task of finding the subjects that are common in the PIE and MultiPIE data sets as PIE $\cap$ MultiPIE problem.

2 PIE $\cap$ MultiPIE

The PIE [6] and MultiPIE [3] data sets were captured at the Robotics Institute, Carnegie Mellon University in 2000 (Oct. to Dec.) and 2004 (Oct. to Mar. 2005)

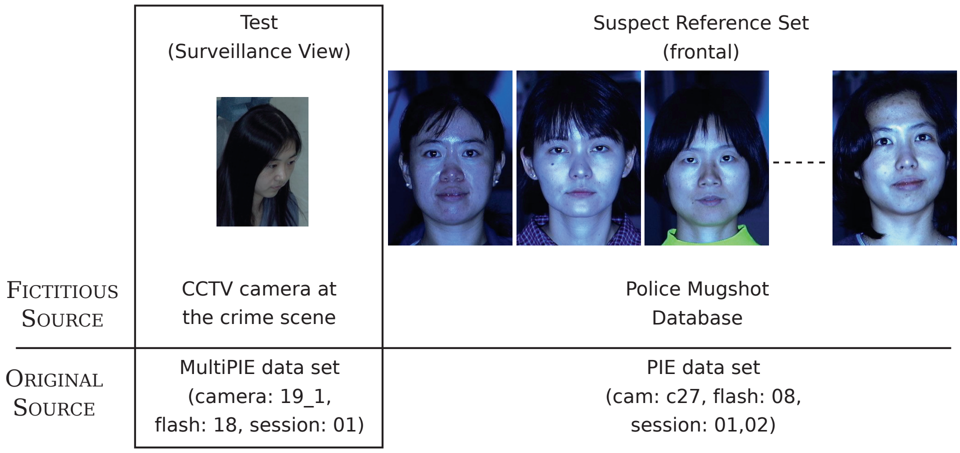


Figure 2: A sample of a fictitious forensic case in PIE $\cap$ MultiPIE problem.

respectively. So there is a high probability that some of the subjects of the PIE data set are also present in the MultiPIE data set.

In 2.1, we establish the ground truth for PIE $\cap$ MultiPIE by comparing frontal view and illumination images in the two data sets using two commercial face recognition system and then visually inspect the results. With the ground truth to hand, we can create a set of forensic evaluation cases involving face recognition. For instance: when the surveillance view test image is taken from the MultiPIE data set and a frontal view suspect reference set is taken from the PIE data set (as shown in Figure2), the forensic evaluation task is to determine if the test image subject is present in the suspect reference set.

Mathematically, the set of forensic evaluation cases in the PIE $\cap$ MultiPIE problem can be defined as follows: Let W_π and $W_{m\pi}$ be the set of all person-ids in the PIE and MultiPIE data sets respectively. This problem requires finding the mapping function $\psi(i)$ such that:

$$\psi(i) = \begin{cases} j \text{ , } W_{m\pi}(i) \text{ and } W_\pi(j) \text{ denote same subj.,} \\ 0 \text{ , otherwise} \end{cases}$$

for $i = 1, \cdots, n(W_{m\pi})$ and $j \in [1, n(W_\pi)]$. We can evaluate the performance of different face comparison schemes and algorithms by comparing their respective mapping function $\psi_k(i)$ with the ground truth $\psi_0(i)$. For the case shown in Figure2, test image contains person `016` $\in W_{m\pi}$ and this subject is not present in the PIE data set. Therefore, $\psi_0(i = 016) = 0$ and all face comparison algorithms reporting otherwise are misleading. In 2.2, we report the performance of 5 face recognition systems applied to the PIE $\cap$ MultiPIE problem.

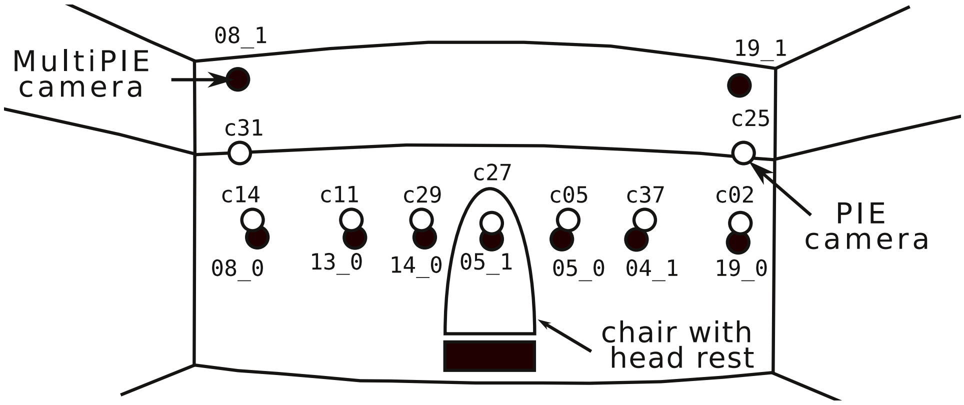


Figure 3: Position of the camera in the image capture environment of the PIE [6] and MultiPIE [3] data set. Note: the separation between cameras at head height is $15°$ and $22.5°$ in the MultiPIE and PIE capture setup respectively.

2.1 Ground Truth for PIE $\cap$ MultiPIE

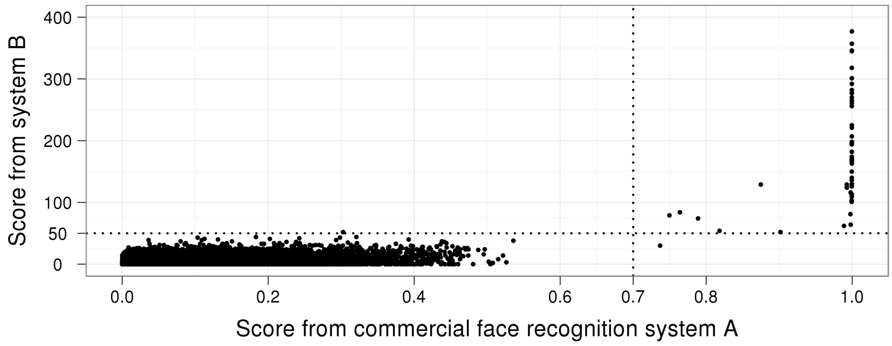
In this section, we establish the ground truth for the PIE $\cap$ MultiPIE problem by comparing images with frontal view and illumination in the PIE (camera c27, see Figure3) and MultiPIE (camera 05_1, see Figure3) using two commercial face recognition systems and then visually inspecting the results. The ground truth refers to the true mapping $\psi_0(i)$ of common subjects in the two data sets. It is required to assess the performance of different forensic face comparison scheme and algorithms.

We create the test and reference set as described in Table 4b and the corresponding similarity scores computed by the two commercial face recognition systems (denoted by A and B) is shown in Figure4a.

Figure4 shows the two clusters formed by the joint score($[x_A, x_B]$) of the system A and B. By visual inspection, it is evident that the following decision threshold $[x_A, x_B] \succeq [0.7, 50]$ can separate the two clusters, where $x \succeq y$ denotes componentwise inequality between vectors x and y. Therefore, the joint scores that satisfy this decision threshold is labeled as the genuine class, otherwise the joint scores are labeled as the impostor class. Based on this decision threshold, the genuine class instances (i.e. positive matches) are shown in Figure1 (only 3 shown for illustration) and the corresponding ground truth (ψ_0) for the PIE $\cap$ MultiPIE problem is tabulated in Table1.

2.2 Performance Results for PIE $\cap$ MultiPIE

In this section, we report the performance of the following five face recognition systems applied to the PIE $\cap$ MultiPIE problem: two commercial face recognition systems denoted by A and B, Local Region PCA (LR-PCA) and LDA - I/Red (LDA-IR) [5], and Local Binary Pattern (LBP) [1] where, PCA and LDA are holis-



(a) Joint similarity scores distribution

	Source	subjects	session	camera	flash	eye*
Test	MultiPIE	337	01-04	05_1	07	manual
Ref.	PIE	68	01,02	c27	f08	manual

(b) Properties of test and reference set. Note : flash f08 and 07 are frontal with respect to the face and manually located eye* coordinates were supplied to both A and B

Figure 4: Similarity scores of two commercial systems (A and B) for test and reference images used for establishing the ground truth (ψ_0)

tic methods while LBP is a local method. These systems are fine tuned for comparing frontal images and therefore direct comparison of surveillance view and frontal view images (as shown in Figure2) results in extremely poor performance. In order to avoid the complexities of a model based approach [2], we report performance results for the view based approach [4]. Therefore, our suspect reference image is chosen such that its pose closely matches the pose in the test set as shown in Figure6a (inset).

First, we determine the rank-1 recognition rate using a test set that only contains the 16 subjects (see Table1) common in the PIE and MultiPIE data set as shown in Figure5. This experiment only reflects the true positive rate. Therefore, in Figure6, we also show the ROC plot for these systems when the test set contains images from all 337 subjects in the MultiPIE data set.

2.3 Discussion

For the PIE $\cap$ MultiPIE problem, the best true positive rate (at false accept rate of 0.01) of ~ 0.68 was achieved by commercial system A. These results clearly indicate that PIE $\cap$ MultiPIE problem is indeed a challenging face comparison problem and current face recognition algorithms are not mature enough to be used in a forensic setting.

PIE $\cap$ MultiPIE problem simulate the following properties of a real forensic

Table 1: Ground truth $\psi_0(i)$ for the PIE $\cap$ MultiPIE

$W_\pi(\psi_0(i))$	$W_{m\pi}(i)$	$W_{m\pi}$ sessions $\in [1, 2, 3, 4]$			
04001	258		$\odot$		$\odot$
04006	001	$\odot$	$\odot$	$\odot$	
04008	007	$\odot$			
04009	104	$\odot$	$\odot$	$\odot$	$\odot$
04015	003	$\odot$	$\odot$	$\odot$	$\odot$
04016	097	$\odot$	$\odot$	$\odot$	$\odot$
04021	154	$\odot$	$\bigcirc$		
04025	013	$\odot$	$\odot$	$\odot$	
04026	002	$\odot$	$\odot$	$\odot$	$\odot$
04030	254		$\odot$		$\odot$
04037	128	$\odot$	$\odot$	$\odot$	$\odot$
04039	079	$\odot$	$\odot$	$\odot$	$\odot$
04041	022	$\odot$	$\odot$	$\odot$	$\odot$
04057	311			$\odot$	
04058	023	$\odot$	$\odot$	$\odot$	$\odot$
04069	085	$\odot$	$\odot$	$\odot$	$\odot$

$\odot$: present in k^{th} MultiPIE session and detected;
$\bigcirc$: present but not detected.

case: *a*) simulates open set recognition scenario (i.e. not all the individuals in the test set are present in the reference set); *b*) test and reference set images were captured about 4 years apart, by different cameras and in a different environment. It is important to mention that, in the proposed PIE $\cap$ MultiPIE problem, the test set always contains surveillance view images (Panasonic AW-E600P camera [3]) taken from the MultiPIE data set while the reference set contains images (Sony DXC 9000 camera [6]) taken from the PIE data set. This strategy ensures that the test and reference set images are always captured by different cameras.

We also report the case of duplicate enrollment under different names in the MultiPIE data set. While establishing the ground truth, we noticed that person-id `120,290` and `301,094` are the same individual appearing under different person-id in the MultiPIE data set. Visual inspection of high resolution photographs of these subjects further confirmed this fact. The MultiPIE data set authors[1] confirmed that "these individuals are indeed same subjects and unfortunately they failed to identify themselves as repeat subjects". Fortunately, these individuals are not among the 16 subjects common in the PIE and MultiPIE data set.

[1]based on email from Ralph Gross dated Mar. 22, 2012

	A	B	LR-PCA	LDA-I/R	LBP
Accuracy	0.62	0.08	0.02	0	0

(a) Rank-1 recognition rate

	source	subjects	# image	session	camera	flash
Test	MultiPIE	16	50	`01-04`	`19_1`	`18`
Ref.	PIE	68	68	`01,02`	`c25`	`f13`

(b) Properties of test and reference set used to determine rank-1 recognition rate.
Note: flash `f13` and `18` are frontal with respect to the face

Figure 5: Rank-1 recognition rate of five face recognition systems for PIE ∩ MultiPIE problem using a view based approach.

3 Conclusion

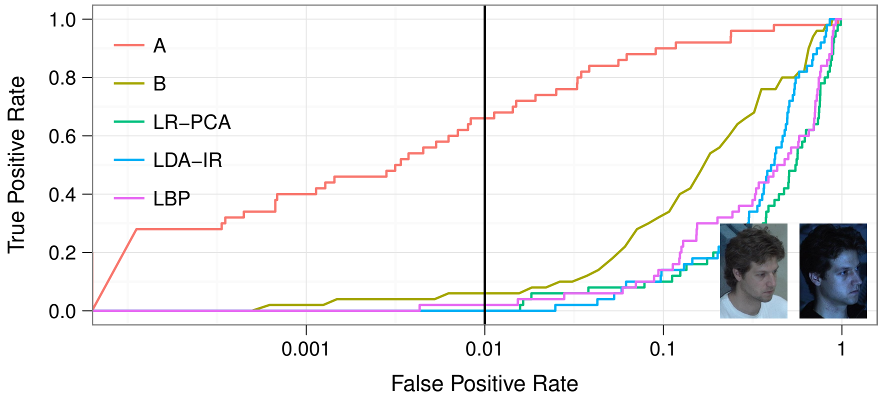
Due to privacy concern and data protection laws, it is very difficult to obtain real forensic data for research purposes. In this paper, we introduced the concept of Database Cross Matching (DCM) and demonstrated its feasibility using two facial data sets captured at the same institution but at different times: the PIE ∩ MultiPIE problem.

Future work needs to be done to establish whether it is possible to apply the concept of DCM in other biometric domains like fingerprint, speech, etc. It would also be interesting to explore other pairs of facial image data sets that fit the requirements of DCM. In addition, future work will also investigate the following non-forensic application of DCM:

- study the effect of database mismatch (difference between the database used to tune a face recognition system and the forensic data that the system has to operate on) in a forensic setting, and

- study the effect of aging and environmental factors on a biometric sample (for example: face appearance variation).

The use of Database Cross Matching as a source of fictitious forensic case has at least two limitations. First, the number of common subjects in two independent data set is usually very small. For example: in the PIE ∩ MultiPIE problem, there were only 16 subjects (with a total of 50 multiple session images for the 16 subjects) common in the two data sets. We require a large number of images to perform statistically significant tests of face comparison algorithms. Second, establishing true ground truth for such problems is very difficult because the authors of the original data sets are often reluctant to share meta data related to the subjects or even the ground truth due to privacy concerns.



(a) ROC for PIE ∩ MultiPIE problem

	Source	subjects	session	camera	flash	eye*
Test	MultiPIE	337	`01-04`	`19_1`	`18`	manual
Ref.	PIE	68	`01,02`	`c25`	`f13`	manual

(b) Properties of test and reference set used to determine the ROC

Figure 6: ROC plot of five face recognition systems applied to the PIE ∩ MultiPIE problem. Note: inset depicts sample test and reference image, x axis of the ROC plot is in log scale.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] V. Blanz. Face recognition based on a 3D morphable model. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 617–624, april 2006.

[3] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *Automatic Face Gesture Recognition, 2008. FG 08. 8th IEEE International Conference on*, pages 1–8, 2008.

[4] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR 94., 1994 IEEE Computer Society Conference on*, pages 84–91, jun 1994.

[5] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. OToole, D. S. Bolme, J. Dunlop, Yui Man Lui, H. Sahibzada, and S. Weimer. An introduction to the good, the bad, amp; the ugly face recognition challenge problem. In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 346–353. `http://www.cs.colostate.edu/facerec/algorithms/baselines2011.php`, 2011.

[6] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 46–51, 2002.