Forensic face recognition based on KDE and evidence theory

Wen Xiao[1,*]

[1]JiangXi Police College, NanChang, JiangXi, China

Abstract. Forensic face recognition (FFR) has been studied in recent years in forensic science. Given an automatic face recognition system, output scores of the system are used to describe the similarity of face image pairs, but not suitable for forensics. In this study, a score-mapping model based on kernel density estimation (KDE) and evidence theory is proposed. First, KDE was used to generate probability density function (PDF) for each dimensional feature vector of face image pairs. Then, the PDFs could be utilized to determine separately the basic probability assignment (BPA) of supporting the prosecution hypothesis and the defence hypothesis. Finally, the BPAs of each feature were combined by Dempster's rule to get the final BPA, which reflects the strength of evidence support. The experimental results demonstrate that compared with the classic KDE-based likelihood ratio method, the proposed method has a better performance in terms of accuracy, sensitivity and specificity.

1 Introduction

In the context of forensic science, face recognition approaches have fallen into two main categories: Subjective-based and objective-based methods. The subjective-based methods are the traditional forensic means in the past few decades, and face biometric features have been commonly used to inspect and compare static images for such methods [1,2]. Mainly four subjective-based methods can be used during the analysis and comparison phase [1]: Holistic Comparison, Morphological Analysis, Photo-anthropometry, and Superimposition. Facial Identification Scientific Working Group (FISWG) recommends morphological analysis by trained examiners as the primary method of comparison [3]. Moreover, in recent years, soft Biometrics, such as gender, age, race, skin colour, spots and other characteristics, have been considered into face recognition procedure so as to improve recognition results [4,5]. But these methods need to be manually carried out by forensic experts, so they heavily depend on the experience and knowledge of the forensic experts. On the other hand, the objective-based methods attempt to identify faces using automatic face recognition [6-10]. Using automatic recognition system to verify faces can not only improve the efficiency of forensic work, but also promote the standardization of the forensic process. In commercial face recognition systems, the similarity or distance between two faces is usually reported in terms of one or several score values, which is so called "Score-based procedures" [11]. In order to

[*] Corresponding author: shauven@126.com

take the differences and typicality of the population into account and allow for comparisons between facial scores from different face recognition systems, there is necessary to convert the score to Likelihood Ratio (LR) value [12]. Some organizations such as the European Network of Forensic Science Institutes (ENFSI), report the certainty of the statement match/nonmatch via a quantifiable amount, that is, verify whether it is the same person/different person or not [13]. To that end, ENFSI enforces the use of a LR value to evaluate the strength of evidence as the degree of supporting the appraisal conclusion, namely, it can be regarded as the mensurable method to express the confidence in the match/nonmatch decision [10]. A suitable approach to achieve this is to append such a score-to-LR mapping in a post-processing step to an existing score-producing facial recognition system [14]. Once a model for score-to-LR mapping has been set up, the strength of evidences can be obtained by plugging the scores into the model.

2 Evidence evaluation and evidence theory

Evidence evaluation has been proposed in recent years as a logical and appropriate way to report evidence to a court of law using a Bayesian probabilistic framework. LR is based on Bayes' rule, it is defined as the ratio of the probabilities of two hypotheses [15]: the null hypothesis of the prosecution (H_p), and the alternative hypothesis of the defense (H_d). The hypothesis of the prosecution H_p means that the evidences are from the same source, and the hypothesis of the defense H_d means that the evidences are from the different source. Then LR is obtained from two conditional probabilities, that is, the conditional probability of the prosecution hypothesis divided by the conditional probability of the defense hypothesis. So, the LR is defined as follow:

$$L\left(H_p, H_d, E\right) = \frac{\Pr\left(E|H_p\right)}{\Pr(E|H_d)} \tag{1}$$

In order to express the strength of evidence support, especially for convenient to communicate evidence values in the courtroom, it would be useful to translate the numerical expression to a verbal counterpart. One of the current frameworks to relate verbal and numerical likelihood ratios is defined as follows [16]:

Table 1. Relation of verbal and numerical LR.

LR range	Evidence to support H_p
1-2	no assistance
2-10	slightly more probable
10-100	more probable
100-10,000	much more probable
10,000 -1,000,000	far more probable
>1,000,0000	exceedingly more probable

2.1 Score-to-LR conversion model

When an automated system is used to calculate the similarity or the distance between the two faces to be compared, it returns a score. This score itself has no forensic relevance and needs to be converted to LR.

Four score-to-LR conversion models have been proposed [17]: Kernel Density Estimation (KDE), Linear Logistic Regression (LLR), Histogram Binning and Pool Adjacent Violators (PAV), where KDE is a commonly used method which is easy to explain. In KDE, a kernel distribution is a non-parametric representation of the probability density function (PDF) of a random variable. It is used when a parameter distribution cannot properly describe the data, or to avoid making assumptions about the data distribution. A kernel distribution is defined by a smoothing function and a bandwidth value h, which controls the smoothness of the resulting density curve. In other words, it is a technique that lets you create a smooth curve given a set of data. It is given by the following equation:

$$f_k(x;h,K)=\frac{1}{n}\sum_{i=1}^{n}K_h(x-x_i)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-x_i}{h}\right) \tag{2}$$

where K is the kernel and h is the bandwidth. The kernel smoothing function K defines the shape of the curve used to generate the probability distribution function, and the bandwidth h steers the smoothness of the resulting approximation. Gaussian distribution is usually used as kernel function, and the bandwidth could be adaptively generated from the sample data. Unlike a histogram, which places the values into discrete bins, a kernel distribution sums the component smoothing functions for each data value to produce a smooth, continuous probability curve.

Use the kernel function to estimate the data from the same source (the prosecution hypothesis H_p) and the data from different sources (the prosecution hypothesis H_d) respectively, LR is calculated with the following:

$$LR(s)=\frac{\Pr\left(s\middle|H_p\right)}{\Pr(s|H_d)}=\frac{f_p(s;h,k)}{f_d(s;h,k)} \tag{3}$$

2.2 Evidence theory

Evidence theory is the generalization of probability theory [18,19], which can handle uncertainty, impreciseness, and unknown information and fuse multi-source information without depending on prior information.

A set of finite mutually exclusive hypotheses or propositions Ω is called the frame of discernment, and the Basic Probability Assignment (BPA) function under the frame of discernment is a function $m\colon 2^{\Omega}\mapsto[0,1]$, and satisfies with $\sum_{A\subseteq\Omega}m(A)=1$ and $m(\Phi)=0$. The BPA m is also called the mass function, and $m(A)$ expresses the proportion of all relevant and available evidence that supports the claim that a particular element of 2^{Ω} belongs to the set A but to no particular subset of A. Suppose H_1 and H_2 are two independent evidence with two mass functions m_1 and m_2 in the same frame of discernment Ω; the Dempster's rule of combination is defined as follow:

$$m_{1\oplus 2}(A)=K^{-1}\cdot\sum_{B_i\cap C_j=A}m_1(B_i)m_2\left(C_j\right) \tag{4}$$

where, $K=\sum_{B_i\cap C_j\neq\Phi}m_1(B_i)m_2\left(C_j\right)$ is referred to as the degree of conflict between the two BPAs. If K is close to 0, the Dempster's rule of combination becomes invalid. One of the methods to solve this problem is to set discount coefficient, which usually represents the unreliability or dependence of the evidences. If m is a BPA and (1-α) is its corresponding discount coefficient, then the BPA after discounting is:

$$m^{\alpha}(A)=\begin{cases}\alpha\cdot m(A) & A\neq\Omega\\ \alpha\cdot m(\Omega)+(1-\alpha) & A=\Omega\end{cases} \tag{5}$$

3 Proposed method

Recently, a new non-parametric method based on KDE has been proposed to determine BPA [20,21]. Inspire by the idea, a score-mapping method for forensic is presented in this paper, in which KDE and evidence theory are used to determine the confidence of whether two face image pairs to compare are from the same source or from the difference source. The flowchart of the proposed score-mapping method, named KDE-DS, is shown as figure 1.

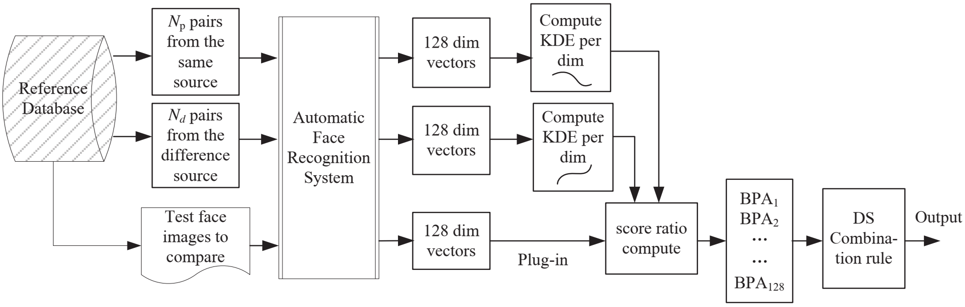

Fig. 1. The flowchart of the proposed score-mapping method.

The steps are described as follows:

Step 1: First, the reference database can be divided into two parts. One is training sample which constructs the model of each feature with its PDF curves. Another one is test sample which is used to verify the constructed model. Each sample includes two types of images: face-pairs from the same source and face-pairs from the difference sources.

Step 2: Automatic face recognition system is used to generate n-dimensional feature vectors (e.g. 128-dim). For each dimensional feature, the similarity/distances of all image pair are calculated and probability density function is generated using those similarity/distances via KDE, which can be regarded as the probability model for the related feature using the training sample.

Step 3: If there are new face image pairs that need to be compared (which could be regard as evidence in a case), the automatic face recognition system is also utilized to generate n-dimensional feature vectors, then the value of each dimensional feature would be plugged in the corresponding PDFs. Thus, we obtain two values k_1 and k_2, one for the prosecution hypothesis H_p and another for the defense hypothesis H_d, as shown in Figure 2.

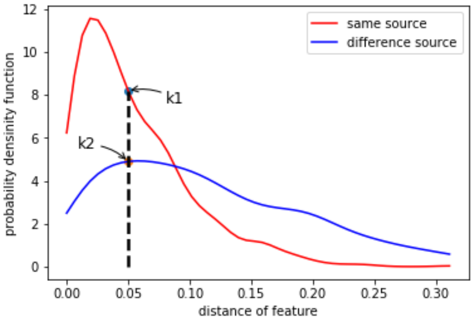

Fig. 2. Generate two pdfs for each feature.

Given a hypothesis H_0, the probability of H_0 for each PDF model is proportional to the specific intersection point value $f(x_0)$ [20]. According to this, a frame of discernment $\Omega = \{H_p, H_d\}$ could be constructed, and the rules about how a membership is assigned to the focal element are as follows:

$$m(\{H_p\}) = c_1 = \frac{k_1}{k_1 + k_2}$$
$$m(\{H_d\}) = c_2 = \frac{k_2}{k_1 + k_2} \qquad (6)$$

Consider that the features of images may not be completely independent of each other, and the source may also be unreliable, the discount efficient is considered to reflect such effects.

$$\begin{aligned} m(\{H_p\}) &= \alpha \cdot c_1 \\ m(\{H_d\}) &= \alpha \cdot c_2 \\ m(\{H_p, H_d\}) &= 1 - \alpha \end{aligned} \tag{7}$$

Generally, (1- α) could be set to 0.3.

Step 4: Finally, Dempster's combination rule is used to combine multiple BPAs to get the final BPA, and the mass function $m(\{H_\mathrm{p}\})$ reflects the strength of evidence support of the prosecution.

4 Experimental results

To Verify the effectiveness of the proposed method, the LFW (Labeled Faces in the Wild) face database is used, which is a popular test set for face recognition [22]. The face images provided are all from natural scenes in life, so the difficulty of recognition will increase. Here we select 1100 pairs of same person and different person respectively as training set, and 500 pairs of same person and different person respectively as test set. After the aligning images step, the available numbers of image-pairs are shown as Table 2.

Table 2. Numbers of face image pairs.

	Same source	Difference source
Training set	1090	1087
Test set	495	495

In forensics, transparency in the methods is essential [10]. Since Openface is an easily available open-source toolkit, which based on the FaceNet algorithm for automatic facial identification that was created by Google [23], it is used in the experiment, and we utilize the commonly known matching indexes of accuracy, sensitivity and specificity, defined as:

$$accuray = \frac{TP + TN}{N} \tag{8}$$

$$sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$specificity = \frac{TN}{TN + FP} \tag{10}$$

With *N* denoting the number of scores, *TP* the number of True Positives, *TN* the number of True Negatives, *FN* the number of False Negatives, and *FP* the number of False Positives. When calculating the accuracy, we first convert the LR values to a grade according to the conclusion scale in Table 2. A LR value counts as a match prediction if it receives a grade +2 or higher, and as a no-match prediction if it receives a grade +0.5 or lower. Similarly, if the ratio of the final BPA supporting the H_p to supporting the H_d is greater than 2, which counts as a match prediction, vice versa. The comparative experimental results are shown as Table 3:

Table 3. Experimental results.

	Accuracy	Sensitivity	Specificity
KDE	91.01%	90.71%	91.31%
KDE-DS	93.84%	94.14%	93.54%

From the experimental results, the proposed method has a better performance in terms of accuracy, sensitivity and specificity than the classic KDE-based likelihood ratio method.

After analysis, the main cause of the error is that multiple face images are extracted incorrectly, or there is a face occluded image in the face image pair, such as wearing glasses, so we should assure that before using automatic face recognition system for face verification, the face image pairs to compare should be carefully checked.

5 Conclusions

The KDE method is a common method for likelihood ratio calculation in the field of face forensics, which utilizes the similarity/distance of face image pairs to generate PDF. N-dimensional feature vectors are integrated into a distance measure, which lose some local information. In this study, each dimensional feature is treated as a random variable, and the distances of each feature between image-pairs are accumulated as PDF via KDE, then the PDFs can be mapped in BPA of the prosecution and the defense, finally the BPAs are combined by Dempster's combination rule. The experiments are verified that the proposed method has a better performance than KDE method. It would be a powerful supplement to traditional likelihood ratio calculation method.

This research was financially supported by Science and Technology Research Project of Jiangxi Provincial Department of Education (NO. GJJ151196) and Collaborative Innovation Center for Economics crime investigation and prevention technology, Jiangxi Province (No. JXJZXTCX-019).

References

1. C. G. Zeinstra, D. Meuwly, A. C. Ruifro, R. N. Veldhuis, L. J. Spreeuwers, Forensic face recognition as a means to determine strength of evidence: a survey. Forensic Sci Rev, **30**, 1, 21-32 (2018).
2. P. Tome, J. Fierrez, R. Vera - Rodriguez, J. Ortega - Garcia, Combination of face regions in forensic scenarios. Journal of forensic sciences, **60**, 4, 1046-1051 (2015).
3. Facial Identification Scientific Working Group: Facial comparison overview and methodology guidelines, https://fiswg.org/fiswg_facial_comparison_overview_and_methodology_guidelines_V1.0_20191025.pdf.
4. M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, M. Tistarelli, On soft biometrics. Pattern Recognition Letters, **68**, 218-230 (2015).
5. P. Tome, R. Vera-Rodriguez, J. Fierrez, J. Ortega-Garcia, Facial soft biometric features for forensic face recognition. Forensic science international, **257**, 271-284 (2015).
6. M. Jacquet, C. Champod, Automated face recognition in forensic science: Review and perspectives. Forensic Science International, **307**, 110124 (2020).
7. A. L. Mölder, I. E. Åström, E. Leitet, *Development of a score-to-likelihood ratio model for facial recognition using authentic criminalistic data*. In 2020 8th International Workshop on Biometrics and Forensics (IWBF), IEEE, 1-6 (2020).
8. D. Meuwly, D. Ramos, R. Haraksim, A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. Forensic science international, **276**, 142-153 (2017).
9. N. Suki, N. Poh, F. M. Senan, N. A. Zamani, M. Z. A. Darus, *On the reproducibility and repeatability of likelihood ratio in forensics: A case study using face biometrics*. In 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), 1-8 (2016).

10. A. Macarulla Rodriguez, Z. Geradts, M. Worring, Likelihood Ratios for Deep Neural Networks in Face Comparison. Journal of Forensic Sciences, **65**, 4, 1169-1183 (2020).
11. G. S. Morrison, E. Enzinger, Score based procedures for the calculation of forensic likelihood ratios–Scores should take account of both similarity and typicality. Science & Justice, **58**, 1, 47-58 (2018).
12. N. Garton, D. Ommen, J. Niemi, A. Carriquiry, *Score-based likelihood ratios to evaluate forensic pattern evidence*. arXiv preprint arXiv:2002.09470, 1-22 (2020).
13. L. McKenna, S. McDermott, G. O'Donell, *ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the evaluation of forensic results across Europe (STEOFRAE)*. Wiesbaden, Germany: European Network of Forensic Science Institutes, 30–41 (2015).
14. T. Ali, L. Spreeuwers, R. Veldhuis, D. Meuwly, *Effect of calibration data on forensic likelihood ratio from a face recognition system*. In 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 1-8 (2013).
15. D. Ramos, R. P. Krish, J. Fierrez, D. Meuwly, *From biometric scores to forensic likelihood ratios*. In Handbook of biometrics for forensic science, Springer, Cham, 305-327, (2017).
16. F. S. Kool, *Feature-based models for forensic likelihood ratio calculation: Supporting research for the ENFSI-LR project*, (2016).
17. T. Ali, *Biometric Score Calibration for Forensic Face Recognition*. Ph.D. Thesis Series, Centre for Telematics and Information Technology, 14–336 (2014).
18. A. Dempster, Upper and lower probabilities induced by multivalued mapping. Annals of Mathematical Statistics, **38**, 2, 325-339 (1967).
19. G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
20. P. Xu, X. Su, S. Mahadevan, C. Li, Y. Deng, A non-parametric method to determine basic probability assignment for classification problems. Applied intelligence, **41**, 3, 681-693 (2014).
21. B. Qin, F. Xiao, A non-parametric method to determine basic probability assignment based on kernel density estimation. IEEE Access, 6: 73509-73519 (2018).
22. G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*. University of Massachusetts, Amherst, Technical Report, 07-49 (2007).
23. A. Fydanaki, Z. Geradts, Evaluating OpenFace: an open-source automatic facial comparison algorithm for forensics. Forensic sciences research, **3**, 3, 202-209 (2018).