Low Rank Analysis of Eye Image Sequence – A Novel Basis for Face Liveness Detection

Chengyan Lin, Yuwu Lu, Jian Wu, and Yong Xu(✉)

Bio-computing Research Center, Shenzhen Graduate School,
Harbin Institute of Technology, Shenzhen, People's Republic of China
CYLin2011@gmail.com, {yuwulu2008,wujianhitsz}@163.com,
laterfall@hitsz.edu.cn

Abstract. The security of the face recognition technology has attracted more and more attention because of the wide applications of this technology. A lot of studies on face liveness detection have been performed. In this paper, we cast the face liveness detection problem as a classification problem to distinguish the images of true faces and photo samples based on the rank analysis of sample matrices. We assume that the rank of the true face sample matrix is much higher than that of the photo sample matrix under an ideal situation. If we denoise the real world samples and convert them into pure samples, we can find a well boundary, that is, a basis for liveness detection. Experiments are conducted on the NUAA imposter database to verify the efficiency of the proposed method.

Keywords: Face liveness · Eye sequence · Low rank · Classification basis

1 Introduction

The face recognition technology is one of the most popular biometrics technologies in pattern recognition field [1-3]. More and more attentions have been paid to the safety of the face recognition system. Face liveness detection has become an important means to identify the cheating behaviors to the face recognition system. The ways of cheating 2D face recognition systems include using the photo and the short video of an authorized user to spoof the system [4]. The photo attack seems to be the easiest and cheapest cheating approach because facial images of users are easily available.

Many methods to resist the photo attack have been proposed. These methods can be classified from different perspectives. Some methods perform the face liveness detection using only one single facial image. Most of such methods detect photo attacks by analyzing the texture features of the captured facial images. The most widely used texture feature is the local binary pattern (LBP) [5]. Määttä J et al. [6] used the multi-scale LBP of one single facial image for face liveness detection. The individual histograms produced by the multi-scale LBP were combined into a concatenated feature histogram. Finally, the obtained feature histogram was used for classification by a nonlinear SVM classifier. LBP can also combine with other features such as frequency [7] and local shape [8]. However, LBP has its inherent weaknesses: noise-sensitive and weak performance in flat regions [9]. Such drawbacks would be

J. Yang et al. (Eds.): CCBR 2015, LNCS 9428, pp. 11–18, 2015.
DOI: 10.1007/978-3-319-25417-3_2

amplified when the attacking photo with high quality. For example, the light reflection on a high resolution photo will not generate obvious reflective stripe which can be captured by frequency or texture analysis.

Other methods make a decision by analyzing a continuous image sequence of one's face or some facial parts. The optical flow field based methods differentiate liveness and photo by analyzing four types of optical flow field generated by 2D and 3D objects. There are some existing optical flow algorithms [10, 11]. In the facial part analysis methods, eye area is the most popular area that has been discussed because of eye-blink. In [12], the authors adopted a motion magnification operation before liveness detection to improve the facial expressions from one's video sequence. LBP is still an important technique for the algorithm in [12]. In [13], the authors constructed a Conditional Random Field (CRF) framework to model eye-blink behaviors. In addition, they used a discriminative measure of eye states to improve the classification performance. The complexity of the image sequences based methods is higher than that of the one single image based methods.

Face liveness detection plays an important role to ensure the security of a face recognition system during the recognition stage. Considering the safety of the whole recognition process and the computation and preprocessing cost, the eye image sequence is a good target for liveness detection research.

In this paper, we cast the face liveness detection problem as a problem of classification for matrices. We argue that the matrices ranks of the liveness and photo samples are different. The rank of the liveness sample matrix is much higher than that of the photo sample matrix under an ideal situation. If we denoise the real world samples and revert them to pure samples, we can find a well boundary, that is, a basis for liveness detection. We conduct a series of experiments to verify the classification performance of the obtained basis.

2 Motivations

One sample is consisted by an eye image sequence. Assume that there is a dataset $X=\{X_1, X_2, \ldots, X_c\}$ which contains c samples from the same class. X_i ($0<i\leqslant c$) includes γ_i eye images. Each eye image in X_i is resized into a vector, and stack these vectors as columns of a matrix

$$X_i = [x_1, x_2, ..., x_{\gamma_i}], \gamma_i > 0 \tag{1}$$

where x_i represents the i^{th} frame of X_i, and the length of X_i, that is, the number of frames in X_i.

In the face liveness detection task, there are two kinds of samples: photo and liveness. In the ideal situation, frames in a photo sample are with little differences, which produces a low-rank of the photo sample matrix. Besides, the rank of a liveness sample matrix is much higher because of the facial movements (such as blink). Fig. 1 takes several images of the liveness and photo samples from the NUAA database for example to explain the motivation visually.

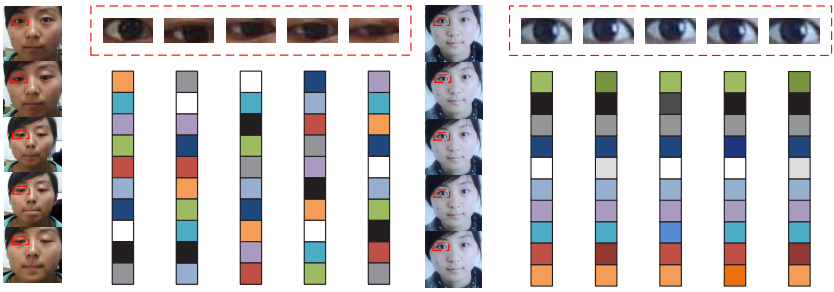

Fig. 1. An example of the representation of liveness and photo samples.

We can obtain the conclusion that there exist a threshold or a boundary θ_1 about rank to differentiate liveness samples from photo samples. Let ω_1 be the set of liveness samples and ω_0 the set of photo samples. The classification rules can be defined as follow:

$$\Phi(X) = \begin{cases} X \in \omega_1, & rank(X) \geq \theta_1 \\ X \in \omega_0, & rank(X) < \theta_1 \end{cases}. \tag{2}$$

If the rank of X_i satisfies the condition that $rank(X_i) \geqslant \theta_1$, we consider X_i as a liveness sample and vice versa. If the lengths of the samples are different, we should divide the rank of each sample by its length for equalization. In this case, $rank(X_i)$ can be rewritten as $rank(X_i) / \gamma_i$.

3 Proposed Method

3.1 Sample Noising Model

We assume that the photo sample matrix has a low rank. However, most of the samples we obtained in reality are not of low rank. Thus, we consider that there exist noises which increase some small eigenvalues to the matrices of the real samples. Suppose that the sample noising model is

$$A = L + N, \tag{3}$$

where A denotes the real world sample matrix, L is a low-rank sample matrix, and N is an additive Gaussian white noise with standard deviation σ. The smaller the norm of N is, the closer A gets to L, which means that the outcome fits the ideal situation.

According to (3), there exists an optimal solution of L. We adopt a matrix denoising scheme to solve the optimal solution of L:

$$L^* = \arg\min_L \frac{1}{2}\|A - L\|_F^2 + \lambda \|L\|_*, \tag{4}$$

where $\|L\|_*$ denotes the nuclear norm of a matrix, i.e., sum of singular values. λ is a hyper-parameter to balance the square loss function and the regulation function.

3.2 Solutions of the Noising Model

Equation(4) can be solved by SVD as follow:

$$L^* = U \begin{pmatrix} \lambda_1^* & & \\ & \ddots & \\ & & \lambda_n^* \end{pmatrix} V^T, \tag{5}$$

where U and V are the left and right singular vector matrices of A, respectively.

Then we solve λ_i^* using the soft thresholding operation:

$$\lambda_i^* = \max(|\lambda_i| - \lambda, 0) sign(\lambda_i), \tag{6}$$

where λ_i is the singular value of A, and λ is the hyper-parameter mentioned in (4). Since λ_i is always positive, Eq.(6) can be simplified as

$$\lambda_i^* = \max(|\lambda_i| - \lambda, 0). \tag{7}$$

After the thresholding operation, $rank(L^*)$ is just the number of non-zero singular values λ_i^*.

The physical interpretation of λ is the standard deviation of the noise σ. To estimate λ, we adopt the following adaptive strategy

$$\lambda = \theta_1 \sum_{i=1}^{r} \lambda_i, \tag{8}$$

or

$$\lambda = \theta_1 \|A\|_*. \tag{9}$$

As the energy weight of the noise in the noisy data, θ_1 can be learned or trained by minimizing the training error.

3.3 Basis for Classification

Once θ_1 is given, the distributions of ranks of two classes of samples are available. Then a classification boundary θ_2 (or decision surface) can be set by either Bayes rule or k-means clustering. Fig. 2 simulates the distribution of two classes and their classification boundary with a given θ_1.

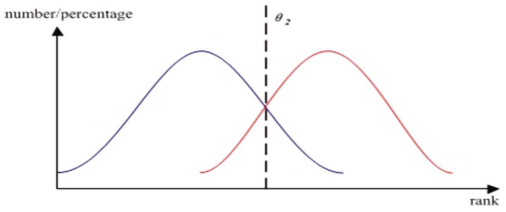

Fig. 2. A simulation of the distribution of two classes and their classification boundary with a given θ_1. The blue bell curve represents the photo samples and the red one the liveness samples. Assume that the samples from one single class form a normal distribution.

A Bayesian minimal error rate can be solved, that is, $ER(\theta_1)$ is essentially dependent on θ_1. Thus, θ_1 can be determined by:

$$\theta_1 = \arg\min_{\theta} ER(\theta) \tag{10}$$

The solution of (10) can be solved by enumerating.

3.4 The Proposed Algorithm

Given an image set of the liveness or photo samples, there are two part of searching for the classification basis. The first part is to preprocess the sample set and obtain the relative data about its singular values which to be used in the second part. The second part is to enumerate an optimal boundary between two types of data with a given θ_1.

The preprocess of a sample set:

- Given a set of one type of sample sequences $A=\{A_1, A_2,\ldots, A_c\}$
- For each $A_i \in A$
  1) Transfer a sequence into a matrix
  $$A_i = [a_1, a_2, ..., a_{\gamma_i}], \gamma_i > 0$$
  2) Do SVD to A_i, and obtain its singular values $\{\lambda_1, \lambda_2, ..., \lambda_r\}$
  3) Sum the singular values of A_i using $s_{A_i} = \sum_{j=1}^{r} \lambda_j$
- Obtain a data set about class A, $S_A = \{s_{A_1}, s_{A_2}, ..., s_{A_c}\}$

The process of training a classification basis between two sample classes with a given ($0 < \theta_1 < 1$):

- Given two preprocessed data set of different classes
  $$S_A = \{s_{A_1}, s_{A_2}, ..., s_{A_c}\} \text{ and } S_B = \{s_{B_1}, s_{B_2}, ..., s_{B_d}\}$$
- Estimate the λ of every sample, then we have $E_A = \theta_1 S_A$ and $E_B = \theta_1 S_B$
- Enumerate an optimal boundary between E_A and E_B: θ_2

Within the range of θ_1, we are able to find an optimal value. The θ_2 connected with the optimal θ_1 value produces a minimum Bayesian minimal error rate. The specific θ_1 and the corresponding θ_2 form a basis for classification of face liveness detection.

4 Experiments

In order to verify the feasibility of our method, we did experiments on the NUAA imposter database [13]. The target of our experiments is the eye areas of the pictures of the database. We detected the eye areas of the pictures in the same size of 32×16. Due to the symmetry of eyes in a face, we just need to consider the left-eye image sequences.

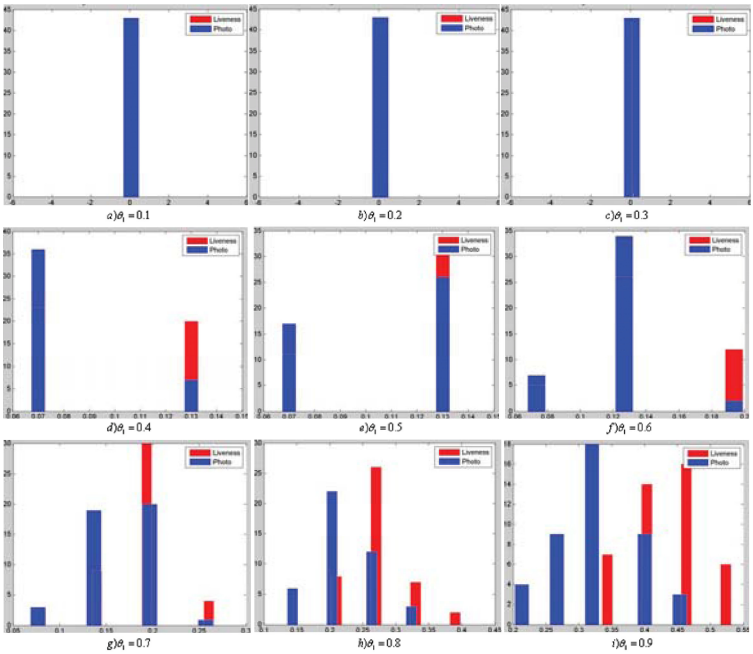

Fig. 3. Histograms of E_A and E_B with the increase of $\theta_1$when the length of samples is 15. The histograms in red are from liveness samples, while those in blue are from photo samples.

The NUAA imposter database includes 15 subjects. The images of this database are collected in three sessions. The first session is the only session which involves all the subjects. This session contains no special location and light conditions which needs extra preprocess such as image alignments. Moreover, our experiments focus on the feasibility of the classification basis in general cases. Thus, we paid most of our attention to the first session in the experiments.

There are 9 subjects involved in the first session of the database. The length of each sequence from a subject in a certain session is about 100 or more. In our tests, we observed the effect of θ_1 under a fixed length of samples. In the experiments, θ_1 is increased by 0.1 from 0.1 to 0.9 under a fixed length.

Let E_A and E_B be the super-parameter set of liveness and photo samples, respectively. Fig. 3 is the histograms of E_A and E_B when the length of the samples is 15. With the increase of θ_1, the difference between liveness and photo samples is becoming more and more apparent. Especially, Fig. 3(i) fits Fig. 2 very well. When the samples of liveness and photo can be well divided, most of the values in E_A is larger than those in E_B.

Due to the Frames per Second (FPS) of the general web camera is 15 at least and 60 at most, we can fix the length of image sequence between 15 and 60 in practice.

Table 1. The error rate for different sequence length (θ_1=0.90).

Sequence length	10	15	20	25	30	35	40	45	50	55	60	65
Error rate	0.28	0.22	0.23	0.16	0.16	0.12	0.11	0.10	0.11	0.12	0.13	0.13

Table. 1 shows the error rate for different sequence length. Given a fixed θ_1, the change of the error rates obtained with different lengths of the samples is steady. When the length of the samples is above 30, the error rate would fluctuate around some value. Thus, we can make a conclusion that the performance of our method is independent with the length of the samples. When the length of a sample is too short, the dimension of the sample matrix is not enough to produces a comparable rank.

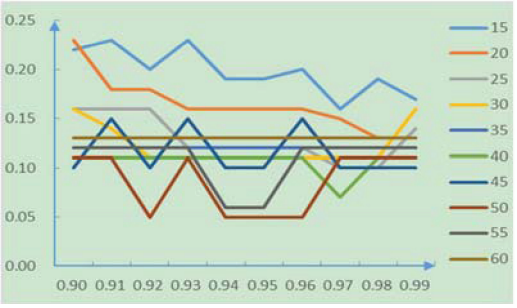

Fig. 4. The error rate for different sequence length when the range of θ_1 is from 0.90 to 0.99.

We narrow the range of θ_1 from 0.9 to 0.99 according to the conclusion of Fig. 3. The error rates for different sequence lengths are shown in Fig. 4. When the length of sequences is small, the error rate drops obviously with the increase of θ_1. The overall error rates tend to be smaller with the increase of θ_1.

5 Conclusion

In this paper, we analyzed the differences between liveness and photo sample matrices, and argued that the matrices ranks of the liveness and photo samples can be used to distinguish the images of true faces and photo samples. We found a boundary which can be seen as a basis for liveness detection during the process of denoising the real world samples. Such a novel basis is related to the matrices ranks of the samples. Through the experiments, we verified that the obtained basis is able to classify liveness and photo samples.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China uder Grant 61370163, and the Shenzhen Municipal Science and Technology Innovation Council under Grant JCYJ20130329151843309, Grant JCYJ20140904154630436 and Grant CXZZ20140904154910774.

References

1. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern. Anal. Mach. Intell. **31**, 210–227 (2009)
2. Zhang, Z., Xu, Y., Yang, J., Li, X.L., Zhang, D.: A survey of sparse representation: algorithms and applications. IEEE Access **3**, 490–530 (2015)
3. Xu, Y., Zhang, D., Yang, J., Yang, J.Y.: A two-phase test sample sparse representation method for use with face recognition. IEEE Trans. Circ. Syst. Video Technol. **21**, 1255–1262 (2011)
4. Nixon, K.A., Aimale, V., Rowe, R.K.: Spoof detection schemes. In: Jain, A.K., Flynn, P., Ross, A.A. (eds.) Handbook of biometrics, pp. 403–423. Springer, US (2008)
5. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern. Anal. Mach. Intell. **24**, 971–987 (2002)
6. Komulainen, Jukka, Hadid, Abdenour, Pietikäinen, Matti: Face spoofing detection using dynamic texture. In: Park, Jong-Il, Kim, Junmo (eds.) ACCV Workshops 2012, Part I. LNCS, vol. 7728, pp. 146–157. Springer, Heidelberg (2013)
7. Kim, G., Eum, S., Suhr, J.K., Kim, D.I., Park, K.R., Kim, J.: Face liveness detection based on texture and frequency analyses. In: IAPR International Conference on Biometrics, pp. 67–72. IEEE Press, New York (2012)
8. Maatta, J., Hadid, A., Pietikäinen, M.: Face spoofing detection from single images using texture and local shape analysis. IET Biometrics **1**, 3–10 (2012)
9. Xu, J., Ding, X.Q., Wang, S.J, Wu, Y.S.: Background subtraction based on a combination of texture, color and intensity. In: International Conference on Signal Processing, pp. 1400–1405. IEEE Press, New York (2008)
10. Barron, J., Fleet, D., Beauchemin, S.: Performance of optical flow techniques. Int. J. Comput. Vis. **12**, 43–77 (1994)
11. Fleet, D., Jepson, A.: Computation of component image velocity from local phase information. Int. J. Comput. Vis. **5**, 77–104 (1990)
12. Bharadwaj, S., Dhamecha, T.I., Vatsa, M., Richa, S.: Computationally efficient face spoofing detection with motion magnification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 105–110. IEEE Press, New York (2013)
13. Pan, G., Sun, L., Wu, Z.H., Lao, S.H.: Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In: Proceedings of the 11th IEEE International Conference on Computer Vision, pp. 1–8 IEEE Press, New York (2007)

http://www.springer.com/978-3-319-25416-6

Biometric Recognition
10th Chinese Conference, CCBR 2015, Tianjin, China, November 13-15, 2015, Proceedings
Yang, J.; Yang, J.; Sun, Z.; Shan, S.; Zheng, W.; Feng, J. (Eds.)
2015, XVIII, 739 p. 377 illus. in color., Softcover
ISBN: 978-3-319-25416-6