A Face Anti-Spoofing Method Based on Optical Flow Field

Wenze Yin, Yue Ming and Lei Tian
Beijing Key Laboratory of Work Safety Intelligent Monitoring
School of Electronic Engineering, Beijing University of Posts and Telecommunications,
Beijing, P.R,China
e-mail: yinwenze562@163.com, myname35875235@126.com, tianlei189@bupt.edu.cn

***Abstract*— Spoofing attack can easily deceive face recognition system. In this paper, we explore the issue of face anti-spoofing with good performance in accuracy by utilizing optical flow vector on two types of attacks: photos and videos shown on high-resolution electronic screens. The key idea is to calculate the displacement of optical flow vector between two successive frames of a face video and obtain a displacement sum of a certain number of frames. Under the circumstance of stable light, the sum of displacement differs from real access and other spoofing attacks. In this situation, we experiment on REPLAY-ATTACK, a common and popular face spoofing database which shows a good performance. We conclude that spoofing attacks and real faces have different optical flow motion trend that our method shows temperate guesstimate when facing with a broad set of face attacks.**

***Keywords*— Face Anti-Spoofing, Optical Flow, Image Processing , kNN**

I. INTRODUCTION

Spoofing attack is the action of deceiving a biometric sensor by showing a counterfeit biometric evidence of a valid user [1]. The attackers do not need to know related knowledge about the recognition mechanism before spoofing the sensorial input of a biometric system. Most of the biometric patterns are unable to guard against spoofing attacks: most face recognition systems are intendedly devised to distinguish between identities without caring about the sign of life of an identity. Although life approval and verification systems are sophisticated today, formulating anti-spoofing proposal for them is still just getting started. Therefore, an effective anti-spoofing method is extremely important, especially in national security authorities.

From current study, anti-spoofing methods can be divided into three categories basing on the different use of clues: analyzing the texture of image [2] [3], multispectral approach basing on the difference of reflection rate between skin and other material [4], analyzing the motion information [5] [6] [7]. Optical flow is a feature of motion information that several optical flow algorithms have been proposed in face field. In [8], the authors mention a non-negative coefficient projection algorithm that maps optical flow onto a subspace of facial expression. In [9], the authors improve the optical flow algorithm based on regularization to calculate precise intensity variations and pixel displacements by forcing limits on a few given point correspondences. In [10], the authors utilize optical flow to construct a likelihood map from the Viola-Jones algorithm?s middle results in order to present a novel face tracking approach. In [11], the authors use optical flow algorithm to segment image in advance for reducing regions of search by using a robot for face detection. In [12], the authors propose an improved optical flow method that brings about a great difference between real faces and face photos by utilizing dense optical flow field during a short period of time. In [13], the authors propose a new liveness detection method by using differentiation brought about by motions of three-dimensional objects and two-dimensional planes in optical flow fields.

The biggest advantage of these methods is insurance of adequate accuracy [12] [13]. However, the limitation of these methods is twofold. On the one hand, the type of spoofing attack studied barely involves face video attack, which is not robust to common spoofing attack. On the other hand, these methods often need user's motion cooperation, which lead to awful user experience. To overcome the mentioned above problems, we propose a DOFV (displacement of optical flow vector) model which bases on dense optical flow instead of sparse optical flow[14]. We consider that for the following reasons.

1) The displacement of optical flow vector between two successive frames varies from spoofing faces and real faces. The mode of spoofing attack is generally divided into photo and video in the real world. Optical flow information in a scene where an attacker hold a pad displaying an photo or a video is a lot more than in a scene where a real person just facing in front normally because of the shake by a hand.

2) Dense optical flow associates a velocity with every pixel in an image, which would have more information that sparse optical flow. That would be easier for training and classification in the case of limited training data, making the judgment more reliable.

3) Compared to other motion based methods, it would not need user's excessive cooperation, which is friendly to user.

4) Input video does not need to be preprocessed before entering into the trainer, which makes the process simple and convenient.

In what follows, we detail the studied measure in Section II. We give our experimental results in Section III. Conclusions and future work are elaborated in Section IV.

II. ANATOMY OF THE STUDIED MEASURE

In this section, we first give a brief introduction to our DOFV (displacement of optical flow vector) model, then

elaborate it in detail, including extraction of DOFV, and kNN(k-Nearest Neighbor) classification.

A. System Overview

The process of our method, which can be divided into training phase and testing phase, is shown in Figure 1 and 2.In the training phase, we extract DOVF between two successive frames of the several frames of every input video and generate a multi-dimensional vector for every training video.

Then we label the DOVF vectors of training samples. We focus on the following three attacks in comparison to real scenes:

1) Handheld video attack (HVA): An attacker holds an attack device with their hands that displaying a master's video.

2) Fixed photo attack (FPA): An attacker sets the attack device fixed that displaying a master's photo.

3) Fixed video attack (FVA): An attacker sets the attack device fixed that displaying a master's video.

In the case of stable light and frame continuity of adjacent video (video for experiment) frames, the SDOVF of FPA will be the least because it almost has no change on the movement. The SDOVF of FVA will be the most because it has changes in facial action such as blink and camera shake, and it has no other movements to offset its camera shake. For the real scene, it only contains small facial action changes so that it has more movement changes than FPA. As for the HVA, it contains some hand movements to offset the camera shake so that the movement changes can be unstable. In summary, the SDOVF of real access is between FPA and other attacks. Verification of the hypothesis above will be expounded in the beginning of experiments in Section III.

In the testing phase, we extract multi-dimensional vector of test vides and use kNN to classify.

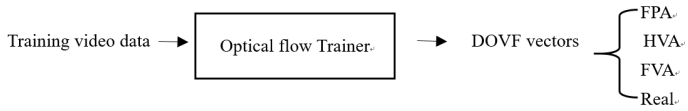

Fig. 1. Training phase.

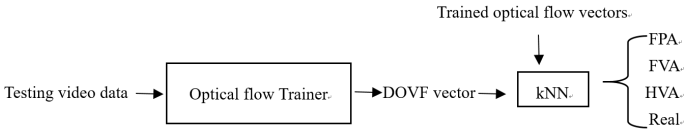

Fig. 2. Testing phase.

B. Extraction of DOFV

In order to estimate the displacement, we use polynomial expansion transform. The core of polynomial expansion is to estimate neighborhood of each pixel with a polynomial. We focus on quadratic polynomials [15], the local semaphore module presented in a local coordinate system is

$$f(x) = x^T Ax + b^T x + c. \tag{1}$$

where **A** is a symmetric matrix, **b** a vector and **c** a scalar. The modulus are reckoned from a weighted least squares suitable to the semaphore values in the neighborhood [16].

We input the video frame by frame and turn the frame image into grayscale image. Each frame is divided into lots of neighborhoods that all neighborhoods form into a rectangle, as Fig 3 shows. Each neighborhood is approximated by a polynomial. Considering the ideal situation, the exact quadratic polynomial is

$$F_1(f_{of}) = f_{of}^T A_1 f_{of} + b_1^T f_{of} + c_1. \tag{2}$$

where $\mathbf{f}_{of}$ is the amounts of optical flow in each neighborhood. Construct a new semaphore $\mathbf{F}_2$ as the neighborhood of next frame by a global displacement by $\mathbf{d}_{of}$,

$$\begin{aligned} F_2(f_{of}) &= F_1(f_{of} - d_{of}) \\ &= (f_{of} - d_{of})^T A_1 (f_{of} - d_{of}) + b_1^T (f_{of} - d_{of}) + c_1 \\ &= f_{of}^T A_1 f_{of} + (b_1 - 2A_1 d)^T f_{of} + d^T A_1 d - b_1^T d + c_1 \\ &= f_{of}^T A_2 f_{of} + b_2^T f_{of} + c_2 \end{aligned} \tag{3}$$

Equating the modulus in the quadratic polynomials yields

$$A_2 = A_1 \tag{4}$$

$$b_2 = b_1 - 2A_1 d_{of} \tag{5}$$

$$c_2 = d_{of}^T A_1 d_{of} - b_1^T d_{of} + c_1 \tag{6}$$

The core suggestion is that by equation (5) we can resolve for the translation $\mathbf{d}_{of}$, at least if $\mathbf{A}_1$ is non-singular,

$$2A_1 d_{of} = -(b_2 - b_1) \tag{7}$$

$$d_{of} = -\frac{1}{2} A_1^{-1} (b_2 - b_1) \tag{8}$$

In view of the above principle, we can get DOVF of all neighborhoods between one frame and its next frame. Then we ignore $\mathbf{d}_{of}$ where its value is two small (we set the minimum value as 0.2) and calculate the sum of DOVF of all neighborhoods as SDOVF. Then we calculate the SDOVFs for several frames, which can be regarded as a judgement value to distinguish whether it is a spoofing attack.

C. kNN Classification

According to the principle above, it is easy to find that if we construct SDOVFs from different frames of a video into an n-dimensional vector, distance between SDOVFs from same type of attack or real access will be smaller than those from different types. Therefore we use kNN to classify.

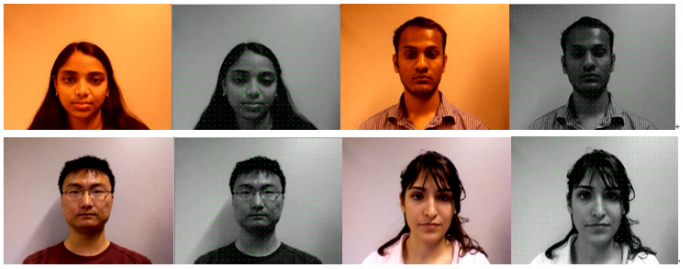

Fig. 3. Original image and grayscale image with optical flow points.

In the training phase, after extraction of the several frames of every input video we generate a multi-dimensional vector for every training video as training tuples. These input video is divided into four categories: FPA, FVA, HVA and real.

In the testing phase, we input a test video we process the same extraction and get a multi-dimensional vector as testing tuple. We set up **k** and maintain a sorted priority queue (size **k**) to store the nearest **k** training tuples. We randomly choose **k** training tuples as initial tuples. We calculate the Euclidean distance between the testing tuple and **k** initial tuples and store the labels of training tuples and distance into the priority queue. Then we iterate the training tuples and calculate the Euclidean distance **L** between every training tuple and testing tuple. If $\mathbf{L} < \mathbf{L}_{max}$ (the biggest distance in priority queue), we delete the $\mathbf{L}_{max}$ and add **L** into our priority queue. After iteration we choose the majority class of **k** tuples in our priority queue as the classification result of the testing tuple.

III. EXPERIMENTS AND ANALYSIS

We evaluate our mentioned DOVF model in this section. At first, we proceed with a previous experiment to verify the hypothesis mentioned in session II. Then we introduce our judgment criterion that evaluate the experiment results before we proceed with the main experiments. Then we experiment DOVF model on REPLAY-ATTACK database. At last we exploit two sets of comparative experiments on the same database for comparison on features and classifiers.

A. Previous Verification Experiment

In order to verify the hypothesis above, we record five videos containing HVA, FPA, FVA and real access,as Fig 4 shows. We use a 720p front camera of a laptop to record. The attacking photo and video are displayed on an iphone 6 cell phone recorded by its front camera. We calculate SDOVFs of first 50 frames of every video, as Table I shows. The result basically confirms our hypothesis.

Table I. SDOVFs(sums of displacement of optical flow vector) of first 50 frames.

Real access	FPA	FVA	HVA
9726	4897	56245	47408

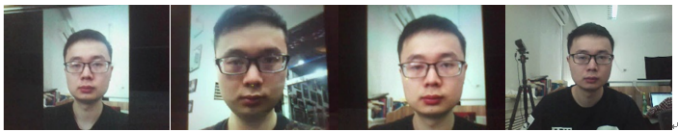

Fig. 4. HVA(Handheld video attack), FPA(Fixed photo attack), FVA(Fixed video attack) and real access.

B. Performance Measure

We use False Rejection Rate (FRR, the real input is refused), the False Acceptance Rate (FAR, the fake input is accepted), Half Total Error Rate (HTER, half of the sum of the FRR and FAR) and Recognition Success Rate (RSR, real input is accepted or an attack is refused) to measure the discriminant accuracy.

C. Experiment of DOVF Model

We use the training set of the REPLAY-ATTACK database to do our training. We choose the section of the controlled (the backdrop of the video is united and illumination environment is stable) and high def (the user shows the high-resolution digital photos and videos on an iPad screen with resolution (1024×768 pixels)) condition [17], as Fig 5 shows. In the

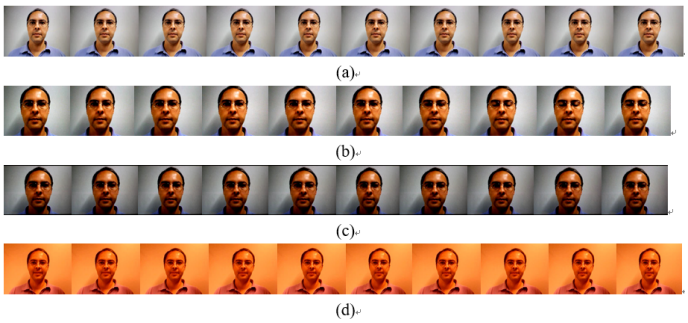

Fig. 5. An example of one client's first 10 frames of the four accesses. (a) FPA (b) FVA (c) HVA (d) real access.

training phase, for the training set we get 57 50-dimensional SDOVFs and label them, which contain 14 FPA, 15 FVA, 13 HVA and 15 real access.

We respectively use the testing and dev set of the REPLAY-ATTACK database to do our testing. The testing set contains 76 videos (19 real access, 19 FPA, 19 FVA, 19 HVA in controlled condition). The dev set contains 60 videos (15 FPA, 15 FVA, 15 HVA and 15 real access in controlled condition).

At first, we explore the optimal value of **k** in kNN classifier on testing set. The RSR for different value of **k** is given in Figure 6 and Table II. The FAR, FRR and HTER of the experiments above are given in Table III.

From the experiment result above, although the HTER of **k**=3 is a little bit small than **k**=5, FVA of **k**=3 is rather smaller than FVA of **k**=5 and the curve of **k**=5 is optimal overall. In summary, we choose **k**=5 as the parameter of kNN in the remaining experiments. The RSR for the dev set is given in

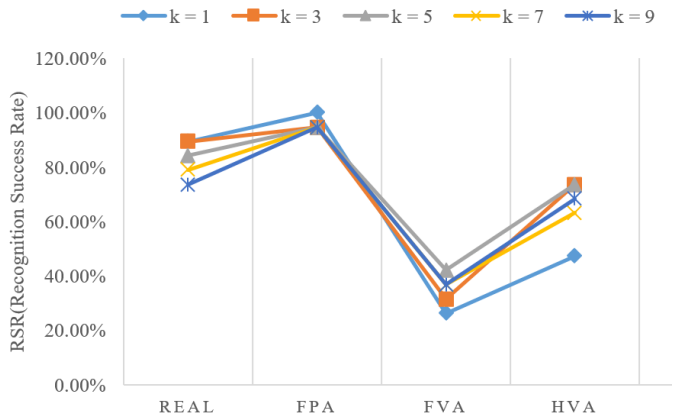

Fig. 6. RSR(Recognition Success Rate) for the testing set with different **k**.

Table II. RSR(Recognition Success Rate) for the testing set(76 videos) with different **k**.

k	Real access	FPA	FVA	HVA
1	89.47%	100%	26.32%	47.37%
3	89.47%	94.74%	31.58%	73.68%
5	84.21%	94.74%	42.11%	73.68%
7	78.94%	94.74%	36.84%	63.15%
9	73.68%	94.74%	36.84%	68.42%

Table III. FAR(False Rejection Rate), FRR(False Acceptance Rate) and HTER(Half Total Error Rate) for the testing set(76 videos) with different **k**.

k	FAR	FRR	HTER
1	10.53%	42.10%	26.32%
3	10.53%	33.33%	21.93%
5	15.79%	29.82%	22.81%
7	21.06%	35.09%	28.08%
9	26.32%	33.33%	29.83%

Table IV. The FAR, FRR and HTER of the experiments above are given in Table V.

Table IV. RSR(Recognition Success Rate) on the testing set(76 videos) and dev set(60 videos) using kNN(**k**=5).

	Real access	FPA	FVA	HVA
Testing	84.21%	94.74%	42.11%	73.68%
Dev	100%	100%	60.00%	60.00%

Table V. FAR(False Rejection Rate), FRR(False Acceptance Rate) and HTER(Half Total Error Rate) on the testing set(76 videos) and dev set(60 videos) using kNN(**k**=5).

	FAR	FRR	HTER
Testing	15.79%	29.82%	22.81%
Dev	0.00%	26.67%	13.33%

D. Comparative Experiments

At first, we choose HOG (Histogram of Oriented Gradient) feature and LBP (Local Binary Patterns) feature for comparative experiments.

1) HOG feature

We randomly choose one of the first 50 frames from the training set and dev set of REPLAY-ATTACK that we get 57 training image samples and respectively extract their HOG feature. The size of each image is 320×320 pixels. The cell size of HOG is set up 64×64 pixels that we get a 288 dimension vector of each image.

For the test data, we randomly choose one of the first 50 frames from the testing set and dev set of REPLAY-ATTACK that we get 76 testing image samples of testing set and 60 testing image samples of dev set. At last we use kNN (**k**=5) to classify.

2) LBP feature

We choose the simplest regular uniform LBP feature for experiment. Training samples and testing samples are same as the experiment of HOG feature above. We get a 59 dimension vector of each image.

The RSR of the experiments above are given in Table VI. The FAR, FRR and HTER of the experiments above are given in Table VII.

From the experiment result above, the HTER of optical flow is rather smaller than other features. What's more, the performance of HOG and LBP is not stable on spoofing attack. On the whole, our DOVF model shows greater performance.

Then we choose SVM (Support Vector Machine) classifier and Softmax Regression instead of KNN for comparative experiments. We use the same optical flow feature and training set and testing set. The RSR for the testing is given in Table VIII. The FAR, FRR and HTER for the testing are given in Table IX.

From the experiment result above, the HTER of kNN(**k**=5) is rather smaller than other classifiers. Although SVM does excellent performance on FPA, the overall effect is not good. As for the Softmax Regression, performance is really ordinary not only on real access but also on spoofing attack. In general, kNN classifier shows greater performance.

IV. CONCLUSION

This paper proposes an optical flow feature based face spoofing algorithm and evaluate its performance against a variety of spoofing attacks on the database of REPLAY-ATTACK. As a result, it shows good performance against other features such as HOG and LBP. Our proposed DOVF model utilizes kNN for classification, which also shows excellent performance against other classifiers such as SVM and SoftMax Regression. In comparison to other optical flow based methods, it does not involve intended motion cooperation of user and preprocessing procedure before training process.

However, the optical flow feature based anti-spoofing method we describe has a limitation that will not show good performance under the circumstance of unsteady light because of the dependency of the accurate computation of the optical flow field. Therefore the future work in the area of optical flow

Table VI. RSR(Recognition Success Rate) for different features on the testing set(76 videos) and dev set(60 videos) using kNN(**k**=5).

	Testing				Dev			
	Real	FPA	FVA	HVA	Real	FPA	FVA	HVA
Optical Flow	84.21%	94.74%	42.11%	73.68%	100%	100%	60.00%	60.00%
HOG	84.21%	42.11%	42.11%	47.68%	73.33%	100%	73.33%	33.33%
LBP	84.21 %	63.16%	57.89%	36.84%	60.00%	80.00%	53.33%	40.00%

Table VII. FAR(False Rejection Rate), FRR(False Acceptance Rate) and HTER(Half Total Error Rate) for different features on the testing set(76 videos) and dev set(60 videos) using kNN(**k**=5).

	Testing			Dev		
	FAR	FRR	HTER	FAR	FRR	HTER
Optical Flow	15.79%	29.82%	22.81%	0.00%	26.67%	13.33%
HOG	15.79%	56.03%	35.91%	26.67%	33.11%	29.89%
LBP	15.79%	47.37%	31.58%	40.00%	42.22%	41.11%

Table VIII. RSR(Recognition Success Rate) for optical flow feature using different classifiers on the testing set(76 videos) and dev set(60 videos).

	Test				Dev			
	Real	FPA	FVA	HVA	Real	FPA	FVA	HVA
kNN(**k**=5)	84.21%	94.74%	42.11%	73.68%	100%	100%	60.00%	60.00%
SVM	57.89%	100%	42.11%	57.89%	73.33%	100%	73.33%	60.00%
Softmax	52.63%	36.84%	10.53%	47.37%	40.00%	53.33%	66.67%	46.67%

Table IX. FAR(False Rejection Rate), FRR(False Acceptance Rate) and HTER(Half Total Error Rate) for optical flow feature using different classifiers on the testing set(76 videos) and dev set(60 videos).

	Test			Dev		
	FAR	FRR	HTER	FAR	FRR	HTER
kNN(**k**=5)	15.79%	29.82%	22.81%	0.00%	26.67%	13.33%
SVM	42.11%	33.33%	37.72%	26.67%	33.34%	33.01%
Softmax	47.37%	68.42%	57.90%	60.00%	44.44%	52.22%

feature based anti-spoofing method would focus on decreasing the influence by external illumination change.

REFERENCES

[1] K. Nixon, V. Aimale, and R. K. Rowe. Spoof detection schemes. *Handbook of Biometrics*, 2007.

[2] Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid. Face anti-spoofing based on color texture analysis. *International Conference on Image Processing(ICIP)*,2015,pp.2636-2640.

[3] Sajida Parveen, Sharifah Mumtazah Syed Ahmad, Marsyita Hanafi, Wan Azizun Wan Adnan. The Design and Compilation of a Facial Spoof Database on Various Textures. *International Conference on Artificial Intelligence with Applications in Engineering and Technology (ICAIET)*,2014,pp.182-186.

[4] Zhiwei Zhang, Dong Yi, Zhen Lei, Stan Z. Li. Face liveness detection by learning multispectral reflectance distributions. *International Conference on Automatic Face and Gesture Recognition and Workshops (FG)*,2011,pp.436-441.

[5] K. Kollreider, H. Fronthaler, J. Bigun. Evaluating Liveness by Face Images and the Structure Tensor. *Automatic Identification Advanced Technologies (AutoID)*,2005,pp.75-80.

[6] K. Kollreider, H. Fronthaler, M.Faraj, J. Bigun. Real-Time Face Detection and Motion Analysis With Application in "Liveness" Assessment. *Transactions on Information Forensics and Security*,2007,pp. 548-558.

[7] K. Kollreider, H. Fronthaler, J. Bigun. Verifying liveness by multiple experts in face biometrics. *Computer Vision and Pattern Recognition Workshops(CVPRW)*,2008,pp.1-6.

[8] Chao-kuei Hsieh, Shang-Hong Lai, and Yung-Chang Chen. Expressional face image analysis with constrained optical flow. *International Conference on Multimedia and Expo(ICME)*,2015,pp. 1553-1556.

[9] Chao-kuei Hsieh, Shang-Hong Lai, and Yung-Chang Chen. Expression-Invariant Face Recognition With Constrained Optical Flow Warping. *Transactions on Multimedia*,2009,Vol.11,pp.600-610.

[10] Andreas Ranftl, Fernando Alonso-Fernandez, and Stefan Karlsson. Face Tracking Using Optical Flow. *Biometrics Special Interest Group (BIOSIG)*,2015,pp.1-5.

[11] Yutong Gao, Xuewei Lv, and Hongyan Jia. Real-time multi-view face detection based on optical flow segmentation for guiding the robot. *Fuzzy Systems and Knowledge Discovery (FSKD)*,2015, pp.2371-2377.

[12] Chia-Ming Wang, Hsu-Yung Cheng, Kuo-Chin Fan, Chih-Chang Yu, and Feng-Yang Hsieh. Distinguishing falsification of human faces from true faces based on optical flow information. *International Symposium on Circuits and Systems(ISCS)*, 2009, pp.2609-2612.

[13] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. *International Conference on Image Analysis and Signal Processing(IASP)*,2009, pp.233-236.

[14] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence 17*,1981,pp.185?203.

[15] Gunnar Farnebäck. Two-Frame Motion Estimation Based on Polynomial Expansion. *in Computer Science*,2003.pp.363-370.

[16] Gunnar Farnebäck. Polynomial Expansion for Orientation and Motion Estimation. *in Linköping University Sweden*,2002.

[17] Ivana Chingovska, André Anjos, Sébastien Marcel. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. *Biometrics Special Interest Group(BIOSIG)*,2012,pp.1-7.