

# Face Liveness Detection by Exploring Multiple Scenic Clues

(Invited Paper)

Junjie Yan    Zhiwei Zhang    Zhen Lei    Dong Yi    Stan Z. Li

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences, Beijing, China

Email: {jjyan, zwzhang, zlei, dyi, szli}@nlpr.ia.ac.cn

**Abstract**—Liveness detection is an indispensable guarantee for reliable face recognition, which has recently received enormous attention. In this paper we propose three scenic clues, which are non-rigid motion, face-background consistency and imaging banding effect, to conduct accurate and efficient face liveness detection. Non-rigid motion clue indicates the facial motions that a genuine face can exhibit such as blink, and a low rank matrix decomposition based image alignment approach is designed to extract this non-rigid motion. Face-background consistency clue believes that the motion of face and background has high consistency for fake facial photos while low consistency for genuine faces, and this consistency can serve as an efficient liveness clue which is explored by GMM based motion detection method. Image banding effect reflects the imaging quality defects introduced in the fake face reproduction, which can be detected by wavelet decomposition. By fusing these three clues, we thoroughly explore sufficient clues for liveness detection. The proposed face liveness detection method achieves 100% accuracy on Idiap print-attack database and the best performance on self collected face antispoofing database.

## I. INTRODUCTION

The aim of face recognition (FR) is to recognize people's identity in a simple but efficient manner as people do in daily lives. Especially at a time when social security is urgently needed, there are sufficient reasons to believe that human identification techniques such as FR will play a more and more important role in areas such as access control, public security monitoring and human-computer interaction (HCI), etc.

Despite the success, it is recently known that FR is vulnerable to fake face attacks. People's faces can be captured easily in these days, even covertly, which makes face no longer a secret biometric feature. Surveillance cameras in the public, digital cameras and mobile phones, or just Google on the Internet, people's faces can be obtained easily. Moreover, with the aid of modern technology, these captured faces can be shown in various mediums: high quality photo, LCD screen, or 3D masks. Fake faces can appear in all kinds of manners which are probably hard to distinguish from genuine ones. For example, the latest Android phones has face recognition system, but unfortunately, it can be easily attacked by photos.

Facing the emerging security need of face biometrics, researchers proposed several methods to tackle the problem of face liveness detection. Generally speaking, there are three categories of solutions for this issue. The first category is multi-spectral approach which utilizes multi-spectral illumination and imaging. As traditional fake methods are only for visible

light [20], fake faces will probably exhibit quite different appearance under other spectrums, which provides clue for liveness detection. Methods of this category firstly selected specific wavelengths for illumination, and then a classification process was adopted for final decision. Although promising results were given in [7], [12], [20], these techniques all required special expensive hardware, making them not as convenient as other visible light based methods in real world applications.

The second category is challenge-response approach based on an intuitive idea that genuine people can respond to certain challenges. The challenge-response manner works as follows: the FR system gives out an order (challenge), and expects the subject to act correspondingly (response); if this response is successfully detected, then the subject is classified as genuine. Responses are usually the biological motions, such as blinking, mouth movement, head rotation, etc. Among them, blinking is perhaps the most commonly used response. In [11], a boosting classifier was trained to describe the eye state and the state transition process was modeled as a conditional random field (CRF). In [19], [10] various features were adopted and boosting was applied as classifiers as well. Besides blinking, mouth movement is also applied as well as the application of voice. In [2], subjects were requested to speak certain words, and by analyzing the mouth movement and the voice waveform, liveness could be determined. Considering that head is a 3D structure, the usage of 3D information can effectively prohibits the attack of 2D planar fake faces such as photos. Using optical flow, [8], [9] calculated the motion vector of head rotation as a supplement to blinking detection. Despite their success, the above methods impose heavy interaction demand on the users, and may fail if only very ambiguous facial motion exists.

The third category is based on texture analysis. [17] applied DoG and LTV algorithm to extract liveness feature, and a very complex classifier is sequentially trained. In [6], an approach of multiple LBP extraction and SVM training was proposed and showed better result than [17]. Recently, [1] released a database on 2D liveness detection for IJCB'11 antispoofing contest, and [6] achieved prefect performance. However in real applications, different imaging conditions may cause serious texture variations, which may fail the texture based approaches. Another problem is that these methods only utilize single

image for liveness detection while the temporal information is discarded, which may lose useful liveness information on more challenging scenes.

In this paper, we focus on the common 2D anti-spoofing issue and propose three novel liveness clues in both temporal and spatial domain. Generally speaking, genuine faces can exhibit non-rigid facial motions, while many fake faces cannot. This difference forms our first liveness clue named non-rigid motion clue. Furthermore, fake faces always reside on certain displaying medium. This reliance makes the fake facial motion is highly consistent with the background motion. This is our second liveness clue named face-background consistency. These two clues are all explored in temporal domain. Since the 2D fake faces are usually printed by a printer or showed through LCD screens, banding effect [5], which is absent in a genuine face, probably will exist on the fake images, due to the quality degradation in reproduction. This is the third liveness clue named banding effect, and is explored in spatial domain. The three clues are fused to give the final decision. Different from previous works which only extract certain feature and train “black box” classifier, our proposed liveness clues have clear semantic definition. Moreover, compared with other methods, our method need fewer training samples, thus it is expected to have better generalization capability.

The remaining of the paper is organized as follows: in Section 2 we briefly introduce the proposed three liveness clues; in Section 3 detailed algorithms of detecting these clues are discussed. In Section 4 experiments on two different databases are given and finally in Section 5 we conclude this paper.

## II. LIVENESS CLUES DETECTION

The liveness detection scenario is described as follows: given an input video of a subject, decide whether or not the target face comes from the genuine human being.

Certain liveness clues are exhibited in the video which provide the information needed for final decision, and the task for liveness detection is to find what they are. Generally speaking, these liveness clues have the following properties:

- Multiple exhibition manners. Many facial phenomena can be adopted as liveness clues, and for a genuine person, it is natural to exhibit one or several of them.
- Expressed at different levels. Some liveness clues, such as blinking, are so evident that observers can immediately recognize them; while some others may be so ambiguous that they are too difficult to be captured, such as the slight movement of facial muscles.

Based on the above observations, we herein propose three clues for liveness detection. These clues reside on both spatial and temporal domain: non-rigid facial motion and face-background consistency possess temporal information while image banding effect involves spatial information.

### A. Non-rigid Motion Analysis

The essence of facial non-rigid motion analysis is to find the non-rigid motion patterns in local face regions. For position-fixed frontal face images, this problem is quite easy since

only non-rigid motion exists. However in real practice, faces probably will also exhibit rigid motions such as head shift or rotation, which makes the problem more difficult. The issue is that we have to extract the facial regional non-rigid motion from the facial global rigid motion.

We adopt the batch image alignment method to separate the non-rigid motion from the rigid ones. Batch image alignment utilizes a series of rigid transformation to align several images to a fixed image, with the residual being the non-rigid motion. Inspired by [13], the low rank matrix decomposition approach is used here.

Specifically for our problem, faces are detected at every frame and roughly aligned according to the detected eye position. Then let them form a single matrix  $I = [I_1, \dots, I_n]$  where  $I_i$  is the 1D vector by raster scanning the original 2D image matrix and  $n$  is the number of frames. The essential appearance of the same face for all these frames should be highly correlated, or even identical. However, considering the existence of facial rigid and non-rigid motion, the observation  $I$  is actually corrupted as follows:

$$I = I^0 \circ \tau + E \quad (1)$$

where  $I^0$  is the essential low rank matrix,  $\tau$  is the affine transformation matrix which is used to model rigid facial motion like face pose and position change, and  $E$  is the sparse residual caused by non-rigid facial motion. Every frame  $I_i$  has a corresponding transformation  $\tau_i$  which is used to transform  $I_i^0$  to  $I_i$ , and the non-rigid motion which cannot be modeled by  $\tau_i$  is consider as residual  $E_i$ . Our goal is to recover these components to find the true non-rigid facial motion information from  $E$ . Since  $I^0$  is essentially low rank and  $E$  is sparse (facial regions always occur in certain local regions such as eye and mouth), the problem can be formulated as the following optimization problem:

$$\{I^0, E, \tau\} = \arg \min \text{rank}(I^0) + \lambda \|E\|_0, I \circ \tau = I^0 + E \quad (2)$$

Since the  $\text{rank}(\cdot)$  and  $\|\cdot\|_0$  are nonconvex and discontinuous operations, Eq. 2 is too difficult to be effectively solved. Herein we follow [13] and solve the following relaxations problem instead:

$$\{I^0, E, \tau\} = \arg \min \|I^0\|_* + \lambda \|E\|_1, I \circ \tau = I^0 + E \quad (3)$$

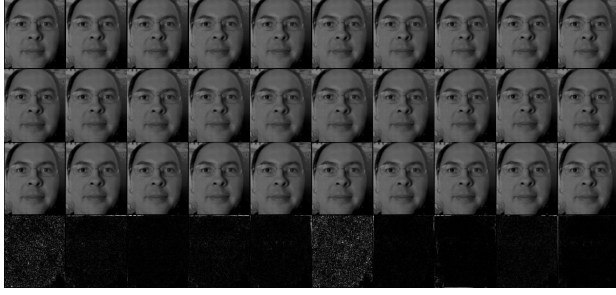
where  $\|\cdot\|_*$  is the nuclear norm and  $\|\cdot\|_1$  is the  $\ell^1$  norm.

Some illustrative examples are shown in Fig.1. We can find that the proposed method can detect the non-rigid motion of genuine face sequence. For genuine face sequence, there are some frames that have large sparse residual in eye regions which corresponds to eye-blinking or in mouth regions which corresponds to mouth movement. There may be some residual in fake face video, such as the 1st and 6th frame in Fig. 1(b), however they are uniformly distributed noises in the whole face region. By defining regional feature, we can distinguish the two kinds of residual easily.

Once the sparse non-rigid motion matrix  $E$  was obtained, we can use it to extract liveness information. We assign  $\Omega_i$  as



(a) genuine face sequence



(b) fake face sequence

Fig. 1. Example of non-rigid motion analysis for liveness detection: the process illumination. 1st row: automatic detected faces,  $I$  in Eq. 2; 2nd row: automatic detected faces after alignment,  $I \circ \tau$  in Eq. 2; 3rd row: low rank face sequence,  $I^0$  in Eq. 2; 4th row: Non-rigid residual,  $E$  in Eq. 2. (a) is a genuine face video while (b) is fake face video.

the interesting facial region. In our experiment left eye region  $\Omega_1$  and right eye region  $\Omega_2$  are selected as two interesting facial regions. Facial non-rigid motion feature  $S_i, i = 1, 2$  are formulated as follows:

$$S_i = \max_{j=1,2,\dots,n} \left\{ \frac{\sum_{(x,y) \in \Omega_i} |E(x,y)|}{\sum_{(x,y) \in Face_j} |E(x,y)|} \times \frac{\sum_{(x,y) \in Face_j} 1(x,y)}{\sum_{(x,y) \in \Omega_i} 1(x,y)} \right\} \quad (4)$$

where  $\sum_{(x,y) \in \Omega_i} |E(x,y)|$  is the residual of the  $j$ th frame in region  $\Omega_i$  and  $\sum_{(x,y) \in Face_j} |E(x,y)|$  is the residual of the  $j$ th frame in the face region.  $\sum_{(x,y) \in (\cdot)} 1(x,y)$  is area of a region which is used to remove the effects of area scale. If a face sequences have a big  $S_i$  like Fig. 1(a), the corresponding face is considered genuine; otherwise fake. As an example, we give the distribution of  $S_1$  and  $S_2$  on Idiap print-attack database [1] in Fig. 2. From this figure we can find that genuine and fake faces are separated. The final feature extracted in non-rigid analysis is  $(S_1, S_2)$ . Since it has only two dimension, it has a better generalization ability and can be well trained by a few number of training samples. By adding additional region, such as mouth, the proposed method can be used to detect other non-rigid facial motions as well.

### B. Face-Background Consistency Analysis

In practice we find that mere non-rigid motion detection is insufficient, since some genuine people may only have

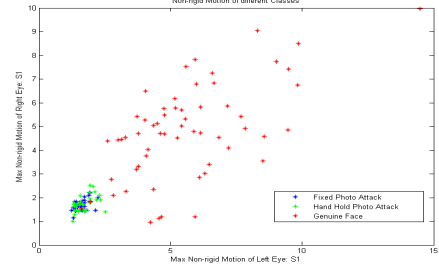


Fig. 2. Non-rigid motion of genuine and photo attacks. Red point are genuine face points while blue and green points are two kind of fake faces.

slight motion that can't be detected through batch image alignment. In this part, we explore and design face-background consistency feature as a complementary to non-rigid motion approach.

The face-background consistency is based on the observation that if the target face is genuine, its motion should be totally independent from that of the background; otherwise the motion relationship between face and background cannot be totally independent because of the exhibition medium's constraint. Our task is thereby to define and explore such motion consistency.

Firstly the scene motion has to be captured. As the exact motion information is not required, we instead adopt the Gaussian Mixture Model(GMM) [16] based background modeling method [16] to describe the motion in the scene. The reason why we use GMM instead of dense optical flow is that GMM is more efficient while robust to illumination and noise. In GMM the scene motion is described in an indirect but efficient manner, which works well in our following experiments. Examples of the genuine face, hand fake face photo, fixed fake face photo and their detected motion by GMM are shown in Fig.3. After some initialization, GMM outputs the foreground-background binary image  $B_i, i = 1 \dots n$  of a given frame  $I_i$ :

$$B_i(x,y) = \begin{cases} 1 & \text{if (x,y) belongs to foreground} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Afterwards the face and background regions are determined automatically according to a Viola-Jones face detector [18]. As face detection has already been applied to every frame, the face and background regions are roughly separated according to the detected facial positions.

Then within the face and background regions, we record the motion trends as time elapses. If the respective motion trends of the face and background regions are very similar to each other, or have high consistency, the face is probably fake; otherwise, if there is low consistency, the face is probably an independent existence, indicating that it is a genuine face. We define  $mt_j, j = 1, 2$  to describe the motion trends.

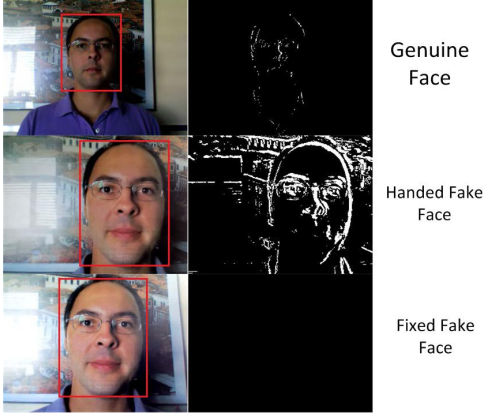


Fig. 3. Face-background consistency analysis for liveness detection: the process illumination. left is a frame from a video while right is detected motion of the frame. 1st row: a frame of the genuine face video; 2nd row: a frame of the handed fake face video; 3rd row: a frame of the fixed fake face video. The red bounding box corresponding to automatically detected face.

$$mt_j(i) = \frac{\sum_{(x,y) \in \Omega_j} B_i(x,y)}{\sum_{(x,y) \in \Omega_j} 1(x,y)} \quad (6)$$

where  $\sum_{(x,y) \in \Omega_j} B_i(x,y)$  is the number of foreground pixels,  $\sum_{(x,y) \in \Omega_j} 1(x,y)$  is the area of region  $\Omega_j$ . In this part  $\Omega_1$  is face region while  $\Omega_2$  is background region. The equation above represents the motion within the scene by measuring the motion area (foreground).

The motion trends are the vectors composed of face region  $mt_1(i)$  and background region  $mt_2(i)$  for every frame, and the divergence between  $mt_1$  and  $mt_2$  can be formulated as the  $\chi^2$  distance, which we call consistency motion distance(CMD) here.

$$CMD = \sum_{i=1}^n \frac{(mt_1(i) - mt_2(i))^2}{mt_1(i) + mt_2(i)} \quad (7)$$

The consistency motion distance of genuine face videos are larger than those of fake face videos since  $mt_1$  and  $mt_2$  are independent. The CMD value is taken as the firstly dimension of the face-background consistency feature.

In Fig. 4 we show three examples. We can find that for the genuine face video, there exists large divergences in motion trends between the face regions and background regions; meanwhile the motion trends of face and background regions are very similar for both handed fake face video and fixed fake face video.

The motion of hand hold photo attack attack videos are global while the motion of genuine videos are local. This inspires us to use the total motion as another measure. The motion entropy  $me(i)$ ,  $i = 1, 2, \dots, n$  is defined as:

$$me(i) = -p_i \log(p_i) - (1 - p_i) \log(1 - p_i) \quad (8)$$

where  $p_i$  is the total foreground radio:

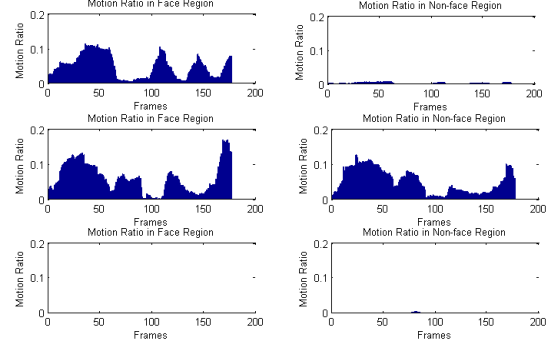


Fig. 4. Example of motion trends. Left is motion trend of face regions  $mt_1$  while right is motion trend of background regions  $mt_2$ . 1st row: genuine face video; 2nd row: handed attacking face video; 3rd row: fixed attacking face video.

$$p_i = \frac{\sum_{(x,y) \in \Omega_1 \cup \Omega_2} B_i(x,y)}{\sum_{(x,y) \in \Omega_1 \cup \Omega_2} 1(x,y)} \quad (9)$$

In some situations,  $p_i = 0$  then  $me(i)$  cannot be calculated, so that we use  $p_i = \min\{p_i, 1/pixelnumber\}$  to force smoothness. To describe all the frames of a video, the motion entropy histogram is used. Fig. 5 shows an example, from which we can see that there's a big difference between genuine and fake face motion entropy histogram. The motion entropy of handed fake face is very large compared with genuine face, while the motion entropy of fixed fake face is close to 0. The motion entropy is taken as the second dimension of the face-background consistency.

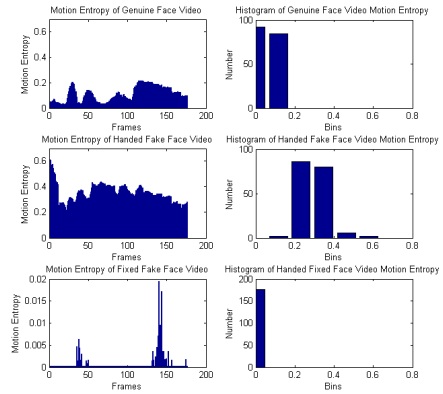


Fig. 5. Example of Motion Entropy. 1st row: genuine face video; 2nd row: handed fake face video; 3rd row fixed fake face video. Left is motion entropy of every frame, right is the histogram of motion entropy in the total video.

### C. Image Banding Analysis

Since fake faces are usually exhibited by certain medium, inevitably some quality degradation will be introduced during the imaging process. For photo based attacks, for instance,

TABLE I  
RESULT OF OUR PROPOSED METHODS ON INDIAP PRINT-ATTACK DATABASE

Methods	Background	Accuracy
Non-rigid Motion Analysis	all background	90%
Face-Background Consistency Analysis	complex background	97.50%
Image banding Analysis	uniform background	97.50%
Fusion	all background	<b>100%</b>
PLS[15]	all background	99.375%
MTA [6]	all background	<b>100%</b>

print banding, jitter, and ghosting artifacts are all common reasons for quality degradation as described in [5], [14]. Here we focus on the banding effect. An example of banding effect is shown in Fig. 6, from which we can find banding noise on fake face image. Note that such kind of noises also exist in video attacks. In a video attack case, the size of each LCD screen emitting pixel is probably different from the sensing pixel of the imaging camera. Furthermore, the flash frequency of the LCD may not match the exposure frequency of the camera. The above two mismatches will cause the strobe/flash effect in the captured image, which is very obvious banding noises.



Fig. 6. Banding effect. Left is a genuine face without banding effects while right is a fake face with banding effects.

Different from methods used in print banding estimation, we find a very simple but effective way to do print banding estimation in face liveness detection by wavelet decomposition. As described in [4], [3], image noise can be estimated by a robust median estimator as follows:

$$\hat{\sigma} = \frac{\text{Median}\{|y(i, j)|\}}{0.6745} \quad y(i, j) \in HH1 \quad (10)$$

where HH1 means the first order wavelet decomposition of the image. Specifically in our problem, we treat the banding as a kind of noise added to original image. We use Haar wavelet in our experiment since the shape of Haar wavelet is similar to the banding caused by printer or LCD screen.

The distribution of the median estimator values of test videos with uniform background on Indiap database[1] is shown in Fig. 7. The first 40 instances in Fig. 7 are genuine faces images while other are fake ones. We find that the median estimator values of fake face images are all near 0.74 while most genuine faces are near 0 with only two exceptions. In banding effect analysis, the only feature is  $\hat{\sigma}$ .

#### D. Fusion of Multiple Clues

The proposed three clues are complementary in real applications. The non-rigid motion analysis can capture the liveness

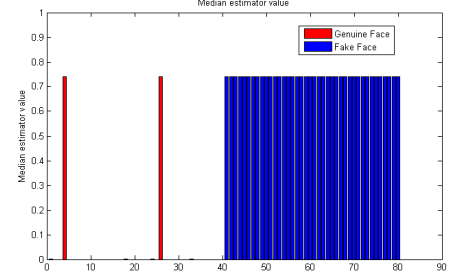


Fig. 7. Median value of Haar wavelet decomposition

clues in videos with facial expressions. The face-background consistency analysis can be used to capture the liveness clues in videos with complex background. For videos with clean background, the face-background consistency is valid, but the proposed image banding effect analysis is suitable since that there's no background influence. In the following experiment, we use the clues based on the image background condition. For the videos with complex background, we use non-rigid motion and face-background consistency clues, while for the videos with clean background, we use the non-rigid motion and the banding effect clues. Since the background condition can be easily estimated by edge detection, our algorithm is fully automatic.

In the experiment, in order to map the feature to score, we learn logistic regression for the three clues respectively on the training set, and the weights of different clues in fusion are learnt on validation using grid search.

### III. EXPERIMENTS

In this section two database are used to verify the proposed algorithm. The first one is the Iadip print-attack database [1], and the second one is a self collected antispoofing database which is more challenging.

#### A. Experiment on Iadip Print-attack Database

The Indiap print-attack database consists of 400 videos, including 200 genuine face videos and 200 attacking videos. The genuine videos are collected in either complex or uniform background with some illumination changes. The attacking faces are printed by color laser printer on A4 papers and are hand held or fixed in front of the camera. In our experiments we follow the protocol proposed in [1], which selects 30% of



the videos as training set, 30% of the videos as development set and other videos as test set.

In the experiment we compare our algorithm with the following two methods:

- **Micro-Texture Analysis(MTA):** The algorithm was proposed in [6], where multiple LBP features are extracted and then fed into a stand SVM classifier.
- **Partial Least Squares(PLS):** The algorithm was proposed in [15], where various features like HOG, CF, GLCM, HSC are extracted and PLS is used to learn the weight of each feature.

The results of these methods are shown in Tab. I. From the table we can see that every single clue can achieve a relative high performance. Non-rigid motion analysis extracts information of local facial movement; face-background consistency analysis reveals the relationship between global face and background motion; image banding analysis examines the image quality. By fusing multiple clues, we can nearly get perfect result. Compared with [6] and [15], all the features have clear semantic meanings and so that has better generalization capability.

#### B. Experiment on Self Collected Antispoofing Database



Fig. 8. Example of Self collected database. 1st row: real facial images; 2nd row: spoofing facial images. 1st and 2nd column: original images; 3rd and 4th column: cropped images.

In this part we collect an antispoofing database to compare our method with MTA. There are 50 subjects in total, with videos recorded in two kinds of imaging qualities. For each imaging quality, every subject has two videos, including one genuine video and one spoof video. Genuine videos are captured in laboratory surroundings, while the spoof facial videos are captured by exhibiting high resolution photos. Compared with Idiap database, this database is more challenging since it includes two kinds of imaging qualities. Since all the videos have very complex background, we only use non-rigid motion clue and face-background consistency clue in this experiment.

The model is trained on randomly selected 20 subjects, and the remaining 30 subjects are used for testing. We implement MTA as [6] except that our cropped face regions are bigger.

The results are shown in Tab. II. We can find that our proposed two clues and their fusion achieve best performance on this database. Although [6] uses radial kernel on Idiap

TABLE II  
EXPERIMENTAL RESULTS ON SELF-COLLECTED DATABASE

Methods	Accuracy
Non-Rigid Detection	81.67%
Face-Background Consistency	90.0%
Fusion of above two clues	<b>91.67%</b>
MTA[6]+Radial Kernel	80.83%
MTA[6]+Linear Kernel	88.83%

database we find that linear kernel is better on our database. From Tab. II we can find that our proposed methods have a better performance than MTA.

#### IV. CONCLUSION

Three novel clues with clear semantic meanings are proposed to handle the face liveness detection problem and we give methods to efficiently detect the three clues. We achieve the state-of-the-art results on two different databases which proves the effectiveness and generalization ability of our methods. Designing more efficient features to represent more robust clues on more challenging situations will be our future work.

#### REFERENCES

- [1] S. M. Andre Anjos. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *IAPR/IEEE IJCB*, 2011.
- [2] G. Chetty and M. Wagner. Liveness verification in audio-video speaker authentication. In *ICSLP*, 2004.
- [3] D. Donoho. De-noising by soft-thresholding. *Information Theory, IEEE Transactions on*, 1995.
- [4] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 1994.
- [5] A. Eid, M. Ahmed, B. Cooper, and E. Rippetoe. Characterization of electrophotographic print artifacts: Banding, jitter, and ghosting. *TIP*, 2011.
- [6] M. P. Jukka, Abdenour Hadid. Face spoofing detection from single images using micro-texture analysis. In *IAPR/IEEE IJCB*, 2011.
- [7] Y. Kim, J. Na, S. Yoon, and J. Yi. Masked fake face detection using radiance measurements. *JOSA A*, 2009.
- [8] Kollreider, K., Fronthaler, H., and J. Bigun. Verifying liveness by multiple experts in face biometrics. In *CVPRW'08*, 2008.
- [9] K. Kollreider, H. Fronthaler, and J. Bigun. Evaluating liveness by face images and the structure tensor. 2005.
- [10] J. Li. Eye blink detection based on multiple gabor response waves. In *Machine Learning and Cybernetics, 2008 International Conference on*, 2008.
- [11] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblick-based anti-spoofing in face recognition from a generic webcam. In *ICCV*, 2007.
- [12] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *CVBVS*, 2000.
- [13] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *CVPR*, 2010.
- [14] N. Rawashdeh, I. Shin, K. Donohue, and S. Love. Printer banding estimation using the generalized spectrum. In *SPIE*, 2006.
- [15] W. Schwartz, A. Rocha, and H. Pedrini. Face spoofing detection through partial least squares and low-level descriptors. In *IJCB*. IEEE, 2011.
- [16] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *CVPR*. IEEE, 1999.
- [17] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. Springer, 2010.
- [18] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 2004.
- [19] C. Xu, Y. Zheng, and Z. Wang. Eye states detection by boosting local binary pattern histogram features. In *ICIP*, 2008.
- [20] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li. Face liveness detection by learning multispectral reflectance distributions. In *FG*. IEEE, 2011.