Face Liveness Detection Using 3D Structure Recovered from a Single Camera

Tao Wang Jianwei Yang Zhen Lei Shengcai Liao Stan Z. Li*

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, China
{twang,jwyang,zlei,scliao,szli}@cbsr.ia.ac.cn

Abstract

Face recognition, which is security-critical, has been widely deployed in our daily life. However, traditional face recognition technologies in practice can be spoofed easily, for example, by using a simple printed photo. In this paper, we propose a novel face liveness detection approach to counter spoofing attacks by recovering sparse 3D facial structure. Given a face video or several images captured from more than two viewpoints, we detect facial landmarks and select key frames. Then, the sparse 3D facial structure can be recovered from the selected key frames. Finally, an Support Vector Machine (SVM) classifier is trained to distinguish the genuine and fake faces. Compared with the previous works, the proposed method has the following advantages. First, it gives perfect liveness detection results, which meets the security requirement of face biometric systems. Second, it is independent on cameras or systems, which works well on different devices. Experiments with genuine faces versus planar photo faces and warped photo faces demonstrate the superiority of the proposed method over the state-of-the-art liveness detection methods.

1. Introduction

Face recognition system, due to its fast development during last decades, has been widely used in our daily life, such as access control, visual surveillance and compute application security. However, most traditional face recognition systems are vulnerable to direct sensory attacks. That is, a simple printed photo or a photo demonstrated on a screen can easily fool the system and an invalid user may gain the access control which may result in severe security problem. Face liveness detection, which aims to judge the face biometric captured from a genuine person or a fake replica, is becoming a critical technique for traditional face recognition system.

*Corresponding author.

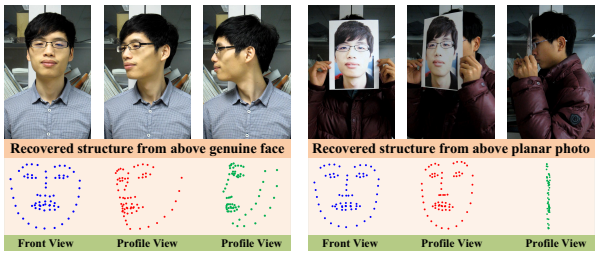


Figure 1. A comparison of recovered sparse 3D facial structures between genuine and photo face. There are significant differences between structures recovered from genuine and photo face.

According to the information (features) used, the existing face liveness detection methods can be roughly divided into three main categories: challenge-response based methods [12, 2, 8, 16], skin property based methods [15, 10, 13, 18] and 3D structure based methods [3].

The challenge-response is a human-computer interaction (HCI) method. The users are asked to response to the specific facial actions given by computer so that the genuine and fake faces can be classified. Eye blinking [12, 16], head rotation [2, 8], and mouth movement [8] are the most commonly used facial actions. The challenge-response methods achieve good results for face liveness detection. However, the users are asked to be highly cooperative with the system, which limits its application in practice.

The skin property based methods classify the genuine and fake face images by analyzing their textures or reflectance properties. Tan *et al.* [15] extracted the latent reflectance information from a captured image, and used a sparse low rank bilinear discriminative model to classify the genuine and fake faces. Maata *et al.* [10] analyzed facial image textures using multi-scale local binary patterns (LBP) to classify the genuine and fake faces. These methods assume the quality of face images is different with that of real ones and more noise and artifacts are included in fake images. However, with the development of printing technology, it is increasingly difficult for these methods to work reliably.

Recently, the multi-spectral methods have been proposed to analyze the reflectance properties of human skin so that the genuine and fake faces can be classified. In [13], Pavlidis and Symosek illustrated how to capture face images at two wavelengths in Near Infrared (NIR), and used a thresholding method to classify the genuine and fake faces. Zhang *et al.* [18] proposed a method for face liveness detection by analyzing the energy under multi-spectral. The shortcoming of multi-spectral method is that it needs additional devices to capture multi-spectral images which is not always applicable in practice.

3D structure based methods make use of 3D structure information to classify the genuine and fake faces. Obviously, a planar photo gives a flat structure whereas a genuine face yields a quite different structure (e.g., nose is convex compared to cheek). To the best of our knowledge, few studies have been devoted to this kind of method, among which the work of Choudhury *et al.* [3] is the most representative one. They mentioned that the depth information can be used for differentiating planar faces and real ones, but no further experiments were conducted in their work.

In this paper, we propose a novel face liveness detection method by analyzing the sparse structure information in 3D space. As shown in Fig. 1, structures recovered from genuine faces usually contain sufficient 3D structure information, while structures recovered from photos are usually planar in depth. We firstly recover the sparse 3D facial structure with the input images and then an SVM classifier is trained based on the 3D structures from the genuine and fake faces. Different from the reflectance based methods, the proposed method is device independent and can work well with various inputs as long as two images from different viewpoints are provided.

The rest of the paper is organized as follows: Section 2 and 3 details our sparse structure recovery method and the classification method, respectively. Experiments compared with recently proposed LBP method are demonstrated in Section 4 and in Section 5, we conclude the paper.

2. Sparse 3D Facial Structure Recovery

The face images with different viewpoints can be captured from a fixed camera with face moving or reversed. But the recovery is conducted under the assumption that the images are derived from a static face and a dynamic camera.

In this work, we firstly use CLM algorithm [14] to locate the sparse facial landmarks. Considering that neighboring frames almost have the same viewpoint, which is useless for structure recovery, we select proper frames using a graph similarity metric in the next step. We call such frames as key frames, which have much diversities in terms of viewpoint.

Once we have obtained two key frames, the camera parameters and initial facial structure are recovered. Notably, general geometry reconstruction algorithm is not applicable to deformable object, such as face. However, face deformations are often caused by local expressions, which occurs most likely in the regions of mouth. Among all the landmarks, only a bit of them locate near mouth. Therefore, we recover facial structures regardless of the deformation, which is proved to be feasible to face liveness problem in our experiments.

Finally, facial structure refinement is taken to refine the initial recovered results and new key frames are added into the bundle adjustment because that the structure recovered from merely two face images may be affected by imprecise detection of landmarks and inaccurate estimation of camera parameters.

2.1. Key Frame Selection

In our work, key frames are defined as those frames which are propitious to recover facial structure. We rely on graph similarity to incrementally extract key frames from an input sequence which are more likely to be those captured in various viewpoints in practice. In the case that a certain relative motion between the face and the camera happens, it is obvious that locations of facial landmarks in the image will be various along with the motion. That is, different viewpoints will lead to different distributions of facial landmarks. Based on this relation, we compute the distribution distance of landmarks to evaluate the difference of viewpoints between current frame and existing key frames in pool. Specifically, we represent the landmarks in a face by an undirected graph $G =< V, E >$, where V represents landmarks and E connects all the nodes. To measure the similarity between two graphs, we build an affinity matrix $W_{68\times 68}$ for each graph, whose entries are the spatial Euclidean distance between two nodes. Assume we have obtained M key frames $\boldsymbol{F} = \{f_1, f_2, ..., f_M\}$, and their corresponding affinity matrixes $\{W_1, W_2, ..., W_M\}$ as well. Then, the maximum similarity (minimum distance) of current frame f_c to previous M key frames is obtained by:

$$S = \max_i exp\left(-\left\| \frac{W_{f_c}}{\parallel W_{f_c} \parallel_*} - \frac{W_i}{\parallel W_i \parallel_*} \right\|_*\right) \quad (1)$$

where $\| \cdot \|_*$ represents the Frobenius norm. If S is smaller than a pre-determined threshold (in this paper, we set it 0.1), we regard current frame f_c as a key frame, and add it to $\boldsymbol{F}$; otherwise, we turn to the next frame.

2.2. Initial Recovery from Two Images

In this subsection, we describe our algorithm for initial facial structure recovery based on facial landmarks from two face images with different viewpoints. Let $\boldsymbol{q} = \{q_1, q_2, ..., q_N\}$ be the facial landmarks in one image and $\boldsymbol{q'} = \{q'_1, q'_2, ..., q'_N\}$ in another image, where q_i denotes a 2D point in the image, N is the number of facial landmarks for a face (N=68 in our experiment). In our work,

we adopt the perspective camera model [6], in which a 2D point q_i projected from a 3D structure point Q_i can be represented as: $q_i = PQ_i$, where P is a 3×4 perspective projection matrix. Both q_i and Q_i are in homogeneous coordinates. Camera projection matrix P can be decomposed as: $P = K(R, t)$. Before initial structure recovery by triangulation [7], we have to estimate camera intrinsic matrix K and the relative pose (R, t) between the two given images.

In order to make our approach work well on different cameras, we propose an auto-calibration method for our face application. In this work, we make the common assumption that the intrinsic matrix K is constant over the whole video sequence. Generally, the intrinsic matrix is formulated as a 3×3 upper triangular matrix [6]:

$$K = \begin{pmatrix} f_x & s & u_x \\ 0 & f_y & u_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

where f_x and f_y are the focal length in terms of pixel dimensions in the direction of x-axis and y-axis, respectively. u_x and u_y are the projection of optical center. s is referred to as the skew parameter. Suppose that the camera sensor pixels are in the shape of square, and the projection center is coincide with the image center, we have $f_x = f_y = f$, and $u_x = u_y = 0$ if the origin of the image coordinate is set at the center of image. The skew parameter s is equal to zero for most normal cameras. As a result, the intrinsic matrix can be simplified with only one parameter f, that is, $K = diag(f, f, 1)$. To derive K, we utilize a predefined normal (upright and frontal) 3D facial structure $\boldsymbol{Q^{model}} = \{Q_1^{model}, Q_2^{model}, ..., Q_N^{model}\}$ recovered from a calibrated camera. $\boldsymbol{Q^{model}}$ is scaled to the physical size of a face in our work (i.e., 10 cm between the center of two eyes). Thus, given $\boldsymbol{Q^{model}}$ and facial landmarks $\boldsymbol{q}$, we can roughly estimate K by optimizing the following equation:

$$K^* = \arg\min_K \sum_{i=1}^{N} d\left(q_i, K(R_m, t_m)_{3\times4} Q_i^{model}\right) \tag{3}$$

where $(R_m, t_m)_{3\times4}$ represents the rigid transformation (rotation matrix $R_{3\times3}$ and translation $t_{3\times1}$) on $\boldsymbol{Q^{model}}$, $d(\cdot)$ computes the Euclidean distance between two given points represented in homogeneous coordinates. In Eq. (3), estimating $\{K, R_m, t_m\}$ simultaneously is intractable. Alternatively, we traversing all possible (R_m, t_m) to find the best matching. Let ω_x, ω_y, ω_z be the rotation angles of $\boldsymbol{Q^{model}}$ whose 2D projection is similar to $\boldsymbol{q}$, in three axes, respectively. First, we rotate the landmarks around z-axis (perpendicular to image plane) to make the pair of eyes horizontal. This rotation can only get an upright face projection ($\omega_z = 0$), whereas ω_x and ω_y may be non-zero for deflecting viewpoint. Therefore, we uniformly discretize the rotations around x-axis and y-axis to n_x and n_y states, respectively. Then, based on Least Square Error (LSE) algorithm, we find the optimal K with minimum residual error by traversing all the $n_x \times n_y$ possible ω_x and ω_y (in this paper, $n_x = n_y = 20$). There is one point need to be known that the above procedure only gives a rough K, which will be further refined in the following step. Clearly, if the camera has been calibrated accurately, we can skip above procedure. However, knowing exact intrinsic matrix is not feasible in many circumstances practically.

Then, we extract the relative pose (R, t) based on landmark correspondences $\{(q_1, q_1'), (q_2, q_2'), ...(q_N, q_N')\}$. According to [6], essential matrix ε has the form of $\varepsilon = [t_\times]R$, where $[t_\times]$ is the skew-symmetric cross-product matrix of $t = (t_x, t_y, t_z)^T$:

$$[t_\times] = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \tag{4}$$

Once the essential matrix is estimated, camera relative pose can be extracted straightforwardly. So, the question is how to compute the essential matrix ε. It can be estimated using the epipolar constraint $q_i^T \varepsilon q_i' = 0$. The general linear algorithm [6] needs eight or more point correspondences and minimize epipolar distance over all points. In our case, however, the facial landmarks may be detected with noises. We adopt a random sample consensus scheme (RANSAC) [4] to robustly estimate the essential matrix ε. For each RANSAC loop, a candidate essential matrix is obtained using an efficient non-linear minimal algorithm [11] which needs at least five correspondences. Upon the completion of RANSAC, we choose optimal ε with minimum average residual error.

After getting camera intrinsic matrix K and relative pose (R, t), the projection matrixes of two images is obtained by the following equation:

$$\begin{aligned} P &= K(I, \mathbf{0}) \\ P' &= K(R, t) \end{aligned} \tag{5}$$

where I is a 3×3 identity matrix. Given P and P', we implement a triangulation algorithm [7]:

$$\min \sum_{i=1}^{N} [d(q_i, \hat{q}_i) + d(q_i', \hat{q_i'})] \tag{6}$$

where $\hat{q}_i$ and $\hat{q_i'}$ are the reprojections of the recovered structure $\hat{\boldsymbol{Q}} = \{\hat{Q_1}, \hat{Q_2}, ..., \hat{Q_N}\}$ on two given images. Though key frame selection is conducted, a special situation that viewpoints of two images are too similar is considered in our paper. In other words, the rotation angle θ, which can be obtained from R directly, between the two images is small. In this case, the triangulated 3D points will be far from the real structure, especially in the depth (z-axis). To address

such problem, we add an extra soft constraint on the triangulation, which is defined as:

$$C = \omega(\theta) \sum_{i=1}^{N} d\left(\hat{Q}_i, Q_i^{model}\right) \tag{7}$$

where, ω is the weight function to the soft constraint, which is defined as $\omega(\theta) = exp(-15\theta/\pi)$ in our experiment. This constraint can be regarded as a prior on the facial structure. The recovered structure which is different from a natural face much is less likely to be a correct estimation.

2.3. Facial Structure Refinement

As illustrated above, the facial structure can be recovered based on two images from different viewpoints. However, such recovered results may not be accurate due to inaccurate detection of landmarks or rough estimation of K. In this subsection, we describe the facial structure refinement step to obtain a more accurate facial structure, where new key frame is added into the refinement one by one. Given the projection matrixes $\boldsymbol{P} = \{P_1, P_2, ..., P_M, P_{new}\}$ of refined M key frames and a new key frame, our goal of refinement is to minimize the reprojection error between predicted points and detected landmarks over all key frames:

$$\min \sum_{j=1}^{M+1} \sum_{i=1}^{N} d^2\left(\hat{q}_{i,j}, q_{i,j}\right) \tag{8}$$

where $\hat{q}_{i,j} = P_j \hat{Q}_i$ is the i^{th} predicted point on image j, $q_{i,j}$ is the i^{th} landmark on image j.

First, the pose (external parameters) (R_{new}, t_{new}) of the new key frame is estimated using Grunert's algorithm [5], rather than the method described in above subsection. According to the estimated (R, t), the distance error between predicted points $\hat{\boldsymbol{q}}$ and landmarks $\boldsymbol{q}$ can be computed. For those frames which have large error, we regard them as useless key frames and skip to refine the next key frame. Next, we group the parameters which will be refined together: $\boldsymbol{\Theta} = \{K, \boldsymbol{T}, \hat{\boldsymbol{Q}}\}$ where $\boldsymbol{T} = \{(R_1, t_1), (R_2, t_2), ..., (R_M, t_M), (R_{new}, t_{new})\}$ is the external parameters for every frame. Subsequently, we use a sparse bundle adjustment algorithm [9] to optimize Eq. 8 with the input of $\boldsymbol{\Theta}$, which solves by a fast Levenberg-Marquardt (LM) optimization algorithm.

With the increasing number of key frames, we can recover a more and more accurate facial structure. This process will stop when the 3D structure difference between current recovered result and the last one is less than a threshold. The metric formula is similar to Eq.(7).

3. Liveness Detection

3.1. Structure Alignment

After we recover the sparse 3D structure, we first align these structures and then structure features are extracted for classification. In this work, we use a predefined 3D facial structure as a reference to align a recovered structure using a coarse-to-fine way. Intuitively, given two sets of 3D points, the reference $\boldsymbol{Q^{model}} = \{Q_i^{model}\}$ described in subsection 2.2, and a sample $\boldsymbol{Q} = \{Q_i\}$ need to be aligned, our goal is to estimate the rigid transformation (scale s, rotation R and translation t) which minimizes the sum of distances between these two sets of 3D points.

At first, we take an initial (coarse) alignment on $\boldsymbol{Q}$ to fit $\boldsymbol{Q^{model}}$. The rigid transformation $\{s_0, R_0, t_0\}$ is obtained by minimizing the following equation using Least Square Error algorithm.

$$\{s_0, R_0, t_0\} = arg \min_{\{s,R,t\}} \sum_{i=1}^{N} d(s(R,t)Q_i, Q_i^{model}) \tag{9}$$

The above coarse alignment assigns each 3D point an equal weight in the minimization procedure. So, the solution $\{s_0, R_0, t_0\}$ will be affected by some outliers which are recovered inaccurately. This case will happen when facial landmarks are detected inaccurately or the number of key frames used for recovery is small. Consequently, we use an iterative algorithm to refine $\{s, R, t\}$ step by step. In each iteration, we compute the distances for every point pair $(s(R,t)Q_i, Q_i^{model})$ given current estimation of $\{s, R, t\}$. Afterward, we choose five point pairs with least distances, which we regard as the most reliable correspondences among all the pairs. These five pairs are utilized for estimating new $\{s, R, t\}$. The iterative process proceeds until the five point pairs with least distances do not change or the overall iteration times exceed a predetermined threshold $(T = 10)$.

3.2. SVM Classifier

After alignment, the 3D coordinates of sparse structure are concatenated to form a feature vector. The SVM classifier is then trained based on the features to classify the genuine and fake face samples.

4. Experiments

We evaluate the proposed 3D structure based liveness detection method on three databases, which is compared with state-of-the-art LBP based anti-spoofing method [10]. The recovered sparse 3D facial structures for genuine and fake faces are also presented.

4.1. Database Description

Recently, diverse face databases designed for face liveness detection have been proposed, such as CASIA database

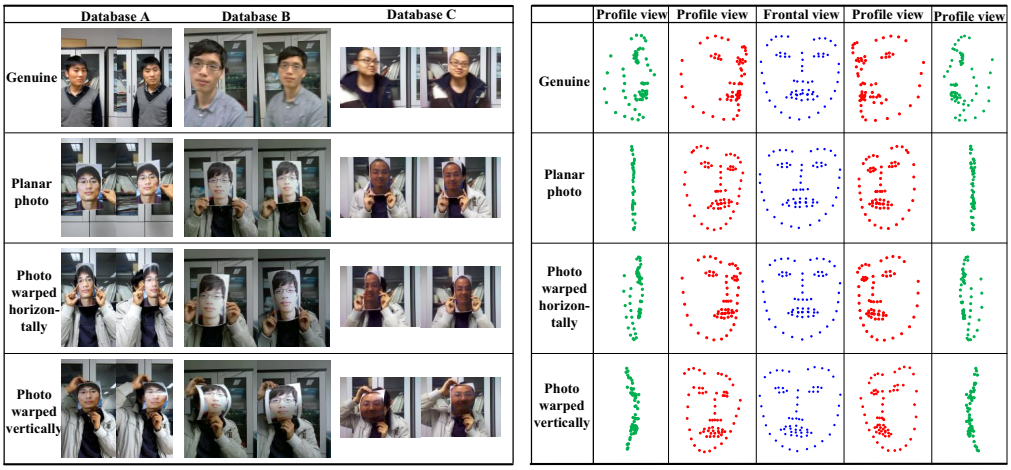


Figure 2. (a) sample frames of a subject from database A, B and C respectively. In each database, we warp the photos horizontally and vertically. These two kind of warping approximate the real face structure much. (b) recovered structures from genuine face, planar photo, photo warped horizontally and photo warped vertically, respectively.

[17], NUAA database [15] and Idiap database [1]. However, these databases do not contain faces with different viewpoints so that they are not proper to evaluate our proposed method.

In this part, we collect three databases using different quality cameras to examine the anti-spoofing performance across different devices. In our experiment, we collect 50 subjects, and both genuine and fake faces are collected 5 times with different motion style. The first one (Database A), second one (Database B) and the third one (Database C) are collected using a high quality camera (Canon IXUS115 HS) with resolution of 1920×1080, a Logitech webcam with resolution of 800×600 and a camera built in NOKIA C6 mobile phone whose resolution is 640×480, respectively. All these three databases records 250 genuine faces and 750 fake photos which include 250 planar photos, 250 photos warped horizontally (warped 1) and 250 photos warped vertically (warped 2). Table. 1 presents the number of video clips in database A, B, C, respectively. Some examples of images sampled from videos for these three databases are shown in Fig. 2 (a).

Table 1. The number of videos in database A, B, C, respectively.

Database	Genuine	Fake photos		
		planar	warped 1	warped 2
Database A	250	250	250	250
Database B	250	250	250	250
Database C	250	250	250	250

4.2. Results and Discussions

Fig. 2 (b) shows the recovered results from genuine faces, planar photo and warped photos. It is easy to see that there are significant differences of 3D structure between the genuine faces and the fake ones, so that they are expected to be well classified based on their 3D structure information.

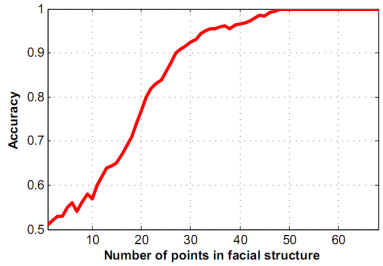
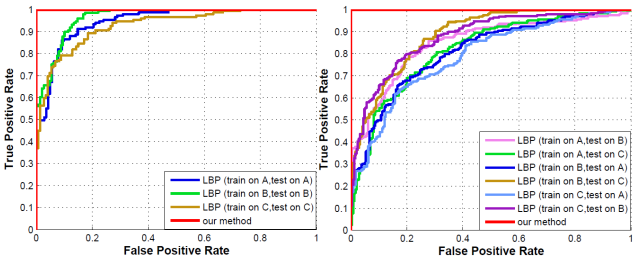


Figure 3. The relation between liveness detection accuracy and the number of 3D points in a structure.

The number of facial landmarks N is a parameter that affects the performance of the proposed method. We examine the affect of N on the anti-spoofing performance on Database A. In this experiment, we select landmarks randomly. For a certain number of landmarks, we conduct 50 trials, and obtain the average accuracy.

Fig. 3 depicts the relation between the number of points N and anti-spoofing accuracy. It is obviously shown that the more facial landmarks, the higher liveness detection accuracy. On the other hand, the more facial landmarks, the more complex of reconstruction with heavier computation cost. Therefore, we set $N = 48$ with the highest detection accuracy (100%) but the least number of landmarks in our experiment.

In this part, we show the face liveness detection performance within and cross devices. In the first experiment, the samples in the training set and testing set are all collected using the same device. For Database A, B and C, 125 subjects are randomly selected to form the training set and the left subjects form the testing set. There is no intersection



(a) (b)

Figure 4. Performance on database A, B, C respectively (a) and on cross database (b).

between the training and the testing sets. Fig. 4 (a) shows face liveness detection results of the proposed method and texture based one (i.e., LBP), whose score is the average value of the face frames of the video. It shows that both the texture and proposed 3D structure based method perform well in this case. The proposed 3D sparse structure method achieves perfect (100%) classification results.

We further test the anti-spoofing performance of different methods in the case of cross devices, which is more common and important in real application. In this part, one of the three databases is selected as the training set and the face liveness detection performance is examined on the other two databases. Fig. 4 (b) shows ROC curves of different methods. The performance of texture based method degraded dramatically in this case. This is because the quality of images captured from different devices varies so much that the model learned from images captured from one camera is not proper any more to other cameras. This is a great problem in real application. In contrast, the proposed method, which takes into account the 3D structure information, is device independent and hence is robust to this variation. As expected, the proposed method also achieves perfect (100%) face liveness detection accuracy, validating its effectiveness in face liveness detection problem.

5. Conclusion

In this paper, a sparse 3D structure based face liveness detection method is proposed. Given at least two face images captured from different viewpoints, the 3D sparse facial structure can be recovered. Based on this, the genuine and fake faces like printed photos can be classified. Different from the popular texture based method, the proposed method is device independent and thus more applicable in practice. Moreover, the requirements of two images from different viewpoints is easy to be met in real application and the proposed method has great potential to be deployed with the existing face recognition system.

Acknowledgement

This work was supported by the Chinese National Natural Science Foundation Project #61070146, #61105023, #61103156, #61105037, #61203267, National IoT R&D Project #2150510, National Science and Technology Support Program Project #2013BAK02B01, Chinese Academy of Sciences Project No. KGZD-EW-102-2, Guangdong-CAS Strategic Cooperation Project #2011B090300054, European Union FP7 Project #257289 (TABULA RASA), and AuthenMetric R&D Funds.

References

[1] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *IJCB*, pages 1–7, 2011.

[2] W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*, pages 233–236. IEEE, 2009.

[3] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland. Multimodal person recognition using unconstrained audio and video. pages 176–181, 1998.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[5] R. M. Haralick, C.-N. Lee, K. Ottenberg, and M. Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision*, 13(3):331–356, 1994.

[6] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2000.

[7] R. I. Hartley and P. F. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[8] K. Kollreider, H. Fronthaler, and J. Bigün. Evaluating liveness by face images and the structure tensor. In *AutoID*, pages 75–80, 2005.

[9] M. I. A. Lourakis, M. I. A. Lourakis, A. A. Argyros, and A. A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Technical report, 2004.

[10] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *IJCB*, pages 1–7, 2011.

[11] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–777, 2004.

[12] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *ICCV*, pages 1–8, 2007.

[13] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *CVBVS*, pages 15–24, 2000.

[14] J. M. Saragih, S. Lucey, and J. F. Cohn. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 91(2):200–215, 2011.

[15] X. Tan, Y. Li, J. Liu, and L. Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *ECCV (6)*, pages 504–517, 2010.

[16] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z. Li. Face liveness detection by exploring multiple scenic clues. In *ICARCV 2012*, 2012.

[17] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li. A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 26–31. IEEE, 2012.

[18] Z. Zhang, D. Yi, Z. Lei, and S. Li. Face liveness detection by learning multispectral reflectance distributions. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 436–441. IEEE, 2011.