Face Spoofing Detection based on 3D Lighting Environment Analysis of Image Pair

Xu Zhang, Xiyuan Hu*, Mingyang Ma, Chen Chen and Silong Peng
Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100190
Email: zhangxu2013@ia.ac.cn, xiyuan.hu@ia.ac.cn, mma8@mmm.com, chen.chen@ia.ac.cn, silong.peng@ia.ac.cn

***Abstract*—In this paper, we present a novel face spoofing detection method based on 3D lighting environment analysis of an image pair collected before and after the lighting environment change. Our idea is inspired from the unimpressive fact that the illumination distributions of the internal spoof face stays stable under the protection of the photo and screen plane, while that of a exposed genuine face changes accordingly to different lighting environment due to a natural response of 3D structure. After estimating two sets of lighting environment coefficients of client's face image pair with the hand of 3D Morphable Model (3DMM) and Sphere Harmonic Illumination Model (SHIM), robust liveness judgement is conducted by hypothesis tests. Experimental results show the effectiveness of proposed method on multiple kinds of face attacks including printed photo, screen photo, and video replay attack, and other advantages such as user cooperation free, loose using conditions, simple equipment demand, easy to camouflage and propitious to face recognition.**

I. Introduction

With the development of face recognition system, face spoofing detection becomes a major issue for urgent solution in the security field. Photo and replay video are often used as spoof face because of their easy access from the web site or surveillance equipment. Based on the different facial clues, the existing methods can be classified into three categories: texture based, motion based, and 3D structure based methods.

Texture based methods [1], [2], [3] assume that texture information on the fake face has been changed due to the shadow, blur and highlight brought by print or screen. Thus these methods distinguish the attack from the genuine access by extracting texture feature representing micro-texture arrangement. However, high resolution camera is needed when capturing the micro-texture change.

Motion based methods assume that the movement of planar faces differs greatly from genuine faces. The real face can blink eyes [4], move lips [5] and change the gaze casually [6]. Furthermore, a real face can response to the system instructions correctly [7]. The attacks can also be detected by analyzing the relative movement between face region and background [8]. However, these motion based methods fail against refusal to cooperate, tilt the papers or video replay attack.

As the essential difference between real and fake face, 3D information provides the most effective protection against spoof attempts. According to the fact that a planar face photo gives a flat structure whereas genuine yields a quite different structure,3D structure based methods [9], [10] make use of structure and depth information to classify real and fake faces.

Among existing face spoofing methods, [2], [11], [12] use the light reflection clue. Tan [2] treats the reflection difference between 2D spoof face and 3D genuine face, which caused by geometry structure and the surface roughness, as one of the features that create different image quality under the same imaging condition. But not all spoof faces have high frequency components, which is easily to be avoided by tilting the photo with a tiny angle, to make the photo looks more like a genuine face. Zhang [11] gives a distance robust and user friendly method by utilizing the difference of surface reflectance properties of skin and non-skin under multi-spectral light. However, the heavy equipment demand restricts its application in practice, especially on the consumer devices. Smith [12] presents a noninvasive anti-spoofing method used on consumer devices by computing the matching degree between the face reflections and the sequence of colors that were displayed on the screen, but it only works in a darkened environment.

It has to be said that in addition to properties mentioned above, there is still an intrinsic but inconspicuous difference, which has been overlooked and under-utilized: under the umbrella of the plane of the photo or the screen, the internal spoof face exists separately from the external environment; on the contrary, the intensities of genuine face change with the lighting environment because it's just exposed to the camera directly. The correlation of face image pair collected before and after the lighting environment changes, is apparently higher for spoof face than that for genuine face, especially when the dominant light direction seriously deviate far from original's. Thus the change of lighting environment can be used as a powerful weapon to distinguish whether the face is genuine or not, and a novel face anti-spoofing method based on 3D lighting environment analysis of image pair is proposed in this paper with the hand of 3DMM and SHIM.

Compared with the existing face spoofing detection methods mentioned above, our work has the following advantages:

1) User cooperation free: the user aren't required to move head, blink eye, smile or keep still deliberately.
2) Loose use condition: it's competent for different light conditions, camera resolutions and skin colors.
3) Extensive application scope: it's effective to tackle variable photo attacks and replay video attacks.
4) Simple equipment demand: only a few extra light sources are needed; change the screen brightness of consumer device also works in relatively dim light.
5) Easy to camouflage: the additional light is easily taken for granted to improve the light condition.
6) Propitious to face recognition: as illumination cones from different subjects are distinctive, face recognition performance itself improved through the active lighting used to change illumination [13].

II. Face Spoofing Detection based on 3D Lighting Environment Analysis of Image Pair

In this section, we first give a reasonable explanation for the analysis foundation that the proposed method relies on in section II-A. Then in section II-B, the adopted 3D lighting environment model is briefly reviewed. At last we present the proposed face spoof detection algorithm based on 3D lighting environment analysis in section II-C.

A. Spoof Faces Have Strong Immunity to The Change of Light

A useful question to ask is: when the environment's lighting condition changes, what's the difference of changes of the illumination distribution between a genuine face and a spoof face? In this part, we use the Phong's illumination model [14] to give the explanation that forms the foundation of our analysis. The principle is illustrated in Fig. 1.

At first, considering a vertex v on the face model with surface normal N lighted by monochromatic light from single direction ω. The corresponding irradiance $E_f(v)$ at the point v under this light model is the sum of ambient term, diffuse term, and specular term:

$$\begin{aligned} E_f(v) &= \rho_{af} L_a + \rho_{df} L_d(\omega) \langle N_f, \omega \rangle_v + \rho_{sf} L_s(\omega) \langle R_f, V \rangle_v^\mu \\ &= \rho_{af} L_a + \rho_{df} L_d(\omega) \cos\phi(v) + \rho_{sf} L_s(\omega) \cos^\mu \theta(v) \end{aligned} \tag{1}$$

where L_a, L_d, L_s are the intensity of ambient light, directed light and specular light; ρ_{af}, ρ_{df}, ρ_{sf} are the ambient reflectance, the albedo and the specular reflectance of human face; ϕ is the angle between surface normal vector N_f and the incident light ω, and θ is the angle between the reflected light R_f and the receiver's viewpoint V; and μ is the Phong's specular exponent. It can be extended to RGB trichromatic light by calculating the irradiance of each channel respectively, and when lights come from multi-direction, the total irradiance is sum of each from specific direction.

Now a face F_0 is appearing before the camera. For simplicity, we suppose that the spoof face was taken in the same lighting environment as the genuine face in advance, then the monochromatic irradiance of a vertex v on F_0 can be written as the form of Eq.(1). One additional light source joins in to change the lighting environment around F_0, and we still take light along direction $\omega 1$ into consideration. As the light shines on the genuine face directly, the irradiance E_G of v on the relight face F_G is:

$$\begin{aligned} E_G(v) &= E_f(v) + E1_f(v) \\ &= \rho_{af} L_a + \rho_{df} L_d(\omega) \langle N_f, \omega \rangle_v + \rho_{sf} L_s(\omega) \langle R_f, V \rangle_v^\mu \\ &+ \rho_{af} L1_a + \rho_{df} L1_d(\omega 1) \langle N_f, \omega 1 \rangle_v + \rho_{sf} L1_s(\omega 1) \langle R_f, V \rangle_v^\mu \\ &= \rho_{af}(L_a + L1_a) \\ &+ \rho_{df}(L_d(\omega) \langle N_f, \omega \rangle_v + L1_d(\omega 1) \langle N_f, \omega 1 \rangle_v) \\ &+ \rho_{sf}(L_s(\omega) \langle R_f, V \rangle_v^\mu + L1_s(\omega 1) \langle R_f, V \rangle_v^\mu) \\ &= \rho_{af} L_a^{'} + \rho_{df} L_d^{'}(v) + \rho_{sf} L_s^{'}(v) \end{aligned} \tag{2}$$

Here we can see that the change of one vertex's intensity depends largely on the surface normal vector N. Unless there is a linear relation between former and latter lights from all directions, that is to say $L_T = \alpha L$, the relight face F_T won't have strong correlation with F_0. Once the dominant direction of the additional light has relatively large deviation with the original, the correlation declines rapidly.

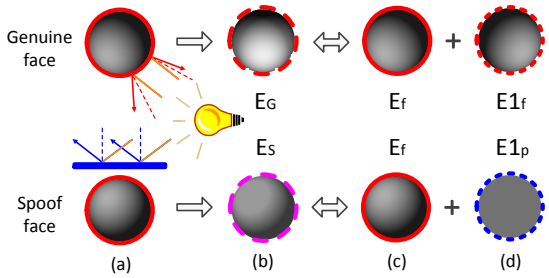

Fig. 1. The principle of the proposed algorithm, with the faces simplified into rendering balls. The first row denotes genuine face and second spoof: (a,c) light balls of faces with irradiance E_f, rendered by the original light; (d) light balls rendered by the additional light $\omega 1$ with huge deviation with original light, for genuine face, the irradiance $E1_f$ differ in different points as the directions of outgoing lights change with the different surface normals on different dots. But for spoof face, the irradiance $E1_p$ is approximate to a constant, as incoming lights are reflected by the smooth plane of photo or screen (blue bar) with similar surface normals in different dots. (b): The observed relighted balls with irradiance E_G = E_f + $E1_f$ for genuine face, and E_S = E_f + $E1_p$ for spoof face. As $E1_f$ has the opposite direction to E_f, the correlation between E_G and E_f is low for genuine face, while E_S has high correlation with E_f for spoof face.

When the surveillance camera encounters the spoof face, will there be some pleasant surprises happen? Different from the genuine face, the spoof face has been implanted into photos and screens. With the protection of the plane, the internal lighting environment exists separately from the external environment. In fact, as every surface normal vector N_p on the plane is nearly parallel, the diffuse part tends to be a constant with very small variation; the specular part can be divided into two zones approximately: highlight zone where θ_n is very small, and non-highlight zone where θ_n is not so small that $\cos^\mu \theta_n$ tends to be zero under the action of μ. The ambient part can still regarded as a constant. Then the irradiance E_S of v on the relight spoof face F_S can be expressed as follows:

$$\begin{aligned} E_S(v) &= E_f(v) + E1_p(v) \\ &= \rho_{af} L_a + \rho_{df} L_d(\omega) \langle N_f, \omega \rangle_v + \rho_{sf} L_s(\omega) \langle R_f, V \rangle_v^\mu \\ &+ \rho_{ap} L1_a + \rho_{dp} L1_d(\omega 1) \langle N_p, \omega 1 \rangle_v + \rho_{sp} L1_s(\omega 1) \langle R_p, V \rangle_v^\mu \\ &= E_f(v) + K_{ap} + K_{dp} + \delta K_{sp} \\ &= E_f(v) + K_p + \delta K_{sp} \end{aligned} \tag{3}$$

where K_{ap}, K_{dp}, and K_{sp} represent the constant ambient and diffuse light irradiance, and the largest specular light irradiance; δ equals to 1 if v falls within highlight zone and 0 in non-highlight zone; ρ_{ap}, ρ_{dp}, ρ_{sp} are the ambient reflectance, the albedo and the specular reflectance of the plane.

Obviously, if there's no large areas of highlight zone within the scope of spoof face, the relight face F_F has strong correlation with F_0 for the reason that the intensity of every vertex is added only with a constant K_p. It's easy to achieve by tilting the plane of photos or screens with a tiny angle. In practice, most attackers devote themselves to avoid the specular reflection so that the spoof face looks like a genuine face to the largest extent. Of course, the specular reflection itself can be treated as a powerful clue to uncover spoof faces.

To summarize, the change of lighting environment can affect the light distribution of genuine faces significantly, while the relight spoof face will have strong correlation with the original one. The fundamental cause lies in the different geometric structure of genuine and spoof faces. This forms the analysis's foundation that the proposed system relies on.

B. Lighting Environment - Sphere Harmonic Illumination Model

As any number of lights can be placed in any positions, the lighting environment of a scene can be arbitrarily complex. Sphere Harmonic Illumination Model (SHIM) [15], [16] is competent to model such complex lighting (other expansion methods such as frames [17] can also model lighting environment, here we choose SHIM for simplicity), under assumptions: (1) lights are distant; (2) surfaces are convex and Lambertian; (3) the surface reflectance is constant; (4) camera response is linear. Although there are some degree of specular reflection exist on oily skins, experiment in [18] has shown that the reflection characteristics of human face are nearly diffuse in many common situations. Along with most parts of human face are convex, especially the nose tip, these assumptions are tenable enough when SHIM is used to analysis human face.

Neglecting the cast shadows and near lights, the lighting environment $L(\omega)$ is a non-negative function on unit spherical surface. Taking light coming from all directions into account, the appearance of vertex v is the convolution of the reflectance function of the surface $R(N,\omega)$, with the lighting environment $L(\omega)$, specifying the intensity of the incident light along the unit vector direction ω. The sum can be written in integral over the surface of sphere:

$$E(v) = \int_{\Omega} L(\omega)R(N,\omega)d\Omega = \int_{\Omega} L(\omega)\max(0,\langle N,\omega\rangle)d\Omega \tag{4}$$

where Ω represents the surface of the sphere. The clamped reflectance fucntion effectively limits the integration to the hemisphere about the surface normal $N = (x,\ y,\ z)^T$.

Using Spherical Harmonics Transformation, the integration can be expressed as a linear sum of spherical harmonics basis, which can be further approximated by the first three orders as the reflectance coefficients $\hat{r}_m$ decay rapidly [15], [16]:

$$E(v) \approx \sum_{m=0}^{2}\sum_{n=-m}^{m} \hat{r}_m l_{m,n} Y_{m,n}(N) \tag{5}$$

where $l_{m,n}$ are the coefficients corresponding to the nth spherical harmonic of order m, $Y_{m,n}(\cdot)$. The first three orders Y are: $Y_{0,0} = \sqrt{1/4\pi}$; $Y_{1,-1} = \sqrt{3/4\pi}y$, $Y_{1,0} = \sqrt{3/4\pi}z$, $Y_{1,1} = \sqrt{3/4\pi}x$; $Y_{2,-2} = \sqrt{15/4\pi}xy$, $Y_{2,-1} = \sqrt{15/4\pi}yz$, $Y_{2,0} = \sqrt{5/16\pi}(3z^2-1)$, $Y_{2,1} = \sqrt{15/4\pi}xz$, $Y_{2,2} = \sqrt{15/16\pi}(x^2-y^2)$. The first three orders $\hat{r}$ are: π, $2\pi/3$, and $\pi/4$. Then the irradiance of a convex Lambertian surface under arbitrary distant lighting can be well modeled by the first nine lighting environment coefficients $l_{m,n}$ ($0 \le m \le 2, -m \le n \le m$).

Assume that a linear relationship between image intensity $I(v)$ and irradiance $E(N(v))$ at vertex v [19], then $I(v)$ can be written in terms of spherical harmonics by expanding Eq.(5):

$$\begin{aligned} I(v) = {} & \hat{r}_1 l_{0,0} Y_{0,0}(N(v)) + \hat{r}_2 l_{1,-1} Y_{1,-1}(N(v)) \\ & + \hat{r}_2 l_{1,0} Y_{1,0}(N(v)) + \hat{r}_2 l_{1,1} Y_{1,1}(N(v)) \\ & + \hat{r}_3 l_{2,-2} Y_{2,-2}(N(v)) + \hat{r}_3 l_{2,-1} Y_{2,-1}(N(v)) \\ & + \hat{r}_3 l_{2,0} Y_{2,0}(N(v)) + \hat{r}_3 l_{2,1} Y_{2,1}(N(v)) \\ & + \hat{r}_3 l_{2,2} Y_{2,2}(N(v)) \end{aligned} \tag{6}$$

We can see Eq. (6) is linear in $l_{m,n}$, so it can be converted to linear equations for q vertices on the face model:

$$\vec{a} = M\vec{e} \tag{7}$$

where $\vec{e}$ is the vector of lighting environment coefficients $(l_{0,0}\ l_{1,-1}\ ...\ l_{2,2})^T$, $\vec{a}$ is the vector of corresponding intensities at the q vertices, $(I(v_1)\ I(v_2)\ ...\ I(v_q))^T$, and M is the corresponding matrix $\hat{r}_m Y_{m,n}$ containing the spherical harmonics.

It's expected that coefficients from faces in different lighting environments should be distinguishable, while similar to the same lighting environment. Given two sets of illumination environment coefficients, denoted as $\vec{e}_1$ and $\vec{e}_2$, we can synthesize two corresponding hemisphere images $\vec{a}_{H1}$ and $\vec{a}_{H2}$ by Eq. (7), with the matrix M_H according to surface normal vectors of hemisphere,

$$\vec{a}_{H1} = M_H\vec{e}_1 \qquad \vec{a}_{H2} = M_H\vec{e}_2 \tag{8}$$

Correlation between two rendered hemispheres is given by:

$$\text{corr} = \frac{\vec{a}_{H1}^T\vec{a}_{H2}}{\|\vec{a}_{H1}\|\|\vec{a}_{H2}\|} = \frac{\vec{e}_1^T Q\vec{e}_2}{\sqrt{\vec{e}_1^T Q\vec{e}_1}\sqrt{\vec{e}_2^T Q\vec{e}_2}} \tag{9}$$

where the matrix Q is equal to $M_H^T M_H$. The correlation value calculated from $\vec{e}_x$ and $\vec{e}_y$ is denoted with Ce_{xy}. As the first coefficient has no contribute to the direction of lighting environment, it can be set to zero when calculate the correlation to maximize the impact of light direction.

C. The Proposed Face Spoof Detection System

The general framework of the proposed method is illustrated in Fig. 2, which consists of four main steps: (1) estimate the original lighting environment coefficients, shown with blue dotted line; (2) decide how to change the lighting environment with purple dotted line; (3) estimate the changed lighting environment coefficients with orange solid line; (4) conduct the liveness judgement with pink solid line. As the first and third steps share the same process of lighting environment coefficients estimation, shown with dark green solid boxes, we will unfold the introduction of our system from the three aspects below:

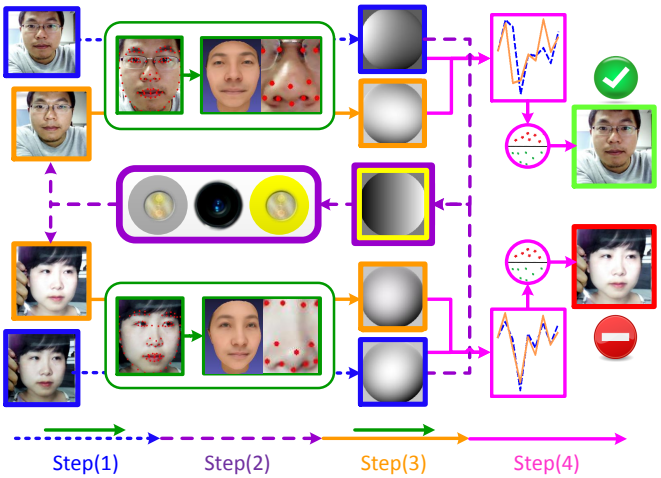

Fig. 2. The general framework of the proposed method, which consists of four steps: (1) estimate the original lighting environment coefficients; (2) change lighting environment; (3) estimate the changed coefficients; (4) conduct liveness judgement by using two hypothesis tests mentioned in Section II-C-3.

1) Estimate Lighting Environment Coefficients of Human Face: In section II-B, we introduce 3D lighting environment model and convert it to linear equations shown as Eq. (7). Now the vector of lighting environment coefficients $\vec{e}$ is unknown. Once given intensities and surface normals of at least nine points on the face model, the lighting environment coefficients can be estimated as the least-squares solution to Eq. (7):

$$\vec{e} = (M^T M)^{-1} M^T \vec{a} \tag{10}$$

Geometry of one arbitrarily object is not always readily obtainable, however, as to human face, it can be comparatively easy to satisfy. Precise measurements of facial shape can be acquired using 3D scanning devices. We can also get approximate face models from 2D images via 3D reconstruction technologies such as 3D Morphable Model (3DMM). In this paper, the lighting robust fitting approach of 3DMM using SHIM [20] are used to reconstruct face models as the preparatory work of our system.

When a client appears in front of the security cameras, he or she will be captured and first processed by algorithms of face detection and recognition to obtain his or her identity. Using the detected feature point coordinates, we can compute the face pose parameters in terms of three rotation angles of yaw, pitch and roll. Then the specific face model of the client is loaded and rotated by the parameters, as a result, surface normal N_0 of vertex v in model coordinate system is transformed into surface normal N with respect to the camera coordinate system. With the corresponding intensities of face skin after aligned, the lighting environment coefficients of the captured face image are estimated immediately.

In practice, the surface normals of human face cheeks change as expression changes, meanwhile the forehead, jaw and eyes are often covered by hair, beard and glasses respectively. Both of these two factors had a significant impact on the estimation accuracy of lighting environment. The human nose, especially the nose tip, by contrast, has a very similar shape which is little affected by expression [21], and exposes in the air without any occlusion in most cases. What's more, the overall shape of noses is relatively stable among different persons, sexes and races. Thus we can use the average nose model to estimate the lighting environment coefficients if the specific face model of one person is not available. The reconstructed faces and rendering balls based on estimated lighting environment coefficients are shown in Fig. 3.

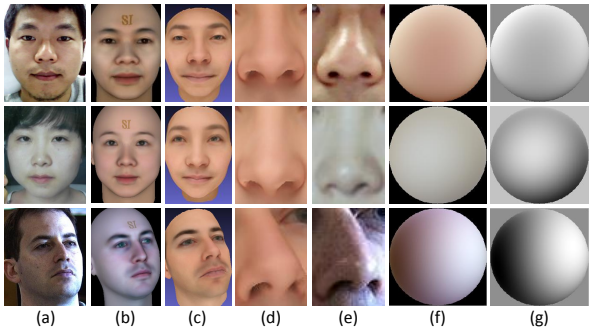

Fig. 3. The reconstructed face and rendering ball of different persons. (a) Faces of three persons with different sex and race [22]; (b) reconstructed faces by FaceGen; (c) reconstructed faces by [20]; (d) aligned shapes of noses; (e) aligned intensities of noses; (f) color balls rendered with the estimated 3-channel lighting environment coefficients from color images; (g) gray balls rendered with the first value of coefficients set to zero from gray images.

2) How To Change The Lighting Environment: We attempt to change the lighting environment by adding additional light sources in front of the client's face, whose dominant direction is prominently differ from the original's. In the ideal situation, a controllable ring light around the camera can adopted to relight the face from the weakest direction of original lighting environment. Of course, the less number of light sources used the better in a practical application. If the dominant light of the original lighting environment is not positive light, a single frontal light can give the solution, like change the brightness of consumer device screen, but not the opposite. So we use two lamps on the left and right of the camera, which can accommodate to most cases with the discriminative power shown in the following experiments. Similar analysis can be extended to other directions.

To judge the dominant light comes from left or right, we give a vector of lighting environment coefficients $\vec{e}_R$ with values (0,0,0,1,0,0,0,0,0) as a reference, whose rendering hemisphere is lighted by lights coming from right absolutely. When original coefficients $\vec{e}_0$ has been estimated, we can make the following decision based on the correlation value Ce_{R0}:

1) $Ce_{R0} > 0.2$, it means the dominant light comes from right with $\vec{e}_r = \vec{e}_0$, then the left lamp is turned on to estimate $\vec{e}_l$;
2) $Ce_{R0} < -0.2$, it means the dominant light comes from left with $\vec{e}_l = \vec{e}_0$, then the right lamp is turned on to estimate $\vec{e}_r$;
3) $|Ce_{R0}| \leq 0.2$, it means the dominant light comes from the front, then the right and left lamps are turned on one by one to estimate $\vec{e}_r$ and $\vec{e}_l$, respectively.

3) Conduct The Liveness Judgement: As analyzed before, Ce_{rl} can be used as a powerful weapon to expose spoof faces with a fixed lighting environment such as photos, and replay face videos recorded in a relative stable lighting environment.

However, videos recorded in dramatically changed lighting environment sometimes also have a very low Ce_{rl} to cloud our judgment. To overcome the effect of this attack, the difference $De_{RrRl} = Ce_{Rr} - Ce_{Rl}$ is used as a meaningful feature based on the fact that the designated right lighting environments $\vec{e}_r$ should has much higher correlation with the reference $\vec{e}_R$ as to genuine faces, and the changes of the recorded lighting environment will be particularly difficult to always match the additional light, from the aspects of direction and time.

One method for detecting spoof face is to train two threshes T_C and T_D: the face with Ce_{rl} below T_C and De_{RrRl} above T_D is genuine while Ce_{rl} above T_C or De_{RrRl} below T_D is fake. However reliable thresholds are difficult to establish. In this paper, we chose a statistical approach (a hypothesis test) for each of these two touchstones to conduct the liveness judgement, as it can report the probability of observing the results and provide more information than approaches with thresholds. Assume that Ce_{rl} and De_{RrRl} are normally distributed, and spoof faces have larger Ce_{rl} meanwhile genuine faces have larger De_{RrRl}. Two hypothesis tests will be conducted:

1) H_{C1}: Ce'_{rl} is equal to or smaller than μ_{C0};
   H_{C2}: Ce'_{rl} is greater than μ_{C0}.
2) H_{D1}: De'_{RrRl} is equal to or smaller than μ_{D0};
   H_{D2}: De'_{RrRl} is greater than μ_{D0}.

These two test statistics are:

$$z_C = \frac{Ce'_{rl} - \mu_{C0}}{\sigma_{C0}/\sqrt{N}} \qquad z_D = \frac{De'_{RrRl} - \mu_{D0}}{\sigma_{D0}/\sqrt{N}} \tag{11}$$

where μ_{C0} (μ_{D0}) and σ_{C0} (σ_{D0}) are the expected mean and standard deviation of Ce_{rl} (De_{RrRl}) and determined empirically from real (spoof) faces. The significance of these two statistic are given in terms of the standard error function:

$$p(z_{C(D)}) = 1 - erf(z_{C(D)}/\sqrt{2}) \tag{12}$$

If $p(z_C)$ is smaller than a level of α_C, H_{C2} is accepted and the face tends to be fake; if $p(z_D)$ is smaller than a level of α_D, H_{D2} is accepted and the face tends to be genuine. If H_{C1} and H_{D2} are accepted at the same time, the client can be deemed genuine. Here α_C and α_D is set to 0.05.

III. Experiments

A. Database

Publicly available databases designed to detect liveness such as CASIA-FASD [23] and REPLAY-ATTACK [24] are all captured at a certain time in the past and don't contain real-time scene information such as lighting environment. They are useful to texture based method for training classifier, but unsuitable to our proposed method as there are no image pairs needed in our algorithm. But all can be used as spoof samples after being printed onto the paper or displayed on the screen to validate our algorithm.

We first collect 500 image pairs each for genuine faces and printed or screen photos under different lighting environments (soft light without significant direction and hard light with significant dominant light directions). Then the corresponding designated $\vec{e}_r^q$ and $\vec{e}_l^q$, along with Ce_{rl}^q and De_{RrRl}^q are calculated. Finally, $\mu_{C0} = 0.3401$ and $\sigma_{C0} = 0.5833$, $\mu_{D0} = 0.0612$ and $\sigma_{D0} = 0.0442$ are calculated and used for the hypothesis test in Section II-C-3.

B. Experimental Results

We then validate our algorithm by collecting 50 subjects with their corresponding real faces (still and moving), printed photos (flat, warped and eyes-cut), screen photos (phone and tablet), replay-attack videos (stable and variable lighting environment) and some 3D head models under soft and hard lighting environment, captured by high and low resolution cameras, together with spoof samples created from the public databases, totally 2000 image pairs. 98.6% recognition accuracy indicates the superior performances in face spoofing detection. Some examples shown in Fig. 4-6 to further illustrate the excellent results, and Fig. 7 gives some failure cases which might be overcame by uniting some other algorithms. First row in each figure is the corresponding image pair of different situations; and seconde row are the corresponding rendering balls of the image pair above; and values below are the corresponding Ce_{rl} and De_{RrRl} respectively.

1) Still Is Not Always Spoof, And Moving Is Not Always Genuine: Motion based methods couldn't handle those clients that do not cooperate, and it's easy to be attacked by twisting photos or replay videos. Things become simple as for our algorithm, as shown in Fig. 4, with the genuine faces' intensities change with the change of lighting environment in Fig. 4(a). but lighting environment is not affected by head movements. Spoof faces, however, have strong immunity against light change as shown in Fig. 4(b) and Fig. 4(c), even the photo is warped or people blink in the video.

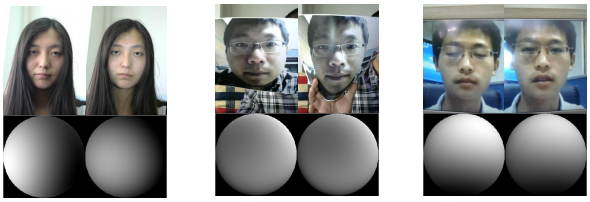

Fig. 4. The illustration of different attacks and corresponding results: (a) Still genuine face, (b) warped photo and (c) moving tablet video with eye blink, all can be correctly detected by our algorithm.

2) Loose Use Conditions: Compared with reflections method [12] which require a darkened environment as the brightness of phone or tablet is poor, our method can competent to a vast majority of lighting environment by adding a few active light sources, as shown in Fig. 5(a) and Fig. 4(a). Of course, our algorithm also works in a not so bright lighting environment compared with [12] by just changing the screen brightness of photos or tablets as shown in Fig. 5(d). Another advantage of our algorithm is free of camera resolution, which severely interfere with texture base methods: high-resolution camera may infer clear photos as genuine in Fig. 5(b), while genuine face may be inferred as spoof for the inherent distortion of low-resolution cameras in Fig. 5(c). Further more, the proposed method is effective against different skin colors (Fig. 5(a) and Fig. 4(a)).

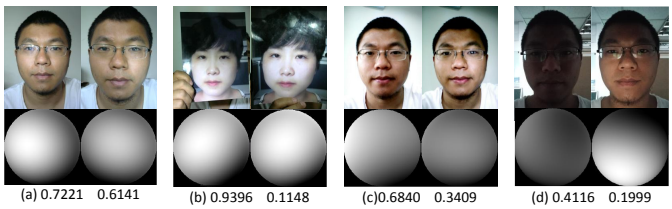

Fig. 5. The illustration of different attacks and corresponding results: (a) Genuine face in soft lighting environment, (b) clear photos captured by high-resolution camera and (c) genuine face captured by low-resolution camera, all can be correctly detected by our algorithm. (d) Change of Screen brightness also works in a not so bright environment.

3) Extensive Application Scope: In addition to handled photos (Fig. 5(b)), screen photos (Fig. 6(a)) and warped photos (Fig. 4(b)), our algorithm also valids for photos whose eyes are cut. Considering videos taken in controlled stable lighting environment (Fig. 4(c)) can't pass H_C, attackers may take videos under change lighting environment, but H_D builds another solid defense when H_C is breached(Fig. 6(c)).

Fig. 6. The illustration of different attacks and corresponding results: (a) Screen photos, (b) photos with eye cut off and (c) video under variable lighting environment, all can be correctly detected by our algorithm.

4) Failure Cases: A few failure cases appear in our test. One reason is over saturated exposure caused by specular reflection of photo and screen exactly right locates on the

nose and covers original nose completely, as shown in Fig. 7(a). This can be excluded by extracting exposure zone and analysis the difference of area and shape between genuine and spoof nose. Another reason is the original light is so strong that the additional light sources contribute little for the change of lighting environment in Fig. 7(b). Fortunately, an overwhelming majority of applications of liveness detection are indoor with soft or slightly hard light around the cameras, and people tend to duck away from the specular reflection by tilting the plane of photos or screens with a tiny angle to pursue the greatest reality. 3D head model (Fig. 7(c)) or 3D face mask of a specific client, however, can pass our system effortlessly because of its 3D structure, and we need unite with other methods to overcome this difficult question in further study. Videos in [23] are all collected in a stable environment, but when a replay attack has variable lighting environment, and the change form exactly matches the form created by our system, it collapses, as shown in Fig.7(d). Utilizing multi-frames of video to validate whether the time and direction of lighting environment changes are consistent with the additional light, might be a possible solution for these kind of attack.

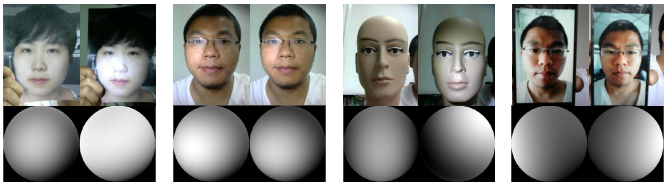

Fig. 7. (a) Photos over saturated exposure in nose zone, (b) genuine face captured in environment with very strong light and (c) 3D head model, can't be correctly detected by our algorithm. (d) Sometimes video under variable lighting environment is also a problem should be solved.

IV. Conclusion and Future Work

In this paper, we propose a novel effective face spoofing method based on 3D lighting environment analysis of image pair. Considering the different manifestations confronted the change of lighting environment for real face versus spoof face, which caused by the intrinsic property that spoof is under the protection of the plane of photos and screens, the correlations of image pair collected before and after the lighting environment change are used to separate the attack from the valid user. Compared with the existing methods, our approach has the following advantages: user cooperation free, loose use condition, extensive application scopes, simple equipment demand, easy to camouflage and propitious to face recognition. We will engage in further study of how to utilize the time and direction information of multi-frames of video, the stability of background and lighting environments of fixed cameras, and combine needful clues such as exposure feature, to improve the performance of our algorithm.

Acknowledgement

This work was supported by the Natural Science Foundation of China (61571438) and Open Project of Trace Science and Technology, Key Laboratory of Ministry of Public Security (2014FMKFKT05).

References

[1] M. Jukka, A. Hadid, and M. Pietikinen, "Face spoofing detection from single images using micro-texture analysis," in *Int. Joint Conf. Biometrics (IJCB)*, 2011, pp. 1–7.

[2] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conf. Computer Vision (ECCV)*, 2010, pp. 504–517.

[3] J. Yang, Z. Lei, and S. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, pp. 797–809, 2015.

[4] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *J. Telecommun. Syst.*, vol. 47, pp. 215–225, 2011.

[5] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, pp. 762–777, 2015.

[6] L. Cai, C. Xiong, L. Huang, and C. Liu, "A novel face spoofing detection method based on gaze estimation," in *Asian Conf. Computer Vision (ACCV)*, 2014, pp. 547–561.

[7] K. Kollreider, H. Fronthaler, M. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in liveness assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, 2007.

[8] A. Anjos, M. Mohan, and S. Marcel, "Motion-based counter-measures to photo attacks in face recognition," *IET Biometrics*, vol. 3, pp. 147–158, 2014.

[9] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, and S. Sridharan, "Liveness detection based on 3d face shape analysis," in *Int. Workshop Biometrics and Forensics (IWBF)*, 2013, pp. 1–4.

[10] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3d structure recovered from a single camera," in *Int. Conf. Biometrics (ICB)*, 2013, pp. 1–6.

[11] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, 2011, pp. 436–441.

[12] D. F. Smith, A. Wiliem, and B. C. Lovell, "Face recognition on consumer devices: Reflections on replay attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 736–745, 2015.

[13] J. R. Beveridge, B. A. Draper, J. M. Chang, M. Kirby, H. Kley, and C. Peterson, "Principal Angles Separate Subject Illumination Spaces in YDB and CMU-PIE," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 351-363, 2009.

[14] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *SIGGRAPH*.NY, USA: ACM, 1999, pp. 187–194.

[15] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[16] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for reflection," *ACM Trans. Graph.*, vol. 23, no. 4, pp. 1004–1042, 2004.

[17] W. Zhao, Y. Zheng, L. Wang, and S. Peng, "Lighting estimation of a convex lambertian object using weighted spherical harmonic frames," *Signal Image Video Process.*, pp. 1–19, 2012.

[18] S. Marschner and S. e. a. Westin, "Image-based brdf measurement including human skin," in *Rendering Techniques*, 1999, pp. 131–144.

[19] M. K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments." *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 450–461, 2007.

[20] M. Ma, X. Hu, Y. Xu, and S. Peng, "A lighting robust fitting approach of 3d morphable model using spherical harmonic illumination," in *IEEE Int. conf. Pattern Recognition (ICPR)*, 2014, pp. 2101–2106.

[21] A. Moorhouse, A. N. Evans, G. A. Atkinson, J. Sunf, and M. L. Smith, "The nose on your face may not be so plain: using the nose as a biometric," in *IEEE Int. Conf. Crime Detection and Prevention (ICDP)*, 2009, pp. 1–6.

[22] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, 2003.

[23] Z. Zhang, J. Yan, S. Liu, and Z. Lei, "A face antispoofing database with diverse attacks," in *Int. Conf. Biometrics (ICB)*, 2012, pp. 26–31.

[24] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *IEEE Int. Conf. the Biometrics Special Interest Group*, 2012, pp. 1–7.