Physical-World Optical Adversarial Attacks on 3D Face Recognition

Yanjie Li,[1] Yiquan Li,[1] Xuelong Dai,[1] Songtao Guo,[2] Bin Xiao[1]*
[1] The Hong Kong Polytechnic University [2] Chongqing University
{yanjie.li,yiquan.li,xuelong.dai}@connect.polyu.hk, guosongtao@cqu.edu.cn,
b.xiao@polyu.edu.hk

Abstract

The success rate of current adversarial attacks remains low on real-world 3D face recognition tasks because the 3D-printing attacks need to meet the requirement that the generated points should be adjacent to the surface, which limits the adversarial example' searching space. Additionally, they have not considered unpredictable head movements or the non-homogeneous nature of skin reflectance in the real world. To address the real-world challenges, we propose a novel structured-light attack against structured-light-based 3D face recognition. We incorporate the 3D reconstruction process and skin's reflectance in the optimization process to get the end-to-end attack and present 3D transform invariant loss and sensitivity maps to improve robustness. Our attack enables adversarial points to be placed in any position and is resilient to random head movements while maintaining the perturbation unnoticeable. Experiments show that our new method can attack point-cloud-based and depth-image-based 3D face recognition systems with a high success rate, using fewer perturbations than previous physical 3D adversarial attacks.

1. Introduction

As 2D face recognition becomes increasingly vulnerable to attacks, researchers are turning to 3D face data for secure user authentication and other tasks. Structured light imaging is a popular method of acquiring 3D face data due to its high precision and superiority in uniform textures [10]. It is widely used in off-the-shelf devices such as Kinect v1 and iPhone [45]. Various 3D face recognition algorithms are proposed based on structured light [11, 15, 29]. FaceID, for example, collects 3D face data using a structured light camera and then uses 3D face data for user authentication.

Although real-world attacks for 2D face recognition have been thoroughly studied [5, 27, 41], studies on physically realizable 3D face recognition attacks are insufficient. Tsai *et al.* [30] first proposed 3D printable adversarial examples for point cloud classification tasks. However, the perturbations must be strictly adjacent to the surface rather than in arbitrary positions in 3D-printed examples, greatly limiting the adversary's attack ability. Previous optical adversarial attacks [6, 12, 19, 20, 36, 47] cannot be applied directly to 3D face recognition either. Face reflection is more complex than opaque objects like the stop sign and the principle of 3D reconstruction is different from 2D imaging.

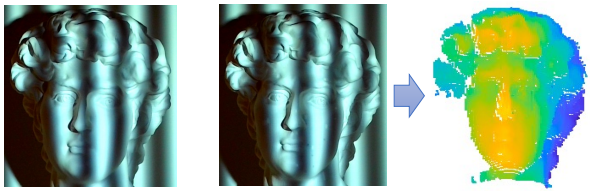


Figure 1. A demonstration of our attack. We project optical noises on the 3D faces to generate adversarial point clouds. Our attack modifies fewer points than previous attacks and does not need the adversarial points to be adjacent to the 3D surface, thus with a high attack success rate.

To address the limitations of existing 3D-printing-based adversarial attacks, we propose a new attack called *Structured-Light Attack* on 3D face recognition systems, which uses adversarial illumination to attack structured-light-based 3D face recognition. The perturbation can be concealed in the normal fringe patterns or superimposed on the original illumination by using the inherent or an additional projector. Figure 1 shows an overview of our attack. Without the optical noises, the 3D reconstructed point cloud is recognized as David. While with the elaborate optical noises, the adversarial point cloud is recognized as Athena.

Our attack achieves a higher success rate than previous attacks by precisely simulating the real-world 3D reconstruction process and taking the random head movements into account. We first incorporate the 3D face relighting process in the attack pipeline through the Lambertian reflectance model. Then, we propose a differential 3D recon-

*B. Xiao is the corresponding author. Our code is available at https://github.com/PolyLiYJ/SLAttack.git.

struction substitution. The adversarial illuminations can result in point shiftings in the point cloud and cause dodging or impersonation attacks. Utilizing the 3D reconstruction principle, our attack is end-to-end and can generate adversarial points in arbitrary 3D positions. It needs fewer perturbations than previous 3D attacks with the average attack success rate of 47% in impersonation attacks and 95% in dodging attacks respectively. Our contributions are summarized as follows.

- We are the first to realize physical 3D face attacks through adversarial illuminations. The optical perturbations are generated end-to-end and are camouflaged by the normal patterns of structured-light projectors.
- Our physical attack is feasible and effective by involving the face relighting process and random 3D transformations in the attack pipeline, which significantly improve the 3D reconstruction precision and adversarial examples' robustness to random head movements.
- We attack both point-cloud-based and depth-image-based 3D face recognition models. Compared with previous attacks, our method needs fewer perturbations with a high success rate in experiments.

2. Related Work

3D Adversarial Attack 3D adversarial attacks focus on generating adversarial examples for 3D deep learning models. Most of them aim for point cloud data. Xiang et al. [39] first proposed an adversarial attack on the point cloud by point perturbation and point generation. Zheng et al. [44] and Wicker et al. [35] spoofed the deep learning model by dropping some critical points based on the saliency map. Some other studies focused on improving the imperceptibility of the adversarial examples based on the geometric properties of the adversarial point cloud [13, 33]. However, these attacks are hard to be realized in the physical world.

For physically realizable attacks, Cao et al. [2] successfully fooled the Lidar sensor on autonomous vehicles by adding fake front-near obstacles through a time-lapse module and a laser emitter. However, the sensors used for face recognition are significantly different from the autopilot. Some studies use 3D-printed objects for physical attacks. Tsai et al. [30] proposed kNN loss to generate 3D printable adversarial examples. Tu et al. [31] proposed utilizing Laplacian loss to improve the mesh's smoothness and 3D printability. However, they did not apply their attacks to the 3D face recognition application, which is a far more complex scenario. Moreover, 3D printing technology can only generate perturbations adjacent to the 3D surface rather than at arbitrary positions in the 3D space.

Optical Adversarial Attack Optical adversarial attacks change the illumination of the target objects to spoof the classifiers. Compared with printing-based attacks, they have better camouflage. But previous works have only considered 2D classifiers. For example, Nicoles et al. [20] generated adversarial illuminations through iteratively capturing and optimizing. Zhou et al. [47] utilized infrared LEDs to attack the face recognition system. Worzyk et al. [36] projected perturbations onto the road stop signs. Nguyen et al. [19] applied the optical adversarial attack to the 2D face recognition and used expectation over transformation to improve the physical robustness. Gnanasambandam et al. [12] improved attack success rate by considering spectral nonlinear. However, these attacks cannot be directly applied to 3D scenarios because of the huge difference in 2D and 3D imaging principles. Moreover, these attacks are too obvious to escape from human eyes. To solve these problems, this work proposes novel 3D optical adversarial attacks for face recognition scenario and consider the face reflection process for end-to-end attacks.

3. Methodology

In this section, we first briefly introduce our attack targets, the structured light profilometry, and then presents our attack methods. We propose phase shifting attack and phase superposition attack for multi-step and single-step technologies, respectively. An overview of our end-to-end attacks is shown in Figure 3.

3.1. Structured light profilometry

Multi-step algorithm Multi-step phase shifting (MSPS) is one of the most important surface profilometry algorithms. We reference Piccirilli's work [21] to build the MSPS-based 3D face data acquisition system. We first calibrate the projector and camera to get their projection matrices A_p and A_c. Then a group of phase-shift images is projected sequentially to encode every unitary position of a surface. Next, the wrapped phase map ϕ_w is computed through captured images $I_0^c, ..., I_{N-1}^c$ by Equation 1,

$$\varphi_w(u_c, v_c) = \tan^{-1} \frac{\sum_{n=0}^{N-1} I_i^c(u_c, v_c) \sin(2\pi n/N)}{\sum_{n=0}^{N-1} I_i^c(u_c, v_c) \cos(2\pi n/N)}. \quad (1)$$

After getting the wrapped phase, it is unwrapped to the absolute phase, $\phi_a(u_c, v_c) = \phi_w(u_c, v_c) + 2\pi K(u_c, v_c)$, where $K(u_c, v_c)$ is the fringe's order. We use cyclic complementary gray code [38] to compute $K(u_c, v_c)$. Then pixels in I_p and I_c are matched according to the absolute phase. Finally, we use Eq.32 in Feng's paper [10] to recover 3D coordinates, which, for the sake of simplicity, can be expressed as the following function,

$$[x, y, z] = h(A_c, A_p, u_c, v_c, u_p), \quad (2)$$

where $[x, y, z]$ is the 3D coordinate. The u_c, v_c are the pixel's horizontal and vertical coordinates in I_c. u_p is the matching pixel's horizontal coordinate in I_p.

Single-step algorithm Multi-step optical metrology faces the problems of error accumulation. Recently, benefiting from the rapid development of deep learning, some researchers proposed to project only one fringe image and then recover the phase map through deep neural networks [4, 9, 24, 48]. Feng et al. [8] achieved SOTA results on single fringe pattern analysis. They rewrote Equation 1 as $\varphi_w(u_c, v_c) = \tan^{-1} M(u_c, v_c)/D(u_c, v_c)$. They first used a CNN and a single modulated fringe image I^c as input to estimate the background image A, and then used A and I^c as the second CNN's input to estimate M and D directly. We refer to this work as SLCNN in this paper.

3.2. Phase shifting attack

For multi-step structured light imaging, we propose a novel attack named phase shifting attack, as shown in Figure 2. The basic idea is to involve the multi-step phase-shifting algorithm into the C&W attack [3]'s optimization process and hide the perturbation into the original patterns. However, because the pixel coordinates are discrete values, the back-propagation of Equation 2 is prevented. Therefore, we first substitute the 3D reconstruction algorithm with a differential one to solve this problem.

Differential 3D reconstruction algorithm Inspired by the natural 3D renderer [14] that involves a differential renderer for 3D mesh reconstruction, we include the 3D reconstruction in the end-to-end attacks. The difference is that Kato's work needs to get $\frac{\partial I(u_c)}{\partial u_c}$, while our problem is to compute $\frac{\partial z(u_p, u_c)}{\partial u_p}$, which can be seen as an inverse rendering process. Specifically, because the pixel coordinate u_p is not differential, we optimize the absolute phase map ϕ_a in the attack iterations instead of optimizing u_p directly. After getting the adversarial phase map, we get u_p though $u_p = round(\frac{w\phi_a}{2\pi n_s})$, where n_s is the fringes' number, w is the width of projected images.

Moreover, simultaneously changing the projector's and the camera's corresponding points' coordinates will influence the reconstruction accuracy. To solve this problem, we project the 3D adversarial point displacements onto the normal vector of the camera imaging plane. Suppose $\boldsymbol{A_c} = K[R\ T] \in \mathbb{R}^{3\times4}$ is the camera's perspective projection matrix, then the projection direction can be expressed as $R^{-1}K^{-1}\boldsymbol{e}$, where $\boldsymbol{e} = (0, 0, 1)^T$. This projection will make the 3D pixel shifting only change the depth in the camera view and leave the corresponding pixel coordinates unchanged when the distance is in a certain range (we put the proof in the appendix). This is also equal to rotating the point cloud before the attack and limiting the perturbation in the depth direction.

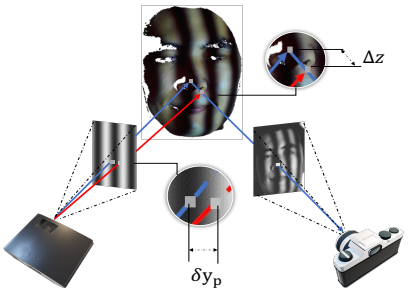


Figure 2. An overview of the phase-shifting attack. We modify the projected patterns to pollute the 3D data indirectly. The point displacements or depth changes are mapped to a flow field in the fringe image through a differential 3D reconstruction.

Sensitivity map and 3D transform invariant loss Noticing that humans are more sensitive to perturbation in the central and flat areas of the face, we propose two sensitivity maps to punish perturbations in areas of high sensitivity. The first is defined as $Sen_1(u_c, v_c) = e^{-\frac{\|(u_c,v_c)-(u_0,v_0)\|_2}{w}}$, where (u_0, v_0) is the face center. For the second map, we noticed the point cloud tends to be uniformly distributed on the surface after the farthest point sampling. When the point cloud is mapped to 2D images, non-zero pixels are concentrated where there is a large change in height. Therefore, the second sensitivity map is defined as $Sen_2(u_c, v_c) = \frac{1}{\|\phi_a(u,v)\|_0}|_{(u,v)\in S(u_c,v_c)}$, where $S(u_c, v_c)$ is a small adjacent area centered by (u_c, v_c) in the absolute phase map. The total sensitivity map is $Sen = Sen_1 + Sen_2$.

When attacking a physical system, the distance and head may move unexpectedly. Expectation over transform loss has been used to solve this problem in 2D scenario [19]. We extend it to the 3D space and propose 3D transform invariant loss to make the adversarial point clouds can generate consistent results when the environment changes. The random spatial transformation function can be expressed as $\mathcal{T}(\mathcal{P}) = (\mathbf{R}_{(\theta_x,\theta_y,\theta_z)}\mathcal{P}^T)^T + \mathbf{M}_{(\eta_x,\eta_y)}$, where $\mathbf{M}$ and $\mathbf{R}$ are spatial translation and rotation matrix and $\eta_x, \eta_y, \theta_x, \theta_y, \theta_z$ are random displacements and rotation angles sampled from Gaussian distributions. Moreover, to simulate the real-world preprocessing process, we involve a resample and renormalization function in the optimization. The final loss function is defined as

$$l_{adv}(\phi'_a) = \mathbb{E}_{\mathcal{T}} l_{logits}\left(\mathcal{M}\left(\mathcal{N}\left(\mathcal{T}(h(\phi'_a))\right)\right)\right), \qquad (3)$$

where $\mathcal{T}$ is random 3D transformation. $\mathcal{N}(\cdot)$ is a resample and renormalization function. $\mathcal{M}$ is the classification model. l_{logits} is a logits loss function [3].

The phase shifting attack algorithm We use the l_1 distance as our distance loss to improve the sparsity of perturbation, which is defined as $\|\phi'_a - \phi_a\|_1 = \sum_{u,v} |\phi'_a(u,v) - \phi_a(u,v)|$. Previous work has shown that l_1 loss function can generate sparser perturbation than l_2 loss [18]. The final loss function is defined as

$$l_{total} = l_{adv} + \lambda \cdot Sen \odot \|\phi'_a - \phi_a\|_1, \tag{4}$$

where $\odot$ is Hadamard products. We optimize this loss function by stochastic gradient descent with very small initial noises (10^{-5}) to skip the non-differential area. Last but not least, at the end of each iteration, we clip ϕ_a to $[max(0, \phi_a - 1/n_s), min(1, \phi_a + 1/n_s)]$. This is because when surfaces have large jumps, the multi-step phase-shift algorithm may unwrap the corresponding pixel into a false cycle. After getting the adversarial phase map, we shift the phase $I_p(u_p, v_p)$ in projected fringe patterns to $I_p(u'_p, v_p)$ to get adversarial illuminations. The final attack algorithm is shown in Algorithm 1.

Algorithm 1 Phase Shifting Attack Algorithm

Input: $I_c = \{I_1^c, ..., I_N^c\}, I_p, A_c = K[R\ T]$, target label t
Output: $I_p^{adv} = \{I_1^p, ..., I_N^p\}$

$\phi_a \leftarrow f(I_c)$
for $i = 0$ to N **do**
  Reconstruct the point cloud $\mathcal{P} \leftarrow h(\phi_a)$
  $\mathcal{P}' \leftarrow K \cdot R \cdot \mathcal{P}$
  Compute the gradient $\nabla_{\mathcal{P}'} l_{adv}(\mathcal{P}', t)$
  $\nabla_{\mathcal{P}'} l_{adv}(\mathcal{P}')[:, 0:2] \leftarrow 0$
  $\nabla_{\phi_a} l_{adv}(\phi_a) \leftarrow \nabla_{\mathcal{P}'} l_{adv}(\mathcal{P}') \cdot \frac{\partial \mathcal{P}'}{\partial \phi_a}$
  Compute the total gradient $\triangle = \nabla_{\phi_a} l_{total}(\phi_a)$
  $\phi_a \leftarrow clip(\phi_a + \alpha \cdot \frac{\triangle}{\|\triangle\|_2})$
end for
$u'_p \leftarrow round(\frac{w\phi_a}{2\pi n_s})$
$I_p^{adv}(u_p) \leftarrow I_p(u'_p)$
return $I_p^{adv} = \{I_1^p, ..., I_N^p\}$

3.3. Phase superposition attack

In the real world, the adversary usually cannot directly modify the inherent projector of a structured light system. Therefore, we propose another end-to-end attack called phase superposition attack. As shown in Figure 4, the adversary uses an additional projector to project adversarial noises on the faces, resulting in dodging or impersonation attacks. We suppose the victim model is SLCNN [8] or any other single-fringe-analysis neural network.

The Lambertian rendering model Our end-to-end attack procedure is shown in Figure 3. In order to generate adversarial noises directly, we involve the projection-and-capture process in our optimization loops. Unfortunately, face reflection is a very complex process because of its translucent quality [28], which hasn't been considered in previous optical adversarial attacks. We use the linear Lambertian rendering model [1] to simulate this procedure, which has been used for illumination-invariant face recognition [46]. The linear Lambertian model is

$$I(\boldsymbol{x}) = a(\boldsymbol{x})\boldsymbol{n}(\boldsymbol{x}) \cdot (\boldsymbol{s}_{p_1}(\boldsymbol{x}) + \boldsymbol{s}_{p_2}(\boldsymbol{x})), \tag{5}$$

where $\boldsymbol{x}$ is the 3D face, $I(\boldsymbol{x})$, $a(\boldsymbol{x})$, $\boldsymbol{n}(\boldsymbol{x})$, $\boldsymbol{s_p}(\boldsymbol{x})$ are image intensity, the face's reflectance, the surface normal and light vector from projector p, respectively. We estimate reflectance and surface normal parameters through SfSNet [26], which can decompose a face image as $a(\boldsymbol{x})$,$\boldsymbol{n}(\boldsymbol{x})$,$\boldsymbol{s}(\boldsymbol{x})$ under Lambertian model through a single forward propagation.

The phase superposition attack algorithm To both fool 3D face recognition models and make the recovered 3D face look realistic, we add a 3D Root Mean Square Error (RMSE) to measure the distance from the adversarial point cloud to the ground truth, which has been used for measuring 3D reconstruction accuracy [16]. RMSE is defined as $RMSE = \frac{1}{N_T}\Sigma_{i=1}^{N_T}(\|\mathcal{P}^* - \mathcal{P}\|/n)$, where $\mathcal{P}^*$ and $\mathcal{P}$ are the adversarial and clean point cloud, respectively, and N_T is the attack batch size. To calculate RMSE, we align the adversarial and real 3D faces with 68 3D landmarks and then crop them into the same radiuses. We also consider the projector's gamma distortion in our attacks, modeled as a *tanh* function. The phase superposition attack algorithm is shown in Algorithm 2.

Algorithm 2 Phase Superposition Attack Algorithm

Input: Test 3D face data P, phase reconstruction model $\mathcal{M}_1$, 3D classification model $\mathcal{M}_2$
Output: Adversarial illumination $\boldsymbol{x}$

Initialization : $\boldsymbol{x} \leftarrow \mathbf{1} \times 10^{-5}$ *//Skip the zero*
for $i = 0$ to N **do**
  Relight the face through the Lambertian model
  Get modulated images I'_p by natural 3D renderer.
  Get phase map through $\mathcal{M}_1$
  Get $\mathcal{P}^*$ through differential 3D reconstruction
  $RMSE \leftarrow \frac{1}{N_T}\Sigma_{i=1}^{N_T}(\|\mathcal{P}^* - \mathcal{P}\|/n)$
  Get the prediction through $\mathcal{M}_2$
  Compute the adversarial loss l_{adv}
  $l_{total} \leftarrow l_{adv} + \lambda_1 \cdot RMSE + \lambda_2 \cdot Sen \odot \|x\|_1$
  Compute gradient $\triangle = \nabla_{\boldsymbol{x}} l_{total}(\boldsymbol{x})$
  $\boldsymbol{x} \leftarrow max\{0, min\{1, \boldsymbol{x} + \alpha \cdot sign(\delta)\}\}$
end for
return Adversarial illumination $\boldsymbol{x}$

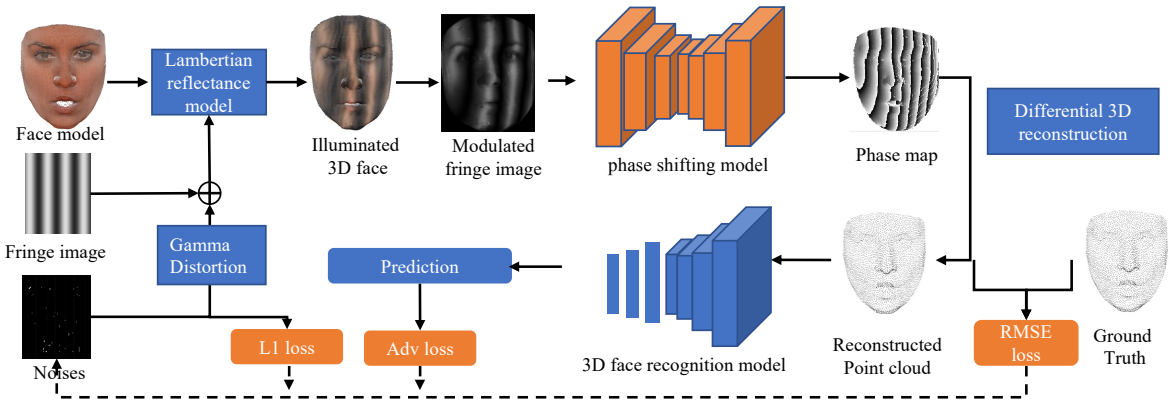


Figure 3. The attack procedure of phase superposition attack. The fringe images and noises relight the human face through the Lambertian model. Then a natural 3D renderer [14] renders the 3D model as modulated fringe images. 3D point clouds are then reconstructed by SLCNN and are classified by 3D face recognition model. Losses include adversarial loss, L_1 loss and RMSE loss that assesses the 3D face reconstruction accuracy. During the attack, the SLCNN and 3D face recognition model are frozen and the noises are updated.

4. Implementation details

Datasets We first train 3D face recognition models on 3D face datasets for attack evaluation. Three datasets acquired by structured light cameras are used: Bosphorus [25], Eurecom [17] and SIAT-3DFE [43]. The Bosphorus dataset is collected by Inspeck Mega Capturor II. It consists of 105 different faces with a rich set of expressions and occlusions. Eurecom consists of 52 faces and is acquired by Kinect. SIAT-3DFE consists of over 400 subjects and has original fringe images. We also collect ten people's faces using our own structured light camera and add them to the above datasets. We downsample these datasets to 4K points through farthest point sampling to train 3D classifiers.

3D face recognition models We evaluate our attacks on state-of-art 3D point cloud classification models, including Pointnet [22], Pointnet++ [23], DGCNN [32] and CurveNet [40]. We also evaluate attacks on the depth-image-based 3D face recognition model FR3DNet [11]. We implement dodging and impersonation attacks on these models. The dodging attack aims to make 3D classifiers classify faces into any classes except the ground truth. The impersonation attack aims to make them classify faces into random labels.

Hyperparameter settings The search space of λ in Equation 4 is set as $[10^{-5}, 10^{5}]$. We use binary search to narrow down λ. We set the binary search step as 10, the iteration number as 100, and the minimum logit difference as 30 to guarantee the attack success rate. We terminate the optimization process after reaching the binary search step and return the adversarial examples with minimum distance loss. We use ASR and RMSE distance between the adversarial and original point clouds as evaluation metrics.

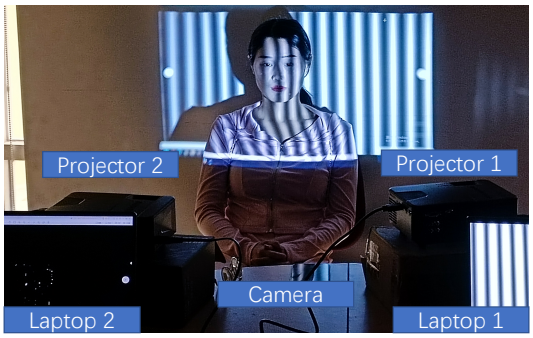


Figure 4. The physical settings of the phase superposition attack. The adversary uses an additional projector to add noises to the original fringe images for 3D reconstructions.

Physical settings As shown in Figure 4, we build a real-world structured light system that refers to Piccirilli's design [21] for physical attacks, which includes an industry camera and home projectors. We capture images at a distance of about 1.5 meters. The projector and the camera's resolutions are both 1600×1200. We linearize the color distortion before the images are projected, then detect and crop the face in the modulated images using the Viola-Jones algorithm and resize it to 128×128 centered by the face. We reconstruct the 3D data using the 12-step phase-shift algorithm or deep-learning-based methods such as SLCNN.

The projected images may be distorted because of manufacturing errors in the projector-camera system. This nonlinear distortion will cause a drop in the attack success rate if not taken into account. The luminance non-linearity usually are formulated as $g(u) = u^{\gamma}$, where $u \in [0, 1]$ is the input intensity of the projector and $g(u)$ is the output intensity of the projector [7]. However, our experiments found that the $tanh$ function can better fit our projectors' distor-

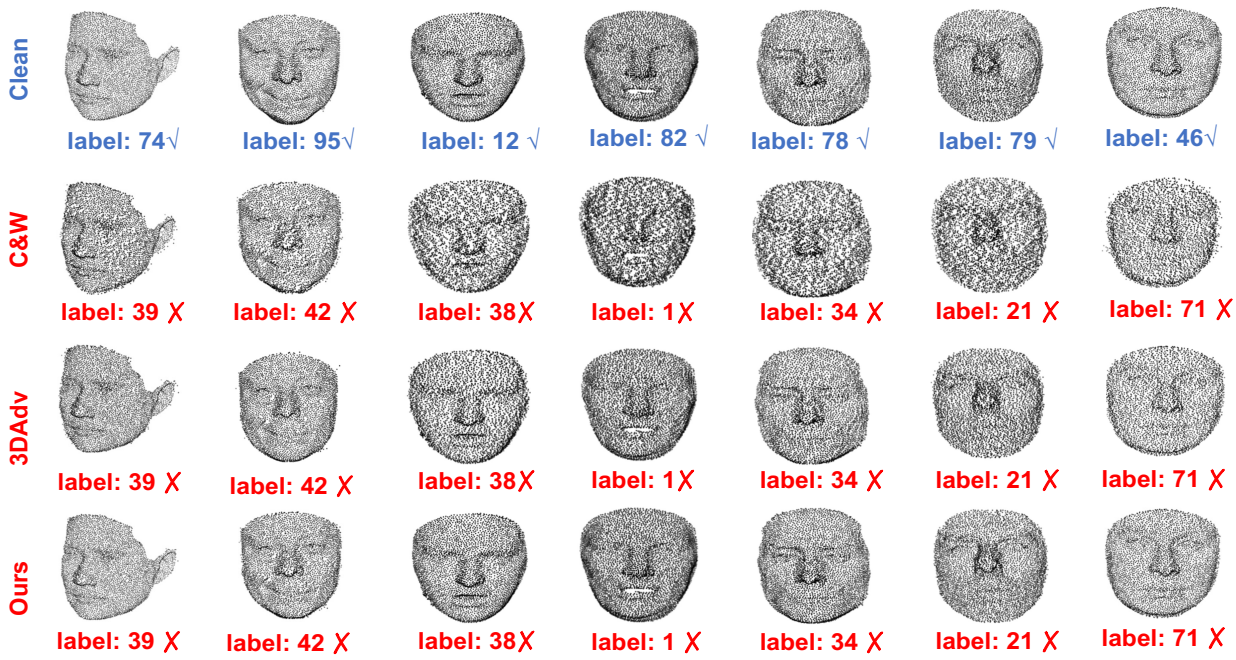


Figure 5. Demonstrations of clean and adversarial face point clouds generated by C&W, 3DAdv and ours attacks.

tion, which is $g(u) = \frac{tanh(\gamma \cdot (2u-1))+1}{2}$, where γ is a hyper-parameter to describe the projector's inherent property. We leave the details about computing γ in the appendix.

5. Experimental evaluation

We implement attacks in both the digital world and the physical world. We present our experiments on 1) targeted and untargeted attacks on various classification backbones, 2) physical attack results, 3) ablation study of different modules, and 4) attack transferability.

5.1. Attack performance

Digital attack results To evaluate the targeted and untargeted attack success rate, we first simulate attacks through the Lambertian rendering model and then attack different models that are trained on three datasets. For targeted attacks, the label is randomly chosen from three datasets. We compared our attacks with several state-of-the-art 3D attacks, including C&W attack [3] that uses l_2 distance, 3Dadv [39] that uses Chamfer distance, KNNadv [30] that use kNN distance and GeoA3 [34] that use local curvature distance. Note that recently shape-invariant attack has been proposed [13], but it is not physically realizable, therefore, is not considered in this paper.

As revealed in Table 1 and Table 2, we achieve an average ASR of 95% on dodging attacks and 47% on impersonation attacks for the pixel shifting attack. At the same time, our attacks have fewer 3D reconstruction errors than previous 3D attacks. This is because our attacks need fewer points to be modified and consider the physical constraint in the 3D reconstruction by projecting the perturbation in specific directions. For the targeted attack, we outperform previous attacks except for KNNadv. This is may because KNNadv can generate adjacent clusters, which are more robust to the downsampling process of Pointnet++. We also visualize the adversarial point clouds in Figure 5.

Attack FR3DNet Because our attacks can generate points at arbitrary positions, they can also be applied to depth-image-based 3D face recognition. FR3DNet [11] is a deep CNN model designed for 3D face recognition which uses depth information as input. The depth image can be easily recovered from the phase map so only minor modifications are required for our attacks.

Physical attack results We attack 10 people in physical attacks and achieved 90% ASR for untargeted attacks and 40% ASR for targeted attacks for both phase shifting and superposition attacks. Figure 6 shows the real-world phase shifting attacks' results compared with other attacks. We show the phase superposition attack's results in the appendix. To compare with benchmark attacks, we use their loss functions to generate the 3D adversarial point clouds and then reverse them to the fringe patterns. We conduct targeted attacks on PointNet with randomly chosen labels. The results show that the reconstructed 3D point clouds by

	PointNet		PointNet++(SSG)		Point++(MSG)		DGCNN		CurveNet	
Metrics	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE
C&W [3]	**0.98**	0.35	0.92	0.45	0.88	0.36	0.94	0.26	0.88	0.43
3Dadv [39]	0.89	0.27	0.86	0.34	0.77	0.25	0.89	0.16	0.95	0.36
KNNadv [30]	0.86	0.15	0.85	0.22	0.79	0.23	0.85	0.08	0.86	0.17
GeoA3 [34]	0.75	0.18	0.91	**0.14**	0.85	0.13	0.84	0.09	0.97	0.16
Ours	0.95	**0.13**	**0.93**	0.15	**0.99**	**0.11**	**0.96**	**0.05**	**0.97**	**0.09**

Table 1. The untargeted attack performance. We evaluate the attack performance by the attack success rate (higher is better) and RMSE [16] (lower is better). RMSE can be used to measure the 3D reconstruction error. Compared with the state-of-art geometry-ware attack GeoA3, our attacks have fewer 3D reconstruction errors while maintaining a high attack success rate. The RMSE is multiplied by 10 for comparison.

	PointNet		PointNet++(SSG)		Point++(MSG)		DGCNN		CurveNet	
Metrics	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE	ASR (%)	RMSE
C&W [3]	0.57	0.76	0.42	0.79	0.32	0.81	0.43	0.69	0.48	0.88
3Dadv [39]	0.58	0.65	0.39	0.66	0.35	0.78	0.37	0.45	0.53	0.65
KNNadv [30]	0.52	0.54	**0.45**	0.51	**0.38**	0.54	0.16	0.34	0.27	0.49
GeoA3 [34]	0.45	0.52	0.35	0.54	0.24	0.45	0.28	0.31	0.25	0.42
Ours	**0.62**	**0.33**	0.37	**0.32**	0.26	**0.44**	**0.55**	**0.25**	**0.57**	**0.39**

Table 2. The targeted attack performance on our method and the other four attacks. The RMSE is multiplied by 10 for comparison.

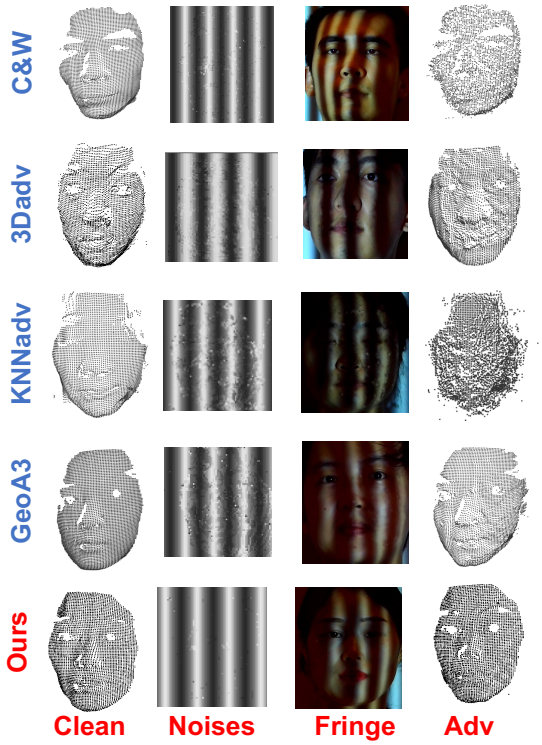


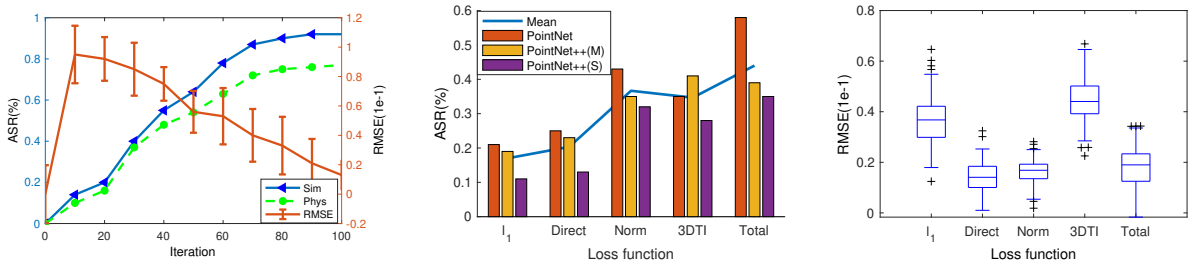
Figure 6. The physical phase shifting attack results. We apply previous attacks in the structured light system for comparison. **The first four rows**: benchmark attacks. **The last row**: Our attack.

our attacks are more similar to real faces, and few perturbations are needed for projection.

5.2. Ablation study

Quantitative results of different modules Figure 7a shows the influence of maximum iteration on ASR. With the increase of iterations for one attack, the ASR gradually increases and the RMSE decreases. To evaluate the effects of different modules, we involve the random transformations of the head and necessary preprocess steps (normalization and down-sampling) in the simulated system, and evaluate the ASR on different models (Figure 7b). The naive l_1 attack without any techniques suffers a low ASR. With the effectiveness of direction constraint, renormalization, and 3D-TI, the targeted ASR increases by 34% on Pointnet, 20% on Pointnet++ SSG, 24% on Pointnet++ MSG, 20% on DGCNN, and 33% on CurveNet. The untargeted ASR also boosted by 27%. Figure 7c shows the influences of different modules on the RMSE. We find 3D-TI may slightly improve the perturbation size. We think this is the necessary cost to resist environmental changes.

Qualitative results of Lambertian model As revealed in Figure 8, the Lambertian model can generate photorealistic rendering results for relighting process. These two faces are from the Bosphorus dataset. From left to right are the face images, the ground truth depth maps, the 3D point clouds rendered by the Lambertian model and the 2D images rendered from the 3D models.



(a) Iteration (b) ASR (c) RMSE

Figure 7. The results of ablation study. l_1 means that only use l_1 loss. $Direct$ means the direction constraint. $Norm$ is the renormalization. $3DTI$ means 3D transformation invariant loss. $Total$ means the loss function that combines all modules.

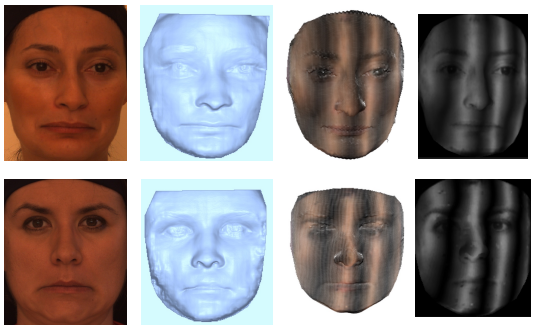
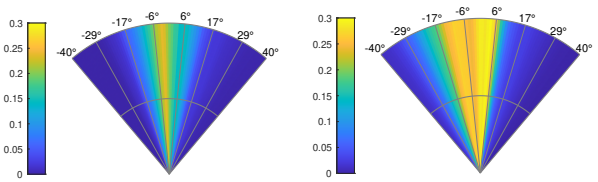


Figure 8. **From left to right**: the face images, the depth maps, the 3D faces rendered by the Lambertian model, and modulated fringe images. **Top row**: normal images. **Bottom row**: attacked images.

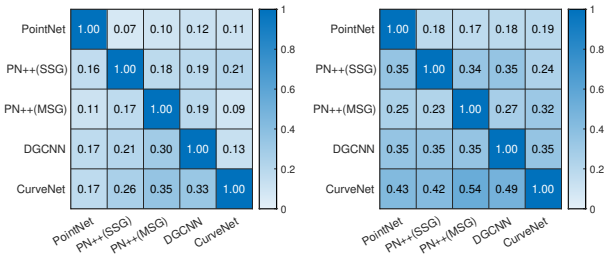


(a) Without 3D-TI (b) With 3D-TI

Figure 9. The adversarial examples' robustness against face rotations. We generate perturbations on Pointnet and then add them to the rotated point cloud. We plot the predictions on the target label without (left) and with (right) 3D-TI module.

Effect of random 3D rotations To evaluate the effectiveness of the 3D transformation-invariant loss (3D-TI), we randomly rotate the original point cloud at different angles and add the same perturbations to the rotated point clouds. Then we compare the predictions on the target label after the softmax layer. As shown in Figure 9, with the 3D-TI module, the prediction is more robust to rotations.

Transferability of adversarial examples Previous work has shown that random transformations can improve the transferability of adversarial images [42]. We find that 3D-TI loss can also improve the transferability of 3D adversarial examples, especially for dodging attacks. We show the transferability of dodging attacks of C&W and ours in Figure 10. Moreover, we find that transferability on face datasets is higher than on object datasets like ModelNet40 [37]. This is may because the latent features of different faces are more similar than different objects.



(a) C&W (b) Ours

Figure 10. The transferability of dodging attacks of C&W and ours. The vertical axis is the shadow model. The horizontal axis is the victim model.

6. Conclusion

In this paper, we propose a novel structured light attack that is practical and effective for 3D face recognition systems in real-world settings. Our attack incorporates the facial reflection process into the attack pipeline, resulting in a more realistic simulation of real-world attacks. Moreover, we enhance the robustness of adversarial examples to random head movements by integrating random 3D transformations into the attack pipeline. We assess the effectiveness of our attacks on state-of-the-art 3D deep learning models. The results of quantitative and qualitative analyses demonstrate that our attacks can successfully deceive real-world systems with fewer perturbations than previous attacks. Furthermore, our approach is not limited to the structured light scanners used in our study and can be extended to other types of scanners, such as infrared scanners.

References

[1] Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2):218–233, 2003. 4

[2] Yulong Cao, Chaowei Xiao, Benjamin Cyr, Yimeng Zhou, Won Park, Sara Rampazzi, Qi Alfred Chen, Kevin Fu, and Z Morley Mao. Adversarial sensor attack on lidar-based perception in autonomous driving. In *Proceedings of the ACM SIGSAC conference on computer and communications security*, pages 2267–2281, 2019. 2

[3] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy (sp)*, pages 39–57. Ieee, 2017. 3, 6, 7

[4] Sam Van der Jeught and Joris J. J. Dirckx. Deep neural networks for single shot structured light profilometry. *Opt. Express*, 27(12):17091–17101, Jun 2019. 3

[5] Yinpeng Dong, Hang Su, Baoyuan Wu, Zhifeng Li, Wei Liu, Tong Zhang, and Jun Zhu. Efficient decision-based black-box adversarial attacks on face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7714–7722, 2019. 1

[6] Ranjie Duan, Xiaofeng Mao, A Kai Qin, Yuefeng Chen, Shaokai Ye, Yuan He, and Yun Yang. Adversarial laser beam: Effective physical-world attack to dnns in a blink. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16062–16071, 2021. 1

[7] H. Farid. Blind inverse gamma correction. *IEEE Transactions on Image Processing*, 10(10):1428–1433, 2001. 5

[8] Shijie Feng, Qian Chen, Guohua Gu, Tianyang Tao, Liang Zhang, Yan Hu, Wei Yin, and Chao Zuo. Fringe pattern analysis using deep learning. *Advanced Photonics*, 1(2):025001, 2019. 3, 4

[9] Shijie Feng, Chao Zuo, Wei Yin, Guohua Gu, and Qian Chen. Micro deep learning profilometry for high-speed 3d surface imaging. *Optics and Lasers in Engineering*, 121:416–427, 2019. 3

[10] Shijie Feng, Chao Zuo, Liang Zhang, Tianyang Tao, Yan Hu, Wei Yin, Jiaming Qian, and Qian Chen. Calibration of fringe projection profilometry: A comparative review. *Optics and Lasers in Engineering*, 143:106622, 2021. 1, 2

[11] Syed Zulqarnain Gilani and Ajmal Mian. Learning from millions of 3d scans for large-scale 3d face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1896–1905, 2018. 1, 5, 6

[12] Abhiram Gnanasambandam, Alex M Sherman, and Stanley H Chan. Optical adversarial attack. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 92–101, 2021. 1, 2

[13] Qidong Huang, Xiaoyi Dong, Dongdong Chen, Hang Zhou, Weiming Zhang, and Nenghai Yu. Shape-invariant 3d adversarial point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15335–15344, 2022. 2, 6

[14] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3d mesh renderer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3907–3916, 2018. 3, 5

[15] Donghyun Kim, Matthias Hernandez, Jongmoo Choi, and Gérard Medioni. Deep 3d face identification. In *2017 IEEE international joint conference on biometrics (IJCB)*, pages 133–142. IEEE, 2017. 1

[16] Feng Liu, Ronghang Zhu, Dan Zeng, Qijun Zhao, and Xiaoming Liu. Disentangling features in 3d face shapes for joint face reconstruction and recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5216–5225, 2018. 4, 7

[17] Rui Min, Neslihan Kose, and Jean-Luc Dugelay. Kinectfacedb: A kinect database for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11):1534–1548, 2014. 5

[18] Apostolos Modas, Seyed-Mohsen Moosavi-Dezfooli, and Pascal Frossard. Sparsefool: a few pixels make a big difference. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9087–9096, 2019. 4

[19] Dinh-Luan Nguyen, Sunpreet S Arora, Yuhang Wu, and Hao Yang. Adversarial light projection attacks on face recognition systems: A feasibility study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 814–815, 2020. 1, 2, 3

[20] Nicole Nichols and Robert Jasper. Projecting trouble: Light based adversarial attacks on deep learning classifiers. *arXiv preprint arXiv:1810.10337*, 2018. 1, 2

[21] Marco Piccirilli, Gianfranco Doretto, Arun Ross, and Donald Adjeroh. A mobile structured light system for 3d face acquisition. *IEEE Sensors Journal*, 16(7):1854–1855, 2016. 2, 5

[22] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 5

[23] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017. 5

[24] Gang Qiao, Yiyang Huang, Yiping Song, Huimin Yue, and Yong Liu. A single-shot phase retrieval method for phase measuring deflectometry based on deep learning. *Optics Communications*, 476:126303, 2020. 3

[25] Arman Savran, Neşe Alyüz, Hamdi Dibeklioğlu, Oya Çeliktutan, Berk Gökberk, Bülent Sankur, and Lale Akarun. Bosphorus database for 3d face analysis. In *European workshop on biometrics and identity management*, pages 47–56. Springer, 2008. 5

[26] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D Castillo, and David W Jacobs. Sfsnet: Learning shape, reflectance and illuminance of facesin the wild'. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6296–6305, 2018. 4

[27] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *Proceedings of the 2016 acm sigsac conference on computer and communications security*, pages 1528–1540, 2016. 1

[28] William AP Smith and Edwin R Hancock. Estimating facial reflectance properties using shape-from-shading. *International journal of computer vision*, 86(2):152–170, 2010. 4

[29] Yang Tan, Hongxin Lin, Zelin Xiao, Shengyong Ding, and Hongyang Chao. Face recognition from sequential sparse 3d data via deep registration. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019. 1

[30] Tzungyu Tsai, Kaichen Yang, Tsung-Yi Ho, and Yier Jin. Robust adversarial objects against deep learning models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 954–962, 2020. 1, 2, 6, 7

[31] James Tu, Mengye Ren, Sivabalan Manivasagam, Ming Liang, Bin Yang, Richard Du, Frank Cheng, and Raquel Urtasun. Physically realizable adversarial examples for lidar object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13716–13725, 2020. 2

[32] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019. 5

[33] Yuxin Wen, Jiehong Lin, Ke Chen, C. L. Philip Chen, and Kui Jia. Geometry-aware generation of adversarial point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2984–2999, 2022. 2

[34] Yuxin Wen, Jiehong Lin, Ke Chen, CL Philip Chen, and Kui Jia. Geometry-aware generation of adversarial point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 6, 7

[35] Matthew Wicker and Marta Kwiatkowska. Robustness of 3d deep learning in an adversarial setting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[36] Nils Worzyk, Hendrik Kahlen, and Oliver Kramer. Physical adversarial attacks by projecting perturbations. In *International Conference on Artificial Neural Networks*, pages 649–659. Springer, 2019. 1, 2

[37] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 8

[38] Zhoujie Wu, Chao Zuo, Wenbo Guo, Tianyang Tao, and Qican Zhang. High-speed three-dimensional shape measurement based on cyclic complementary gray-code light. *Optics express*, 27(2):1283–1297, 2019. 2

[39] Chong Xiang, Charles R Qi, and Bo Li. Generating 3d adversarial point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9136–9144, 2019. 2, 6, 7

[40] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 895–904, 2021. 5

[41] Zihao Xiao, Xianfeng Gao, Chilin Fu, Yinpeng Dong, Wei Gao, Xiaolu Zhang, Jun Zhou, and Jun Zhu. Improving transferability of adversarial patches on face recognition with generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11845–11854, 2021. 1

[42] Cihang Xie, Zhishuai Zhang, Yuyin Zhou, Song Bai, Jianyu Wang, Zhou Ren, and Alan L Yuille. Improving transferability of adversarial examples with input diversity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2730–2739, 2019. 8

[43] Yuping Ye, Zhan Song, Junguang Guo, and Yu Qiao. Siat-3dfe: A high-resolution 3d facial expression dataset. *IEEE Access*, 8:48205–48211, 2020. 5

[44] Tianhang Zheng, Changyou Chen, Junsong Yuan, Bo Li, and Kui Ren. Pointcloud saliency maps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1598–1606, 2019. 2

[45] Song Zhou and Sheng Xiao. 3d face recognition: a survey. *Human-centric Computing and Information Sciences*, 8(1):1–27, 2018. 1

[46] Shaohua Kevin Zhou, Gaurav Aggarwal, Rama Chellappa, and David W Jacobs. Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):230–245, 2007. 4

[47] Zhe Zhou, Di Tang, Xiaofeng Wang, Weili Han, Xiangyu Liu, and Kehuan Zhang. Invisible mask: Practical attacks on face recognition with infrared. *arXiv preprint arXiv:1803.04683*, 2018. 1, 2

[48] Chao Zuo, Jiaming Qian, Shijie Feng, Wei Yin, Yixuan Li, Pengfei Fan, Jing Han, Kemao Qian, and Qian Chen. Deep learning in optical metrology: a review. *Light: Science & Applications*, 11(1):1–54, 2022. 3