

Construction and Calibration of a Stereo Vision Acquisition Platform for Multimodal Face Antispoofing

Zuhe Li¹, Yuhao Cui^{1,*}, Weihua Liu², Yongshuang Yang¹

¹*School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China*

²*China Mobile Research Institute, Beijing, 100053, China*

**Corresponding author*

Keywords: Face anti-spoofing, multispectral data, multidimensional information, depth alignment, binocular matching

Abstract: In recent years, face antispoofing (FAS) has played an important role in protecting face recognition systems from various types of attacks, and with the emergence of various large-scale face antispoofing datasets, multimodal face antispoofing algorithms have become the mainstream method in the field of FAS. Therefore, how to efficiently collect high-precision and high-resolution multimodal face images is also an important issue in the field of FAS. This paper uses the feature that multispectral data have increasingly subdivided bands compared to visible light data to analyse and identify the essential properties of objects. At the same time, we consider the real-time requirements of the human face detection task. The mode of combining three channels of visible light and one channel of infrared 960 nm is adopted to form four-channel multispectral data to obtain multidimensional information of the target and minimize band redundancy in terms of the data construction mode. Finally, we use the principle of speckle structured light to obtain the spatial three-dimensional point cloud data of the scene, and a complete scene data type is further constructed, which provides strong support for the application of multimodal face antispoofing technology.

1. Introduction

At present, face recognition technology is widely used in fields such as information security, financial services, cultural and sports entertainment, smart cities, and social media and is developing towards the trend of automation and unsupervised. Automatically and efficiently identifying the authenticity of faces and resisting spoofing attacks have become urgent problems in the field of face recognition. Face antispoofing aims to judge whether the face data collected by the face recognition system are a real face to prevent fake face attacks. Its essence is to distinguish the signals sent by the biological face and the nonbiological face to achieve the purpose of distinguishing the truth from the false face [1, 2].

Reasonable description, efficient extraction and accurate discrimination of facial biometric features are important issues for face anti-spoofing [3]. As shown in Figure 1, the physical forms and attack methods of fake faces are different, and the main methods can be paper, silicone and

resin. Therefore, in some systems with high security requirements, relying only on traditional visible light sensors to collect image signals and extract features in the spatial dimension cannot essentially fully express the biological signal characteristics of real faces. This makes the recognition accuracy of the face antispoofing system unable to meet the actual application requirements.

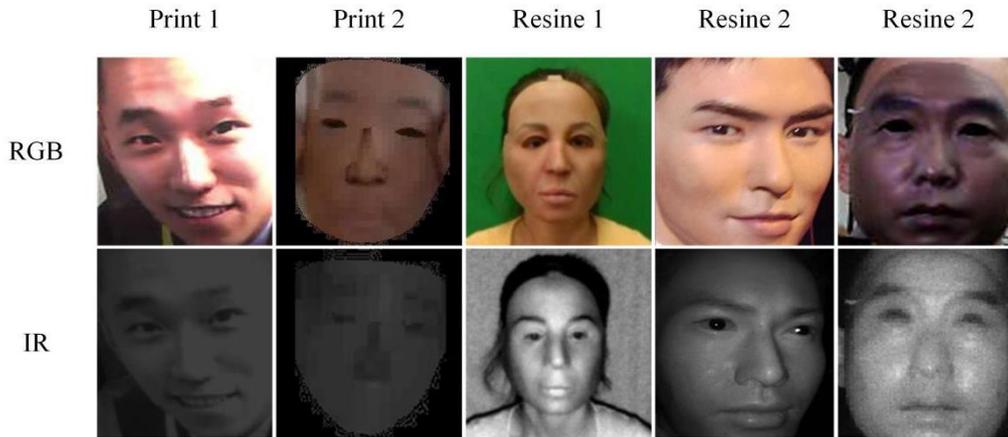


Figure 1: Spoof face attack methods in different modalities.

In recent years, many scholars have used multimodal face data fusion to address various attacks in face recognition systems and realized the detection of real and fake faces by fusing feature information from multiple modalities. Spectral data contain rich triple information of space, radiation and spectrum, which brings opportunities for face antispoofing. The spectral characteristic curve at a higher spectral resolution can provide an accurate description of the material properties of the face.

At present, few studies on face anti-spoofing comprehensively utilize effective spectral bands, depth information and visible light information. Some previous studies on multimodal anti-spoofing only used near-infrared bands to resist light interference and did not analyse in depth the effective discrimination between live and spoof features of faces under a certain band and the impact of band redundancy on the mining of live features. Therefore, how to collect multimodal face image data and provide accurate "source information" for deep neural network mining of live features is also an important task in face antispoofing. Starting from the problem of data source acquisition, this paper builds a stereo vision acquisition platform for multimodal face antispoofing, which provides strong support for the application of multimodal face antispoofing technology.

2. Related Works

At present, with the emergence of various realistic and fake face attack methods, face detection methods based on visible light cannot meet some systems with high security requirements, so multidimensional information is adopted and used as the source input of the anti-spoofing system [4-7].

Among them, the most common multidimensional data information is the combination of visible light and infrared information. Zhang et al. [8] introduced near-infrared images into the human face detection task and distinguished real and fake faces by distinguishing material spectral properties. The method judges real and fake faces based on the difference in reflectance between real and fake faces in certain spectral bands. This method selects two groups of effective bands from the visible light to the near-infrared spectrum to carry out the task of antispoofing, which provides a new research direction for multimodal face antispoofing.

In recent years, the maturity of multisource fusion sensing technology has provided a new development direction for face anti-spoofing. Multimodal face antispoofing research combines depth information, visible light information, and near-infrared information so that the algorithm model under this data form has stronger generalization ability and better detection performance [9-14].

Zhang et al. [15] constructed a face antispoofing benchmark dataset, CASIA-SURF, which divided fake faces into six different attack forms and simulated fake face attacks by using paper remakes of various materials. Based on the CASIA-SURF dataset, Liu et al. [16] established a large-scale cross-ethnic face anti-spoofing dataset CASIA-SURF CeFA. This dataset covers three races, three modes, 1607 subjects and 2D plus 3D attack types and adds live samples of different skin colors and face model samples such as silicone and resin. Boulkenafet et al. [17] constructed a face anti-spoofing database OULU-NPU that includes two types of print attacks and video attacks. The database consists of 4950 live and spoof videos and contains three different lighting conditions and backgrounds. Liu et al. [18] constructed a novel face antispoofing database that contains a large amount of video data from live broadcasts and spoof videos of 165 subjects, each with variations in distance, pose, lighting, and expression.

3. Methods

3.1. Construction and Calibration of a Stereo Vision Acquisition Platform for Multimodal Face Antispoofing

Compared with visible light data, multispectral data have increasingly subdivided bands, so they can quantitatively and qualitatively analyse and identify the essential properties of objects. Considering the real-time requirements of the behavior detection task, this study adopts the mode of combining the three channels of visible light and the single channel of infrared 960 nm to form four-channel multispectral data to obtain the multidimensional information of the target and minimize the band redundancy in terms of the data construction mode. Remain. While the multispectral module acquires the structure and texture information of the target, its active infrared light source band can acquire target images that are not interfered with by external ambient light, thus bringing convenience to the detection and identification of targets from the data layer.

At the same time, the principle of speckle structured light [19-20] is used to obtain the spatial three-dimensional point cloud data of the scene and further construct a complete scene data type. This chapter focuses on how to construct aligned 4-channel multispectral data and the principle, method and calibration scheme of matching depth data in this study. The basic calibration process is shown in Figure 2.



Figure 2: Flow chart of the multispectral/depth module calibration scheme.

3.2. Depth Calculation Based on Speckle Matching

There are many ways to obtain the scene depth map, which can be divided into monocular and binocular methods in terms of the implementation system. The binocular camera obtains depth through binocular stereo vision. This method relies on two pictures (color or grayscale images) taken to calculate the depth, so it is sometimes called a passive binocular depth camera. Depth

cameras based on the principle of time-of-fit (TOF) and structured light are different. This system uses an image sensor to complete depth measurement by means of laser projection. The scene object depth is calculated by comparing the object image with calibration patches already stored in hardware.

In this study, structured light speckle projection is used to increase the texture characteristics of the object through laser projection, and then two image sensors are used to collect speckle data projected on the target surface to calculate depth information. In the depth acquisition calculation, regardless of the monocular and binocular systems, the most common similar triangle features in geometry are used to calculate the disparity, and the depth information is further calculated through the original disparity. The basic geometric model is shown in Figure 3.

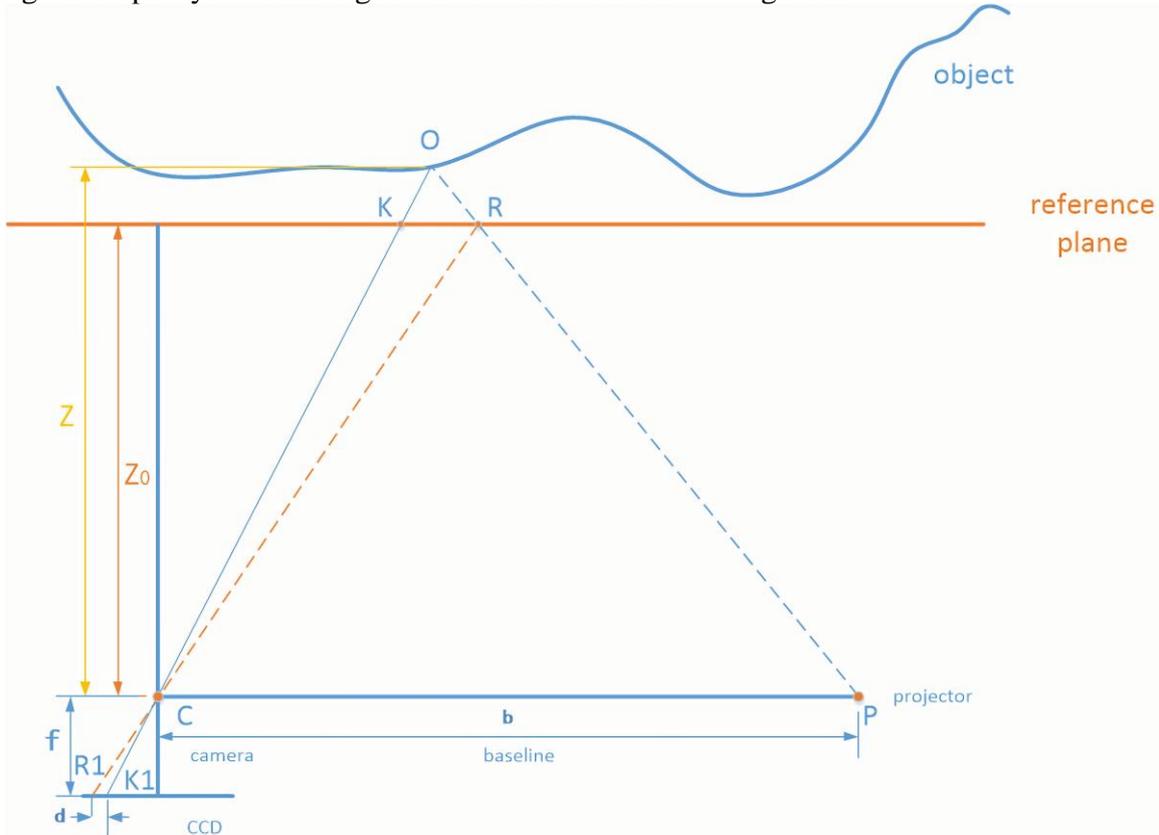


Figure 3: Schematic diagram of depth camera calibration.

Similar to $\Delta R1K1C$ according to $\Delta RK1C$:

$$\frac{R_1K_1}{RK} = \frac{f}{z_0} \quad (1)$$

Similar to ΔOCP according to ΔOKR :

$$\frac{RK}{b} = \frac{z - z_0}{z} \quad (2)$$

R_1K_1 is the parallax d ; the depth formula is as follows:

$$z = \frac{z_0}{1 + \frac{z_0}{f} * d} \quad (3)$$

Among them, b is the baseline, which refers to the distance between the center of the laser and the center of the infrared camera, and this parameter depends on the physical structure design of the camera; d is the original parallax, which refers to the pixel distance between the pixel position of the imaging point of the measured target on the infrared camera pixel array and the distance from the reference point; f is the focal length, which refers to the focal length of the infrared camera lens, and this parameter depends on the optical design of the infrared camera lens; z_0 is the calibration distance, which refers to the image acquisition distance when the camera is calibrated, and this parameter depends on the calibration design; z is the distance or depth from the target point to the baseline; from the calculation formula (3) of z , we know that, except d , other variables are fixed values, which can be obtained by optical-mechanical structure measurement, so the depth value can be calculated by obtaining d . For the depth parallax calculation method of speckle structured light matching, the acquisition of d uses the calculation method of the sum of absolute errors to match the speckle.

3.3. Camera Internal and External Parameter Calibration

To obtain the focal length f of the camera, it is necessary to calculate the internal parameter matrix of the camera through camera calibration. To subsequently align the color image with the depth and infrared images, it is necessary to calculate the rotation and displacement matrix between the color camera sensor and the infrared camera sensor. First, establish the rigid body transformation relationship of the object point from the world coordinate system $M_w(x_w, y_w, z_w)$ to the camera coordinate system $M_c(x_c, y_c, z_c)$:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (4)$$

Second, we establish the coordinate transformation relationship from the camera coordinate system $M_c(x_c, y_c, z_c)$ to the camera plane coordinate system $M_u(x_u, y_u)$:

$$\begin{cases} x_u = f \frac{x_c}{z_c} \\ y_u = f \frac{y_c}{z_c} \end{cases} \quad (5)$$

where f is the focal length of the camera (the focal length is the distance from the optical center of the camera to the imaging plane), and the form of formula (4) in homogeneous coordinates can be expressed as:

$$\begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (6)$$

In the calibration process, we have not considered the lens distortion model (linear model) and combine formulas (5) and (6) to obtain the following expression:

$$\begin{aligned}
z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} 1/d_x & -\cot\theta/d_x & u_0 \\ 0 & 1/d_y \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \alpha & \gamma & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}
\end{aligned} \tag{7}$$

For the camera system used in this study, the M1 and M2 matrices mentioned above need to be determined. M1 is recorded as the camera internal parameter matrix, which is related to the camera and not related to the external calibration object. M2 is the external parameter transformation matrix of the camera and the calibration object, which is generally a rigid body transformation. Using Zhang Zhengyou's calibration method [21] to calibrate the internal parameters of the camera, the specific process is as follows:

- Shoot the image of the calibration board, shoot 3 frames of plane targets with different attitudes, and DLT only needs one frame to shoot 3D calibration targets;
- After shooting the calibration board, identify the feature points on the calibration board, and encode the feature points on the calibration board to establish the 2D-3D correspondence in the image;
- After finding the 2D-3D corresponding point pairs, perform a linear initial value solution;
- After the initial value solution, the nonlinear optimization iterative solution is used to solve the exact solution and distortion coefficient.

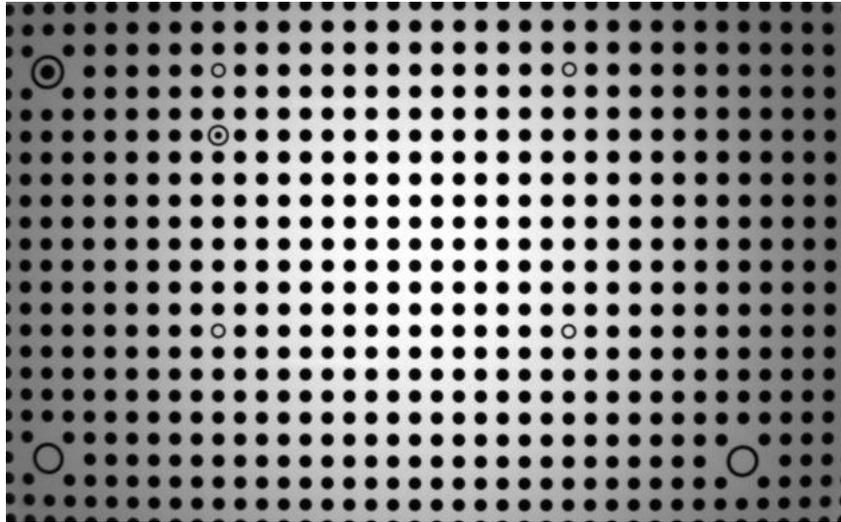


Figure 4: Schematic diagram of the ring calibration board.

From the above calibration steps, shooting the calibration target is the first step of calibration. The design principle of the calibration board is easy to distinguish from the background, accurate positioning, anti-camera perspective transformation, and routines such as line information and surface information calibration points. Conventional calibration targets include 3D calibration targets and 2D calibration plates, among which there are checkerboards in the 2D plane target plates and circle center target plates. Among them, the planar calibration plate has a planarity requirement

for the planarity of the calibration plate. If the planarity is not sufficient, it will seriously affect the subsequent calibration accuracy. The feature points on the plane calibration board generally have specific positioning codes to sort the feature points, as shown in Figure 4.

3.4. Depth Alignment (D2C) and Binocular Matching

The purpose of depth alignment is to use known calibration information to establish a simple mapping of depth and color images on 2D coordinates so that a 2D coordinate can be associated with depth and color pixels to obtain an RGBD image. After calibration and alignment, the internal and external parameters of each camera will be obtained. According to the known external parameter information \mathbf{T} (including rotation \mathbf{R} and translation \mathbf{t}), the following relationship can be established:

$$\begin{cases} P_L = \mathbf{T}_{LW} P_W \\ P_C = \mathbf{T}_{CW} P_W \end{cases} \quad (8)$$

where L and C represent the infrared camera and color camera coordinate systems, respectively, $P = [X_L, Y_L, Z_L]^T$ represents the coordinates of any vector in R^3 (3-dimensional Euclidean space) in the infrared camera coordinate system (referred to as infrared camera coordinates), and $P = [X_W, Y_W, Z_W]^T$ is the world coordinate of the vector and the color camera coordinate of the vector. \mathbf{T}_{LW} and \mathbf{T}_{CW} are the transfer matrices from the world coordinate system to the infrared camera and color camera coordinate systems respectively.

For monocular, the infrared camera is the depth camera (indicated by D). It is easy to obtain the conversion relationship of the space vector from the depth camera to the color camera coordinate system:

$$P_C = \mathbf{T}_{CW} \mathbf{T}_{LW}^{-1} P_L = \mathbf{T}_{CW} \mathbf{T}_{LW}^{-1} P_D \quad (9)$$

If uv is used to refer to the image pixel coordinate system and K is used to refer to the camera internal reference matrix, then the coordinate relationship of a certain pixel position $P_{D_{uv}}$ on the depth map mapped to the color image is:

$$Z_C \mathbf{K}_C^{-1} P_{C_{uv}} = Z_D \mathbf{T}_{CW} \mathbf{T}_{LW}^{-1} \mathbf{K}_L^{-1} P_{D_{uv}} \quad (10)$$

Among them, \mathbf{K}_L (internal reference of infrared camera) and \mathbf{K}_C (internal reference of color camera) are the known information obtained after alignment and calibration, Z_D is the depth value, and $P_{D_{uv}} = [u_D, v_D, 1]^T$ is the alignment value in the pixel coordinate system of the depth map. The calculation on the right side of the equation is P_C . After normalizing P_C (dividing both sides of the equation by Z_C), $[X_C/Z_C, Y_C/Z_C, 1]^T$ is obtained. After the internal reference of the color camera (both ends of the equation are multiplied by \mathbf{K}_C), the homogeneous coordinate $P_{C_{uv}} = [u_D, v_D, 1]^T$ in the pixel coordinate system of the color image can be obtained.

If f_{dist} and f_{undist} are the functions of adding distortion and dedistorting, respectively, then the alignment process after considering distortion is:

$$Z_C P_{C_{uv}} = Z_D \mathbf{K}_C f_{C_{dist}} \mathbf{T}_{CW} \mathbf{T}_{LW}^{-1} f_{L_{undist}} \mathbf{K}_L^{-1} P_{D_{uv}} \quad (11)$$

4. Experiments

4.1. Calibration Experiment

Using the above calibration scheme, carry out internal and external parameter calibration for the following module cameras. The camera module is a Deeya series camera produced by Orbi Zhongguang Technology Co., Ltd. The parameter settings of this model are shown in Table 1.

Table 1: Camera module parameters.

Sensor information	Target size	Pixel size	Resolution	Shutter mode
OV9282	1/4 inch	3.0 μ s \times 3.0 μ m	1280 \times 800	Global Shutter

Since the experimental module uses structured light technology to obtain the depth information of the scene, the laser speckle it actively emits will cause the module to overheat and cause local depth information to be lost. To ensure the stability of the module during the long-term video acquisition process, after installing a heat sink behind the module, the problem of long-term heating can be solved. A photo of the hardware device of the camera module is shown in Figure 5.



Figure 5: The camera module used in this research institute.

The module is embedded with an RGB camera, a 960 nm band infrared camera, and a laser transmitter. We calibrated the internal and external parameters of the RGB camera and the infrared camera, and the calibration results are shown in Table 2 and Table 3 below. We first specify the internal parameters of the depth camera and color camera as follows:

(f_u, f_v) is the focal length of the camera in the two image coordinate systems, in pixels, and the approximate conversion relationship with the length focal length (in mm) in the optics is:

Length focal length f (pixels) = (mm)/pixel size (mm/pixels), where the length focal length and pixel size are generally provided by the camera supplier;

Table 2: Internal parameters of the depth camera.

f_u^{ir}	f_v^{ir}	γ	c_x^{ir}	c_y^{ir}
905.832	951.283	0	646.825	384.165

Table 3: Internal parameters of the color camera.

f_u^{ir}	f_v^{ir}	γ	c_x^{ir}	c_y^{ir}
905.832	905.832	0	654.887	480.571

Table 4: Depth camera to color camera calibration extrinsic parameters rotation matrix.

Rotation matrix R				
R_x	0.999995	-0.00240419	0.00213467	0.999995
R_y	0.00239793	0.999993	0.00293193	0.00239793
R_z	-0.0021417	-0.0029268	0.999993	-0.0021417

(c_x, c_y) is the point on the main optical axis imaging and the camera sensor. Generally, the error of the camera is not large in the center of the sensor, or half of the image resolution;

Depth camera to color camera calibration external parameter rotation matrix and displacement matrix are shown in Table 4 and Table 5, where R is the rotation matrix; based on the Euler angle conversion formula, under the xyz rotation order, $R=R_z(\theta)R_y(\theta)R_x(\theta)$.

The distortion parameters of depth camera and color camera are shown in Table 6 and Table 7, where $(k_1, k_2, k_3, p_1, p_2)$ is the distortion coefficient, describing the Brown distortion model of the camera.

Table 5: Depth camera to color camera calibration extrinsic parameter displacement matrix.

t	
t_x	-9.84032
t_y	0.00115723
t_z	-0.568239

Table 6: Depth camera distortion parameters.

k_1	k_2	k_3	p_1	p_2
0	0	0	0	0

Table 7: Color camera distortion parameters.

k_1	k_2	k_3	p_1	p_2
0.0595024	-0.0776618	6.60556e-5	-0.000618483	0

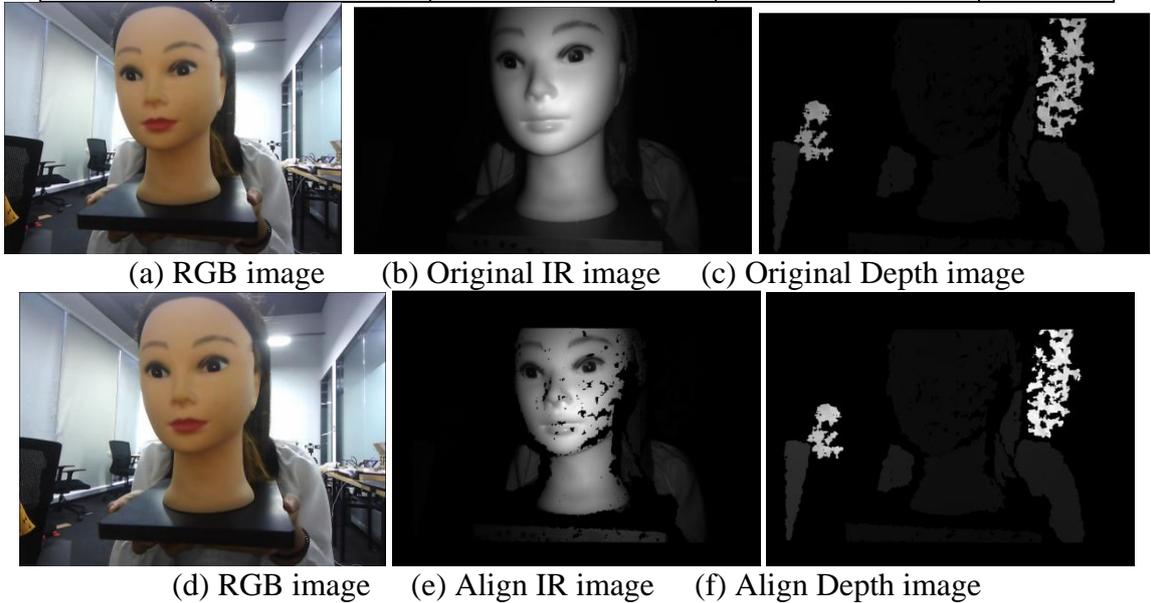


Figure 6: Schematic illustration of the alignment of three-way multimodal data before and after camera calibration.

Figure 6 shows the effect comparison before and after the depth-to-color alignment of the collected RGB, IR and depth fake face images. Among them, the first row is an image in the original acquisition state. The resolution of the RGB image is 640×480 , and the original resolution of the IR and depth images is 640×400 ; after depth alignment, the resolution of the IR and depth

images is consistent with that of the RGB image, which is 640×400 . It can be seen from Figure 6 (e) that the aligned IR image has missing gray values of some pixels. This is because the target surface may have specular reflection, resulting in a lack of depth, which causes the corresponding loss of IR gray.

5. Conclusions

In this paper, we built a video acquisition platform based on multimodal information flow to simultaneously acquire visible light, near-infrared and depth video image frames. To maximize the target public information based on multimodal data, the camera calibration work is carried out on the video acquisition platform to obtain matching multimodal images, which effectively collects multimodal align face images, and further provides strong support for the application of multimodal face anti-spoofing technology.

Acknowledgements

This paper was supported by the Henan Provincial Science and Technology Research Project under Grants 222102210010, 222102210039, 222102210103 and 222102210064, the Research and Practice Project of Higher Education Teaching Reform in Henan Province under Grants 2019SJGLX320 and 2019SJGLX020, the Undergraduate Universities Smart Teaching Special Research Project of Henan Province under Grant Jiao Gao [2021] No. 489-29, and the Academic Degrees & Graduate Education Reform Project of Henan Province under Grant 2021SJGLX115Y.

References

- [1] Li H., Li W., Cao H., Wang S., Huang F. and Kot A. C. (2018). *Unsupervised domain adaptation for face anti-spoofing*. *IEEE Transactions on Information Forensics and Security*, 13(7), 1794-1809.
- [2] Chen B., Yang W., & Wang S. (2021). *Generalized face antispoofing by learning to fuse features from high-and low-frequency domains*. *IEEE MultiMedia*, 28(1), 56-64.
- [3] Cai R., Li H., Wang S., Chen C. and Kot A. C. (2020). *DRL-FAS: A novel framework based on deep reinforcement learning for face anti-spoofing*. *IEEE Transactions on Information Forensics and Security*, 16, 937-951.
- [4] Rehman Y. A. U., Po L. M., Liu M., Zou Z., Ou W. and Zhao Y. (2019). *Face liveness detection using convolutional-features fusion of real and deep network generated face images*. *Journal of Visual Communication and Image Representation*, 59, 574-582.
- [5] Yang Q., Zhu X., Fwu J. K., Ye Y., You G. and Zhu Y. (2020). *PipeNet: Selective modal pipeline of fusion network for multi-modal face anti-spoofing*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 644-645).
- [6] Alotaibi S. and Smith W. A. (2019, September). *Decomposing multispectral face images into diffuse and specular shading and biophysical parameters*. In *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 3138-3142). IEEE.
- [7] Liu A., Li X., Wan J., Liang Y., Escalera S., Escalante H. J., Madadi M., Jin Y., Wu Z., Yu X., Tan Z., Yuan Q., Yang R., Zhou B., Guo G. and Li S. Z. (2021). *Cross-ethnicity face anti-spoofing recognition challenge: A review*. *IET Biometrics*, 10(1), 24-43.
- [8] Zhang Z., Yi D., Lei Z. and Li S. Z. (2011, March). *Face liveness detection by learning multispectral reflectance distributions*. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 436-441). IEEE.
- [9] Chen H., Chen Y., Tian X. and Jiang R. (2019). *A cascade face spoofing detector based on face anti-spoofing R-CNN and improved Retinex LBP*. *IEEE Access*, 7, 170116-170133.
- [10] Wild P., Radu P., Chen L. and Ferryman J. (2016). *Robust multimodal face and fingerprint fusion in the presence of spoofing attacks*. *Pattern Recognition*, 50, 17-25.
- [11] Arashloo S. R., Kittler J. and Christmas W. (2015). *Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features*. *IEEE Transactions on Information Forensics and Security*, 10(11), 2396-2407.

- [12] Kim W., Suh S. and Han J. J. (2015). Face liveness detection from a single image via diffusion speed model. *IEEE transactions on Image processing*, 24(8), 2456-2465.
- [13] Zhang J., Di L. and Liang J. (2021). Face alignment based on fusion subspace and 3D fitting. *IET Image Processing*, 15(1), 16-27.
- [14] Saha S., Xu W., Kanakis M., Georgoulis S., Chen Y., Paudel D. P. and Van Gool L. (2020). Domain agnostic feature learning for image and video based face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 802-803).
- [15] Zhang S., Liu A., Wan J., Liang Y., Guo G., Escalera S., Escalante H. J. and Li S. Z. (2020). Casia-surf: A large-scale multi-modal benchmark for face anti-spoofing. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2), 182-193.
- [16] Liu A., Tan Z., Wan J., Escalera S., Guo G. and Li S. Z. (2021). Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 1179-1187).
- [17] Boulkenafet Z., Komulainen J., Li L., Feng X. and Hadid A. (2017, May). OULU-NPU: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)* (pp. 612-618). IEEE.
- [18] Liu Y., Stehouwer J., Jourabloo A. and Liu X. (2019). Deep tree learning for zero-shot face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4680-4689).
- [19] Wanner S. and Goldluecke B. (2012, June). Globally consistent depth labeling of 4D light fields. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 41-48). IEEE.
- [20] Li B., Karpinsky N. and Zhang S. (2014). Novel calibration method for structured-light system with an out-of-focus projector. *Applied optics*, 53(16), 3415-3426.
- [21] Zhang Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11), 1330-1334.