

PAPER • OPEN ACCESS

Acquisition and Preprocessing of Face Recognition Data

To cite this article: Zihao Zhang 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **252** 042089

View the [article online](#) for updates and enhancements.

Acquisition and Preprocessing of Face Recognition Data

Zihao Zhang

People's Public Security University of China

615791251@qq.com

Abstract. Face has always been the research object in the fields of biometric recognition, computer vision, computer graphics and animation generation. In face recognition research, the acquisition of three-dimensional face data is the first step, and the quality of the acquired data directly affects the follow-up research. With the further development of computer vision technology, there are some tools and new methods for data acquisition.

1. Face recognition data acquisition

Data acquisition technology can be divided into two categories: Active Range Sensors and Passive Range Sensors, that is, data acquisition based on physical devices and data acquisition based on multiple images [1].

Active Ranging Sensing is that the vision system first emits energy to the scene, and then receives the energy emitted by the scene to calculate the location information of each point in the scene. Passive ranging sensor is a visual system that receives the light energy emitted or reflected from the scene, forms the light energy distribution function (i.e. gray image) of the scene, and then restores the depth information of the scene on the basis of these images. The most common method is to use two cameras at a certain distance to obtain scene images simultaneously to generate depth maps. Another method similar to this method is that a camera obtains two or more images in different spatial positions, and generates depth maps by gray information and imaging geometry of multiple images. Depth information can also be indirectly estimated by using the gray and dark features, texture features and motion features of grayscale images.

1.1. Data acquisition based on physical device structured light ranging

(1) Structured light

Structured Light is a group of known geometrical modes of irradiated light. Structured light ranging is one of the most widely used and promising classical active vision methods. It uses the principle of triangulation to calculate depth. It consists of a projection light source and a camera. As shown in Figure 1, the distance between the projection light source and the camera is a baseline distance b , f is the focal length. The projector emits a plane light which is perpendicular to the plane XZ and the angle between the planes XY is called θ . The plane of light intersects the scene surface into a plane curve, called Stripe. In such a system, the relationship between a point coordinate $P(X, Y, Z)$ of an object and the projected image coordinate $P(x, y)$ and the projection angle θ can be given by formula (1):



$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \frac{b}{f \cos \theta - x} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (1)$$

In this way, all the scene points on the stripes can get the depth information according to this formula. The range resolution of a triangulation system can be determined by the measurement accuracy of the projection angle theta and the horizontal position X of the image point.

In the 1980s, Posdamer et al. [2] and Boyer et al. [3] proposed this method and achieved good results. Beumier et al. [4]-[5] further developed this idea, established a fast acquisition system based on structured light, and used this system to collect a three-dimensional face database containing 120 people. This database takes into account the attitude (left-right rotation and up-down rotation), slight facial expressions (such as smiles) and affiliations of the subjects under test. The influence of products (such as glasses) is the largest capacity 3D face database published. Based on this database, they developed a three-dimensional face recognition method based on face matching and contour matching.

The data acquisition system based on structured light is fast, and the whole three-dimensional information of the object can be obtained by taking a single image of the camera, so the movement of the measured object does not affect the measurement accuracy. The illumination of the projector overcomes the influence of ambient light, so the influence of light intensity is not obvious. At the same time, the system can obtain three-dimensional data and texture information respectively by opening and closing the projector. The disadvantage of this system is large volume, and the accuracy and scope of acquiring 3D data are limited.

(2) Laser scanning

Based on the same or similar principle, visible structured light can also be replaced by laser, called Laser Scanner. Laser scanning is an accurate and effective method to obtain distance data, which can simultaneously obtain a large number of three-dimensional data and texture information. At present, many international research institutes have developed a commercial three-dimensional data acquisition device based on laser scanning. The three-dimensional laser scanner is widely used in industry and commerce. The most famous one is Cyberware's three-dimensional scanner and Metris, Minolta-Europe's three-dimensional laser scanner. Instrument products are based on triangulation principle. According to the different measuring range, the product can be divided into different types, such as short distance, medium distance and long distance. The mid-distance scanner is more suitable for the acquisition of head data. The data acquired by laser scanner is of high accuracy and high speed, but it requires high performance computers and expensive hardware devices.

Lee of University of Toronto [6], uses this device to get complete 3D facial information. Waters et al. [7] of Cambridge graduate students used the three-dimensional data acquired by the system to build a three-dimensional head model. Blanz and Vetter, German scholars, established a head database including three-dimensional distance information and surface texture information.

1.2. Based on data of multiple images [9].

The essence of active vision-based three-dimensional data acquisition method is to reduce the processing difficulty of stereo vision by means of artificial light source. In addition, researchers have been working on another challenging issue, that is, how to use two or more images captured by ordinary cameras to obtain stereo information in natural environment. The hardware equipment needed by this idea is relatively common, but the computer vision algorithm is required to have high robustness and practicability. According to the different acquisition devices, these algorithms can be divided into two categories: single camera-based method and multi-camera-based method.

Stereo vision refers to the process of inferring the three-dimensional structure information of a scene from two or more images of different perspectives. Its basic working principle is based on the three-dimensional physiological perception mechanism of human eyes. When people's two eyes are observing things, because they are in different angles, the images they see are also different. By using this difference information, the distance between objects and eyes can be obtained, that is, the three-

dimensional structure information of objects. The basic binocular geometry relationship is shown in Figure 1.

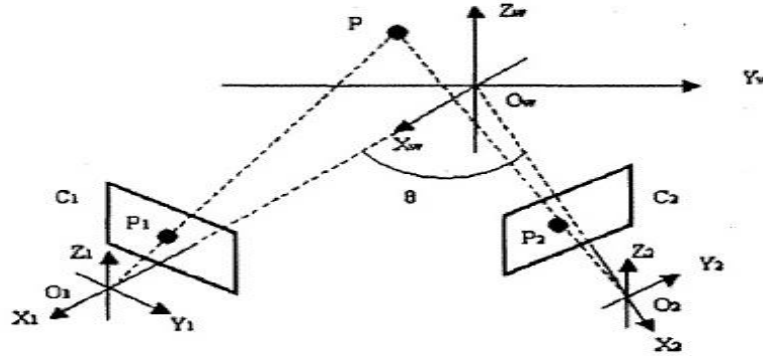


Figure 1. Binocular stereo relationship.

We suppose that make the optical axis of the camera C1 coincide with the X0 axis of the object coordinate system, and the Z1 axis in the coordinate system of the camera C1 and the Z0 axis in the object coordinate system are parallel to the camera C2, which is obtained by the rotation angle of the camera C1 around the Z0 axis in the object coordinate system. Let the coordinates of point P in the object coordinate system be (X, Y, Z), the distance between the optical center O1 and Ow of the camera C1 is D, the parameters of the camera C1 and 22 are the same and the focal length is L, and the coordinates of the image points P1 and P2 in the respective camera image plane are (u1, v1) and (u2, v2), respectively. According to the linear camera model (i.e. pinhole model), we can get the following results: In order to obtain:

$$\left\{ \begin{array}{l} u_1 = \frac{L}{D-X} Y \\ v_1 = \frac{L}{D-X} Z \\ u_2 = \frac{L}{D-(X \cos \theta + Y \sin \theta)} (Y \cos \theta - X \sin \theta) \\ v_2 = \frac{L}{D-(X \cos \theta + Y \sin \theta)} Z \end{array} \right. \quad (2)$$

When the shooting distance D, focal length L and rotation angle theta of the camera are known, the three-dimensional coordinates of the space point P can be obtained by the equations in equation (2). These three parameters can be calibrated by stereo vision camera.

2. Preprocessing of face recognition data

This paper summarizes the pretreatment method of Artec Spider 3D scanner [10] acquisition data. The original data collected by it is a set of curved surface frames, which contains 400-1500 curved surface frames according to the acquisition time and frame rate. A series of model processing is needed before synthesizing these data into a complete 3D face image. It mainly includes: whole matching, fairing synthesis, hole filling, small object filtering, mesh simplification and texture mapping. However, due to the slight change of illumination conditions, the different smoothness of face skin, or the mutual occlusion of surface structures, the results of automatic processing of some models are not ideal. As shown in Figure 2, automatic registration results in obvious cracks. This will not only affect the visual effects, but also interfere with the late application. This requires manual intervention to achieve ideal registration results.

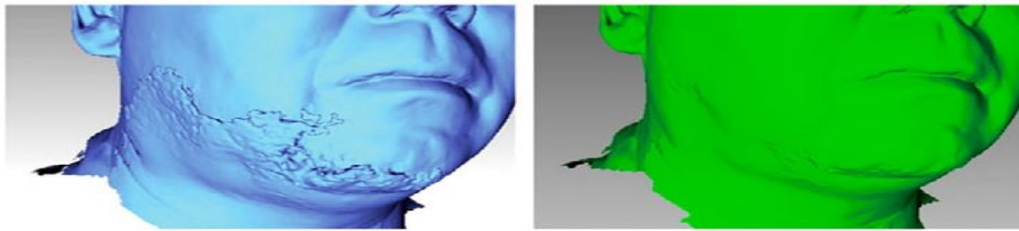


Figure 2. Left: automatic pretreatment for cracks; right: after artificial intervention to eliminate cracks.

Manual intervention registration mainly consists of three steps:

The first step is to find out the undesirable frame or frame set for automatic registration, and remove the selected frame into a new frame set. Some computers have large deviations in automatic registration of scanned images, so it is necessary to screen the deviated surfaces manually to generate a new set of frames.

In the second step, feature alignment is performed between the separated frame sets, i.e. marking the same parts (same name) with more obvious features on the two parts of the frame sets that need to be combined, such as the corners of the eyes, the corners of the mouth, the auricles, the tip of the nose, the alae of the nose and so on. If the segregated images do not share the same obvious features, they can also be aligned and coincided with the reference frame set by manual interaction. The coincidence accuracy mainly depends on the human eye.

The third step is automatic stitching mesh. If "feature alignment" is understood as "rough mosaic" of separated images, then "mosaic image" is "fine mosaic" of separated images. On the basis of automatic mesh splicing, "image texture splicing" is allowed again, so that the artificially fitted three-dimensional image can be obtained, and subsequently the whole registration and smoothing process can be repeated to achieve better processing effect.

References

- [1] Tian Wenjun. 3D face feature extraction based on depth image [D]. Beijing Jiaotong University, 2009.
- [2] Posdamer J L, Altschuler M D. Surface measurement by space-encoded projected beam systems [J]. *Computer Graphics & Image Processing*, 1982, 18 (1): 1-17.
- [3] Boyer K L, Kak A C. Color-encoded structured light for rapid active ranging. [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1987, 9 (1): 14-28.
- [4] Beumier C, Acheroy M. 3D facial surface acquisition by structured light [A]. In *Proceeding International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging* [C], Santorini, Greece, 1999.103-106.
- [5] Beumier C, Acheroy M. Automatic 3D face authentication [J]. *Image and Vision Computing*, 2000, 18 (4): 315-321.
- [6] Lee S Y, Chwa K Y, Shin S Y. Image metamorphosis using snakes and free-form deformations [C] // *Conference on Computer Graphics and Interactive Techniques*. 1995: 439 - 448.
- [7] Waters K, Terzopoulos D. Modelling and animating faces using scanned data [J]. *Computer Animation and Virtual Worlds*, 1991, 2 (4): 123-128.
- [8] Blanz, Vetter V, Thomas. A Morphable Model for the Synthesis of 3D Faces [J]. *Computer Graphics Proceedings Siggraph*, 1999: 187-194.
- [9] Xu Chenghua, Wang Yunhong, Tan tie Niu. 3D face modeling and application [J]. *Chinese Journal of image and graphics*, 2004, 9 (8): 893-903.
- [10] Wang Junjuan, Xu Lei, Li Zhi Hui, et al. Data acquisition and preprocessing of 3D face images [J]. *criminal technology*, 2015 (2): 98-101.