中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

A Food Package Recognition and Sorting System Based on Structured Light and Deep Learning

Xuanzhi Liu, Jixin Liang, Yuping Ye, Zhan Song and Juan Zhao

Shenzhen Institutes of Advanced Technology (SIAT)

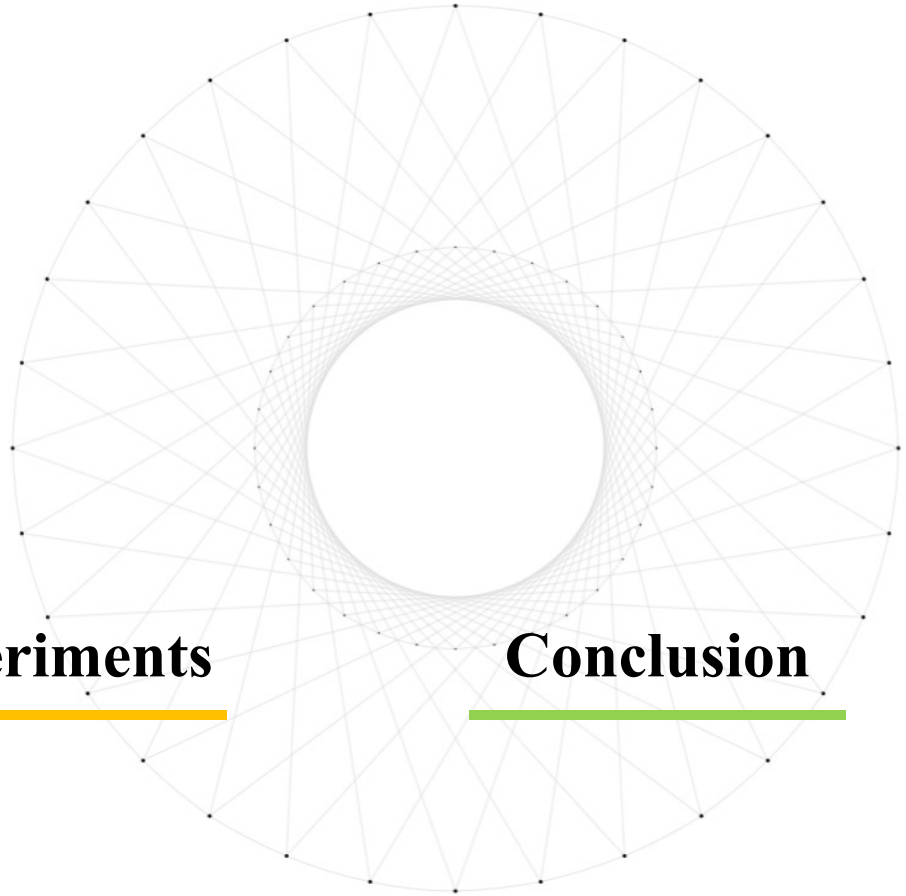
Chinese Academy of Sciences (CAS)

JCRAI 2023

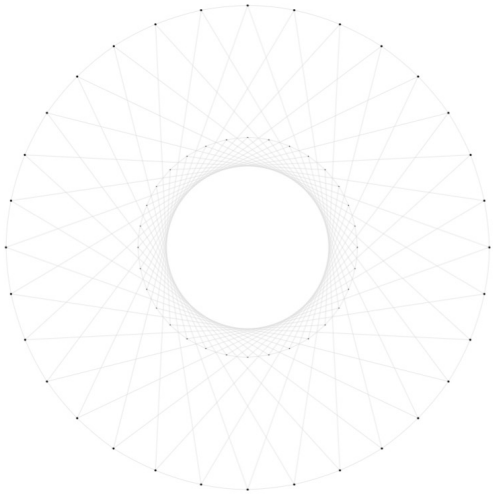
Shanghai, China

July. 22, 2023

Outline



Background **Methodology** **Experiments** **Conclusion**



Background

- In industry, some objects, like food packages, have **reflective or light-transmitting** properties on their surfaces, and will be affected by changes in external environment light, creating challenges for automated production.

- **Deep learning** vision algorithms have been widely used in many area, like face recognition and automatic driving. However, the existing object detection algorithms can only obtain the 2D coordinate information **(X and Y)** of the target. For a robotic arm grasping task, the key **Z-axis** coordinate values are missing.

Figure 1. The samples we use in our experiments.

Process

We have designed an automated grasping system combining object detection and 3D reconstruction. The object detection algorithm is used to identify the class and location of the target in the image, and the 3D reconstruction can get the 3D coordinates corresponding to the 2D coordinates. The flow of the system is shown in Fig. 2.

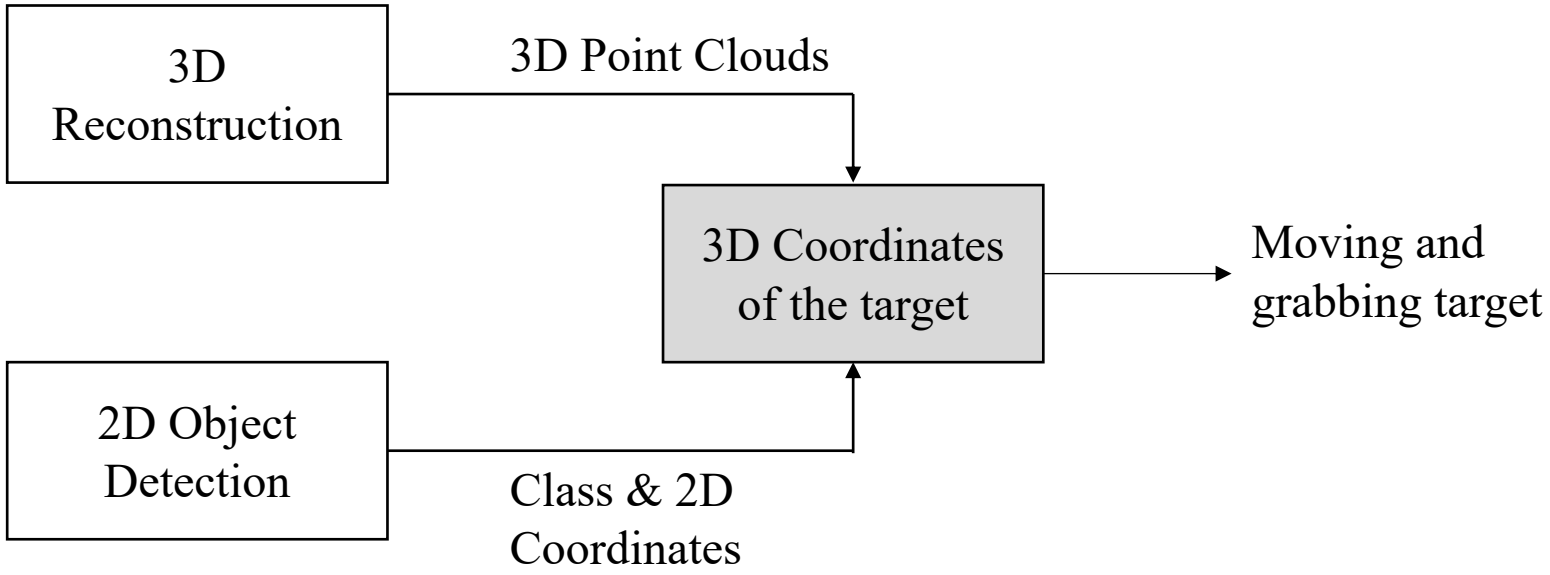


Figure 2. The complete process of our method.

Dynamic Structured Light System

A typical structured light system consists of a projector and a camera, the specific mathematical model representation is shown in Fig.3. From the camera imaging model, we can obtain:

$$s\begin{bmatrix} m^{c/p} \\ 1 \end{bmatrix} = \begin{bmatrix} f_u^{c/p} & \gamma^{c/p} & u_0^{c/p} \\ 0 & f_v^{c/p} & v_0^{c/p} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I_3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M^{c/p} \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} M^{c/p} \\ 1 \end{bmatrix} = \begin{bmatrix} R^{c/p} & T^{c/p} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} M^{w} \\ 1 \end{bmatrix}$$

By the triangulation principle we can get the depth information:

$$z_c = \frac{(R\tilde{m}_c\tilde{m}_p)(\tilde{m}_c T) - \|\tilde{m}_c\|^2 (R\tilde{m}_c T)}{\|R\tilde{m}_c\|^2 \|\tilde{m}_p\|^2 - (R\tilde{m}_c\tilde{m}_p)^2}$$

By these main equations and calibration we can achieve one-to-one correspondence from 2D pixel points to 3D spatial points. With the dynamic structured light acquisition device and the above algorithm, we can obtain the texture images, depth images and point clouds sequences from different views of the target scene.

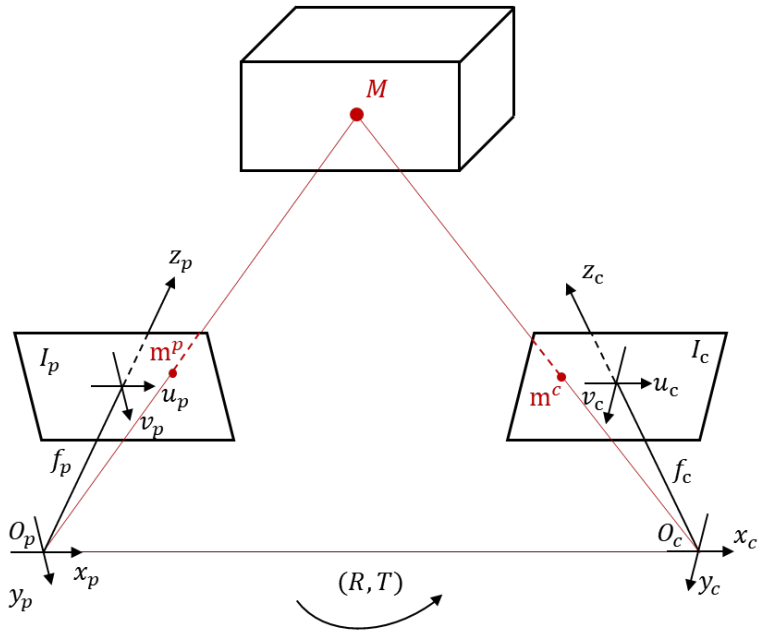


Figure 3. Mathematical model of structured light system.

MASK R-CNN Model

MASK R-CNN[16] is a convolution-based network model that is able to output the class and location information of targets in an image by performing a series of convolution calculations on the image. Fig. 4 shows the network structure of this model.

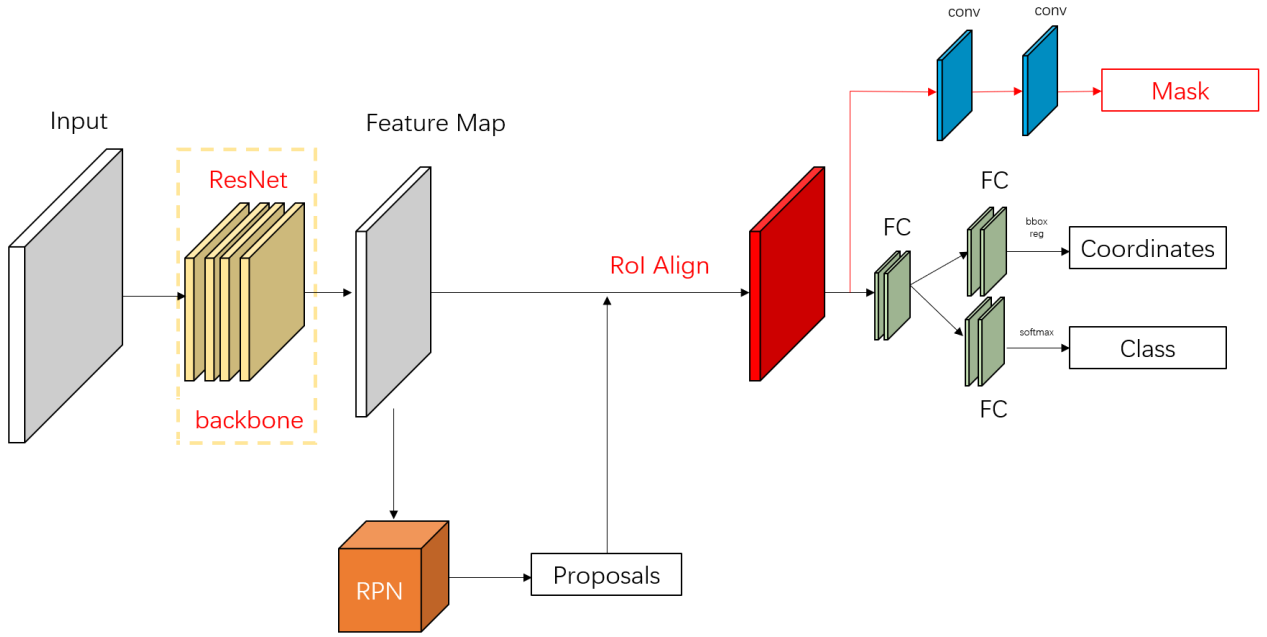


Figure 4. MASK R-CNN network structure.

[16] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

Experiments

We use a depth camera based on structured light technology as a 3D sensor. With a single scan, a 3D point cloud map, depth map and texture map of the image within the field of view can be obtained directly. Here is some information about the parameters of the camera:

- Working distance: $650mm$

  Scanning speed: $6Hz$

  Working range (X,Y direction): $400mm * 300mm$

  Working range (Z-direction): $\pm 150mm$

  Resolution (X, Y direction): $3.2M$

  Repeatability accuracy (Z-direction): $5\mu m$

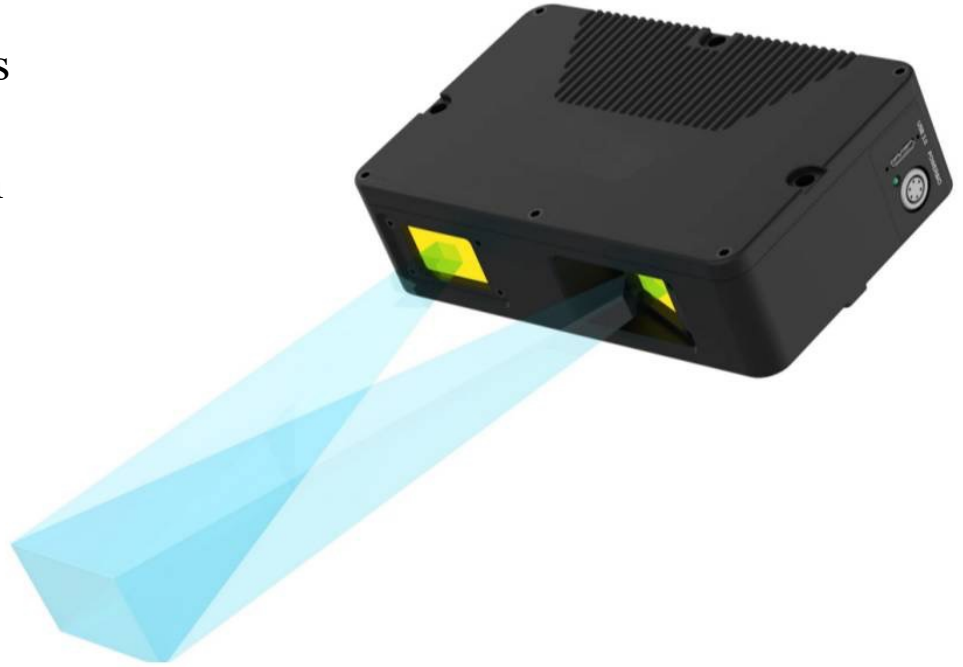
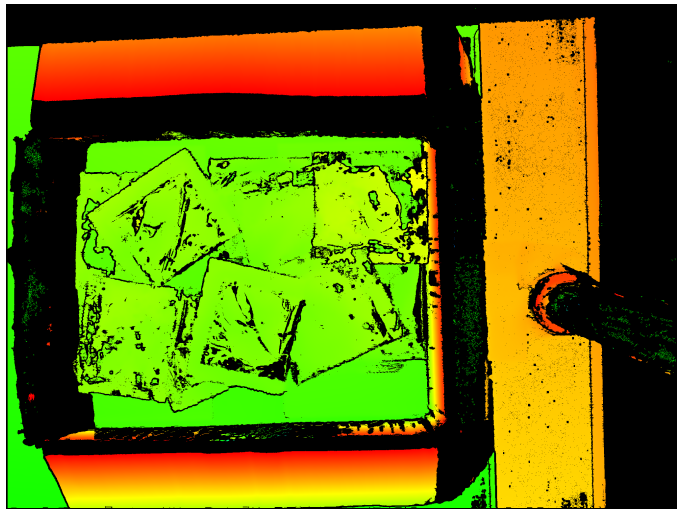
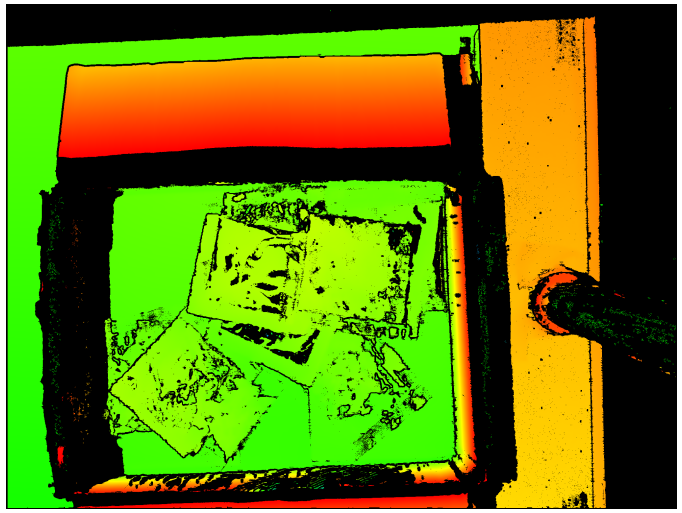


Figure 5. The structure light depth camera we used.

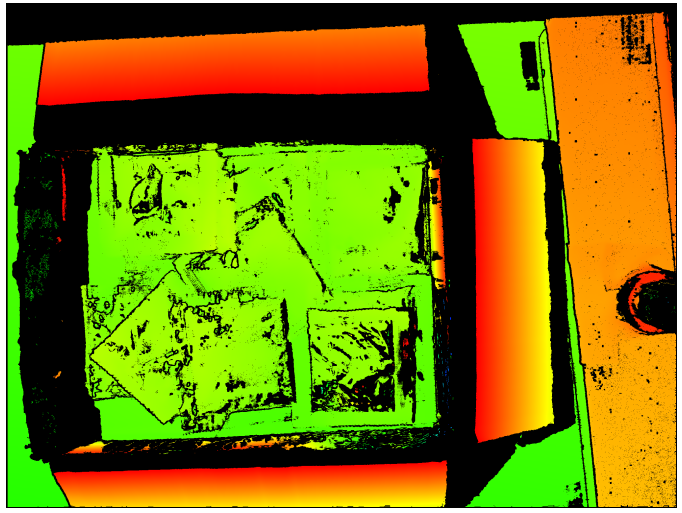
Experiments



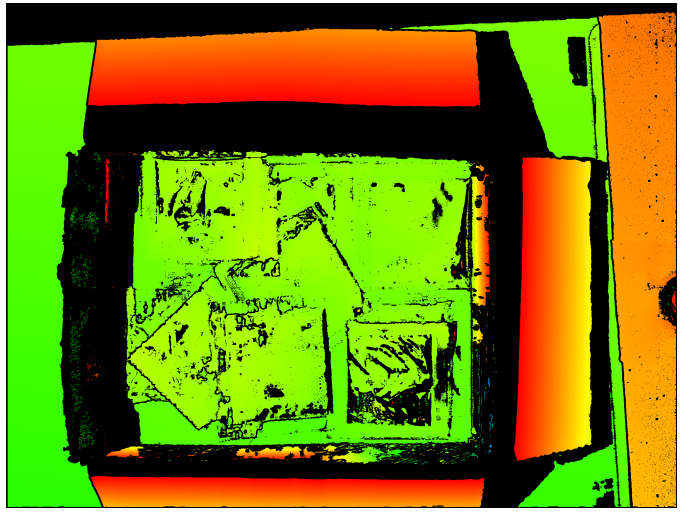
(a)



(b)



(c)



(d)

Figure 6. Depth map of the 3D reconstruction.

Experiments

- GPU: GeForce RTX 3090*8

- Training set: 210 images
  Validation set: 90 images
  Image size: 2044*1536
  Image type: single channel grayscale image (texture image)

- Hyper Parameters:
  backbone: ResNet101
  optimizer: SGD
  epoch: 100

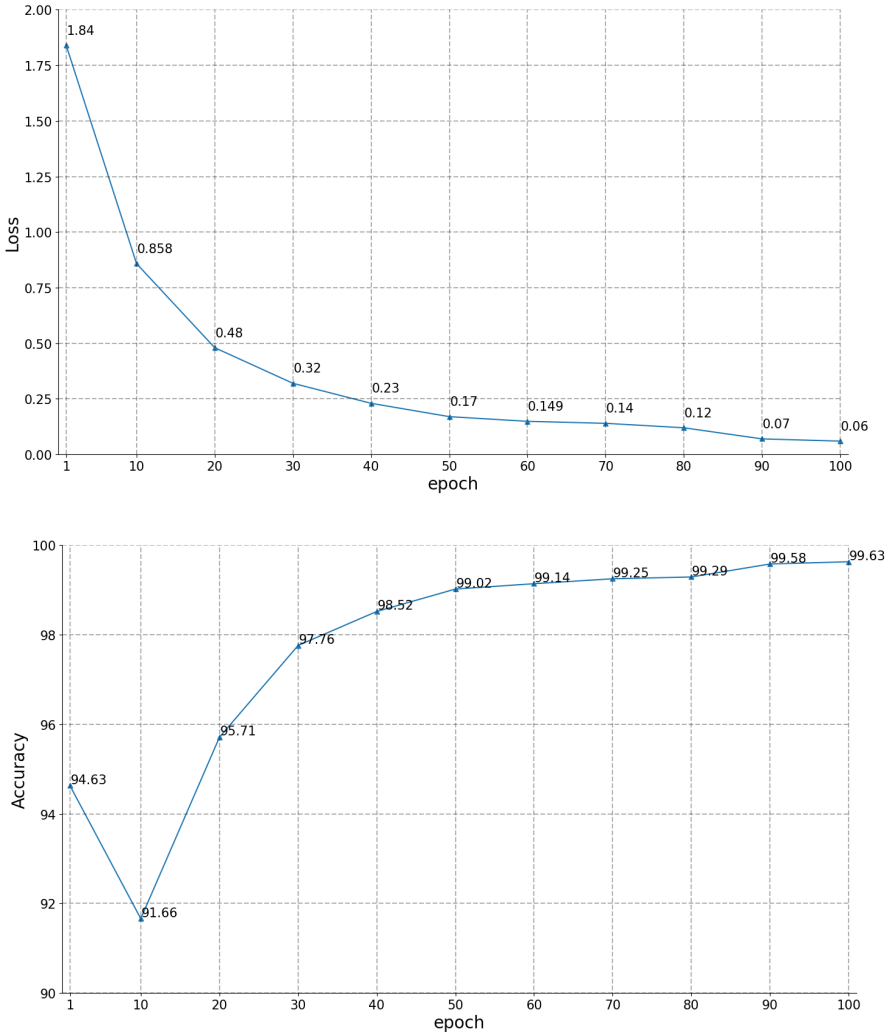


Figure 7. Parameter changes during model training.

Experiments

(a) (b) (c)

Figure 8. Recognition results of the model on the texture maps.

Conclusion

- **Methodology:**
  - The R-CNN algorithm is deployed to detect the target and get the class and 2D coordinate information.
  - The 3D reconstruction technique using structured light reconstructs the 3D point cloud of the target and obtains the 3D coordinates corresponding to each 2D point.

Highlight：

- Using deep learning algorithm instead of traditional vision algorithm for transparent and reflective material object recognition, the recognition accuracy is improved.
- Combine 2D vision algorithms with 3D point cloud reconstruction to compensate for each other's shortcomings

Future work:

- We will subsequently try to install polarized lenses outside the camera, using the properties of polarized light to eliminate specular reflection in order to improve the quality of the image.

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

Thank You