INFOMCV Assignment 2
Ana San Román Gaitero and José Luis (Pepe) Cano-Manuel Claver (Group 28)

Summary

The calibration of the video is assessed so that the same corner of the chessboard is visualized in all the images. Also, it is considered that the z axis is flipped and changed with the y axis, and that the camera rotation matrices are different when using glm.
The background subtraction method chosen is Gaussian mixture distribution, although an average method is also provided. Finding the contours of the shape and applying closing (dilation+erosion) is used to optimize the procedure.
To create the voxel model, we have considered that each square in the chessboard is of 115mm, to get a 3D space that ensures containing the subject but not uses the whole volume of the room. Then, a look-up table is creating from 3D space to 2D, and if the pixel values are of 255 on the four cameras, it is activated in the 3D space.

Extrinsic parameters

The translation matrices are in mm and the rotation matrices in radians.
Camera 1:

$$R_1 = \begin{bmatrix} -0.69 & -0.72 & 0.0051 \\ 0.10 & -0.094 & 0.99 \\ -0.71 & 0.69 & 0.14 \end{bmatrix} t_1 = \begin{bmatrix} 813.16 \\ 877.35 \\ 4013.36 \end{bmatrix}$$

Camera 2:

$$R_2 = \begin{bmatrix} -3.92 \cdot 10^{-2} & -9.99 \cdot 10^{-1} & 7.72 \cdot 4 \\ -1.46 \cdot 10^{-2} & 1.34 \cdot 10^{-3} & -9.99 \cdot 10^{-1} \\ -9.99 \cdot 10^{-1} & -3.91 \cdot 10^{-2} & -1.46 \cdot 10^{-2} \end{bmatrix} t_2 = \begin{bmatrix} 553.48 \\ 1484.15 \\ 3595.48 \end{bmatrix}$$

Camera 3:

$$R_3 = \begin{bmatrix} 0.97 & 0.23 & 0.021 \\ -0.033 & 0.052 & 1.00 \\ 0.23 & -0.97 & 0.058 \end{bmatrix} t_3 = \begin{bmatrix} -906.16 \\ 1200.88 \\ 3222.93 \end{bmatrix}$$

Camera 4:

$$R_4 = \begin{bmatrix} 0.78 & -0.63 & -0.011 \\ 0.079 & 0.080 & 0.99 \\ -0.62 & -0.77 & 0.11 \end{bmatrix} t_4 = \begin{bmatrix} -455.85 \\ 846.37 \\ 4631.30 \end{bmatrix}$$

They are used, along with the intrinsic parameters (distortion and K), to locate the voxels and the cameras.
To locate the voxels, a 3D space of 60x60x60 (which correspond to 6.9x6.9x6.9 m) is utilized, to make sure the body of the subject is captured, but not taking the entire room size. Also, the subject is located considering that each square of the board is of 115mm in size, and that it looks like it is very sitting on top of it, also using the translation vector.
The x and y axis are centered with respect to half the width and depth of the voxel grid, and the height starts from 0 until the total height. Also, the voxels' locations are multiplied by a factor of 40 to prevent all pixels to locate in the same central voxel, that is why then x and y is translated 40 voxels to locate them in the center again. With each position on the voxel space, its corresponding position on each image is captured using the intrinsic and extrinsic parameters (look-up table), using projectPoints. Then, depending on whether that pixel is white (1) or black (0) in the 2D masks of the four cameras, the 3D pixel is activated or deactivated respectively.

Background subtraction

In relation to the background subtraction method, we made used of the gaussian background subtraction model. First, a background subtraction object model is created using the function createBackgroundSubtractorM0G2, which has a gaussian nature. It is trained with each one of the frames with only background.

Then, the model is applied to the frames of the videos containing the subject. To select an automatic threshold otsu-multithresholding is utilized, since can be appreciate that the peak in the histogram to which most of the pixels of the subject is the rightest one (Figure 2). Also, a findContours algorithm is used to improve the delimitation of the subject shape, it is a fully automated process. Alternatively, a background subtraction model based on the average of all the frames is create. In this one, a closing (dilation+erosion) is applied to remove holes, as an alternative to findContours.

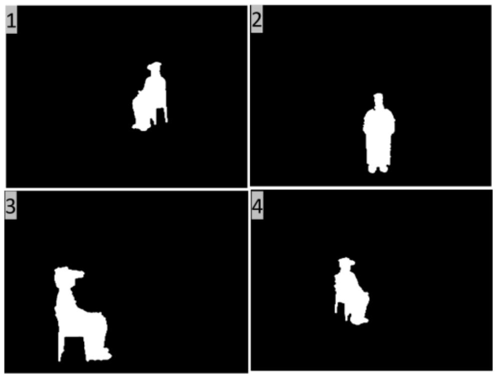

Figure 1: masks for cameras 1-4 on one frame after background substraction

Choice tasks

- *Choice 2: Automating the way the thresholds are determined.*

Looking at the model and its corresponding histogram (figure 2), can be appreciated that the peak corresponds to the subject. Therefore, by applying Otsu with multiple thresholds the separation between the middle and the highest peak is obtained, which separates the foreground from the background.

This process is optimized by finding the contours of the image, preventing other pixels with similar values (contained in the rightest peak) from been chosen as foreground too.

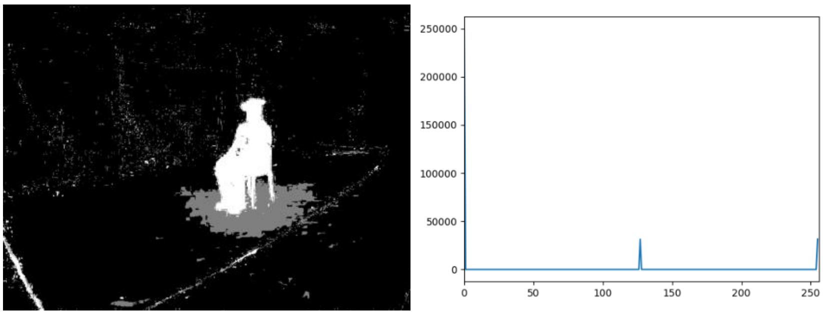

Figure 2: background removing model object and corresponding histogram.

- *Choice 3: coloring the voxel model*

As can be seen in the [link to the video](), by creating a new list with the containing colors of each pixel the 3D object of the subject is colored with its corresponding color in the 2D space, considering that only the voxels of the outer part that are projected by the camera should be colored.

- *Choice 4: Implementing the surface mesh*

To implement the surface mesh, the algorithm that showed the best result, and proved to maintain the original shape of the subject, is the one from *PolyData* package from *pyvista*. A cloud of points is calculated, after which the volume is extracted, selecting an alpha of 0.7, and extracting the geometry. The result is shown in figure 3.

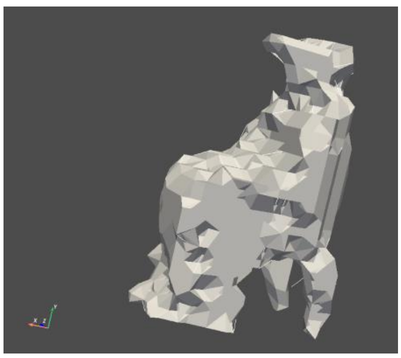

Figure 3: mesh of the subject.