3D RECONSTRUCTION USING EVERYDAY CAMERAS

Al Medinah, Areej
Nadeem, Syed Sameer

Electrical Engineering
Habib University

ABSTRACT

The paper presents a 3D reconstruction project, using stereo images captured with an everyday phone camera. The methodology used reciprocates the working of the human eye and the brain, where epipolar lines are used to create depth perception between two images, slightly rotated on the horizontal axis. Since the project is built on MATLAB 2016, some stereo vision tools from the computer vision toolbox of MATLAB 2019 were inaccessible which limited the project in creating a well-defined final version of the 3D mesh. However, with a few additional steps, the project can be built for use in e-commerce sites, interior design applications, etc.

***Index Terms*—** stereo images, 3D reconstruction, computer vision, digital image processing

1. INTRODUCTION

Mixed reality, an environment which integrates and allows the interaction of real and virtual objects is gaining popularity in multiple applications. Some of these include augmented reality applications, especially in the online shopping sector and gaming applications. For most applications of mixed reality, 3d models are created which is a specialized skill not acquired by all individuals. Another option is 3d scanning, which is currently enabled by complex hardware and software solutions. The motivation behind this project is to find a way to create 3d models of real scenarios through everyday devices for a user with no specialized knowledge of 3d modelling. [1]
The significance of this project is that it can lead to augmented reality integration of 3d models of humans in different applications. A few examples are listed below:

- A person can take a few images from different angles, create a 3d image of themselves, and try on different accessories before online shopping. Currently, this is only used widely for 2d images, similar to Snapchat filters which puts on accessories on 2d images.
- Interior designers can 3d scan spaces, and 3d scan decoration objects and furniture to customize the interior of a space and visualize it. Currently, this is only possible for objects which are pre-loaded as 3d models from their respective companies, such as IKEA. With easier 3d scanning using phone cameras, this will be possible for all objects and all spaces without pre-existing 3d models of the inventory.
- For the gift industry, 3d models of humans and families can also be 3d printed. Other objects can also be 3d printed. For example, 3d modelling of the jaw can allow 3d printing of replacement teeth by knowing the dimensions that fit in the jaw

Thus, the project focuses on using everyday phone camera to capture images of the subject from different angles, and create a 3d model through depth estimation that can be used for further applications. The scope of this project is limited to image capturing along the horizontal axis (rotationary axis may be included), and creating a MATLAB point cloud of the 3d object.

2. LITERATURE REVIEW

For 3D reconstruction, some existing solutions that are already in use are:

- Keeping the camera/capturing device stationary and rotating the object to be rendered in 3D.
- Using Kinect to map out the object and render a 3D model.
- Utilizing 2 DSLRs with comparatively complex equipment.
- Using LiDAR or infrared sensors for depth estimation.

Most of the current solutions utilize techniques that are not feasible for people to execute on their own, thus we explore 3D reconstruction from 2D images.

In the recent research developments, there are multiple approaches used for 3D reconstruction from 2D images. The first one is 3D Morphable Models (3DMM), which is a statistical model consisting of a Principle Component Analysis (PCA) model of shape, color, camera and a lighting model. This method requires a lot of training data, which can be inaccessible for daily use. Shape and texture coefficients describe the model. These coefficients are determined from 2D images. Features using scale-invariant feature transform (SIFT), which is a feature detection algorithm in computer vision to detect and describe local features in images is used to fit 3D Morphable Model. [2]

Another technique used is warping. In this, a reference shape comprising of the pre-defined generic model is warped by using model landmarks. This uses stereo images to detect 3D model features and deform the surface of the model using Thin-Plate Spline method. Thin plate spline method is a spline-based technique for data interpolation and smoothing. [3]

Another technique that is used in this project is identifying features using the minimum eigenvalue algorithm. The techniques mentioned are more accessible than using sensors, however 3D Morphable Model requires a lot of training data and warping produces generic models. Therefore in this project, we have attempted to tackle 3D face reconstruction based on stereo matching. The rest of the report includes the methodology and discusses the testing environment. This is followed by a description of the results and conclusion, including the way forward to improve accuracy.

3. METHODOLOGY

The aim for the design is to ensure minimal external hardware. In essence, a normal camera phone is used to capture two images, with a few degrees of horizontal rotation. These images are used to perform stereo reconstruction. Stereo reconstruction uses the same principle the brain and eyes use to actually understand depth.

The main principle of stereo reconstruction consists of looking at the same picture from two different angles, look for the same thing in both pictures and infer depth from the difference in position. This process is called stereo matching. This is demonstrated in Figure 1. Ideally, when there are two fully aligned cameras, it requires that the baseline is oriented horizontally (x-direction), the horizontal directions of both image planes are parallel to the baseline, and the two optical axes are parallel along z-direction and orthogonal to the baseline. Then, the depth of 3D point 'p' can be estimated according to similar triangles' principle.

The disparity value, which is spatial shift of the point in the two images, increases when the distance between the point 'p' and the camera decreases. To estimate the disparity, the process includes finding pixels in both images that have a correspondence with each other. This process is called stereo matching. Once the disparity value and the camera parameters ('b' and 'f') are known, the depth can be estimated.

With the depth, we have the location of points in the scene in all three dimensions, which can be plotted to visualize the 3D scene.

For stereo matching, it is essential that both pictures have the exact same characteristics, which means that they should not have any distortion. However, the lens in most cameras causes distortion. Thus, for accurate stereo matching, the optical centers and focal length of the camera need to be known. For the purpose of this project, default camera parameters from the MATLAB toolbox are used. However, to minimize error and improve accuracy, camera calibration can be done by using a bunch of images to infer the focal length and optical centers of the camera.

These parameters are used to undistort images, that is to get rid of lens distortion in the pictures used for reconstruction. Next, during feature matching, similar features between both pictures are identified using the minimum eigenvalue algorithm. This algorithm, developed by Shi and Tomasi is used for corner detection through eigenvalue calculation.

Then, a depth map is built for these features. Using the depth map, pixels are re-projected into the 3D space. Finally, a point cloud is built that contains points in 3D space for visualization.

A few constraints were set in the working environment to test this project which are discussed.

3.1. Constraints in the Testing Phase

- Object placement can be in any arbitrary solid white background. (The rationale is to not restrict the use to green background which is used for professional photography, but also allow for acceptable solid background in the object's picture. White background is easy to access for most users in everyday environments, such as by using A4 size sheets or white walls.)

- Object capturing through multiple images along the horizontal axes through a phone camera. Since some percentage of vertical distortion is inevitable, the test

.3

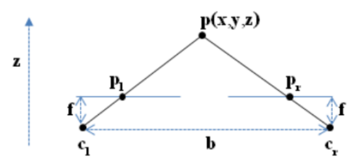


Fig. 1. Geometric model of calibrated stereo vision system. f:focal length, b:baseline, p(x; y; z):3D point, pl; pr:pixel values of projected point on left and right image respectively.[5]

image will be captured with a minimal vertical distortion error.

- MATLAB 2016 stereo vision toolbox will be used for this project.

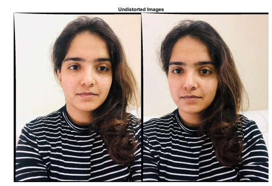
4. RESULTS

The results of a sample stereo image pair are presented in Figure 2. It can be observed that the original images in Figure 2(a) have a horizontal rotational difference, however the vertical rotation is minimally present to replicate human error in practical situations. The background in the original images is mostly white, which is easily accessible in daily life.
In Figure 2(b), radial distortion in images due to lens is removed and lines are straightened. The step brings some uniformity to the captured image.
A hundred and fifty strongest corners are identified in the first image, as seen in Figure 2(c). This is done with the minimum eigenvalue algorithm. These features are then tracked in the second image, as seen in the overlapping images in Figure 2(d).
Using internal camera parameters that highlight epipolar characteristics and the tracked features, the main features are highlighted in Figure 2(e). Finally, using these features, a 3D point cloud is generated in Figure 2(f). The camera positions are also shown through red and blue camera icons. Finally, a scaled reconstruction is demonstrated in Figure 2(g).

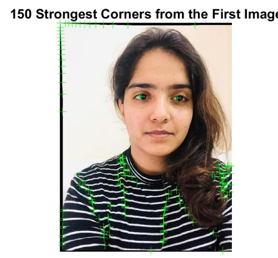
5. CONCLUSION

The methodology presented simplifies 3D reconstruction using stereo images for every day applications. Additionally, the project can be developed to include 3D mesh from the point cloud. The default camera parameters can be replaced my customized camera parameters by calibrating them. Calibration is usually done by capturing multiple stereo images of a check-board and identifying the corners in each image, to create multiple epipolar lines. With a customized calibration, the results would be more accurate.

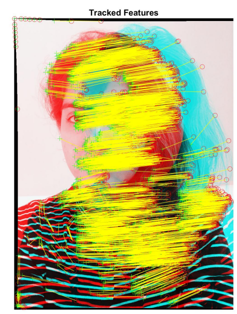
(a) Original Stereo Images



(b) Lens Distortion Removal

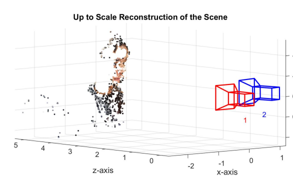


(c) Strongest Corners on 1st Image

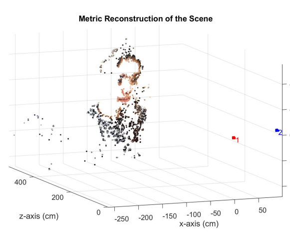


(d) Tracking Features on Both Images

(e) Highlighted Features



(f) Reconstructed 3D point cloud



(g) Scaled Reconstruction

Fig. 2. Results with an iPhone 8plus to capture images

6. REFERENCES

[1] H. Kim, K. Sohn. 3D reconstruction from stereo images for interactions between real and virtual objects. Signal Processing: Image Communication 20 (2005) 61–75.

[2] Huber, Patrik, et al. ”Fitting 3D morphable face models using local features.” Image Processing (ICIP), 2015 IEEE International Conference on. IEEE, 2015.

[3] Choi, Jongmoo, et al. ”3D face reconstruction using a single or multiple views.” Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[4] Y. Zhang, G. M. Gibson, R. Hay, R. W. Bowman, M. J. Padgett, and M. P. Edgar, “A fast 3D reconstruction system with a low-cost camera accessory,” Scientific Reports, vol. 5, no. 1, Sep. 2015.

[5] Hartley, Richard, and Andrew Zisserman. Multiple View Geometry in Computer Vision. Second Edition. Cambridge, 2000.