Zhengyou Zhang, "Camera Calibration: a Personal Retrospective", *Machine Vision and Applications*, Vol.27, No.7, pages 963–-965, 2016.

Camera Calibration: A Personal Retrospective

Zhengyou Zhang

Computer vision is a key branch of artificial intelligence, aiming at understanding the surrounding environment from the visual information captured by cameras. 3D computer vision is a subfield of computer vision that focuses on extracting 3D metric information from 2D images. This is a very challenging task since what we have in an image is the projection of a 3D scene, whose depth information is lost during image formation (image projection). However, it has many important applications such as the following:

- Robot/Car/Drone/Visually-impaired Navigation. 3D information is needed in this application for obstacle avoidance, free space determination, and pose estimation.
- Virtual/Augmented Reality (VR/AR). 3D computer vision is crucial for VR/AR. This includes camera pose estimation for head-mounted displays, 3D modeling for VR/AR contents, and seamless integration of virtual objects into the real world.
- Human-Computer/Robot Interaction. 3D computer vision allows the computer/robot to see the user, understand the user's non-verbal body language (gesture, body posture, facial expression, etc.), and interact in natural ways such as gesture and speech, instead of the traditional keyboard and mouse.

When I joined Microsoft Research early 1998, Microsoft's mission was to put a computer on every desk and in every home; my mission was to empower every computer with the visual capability of a human. I created a vision document called "Desktop Vision System" (DVS), with "vision" implying both "computer vision" and my vision for the future of desktop computers. In my DVS document, I set up my agenda along three lines of research:

- 3D object modeling. I imagined that products such as furniture would all have 3D photorealistic models. Before I bought them, I would use the DVS to build a 3D model of my house, import the furniture model, conduct lighting simulations, and determine whether it fits the house and where it fits the best.
- 3D face modeling. I imagined the DVS would create a 3D digital replica of each user, and this digital avatar would be a faithful representation of the user in 3D geometry, appearance, and dynamics, to be used when playing online games with others or when interacting with others in an online world.
- Human emotion understanding. I imagined the DVS would understand the user's facial expression, and, depending on the user's mood, the computer would react differently while interacting with the user. For example, when the user approached, the DVS would recognize that the user is not in a happy mood; it would then play soothing music, and chat with sympathy to improve the user's mood.

I have carried out all three lines of research to certain degrees of success. An early piece of work on 3D object modeling appeared at ICCV 2001 (Nishino, Zhang and Ikeuchi 2001). A live demo on 3D face modeling was shown at Bill Gates' keynote in launching Xbox in 2001, and the work was summarized in a book published by the Cambridge University Press in 2011 (Liu and Zhang

2011). The emotion-understanding work was shipped as part of Microsoft Project Oxford (https://www.projectoxford.ai/emotion), and an initial work was published in the IJPRAI journal (Zhang 1999a).

Having presented the context, I will now describe my invention on camera calibration, now known as *Zhang's method*. In order to fulfill the tasks that we described above with desired performance, the DVS needs to know the characteristics of its visual sensor (camera). That's the task of camera calibration, which is a process of determining the camera's internal parameters (focal length, aspect ratio and rectangularity of a pixel, the center of the image sensor) and the camera's external parameters (position and orientation). And that was my first project after joining Microsoft Research, working toward my *Desktop Vision System* vision.

Before I attacked the problem, there was already a technique available, which consisted of using a precisely fabricated 3D apparatus with painted patterns, such as two orthogonal planes hinged together, a cube, or a plane translated an exact amount in the normal direction. This makes sense since a camera records the projection of a 3D scene onto a 2D image plane, and if we know exactly what the 3D scene is, the camera parameters are readily determined. However, those 3D apparatuses are expensive to make and inconvenient to use. I wondered whether there could be a more flexible calibration technique.

With the knowledge of projective geometry I gained through my previous research career at INRIA in France, I discovered that we could calibrate a camera by just showing a planar pattern at a few different orientations (at least two) to the camera. The motion did not need to be known. Either the camera or the planar pattern could be freely moved, and the pattern could be printed on paper, then attached to a planar surface such as cardboard, making it truly flexible and easy to use.

This is rather counter-intuitive at first. We are only using a 2D object to calibrate a sensor which performs the projection from 3D to 2D. Without going into the details, I will explain why this is possible. Algebraically, a 2D plane in 3D space and its projection in the image define a homography mapping which has 8 parameters. The geometric relationship between the camera's 3D coordinate system and the plane's coordinate system has 6 degrees of freedom, which are unknown. Thus, each observation of a plane in space provides two constraints (8 - 6 = 2) on the camera's internal parameters. We therefore need two (for a simplified camera model with rectangular pixels) or more (for a general camera model) observations of a plane in space.

My calibration technique can also be explained geometrically. This requires a concept called *absolute conic* in projective geometry. The absolute conic is a conic on a plane at infinity. Its projection on an image plane, irrespective of the camera position and orientation, remains the same and depends only on the camera's internal parameters. A plane in 3D space intersects with the absolute conic at a pair of conjugate points. Their projection in the image plane is also a pair of conjugate points and is determined by the homography matrix. With N ($N \geq 3$) observations, we can determine the equation of the projection of the absolute conic, and in turn, the camera's internal parameters can be readily recovered from the conic equation.

Because of the ease of producing a planar pattern and the simplicity and robustness of the closed-form solution, my calibration technique is employed in computer-vision research labs the

world over, as well as by many companies. A version of this method calibrated the vision system in NASA's Mars Rover, while another version is used to recalibrate Kinect sensors to address the issue of sensor drift, saving several hundred million dollars of potential product returns. In fact, it could well be the most widely used camera-calibration algorithm extant.

This camera calibration technique was first described in an ICCV paper (Zhang 1999b), and was later published in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Zhang 2000). As of February 12, 2016, these two papers have been cited 10,515 times according to Google Scholar.

I developed the technology initially for my own need, and I did not have any idea that my work would have this sort of lasting impact. In 2013, I was rewarded with an IEEE Helmholtz Test of Time award. I am very happy that it has been used by so many people and companies. In retrospect, I did write in my paper, nearly two decades ago, that my algorithm "advances 3D computer vision one step, from laboratory environments to real-world use."

References

Liu, Z., and Z. Zhang. 2011. *Face Geometry and Appearance Modeling.* The Cambridge University Press.

Nishino, K., Z. Zhang, and K. Ikeuchi. 2001. "Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis." *International Conference on Computer Vision (ICCV).* IEEE. 599-606.

Zhang, Z. 2000. "A Flexible New Technique for Camera Calibration." *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 22 (11): 1330-1334.

Zhang, Z. 1999a. "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments With a Multi-Layer Perceptron." *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)* 13 (6): 893-911.

—. 1999b. "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations." *Int.ernational Conference on Computer Vision (ICCV).* IEEE. 666-673.