Acquisition of 3-D Scenes with a Single Hand Held Camera

André Redert and Emile Hendriks

Information and Communication Theory Group, Department of Electrical Engineering
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
email: {P.A.Redert,E.A.Hendriks}@its.tudelft.nl
http://www-ict.its.tudelft.nl
phone: +31 15 278 6269, fax: +31 15 278 1843

ABSTRACT

We investigate the acquisition of 3-D scenes by a single hand-held camera. The camera is mounted on a special device with four mirrors, enabling stereo capturing of the scene. We will discuss the signal processing tasks involved, camera calibration and correspondence estimation, and show that both of them benefit from the use of the device. Specifically, we will show that the device enables the full self-calibration of the camera, without loss of absolute scale as in general stereo self-calibration methods. Our experiments show that good 3-D models can be obtained.

1. INTRODUCTION

In the area of 3-D scene acquisition by stereo equipment, two signal processing tasks are involved: calibration of the camera pair [1] and estimation of corresponding pixels in the image pair [5]. Using the camera calibration parameters, we can construct the two light rays originating from a pair of corresponding pixels. The intersection of the two rays then provides the 3-D coordinates of a scene point.

In this paper, we examine the acquisition of 3-D scenes with a very specific stereo camera shown in Fig. 1. A single hand held camera is mounted on a patented apparatus with mirrors [7]. Imagine that the directions of the incoming light rays are reversed, then the two center mirrors split the bundle of light rays from the camera in two parts. The two side mirrors redirect each bundle towards the scene. The convergence point of the two bundles can be adjusted by rotation of the side mirrors.

The small size and low weight of this stereo camera provide high user mobility. In addition, the use of a single camera is economical and does not require shutter synchronisation of a camera pair. For storage, only one conventional recorder is needed.

Both the calibration and correspondence estimation tasks benefit from this particular setup. Correspondence estimation is based on photometric similarity of corresponding pixels. Photometrically unbalanced stereo cameras are a cause of errors, which is avoided to a large extend by the use of a single camera. In the area of camera calibration, there are two different techniques: fixed and self-calibration, which both benefit from the apparatus.

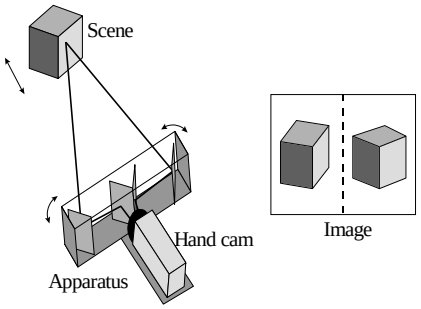

Figure 1: The scene, apparatus, hand cam and image

In fixed calibration, all camera parameters are extracted off line by placing a special object with known geometry in front of the cameras and processing of the camera images [1,2]. This method provides very accurate and complete results (all parameters can be obtained). Additionally, calibration reduces correspondence estimation from a 2-D search problem to a more efficient and reliable 1-D search [5]. With the device, the use of a single camera simplifies the stereo camera model without loss of generality.

Fixed calibration suffers from a number of disadvantages. A special calibration object and user interaction is required. Each time the camera parameters change, e.g. due to zooming or change of convergence angle, the calibration has to be repeated.

Self-calibration circumvents these disadvantages. In this method, the correspondences are estimated first, in an image pair of the scene. After this, the camera parameters are extracted from the found correspondence field [4]. The price to be paid is two-fold. First correspondence estimation is a more demanding task since no reduction from a 2-D to a 1-D search can be applied. Secondly, in self-calibration methods with normal stereo camera pairs, we do not have any reference to the standard SI meter. Thus the scale of the 3-D models can not be

obtained [3]. We will show that the known geometry of our device provides the scale.

The paper is organized as follows. In the next section, we give the camera model for our device and explain how absolute scale can be determined. We outline the complete scheme for 3-D acquisition in section three, which involves both fixed and self-calibration of the camera. Section four describes the experimental results. Finally section five summarizes the paper.

2. STEREO CAMERA MODEL

In this section we describe the camera model for our device. It is a specific version of the general model for stereo cameras in [4].

Figure 2 shows the function of the mirrors in the apparatus. The single real camera is split into two virtual cameras, each with half of the original CCD chip. The half CCDs are not centered on the optical axes of the virtual left and right cameras. The rotation of the two side mirrors is mechanically coupled. To have any overlap in the two virtual camera images, we must have $\alpha = 45° + \Delta\alpha$, with $\Delta\alpha > 0$. If the side mirrors are rotated around point *P* and *Q*, the two virtual cameras rotate around the same points with double speed.

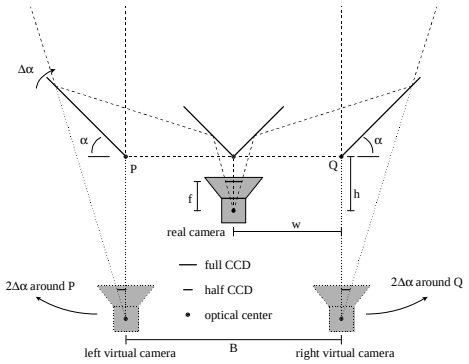

Figure 2: Geometry of the apparatus and camera

Figure 3 shows the general stereo camera model from [4]. Five reference frames are defined, the stereo frame, the left/right lens frames and the left/right projection frames. The camera baseline is *B*. The frame SF is defined to be a right handed frame in which the two optical centers lie on the x-axis symmetrically around the origin, at (-½*B*,0,0) for the left camera and (+½*B*,0,0) for the right camera, in SF coordinates. From Fig. 2 we can deduce:

$$B = 2w + 2(h + w)\sin 2\Delta\alpha + \varepsilon_0 \qquad (1)$$

relating meters to angles. This provides a means for self-calibration in meters, instead of unknown units. The ε_0 models remaining imperfections, and is assumed to be small.

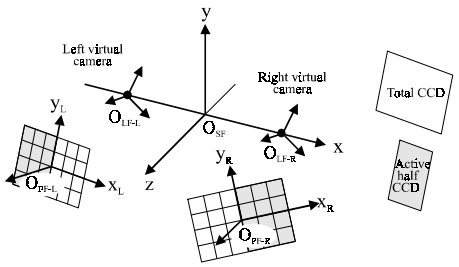

Figure 3: The stereo camera model

The orientations of the left and right lens frames are defined by two sets of Euler angles $(\varphi_x,\varphi_y,\varphi_z)$. The lens is present in the origin of the lens frames LF-L and LF-R, oriented in their *xy* planes. We assume radial symmetry in the lenses. Therefore, at this point, the angle φ_z has no meaning. We will not discard φ_z but use it for the orientation of the CCD. The reference frame *SF* is defined up to a rotation around the *x*-axis. We can therefore introduce an arbitrary equation that eliminates either $\varphi_{x;L}$ or $\varphi_{x;R}$, such as $\varphi_{x;L} + \varphi_{x;R} = 0$. Ideally, both are zero, but due to imperfections in the apparatus and the hand-cam this might not be the case:

$$\varphi_{x;L} = \varepsilon_1 \quad , \quad \varphi_{x;R} = -\varepsilon_1 \qquad (2)$$

For the $\varphi_{y;L}$ or $\varphi_{y;R}$ we have ideally $\varphi_{y;L} = 2\Delta\alpha$ and $\varphi_{y;R} = -2\Delta\alpha$. Allowing for small imperfections we have:

$$\varphi_{y;L} = 2\Delta\alpha + \varepsilon_2 \quad , \quad \varphi_{y;R} = -2\Delta\alpha + \varepsilon_3 \qquad (3)$$

We assume the CCD to be perfectly flat and have perfectly perpendicular image axes. The image formation is invariant for scaling of the triplet focal length, horizontal and vertical pixel size. Therefore we choose without with loss of generality the horizontal size of the pixels equal to 1 and the vertical size equal to *R*, the pixel aspect ratio.

The positions of the projection frames $PF_{L/R}$ (total CCD chip) relative to the lens frames $LF_{L/R}$ are defined by a *single* vector $(O_{PF}^{X_{LF}}, O_{PF}^{Y_{LF}}, O_{PF}^{Z_{LF}})$, since they refer to the same physical camera. The first two numbers define the intersection of the lens optical axis with the total CCD (mis-positioning) and the third is the focal length *f*:

$$O_{PF}^{X_{LF}} = \varepsilon_4 \quad , \quad O_{PF}^{Y_{LF}} = \varepsilon_5 \quad , \quad O_{PF}^{Z_{LF}} = f \qquad (4)$$

Since a change of focal length in cameras is usually performed by movement of the lens rather than the CCD chip, we model *h* in (1) as a linear function of *f*:

$$h = a + bf \qquad (5)$$

The orientation of the projection frames $PF_{L/R}$ (total CCD chip) relative to the lens frames $LF_{L/R}$ is defined by a single set of Euler angles $(\theta_x,\theta_y,\theta_z)$.

θ_z relates to the rotation of the projection frame. This is already modeled with φ_z and thus we use $\theta_z = 0$. For the φ_z we have:

$$\varphi_{z;L} = \varepsilon_6 \quad , \quad \varphi_{z;R} = \varepsilon_7 \tag{6}$$

The θ_x and θ_y model the non-orthogonal CCD placement with respect to the optical axis. Thus:

$$\theta_x = \varepsilon_8 \quad , \quad \theta_y = \varepsilon_9 \quad , \quad \theta_z = 0 \tag{7}$$

Since mispositioning and misorientation of the CCD is incorporated in (4) and (7), lens distortion can be modeled simpler than in [6]. We use only the radial distortion parameter K_3:

$$K_3 = \varepsilon_{10} \tag{8}$$

We have now defined a stereo camera model that contains the following parameters. For fixed calibration, we have baseline *B*, convergence angle $\Delta\alpha$, focal length *f*, pixel aspect ratio *R* and ten error parameters ε_1 .... ε_{10} which are assumed to be small. For self-calibration, the baseline *B* is discarded from the model by setting it to 1 during the calibration [4]. Afterwards, it can be obtained by (1), provided that *w*, *a* and *b* have been determined before hand.

3. ACQUISITION SCHEME

Figure 4 shows the complete scheme of acquisition. First, the hand-cam is mounted on the apparatus. Then we perform a fixed calibration for several values of convergence angle $\Delta\alpha$ and focal distance (zoom) *f* in order to obtain *a*, *b* and *w*. In addition, we obtain life-time constants such as the pixel aspect ratio *R*. The constants obtained will be invariant during the recording of the actual scene.

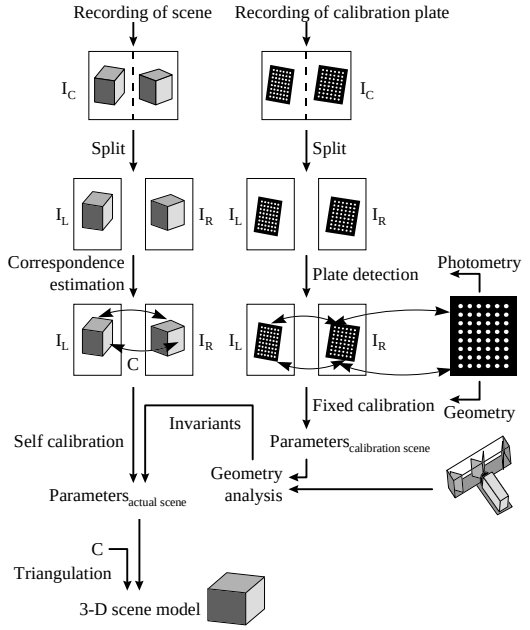

Figure 4: The complete scheme of acquisition

Then we record the scene, during which any change in convergence angle and zoom are allowed. Afterwards, we process the sequence according to the left route in Fig. 4. After correspondence estimation, we apply self calibration [4]. The invariant (1) then enables 3-D model acquisition with the correct scale.

4. EXPERIMENTS

In our experiments we used a digital photocamera that takes 1024x768 images in JPEG format. Figure 5 shows images of the calibration plate and actual scene.

Figure 5: Stereo views in a single image, calibration plate and actual scene

We have performed several fixed calibrations, with different values of the convergence angle and focal distance. With the different values for *B*, *f* and $\Delta\alpha$, we applied a least squares technique to estimate *w*, *a* and *b*. However, we could not find a good fit with relations (1) and (5), i.e., the values for ε_0 were not small compared to *B*. After inspection of the camera parameters obtained, we found that they are capable of explaining the measured features of the calibration object up to sub-pixel accuracy. Unfortunately, we found that there are multiple sets of camera parameters for which this holds, i.e. the parameters obtained are good but not unique. For the reconstruction of a 3-D model, each parameter set yields the same model but at a different absolute position in space. At this moment, the self-calibration method in Figure 4 remains open, until the fixed calibration method is more accurate or an alternative to (1) and (5) is found.

With the parameters from the fixed calibration, we were able to obtain a good 3-D model from the scene image in Figure 5. After splitting the scene image in a left and right image pair, we rectified the images [5], see Fig.6. All correspondences lie now on equal scanlines and 1-D disparity estimation can be performed.

Figure 6: Rectified image pair

We used a Markov Random Field disparity estimator [5] to obtain the disparity field shown in Figure 7.

Figure 7: The disparity field, denoting for each pixel in the right image what is the displacement to the corresponding pixel in the left image

After triangulation of all corresponding pixel pairs, we obtain the 3-D model. Figure 8 shows the details in the facial area.

Figure 8: The 3-D model

5. CONCLUSIONS

We have investigated the acquisition of 3-D scenes with a special device [7] that enables stereo vision by a single hand-held camera (see Fig. 1). This system has several advantages. It is small and thus provides high user mobility, it needs only a single conventional recorder for storage and the use of a single camera is economical.

The processing of a stereo image to obtain 3-D models involves camera calibration and correspondence estimation. Both these tasks benefit from the device. Correspondence estimation relies on photometric similarity between corresponding pixels. With this device there are no photometrical differences between left and right cameras. Further, left and right shutter synchronisation is guaranteed by definition.

For camera calibration, we showed that a simpler stereo camera model can be used since the virtual left and right cameras share some physical properties from the single real camera. In addition, we have shown that the device allows for self-calibration methods, while still providing a means for the capturing of absolute scale.

Our experiments showed that good 3-D models can be obtained with the device. We used a fixed calibration method at this moment. Although the camera parameters obtained are well suited for the acquisition of a 3-D model, the parameters are currently not accurate enough to serve as input for the self-calibration method.

Currently we are pursuing the improvement of the fixed calibration method, in order to apply the self-calibration method. In the presentation we will elaborate on these results.

ACKNOWLEDGEMENT

We would like to thank Olivier Zanen for the insightful discussions that contributed to this paper and for providing us with his patented device [7].

REFERENCES

[1] O. Faugeras, “Three-dimensional computer vision, a geometric viewpoint”, MIT Press, 1993

[2] F. Pedersini, D. Pele, A. Sarti and S. Tubaro, “Calibration and self-calibration of multi-ocular camera systems”, in proceedings of the *International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging* (IWSNHC3DI’97), Rhodos, Greece, pp. 81-84, 1997

[3] M. Pollefeys, R. Koch, M. Vergauwen and L. van Gool, “Flexible acquisition of 3D structure from motion”, in proceedings of the *IEEE Image and Multidimensional Digital Signal Processing (IMDSP) Workshop ’98*, pp. 195-198, 1998

[4] P.A. Redert and E.A. Hendriks, “Self calibration of stereo cameras with lens distortion”, *Proceedings of the IEEE Image and Multidimensional Digital Signal Processing (IMDSP) Workshop* ’98, pp. 163-166, 1998

[5] P.A. Redert, E.A. Hendriks and J. Biemond, “Correspondence estimation in image pairs”, *IEEE Signal Processing Magazine*, special issue on 3D and stereoscopic visual communication, Vol. 16, No. 3, pp. 29-46, May 1999

[6] J. Weng, P. Cohen and M. Herniou, “Camera calibration with distortion models and accuracy evaluation”, in *IEEE Transactions on PAMI*, Vol. 14, No. 10, pp. 965-980, 1992

[7] P.O. Zanen, “Single lens apparatus for three-dimensional imaging having focus-related convergence compensation”, US Patent #5,532,777, July 2, 1996. Related patents US 5828913, October 27, 1998 and US 5883662, March 16, 1999