Fakultät für Physik und Astronomie

Ruprecht-Karls-Universität Heidelberg

Masterarbeit

Im Studiengang Physik

vorgelegt von

(Vor- und Zuname)

geboren in (Geburtsort)

(Jahr der Abgabe)

(Titel)

(der)

(Masterarbeit)

Die Masterarbeit wurde von (Vorname Name)

ausgeführt am

(Institut)

unter der Betreuung von

(Frau/Herrn Prof./Priv.-Doz. Vorname Name)

Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

(name and surname)

born in (place of birth)

(year of submission)

(Title)

(of)

(Master thesis)

This Master thesis has been carried out by (Name Surname)

at the

(institute)

under the supervision of

(Frau/Herrn Prof./Priv.-Doz. Name Surname)

(Titel der Masterarbeit - deutsch):

(Abstract in Deutsch, max. 200 Worte. Beispiel: **?**)
Lorem ipsum dolor sit amet, consectetur adipisici elit, sed eiusmod tempor incidunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquid ex ea commodi consequat. Quis aute iure reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint obcaecat cupiditat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.
Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.
Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

(Title of Master thesis - english):

(abstract in english, at most 200 words. Example: **?**)
Lorem ipsum dolor sit amet, consectetur adipisici elit, sed eiusmod tempor incidunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquid ex ea commodi consequat. Quis aute iure reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint obcaecat cupiditat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.
Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.
Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Contents

1 Theory

1.1 Depth from focus

One advantage of using lightfields for depth measure is its ability to get a two-dimensional image of the scene at any depth. Integrating the views of the light field camera array has the same effect as the integration of a focussed lense camera, as the lense is simply integrating slightly different viewpoints of the same scene point when focussed on the correct depth.
Obtaining the refocussed integrated image is a synthetic process that only requires shifting the view coordinates artificially. Given a full four-dimensional light field $L(u, v, x, y)$ we can refocus the light field as described in Ng et al.:

$$L'(u, v, x, y) = L(u(1 - d'), v(1 - d'), x, y), \tag{1.1}$$

where d' describes the relative pixel shift. The disparity is directly related to the absolute depth of the focus (relate to PICTURE) if the relevant camera parameters are known. Given the baseline b in meters and the focal length f in pixels, the depth Z is given as

$$Z = \frac{f \cdot b}{d}. \tag{1.2}$$

We obtain

$$\bar{L}(x, y) = \frac{1}{N_{u,v}} \int \int L'(u, v, x, y) du dv = \frac{1}{N_{u,v}} \sum_u \sum_v L'(u, v, x, y) \tag{1.3}$$

Once we can focus at any range, one can adopt *depth-from-focus*-techniques as described in Watanabe and Nayar [1998] for depth measure. If the scene point at a given image coordinate (x, y) in the center view is in focus, the contrast in the integrated image $\bar{L}(x, y)$ is high, thus a contrast measure at each pixel combined with stepwise refocussing yields a depth map.
For measuring the contrast, one has different options: The most straight forward approach is calculating the first derivative of the grey-value image. At high contrast structure the local intensity changes are expected to be high. Alternatively one could measure the second derivative laplacian that eventually results in higher robustness. The implementation and tests of those techniques for the benchmark dataset can be found in section **??**.
Using a pinhole camera array allows us to go further and find a response value that shows higher consistency. Taking the absolute difference between the center view of

the camera array and the refocussed image yields to promising results as shown in Tao et al. [2017]. Under the assumption of lambertian surfaces the RGB- value of any scene point should be the same under all angles. Thus when refocussed on the correct depth, summing over all angles should result in a value that ideally is the same as in the center view alone. This is referred as *photo consistency*; for more information read Tao et al. [2017]. The response value at a given depth is obtained from

$$D'(x,y) = \frac{1}{|W_D|} \sum_{x',y' \in W_D} \left| \bar{L}(x',y') - P(x',y') \right|, \tag{1.4}$$

where $P(x,y)$ is the center view. For more robustness, it is averaged over a small window. We refer to this measuring technique as *photo consistency* in the following. Note that calculating the absolute results in a 1-channel-image while the input images are RGB-images.
Tao et al. propose another measure that they refer to as $angular correspondence$. It follows the same principle, but instead of integrating the refocussed lightfield followed by comparing it to the center view, they directly take the difference of each viewpoint to the center view and sum up those differences:

$$D'(x,y) = \frac{1}{N_{u,v}} \sum_u \sum_v |L'(u,v,x,y) - P(x,y)|. \tag{1.5}$$

We tested those methods against the common contrast measures mentioned above, the results are found in section results.

1.2 Semi-Global Matching

1.2.1 Semi - Global Matching for Stereo Vision

In contrast to Light field depth estimation techniques Stereo systems often suffer from mismatching pixels between the left and right images. Many attemps have been made to smoothen bad pixels, resulting in blurred edges or long calculation times. One promising attempt was published in 2005 by Heiko Hirschmüller (Hirschmuller [2005]) that was described as „a very good trade off between runtime and accuracy “by himself (Hirschmüller [2011]): we speak of Semi-Global Matching.
In general, matching of two stereo images means shifting the disparity over the predefined disparity range and comparing both images until we have a cost value at each image point for each discrete disparity. We assign to each pixel $\vec{p}$ the disparity value $D_{\vec{p}}$ which resulted in the lowest cost $C(\vec{p}, D_{\vec{p}})$. This matching does not have to be unique, resulting in errorneous pixel disparities. To overcome this one wants to minimize a global cost function of the form

$$E(D) = \sum_{\vec{p}} \left(C(\vec{p}, D_{\vec{p}}) + \sum_{q \in N_p} \begin{cases} P1 & \text{if } |D_{\vec{p}} - D_{\vec{q}}| = 1 \\ P2 & \text{if } |D_{\vec{p}} - D_{\vec{q}}| \geq 1 \\ 0 & \text{else} \end{cases} \right). \tag{1.6}$$

The fist term sums all matching costs over the whole image, while the second term forces continuity by comparing the disparity of all neighbour pixels N_q to the disparity D_p; if a small discontinuity is detected ($D_{\vec{p}} - D_{\vec{q}} = 1$), a small penalty is added to the global cost function. Since a small discontinuity can be found essentially at any tilted plane, only a small error is added. A bigger disparity difference indices a clear discontinuity in the disparity map.
However, minimizing the global cost function involves computational cumbersome algorithms as it is a NP-complete Problem (Hirschmüller [2011]). Semi-Global Matching however chooses another approach by minimizing the global cost function along one-dimensional lines – this can indeed be calculated in polynomial time. The new smoothed cost function at pixel $\vec{p}$ is then given as the sum of all 1D minimum cost paths that are ending in $\vec{p}$. The minimal cost L'_r along the path r is defined recursively as

$$L'_r(\vec{p}, D) = C(\vec{p}, D) + \begin{cases} L'_r(\vec{p_{\text{before}}}, D) \\ L'_r(\vec{p_{\text{before}}}, D+1) + P1 \\ L'_r(\vec{p_{\text{before}}}, D-1) + P1 \\ L'_r(\vec{p_{\text{before}}}, D) + P2 \end{cases} \tag{1.7}$$

2 results

2.1 Depth from focus

The depth measure using epipolar plane analysis requires iterative calculation of the structure tensor for each EPI at each disparity. A way to overcome this is to generate a preestimate of the depth before actually calculating the correct depth. This could also help to prevent possible errors due to periodic scene characteristics which can lead to mismatch errors when calculating the structure tensor. Therefore the depth pre-estimate should fulfil the following criteria:

1. It should be *consistent*, meaning that the number of pixels with low confidence should be the lowest possible.
2. It should result in a *fast* measure, ideally faster then it would take to do the full iterative structure tensor algorithm.
3. It does not have to be subpixel accurate, since it only serves as a pre-estimate.

The methods that are tested are described in section 1.1. We test four different ways to obtain a depth map using depth from focus:

Photo consistency This measure takes advantage of the fact that the difference between the refocussed two-dimensional image and the center view is close to zero when refocussed to the correct depth. Response value:

$$D'(x,y) = \frac{1}{|W_D|} \sum_{x',y' \in W_D} \left| \bar{L}(x',y') - P(x',y') \right|, \tag{2.1}$$

Angular correspondence In contrast to the *Photo consistency* - measure, it first calculates the absolute difference between each camera array view and the center view followed by the summation of those deviations. The response value is given as in equation (1.5)

$$D'(x,y) = \frac{1}{N_{u,v}} \sum_u \sum_v |L'(u,v,x,y) - P(x,y)| \tag{2.2}$$

First derivative The first derivative is calculated for contrast measure by applying the sobel filter onto the refocussed image I:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \cdot I \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \cdot I \tag{2.3}$$

The directional gradients are simply added up to the response value

$$D'(x,y) = |G_x(x,y)| + |G_y(x,y)| \tag{2.4}$$

Laplace Here we calculate the second derivative laplacian by appling the sobel operator twice:

$$D'(x,y) = \text{Laplace}(I)(x,y) = \frac{\partial^2 I}{\partial x^2}(x,y) + \frac{\partial^2 I}{\partial y^2}(x,y) \tag{2.5}$$

Part I

Appendix

A Lists

A.1 List of Figures

A.2 List of Tables

B Bibliography

Heiko Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 807–814. IEEE, 2005.

Heiko Hirschmüller. Semi-global matching-motivation, developments and applications. *Photogrammetric Week 11*, pages 173–184, 2011.

Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera.

Michael W Tao, Pratul P Srinivasan, Sunil Hadap, Szymon Rusinkiewicz, Jitendra Malik, and Ravi Ramamoorthi. Shape estimation from shading, defocus, and correspondence using light-field angular coherence. *IEEE transactions on pattern analysis and machine intelligence*, 39(3):546–560, 2017.

Masahiro Watanabe and Shree K Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, 1998.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den (Datum) ....