NOTES^{3D}: ENDOSCOPES LEARN TO SEE 3-D
Basic Algorithms for a Novel Endoscope

J. Penne, K. Höller
Friedrich-Alexander-University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany

S. Krüger
Surgical University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg
Krankenhausstr. 1, 91058 Erlangen, Germany

H. Feußner
University Hospital Rechts der Isar, Technical University Munich, Ismaninger Str. 22, 81675 Munich, Germany

Keywords: NOTES, Time-Of-Flight, Endoscopy, Minimally Invasive Surgery, NOTES^{3D}, Reconstruction, Calibration, MUSTOF endoscope.

Abstract: TOF chips enable the acquisition of distance information via the phase shift of a reference signal and the reference signal reflected in a scene. Transmitting the reference signal via the light conductor of an endoscope and mounting a TOF chip distally enables the acquisition of distance information via an endoscope. The hardware combination of TOF technology and endoscope optics is termed Multisensor-Time-Of-Flight endoscope (MUSTOF endoscope). Utilizing a MUSTOF endoscope in the context of NOTES procedures, enables the direct endoscopic acquisition of 3-D information (NOTES^{3D}). While hardware issues are currently under investigation, an algorithmic framework is proposed here, dealing with the main points of interest: calibration, registration and reconstruction.

1 INTRODUCTION

The expanding possibilities in endoscopic and laparoscopic surgery have reached the point where many procedures can now be performed in a minimally invasive manner. This is enforced by the rapid development of computer assisted surgery (Feußner, 2003). It could possibly be performed without skin incisions. The natural orifices may provide the entry point for surgical interventions in the peritoneal cavity. Entering through an incision in the digestive tract such as the stomach or colon will avoid abdominal wall incisions. 14 leaders from the American Society of Gastrointestinal Endoscopy (ASGE) and the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) met on July 2005 and agreed that Natural Orifice Translumenal Endoscopic Surgery (NOTES) could offer significant benefits to patients such as less pain, faster recovery, and better cosmesis than current laparoscopic techniques (Rattner et al., 2006).

A potential barrier to clinical practice is the missing knowledge of spatial orientation and exact endoscope position during the operation. Many NOTES procedures will be performed with the endoscope in a retroflexed position and require secondary access sites creating situations in which the image is upside down and an offaxis manipulation is required. Potential solutions include incorporating visualization, reconstructed images, registeres volumes, Augmented Reality or the use of multiple cameras to achieve the appropriate inline view of the working area. If the principles learned in advanced laparoscopic operations are applicable to NOTES, orientation and knowledge of precise position will be a fundamental requirement for any NOTES surgical system. Online 3-D information registered with preoperative CT or MR data may provide information on position and orientation of the endoscope.

One possibility to face this challenge is to acquire 3-D information from endoscopic images. Or even better: to acquire the 3-D information directly via the endoscope. Making an endoscope actually *see* three-dimensionally will constitute a significant step towards the information necessary to meet the re-

Penne J., Höller K., Krüger S. and Feußner H. (2007).
NOTES3D: ENDOSCOPES LEARN TO SEE 3-D - Basic Algorithms for a Novel Endoscope.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 134-139
DOI: 10.5220/0002068401340139

quirements of NOTES procedures. Considering this new option of acquiring data via an endoscope, the term NOTES^{3D} was chosen.

2 STATE OF THE ART

Acquiring 3-D information from endoscopic images is usually done during an off-line routine, i.e. an image sequence is recorded and subsequently processed to get the desired information about the operation area (Vogt et al., 2005; Thormählen et al., 2002).

Some systems can provide 3-D information which can be used to registrate it with preoperative data and subsequently provide Augmented Reality (Milgram and Colquhoun, 1999; Vogt et al., 2004; Olbrich et al., 2005; Vogt, 2005).

TOF technology enables the direct acquisition of the distance information about a world point which is projected on a sensor element (Schwarte et al., 1999). Currently, framerates $\geq$12 fps are achieved by TOF cameras.

TOF cameras illuminate the scene actively with an optical reference signal. By Smart Pixels, which are integrated into the TOF camera chip (TOF chip), the reflected optical wave is analysed and for each pixel the phase shift compared to the reference signal is estimated. Assuming a constant speed for the spread of the signal the phase shift is directly proportional to the distance of a point in the recorded scene. Currently, lateral resolutions of up to 140$\times$170 pixel and z-resolutions of 1 mm are available. Simultaneously to the phase delay, the amplitude of the reflected optical wave is estimated. This information provides a gray-scale image of the scene, with the reflectivity of the material being encoded in the gray-values. We will reference the information of amplitude and distance data as one *frame* acquired with a TOF camera. It will become clear from the context whether we use the distance and/or amplitude data of the current frame.

Minimally invasive surgery and especially NOTES extensively uses endoscopes as image acquisition devices. An endoscope fulfills two main tasks: it enables the transmission of optical signals (for illumination purposes) into the operation area via fiber optic cables; and it enables the transmission of optical signals from the operation area to the distally mounted endoscope camera (equipped with a CCD chip) using a system of only lenses (rigid endoscopes) or lenses and fiber optic cables (non-rigid endoscopes). Based on these image acquisition devices and benefitting from the development of adequate surgical devices the field of minimally invasive surgery subsummarizes a wide field of surgical operations today.

The image acquisition technique applied in endoscopes fits the requirements for a TOF based distance measurement. By transmitting the optical reference signal via the light fibre cables as well as the cold daylight which is commonly used for illumination purposes and distally mounting a TOF chip as well as a CCD chip the acquisition of image and distance data is possible. The TOF chip evaluates the reflected reference signal and provides distance information; the CCD chip provides a color image of the operation area based on the cold daylight. Considering the basic techniques applied in this endoscope, the name Multisensor-Time-Of-Flight endoscope (MUSTOF endoscope) will be used further.

Considering this hardware configuration, which is currently intensively investigated for optimization purposes, an algorithmic framework for the handling of this multisensor system is required. First, a model of the multisensor system is presented. Second, algorithms for the basic requirements (calibration, registration, reconstruction) are proposed.

3 MODEL DESCRIPTION

The amplitude value acquired with a TOF camera at pixel (i, j) is denoted with $a_{i,j}$. The corresponding distance value measured in *mm* is denoted with $d_{i,j}$. The color value in the CCD camera at position (i, j) is denoted with $p_{i,j}$. TOF camera and CCD camera are modeled as pinhole cameras. The intrinsic parameters (f: focal length, $(c_x,\ c_y)$: principal point) and the extrinsic parameters ($R \in \mathbb{R}^{3\times3}$: rotation matrix, $t \in \mathbb{R}^3$: translation vector) are denoted as given for the CCD camera and denoted with a $'$ for the TOF camera.
Using homogenous coordinates (indicated by $\underline{\ }$) a 3-D point q is projected to the CCD camera pixel p, according to

$$p = K[R|t]\underline{q} \tag{1}$$

and to the TOF camera pixel

$$p' = K'[R'|t']\underline{q}, \tag{2}$$

where K is the calibration matrix containing the intrinsic camera parameters according to (Tsai, 1987).
Let $(p_x,\ p_y)$ and $(p'_x,\ p'_y)$ denote the physical dimensions of a sensor element of the CCD or TOF camera in *mm*. Given a distance value $d_{i,j}$, let $d_x = (i - c'_x)p'_x$ and $d_y = (j - c'_y)p'_y$ be the distance of the pixel from the principal point measured in *mm* and $d_z = \sqrt{d_x^2 + d_y^2 + f'^2}$ be the distance of the optical

centre to the pixel measured in *mm*, the corresponding 3-D point $q = (q_x, q_y, q_z)^T$ may be computed by

$$q_x = \frac{d_x d_{i,j}}{d_z},\ q_y = \frac{d_y d_{i,j}}{d_z},\ q_z = \frac{f' d_{i,j}}{d_z}. \qquad (3)$$

Assuming the CCD and TOF camera being rigidly mounted (at the distal end of the endoscope), implies that the spatial relation between the optical centres of both cameras can be described by a relative rotation $R_r \in \mathbb{R}^{3\times3}$ and translation $t_r \in \mathbb{R}^3$, with

$$R_r = RR'^{-1},\ t_r = t - R_r t', \qquad (4)$$

where R, R', t and t' describe the pose of the corresponding camera in a common world coordinate system.

4 METHODS

4.1 Calibration

Two cameras are to be calibrated: the CCD and the TOF camera. Tsai's widely used algorithm (Tsai, 1987) is in principle applicable to both cameras:

1. Capture image of a calibration pattern with N calibration points.
2. Determine 2-D calibration points $c_i,\ 1 \leq i \leq N$.
3. Assigning 3-D world points $w_i,\ 1 \leq i \leq N$ to 2-D calibration points.
4. Estimation of an intrinsic (K) and extrinsic ($R,\ t$) camera parameters involving Levenberg-Marquardt non-linear optimization (Dennis and Schnabel, 1983).

For the CCD camera the algorithm does not have to be modified. For calibrating the TOF camera the standard algorithm was modified due to the low lateral resolution of the TOF cameras which leads to a very unspecific localization of the points of the calibration pattern in the acquired data. The non-linear optimization usually aims at minimizing the squared backprojection error

$$\sum_{i=1}^{N} \|c_i - proj(w_i, K', R', t')\|^2 \qquad (5)$$

, where $proj(w_i, K', R', t')$ is the projection of the world point w_i into the image plane specified by K', R' and t'. This functional was extended by a term which describes the deviation of the 3-D reconstructed calibration points from the plane which they are lying on. Let $\hat{c}_i$ be the 3-D point reconstructed from c_i specified in world coordinates (of the calibration pattern) using the distance information available from the TOF camera and the extrinsic parameters. Furthermore, ε_c denotes the regression plane calculated using all $\hat{c}_i,\ 1 \leq i \leq N$. The extended functional, which is minimized for a TOF camera in the calibration routine, is

$$\sum_{i=1}^{N} (\|c_i - proj(w_i, K', R', t')\| + \alpha\|\hat{c}_i - w_i\| + \beta d(\hat{c}_i, \varepsilon_c))^2, \qquad (6)$$

where $d(\hat{c}_i, \varepsilon_c)$ is the distance of $\hat{c}_i$ to the regression plane ε_c and $\alpha,\ \beta$ are scaling parameters. The term $\|\hat{c}_i - w_i\|^2$ penalizes wrong intrinsic and extrinsic camera parameters which lead to a wrong reconstruction of the calibration points. The term $d(\hat{c}_i, \varepsilon_c)$ only penalizes wrong intrinsic camera parameters as only those are relevant for the reconstruction of all $\hat{c}_i$ on a plane (wrong extrinsic parameters only imply a rotation and translation of the plane).

4.2 Registration

CCD and TOF camera have to be registrated as it is for medical applications necessary to relate the acquired 3-D information with the acquired color information of the operation area. Using the calibration routine described above the extrinsic parameters ($R',\ t',\ R,\ t$) and intrinsic parameters ($K,\ K'$) are known for each camera when capturing simultaneously an image of the calibration pattern. This enables the calculation of R_r and t_r as described by equation (4) and (assuming a parallel acquisition of data) a computationally inexpensive assigning of 3-D points to color information: by using formula (3) a 3-D point can be reconstructed which is specified in TOF camera coordinate system and by applying formulas (4) and (1) the 3-D point is projected into the CCD camera image plane.

4.3 Reconstruction

The TOF camera enables the reconstruction of the scene visible in the actual frame. For medical applications it is of utmost importance to provide a reconstruction of the whole operation area or at least an arbitrary defined region. This requires the registration of multiple reconstructions r_i, containing a certain number of 3-D points, available from a TOF camera. Two reconstructions r_i and r_j (with TOF camera poses (R'_i, t'_i) and (R'_j, t'_j)) are under the assumption of constant intrinsic camera parameters related via a rotation $R'_{i,j}$ and translation $t'_{i,j}$ of the optical centre of the TOF camera. The great amount of 3-D points to be registrated (approx. 20000 points for each frame) at a high framerate requires an efficient determination of

$R'_{i,j}$ and $t'_{i,j}$. The framerate of $\geq$12 fps allows the assumption of only relatively small camera movements between consecutive frames. Thus, the following algorithm for determining $R'_{i,j}$ and $t'_{i,j}$ is proposed:

- Initialization:
  1. Acquire the first frame r_1 of the TOF camera.
  2. Detect edges in the amplitude and distance data (number of found points: N_1). The 3-D coordinates $q_i,\ 1 \leq i \leq N_1$ of points detected as lying on an edge are used as world decription data $W = \{q_1, ..., q_{N_1}\}$.
  3. Initialize the TOF camera position with $R'_1 = I$ and $t'_1 = \underline{0}$, where I is the identity matrix and $\underline{0}$ is a 3×1 zero-vector.
- Processing of subsequently acquired frames:
  1. Acquire a frame r_i of the TOF camera.
  2. Detect edges in the amplitude and distance data (number of points N_i). The 3-D coordinates $q_i,\ 1 \leq i \leq N_i$ of points detected as lying on an edge are used as current world decription data $W_{cur} = \{q_1, ..., q_{N_i}\}$.
  3. Set the initial solution of the current camera pose to $R'_i = R'_{i-1}$ and $t'_i = t'_{i-1}$.
  4. Apply a Levenberg-Marquardt non-linear optimization on (R'_i, t'_i) to maximize
  $$\rho(proj(W, K', R'_i, t'_i), W_{cur}) \qquad (7)$$
  , where $proj(W, K', R'_i, t'_i)$ describes the projection of the world description data W on the image plane whose pose is described by (R'_i, t'_i) and whose intrinsic parameters are given by K' (known from the calibration routine). Furthermore, ρ describes the correlation coefficient according to Neyman-Pearson. Spoken briefly: the current extrinsic camera parameters are estimated by maximizing the correlation coefficient between edges detected in the current frame and edges already detected in the preceding frames.
  5. Update W: add the edges found in the current frame (W_{cur}) to W.
  6. The relative camera pose compared to the preceding frame is described by $R'_{i-1,i} = R'_i R'^{-1}_{i-1}$ and $t'_{i-1,i} = t'_i - R'_{i-1,i} t'_{i-1}$.

The algorithm substitutes the task of registrating two point clouds (each containing thousands of points) by the task of registrating edges detected in the current frame with a world description of the world containing only edges found in previous frames. This enables an impressive on-the-fly-registration of consecutively acquired frames.

5 RESULTS

For evaluating the effects of the extended functional in the non-linear optimization step of the calibration, the calibration routine was performed 11 times with the standard functional (backprojection error) and 11 times with the extended functional (using $\alpha = 0.0001,\ \beta = 0.001$). The TOF camera used was a PMD19k (lateral resolution: 120×160 pixel). The average (indicated by $_\mu$) and standard deviation (indicated by $_\sigma$) of the calculated intrinsic camera parameters ($f',\ c'_x,\ c'_y$) are given in table 1. The results point

Table 1: Evaluation of calibration routines (I) - average and standard deviation of intrinsic camera parameters.

	standard calibration	extended calibration
$f'_\mu[mm]$	11.526	12.604
$f'_\sigma[mm]$	0.732	0.260
$c'_{x_\mu}[pix]$	80.000	80.009
$c'_{x_\sigma}[pix]$	0.0	0.342
$c'_{y_\mu}[pix]$	60.000	60.525
$c'_{y_\sigma}[pix]$	0.0	0.152

out that the stability of the calibration is improved by the extended functional as the standard deviation of the calculated focal length is decreased. Note, that in the case of the standard calibration the principal point was not changed: the initial pixel coordinates (80,60) were not altered by the routine. By using the extended functional the parameters describing the position of the principal point became more sensitive.

For evaluating the effects of the different calibration routines on the reconstruction planes at different distances were reconstructed: first a white wall at a distance of approx. 1 m (Scene A), approx. 50 cm (Scene B) and approx. 20 cm (Scene C). Each scene was reconstructed using first standard assumptions for the intrinsic parameters ($f' = 12\ mm,\ c'_x = 80,\ c'_y = 60$; focal length known from datasheet of the TOF camera), then the average intrinsic parameters computed by the standard calibration routine ($f' = 11.526\ mm,\ c'_x = 80,\ c'_y = 60$) and finally using the intrinsic camera parameters computed by the extended calibration routine ($f' = 12.604\ mm,\ c'_x = 80.009,\ c'_y = 60.525$). For each reconstruction the average distance of all points to the regression plane was calculated (table 2). The results indicate that the intrinsic camera parameters, which were calculated using the extended functional, reduce the error, i.e. they lead to a reconstruction closer to the world geometry (a plane). It is important to notice that the camera used in the tests has the reference light sources mounted to the left and right of the ocular. Thus, the

difference of the travelled distance of signals emitted from the left light sources compared to signals emitted from the right light sources can not be neglected for objects very close to the ocular of the camera. This effect can be expected to be reduced by the proposed MUSTOF endoscope as it uses the optical fibres of a standard endoscope: the reference light is emitted from the endoscope tip and fits much more to the characteristics of a point light source than the light sources of the camera used in the tests. To illustrate the possibilities of a stable calibration of a TOF camera example images of the reconstruction of a liver model are given: after calibrating a system consisting of a Webcam rigidly mounted on a TOF camera the relative pose of the image planes was computed. Consequently, it was possible to project TOF camera pixels onto corresponding Webcam pixels. The images show different views of a reconstruction of a liver model: figure 1 shows the original liver model; figure 2 shows the 3-D reconstruction of the liver model (the gray values correspond to the amplitude data); figure 3 uses the color information of the registrated Webcam image plane. The achieved level of detail in the reconstruction is noticable: the gall bladder is well distinguishable; and the structure of the liver can well be recognized.

Table 2: Evaluation of calibration routines (II) - average distance of points to regression plane in mm.

	standard assumption	standard calibration	extended calibration
Scene A	91.8	83.1	73.9
Scene B	60.5	54.7	51.8
Scene C	40.9	41.9	35.9

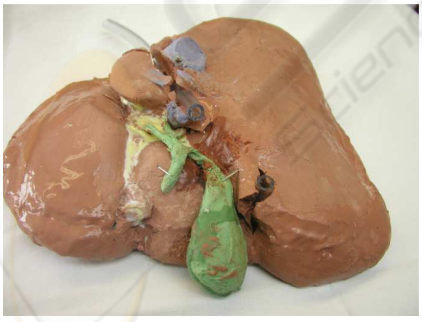

Figure 1: Silicon model of the liver with gall bladder.

The registration of multiple views is illustrated in figures 4-7. Figure 4 shows the scene which had to be reconstructed. Figure 5 visualizes the first frame which was reconstructed (left part of the chair). Then a camera movement to the right and upwards was performed and the acquired frames were registrated on-the-fly. Figure 6 shows the state of the reconstruction during the camera movement. Figure 7 visualizes the final 3-D reconstruction: the virtual camera was moved upwards to allow the investigation of the scene from a higher viewpoint. This possibility might be very useful for intraoperative usage: while being able to compute a 3-D reconstruction of the scene/operation area the already reconstructed world geometry can be observed and explored by a virtual camera. Consequently, the endoscope does not have to be moved while the reconstructed operation area might be explored three-dimensionally.

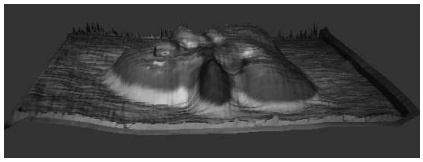

Figure 2: 3-D reconstruction of the model (gray values correspond to amplitude data).

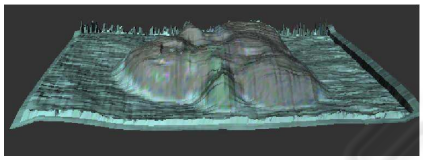

Figure 3: 3-D reconstruction of the model with registrated color information from a Webcam.

Figure 4: Scene to be reconstructed.

6 CONCLUSION

Motivated by the requirement to provide a 3-D reconstruction for NOTES^{3D} procedures the idea of a MUSTOF endoscope has been proposed. The main

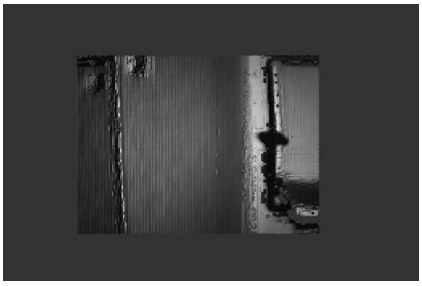

Figure 5: Initial 3-D reconstruction.

Figure 6: Expanding the 3-D reconstruction.

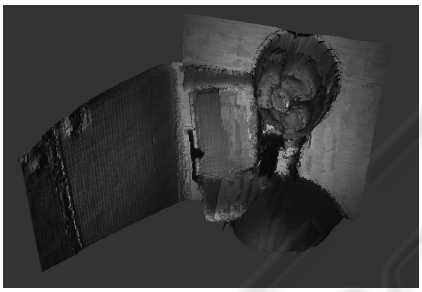

Figure 7: Final 3-D reconstructed scene.

contribution of this work is the description of basic algorithms for calibration, registration and reconstruction using TOF cameras (in combination with CCD cameras). These algorithms are not explicitly limited to medical applications, but constitute a proof of concept considering the applicability of this technology in minimally invasive surgery and especially NOTES^{3D}, i.e. NOTES procedures utilizing a MUSTOF endoscope. Future investigations will address extensions and modifications of the proposed algorithms to ensure an optimal performance if the TOF camera acquires its data via an endoscope optic. Furthermore, the achievable accuracy of the 3-D reconstruction will be investigated: currently, an evaluation of the accuracy of a TOF camera has only very limited relevance for a MUSTOF endoscope as the characteristics of the used optical system (standard lense vs. endoscope optic) vary significantly.

REFERENCES

Dennis, J. E. and Schnabel, R. B. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, New Jersey.

Feußner, H. (2003). The operating room of the future – A view from Europe. *Seminars in Laparoscopic surgery*, 10:3:149–156.

Milgram, P. and Colquhoun, H. (1999). *Mixed Reality – Merging Real and Virtual Worlds*, chapter A Taxonomy of Real and Virtual World Display Integration, pages 5–30. Springer, Berlin, Heidelberg, New York.

Olbrich, B., Traub, J., Wiesner, S., Wiechert, A., Feußner, H., and Navab, N. (2005). Respiratory motion analysis: Towards gated augmentation of the liver. In Lempke, H. U., Inamura, K., Doi, K., Vannier, M. W., and Farman, A. G., editors, *Computer Assisted Radiology and Surgery (CARS), Proceedings of the 19th International Congress and Exhibition*, pages 248–253. Elsevier, Amsterdam.

Rattner, D., Kalloo, A., and the SAGES/ASGE Working Group on NOTES (2006). ASGE/SAGES Working Group on Natural Orifice Translumenal Endoscopic Surgery. *Surgical Endoscopy*, 20:329–333.

Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T., Xu, Z., and Hartmann, K. (1999). *Handbook of Computer Vision and Applications*, volume 1. The Academic Press.

Thormählen, Broszio, T., Meier, H., and P.N. (2002). Three-dimensional endoscopy. *Falk Symposium No. 124, Medical Imaging in Gastroenterology and Hepatology, Hannover, Germany, September 2001, Kluwer Academic Publishers, 2002, ISBN 0-7923-8774-0*, 0(0).

Tsai, R. Y. (1987). A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, Ra-3(3):323–344.

Vogt, F. (2005). *Augmented Light Field Visualization and Real-Time Image Enhancement for Computer Assisted Endoscopic Surgery*. PhD thesis, Friedrich-Alexander-University Erlangen-Nuremberg.

Vogt, F., Krüger, S., Niemann, H., Hohenberger, W., Greiner, G., and Schick, C. (2005). Erweiterte Realität und 3-D Visualisierung für minimal-invasive Operationen durch Einsatz eines optischen Trackingsystems. In Meinzer, H.-P., Handels, H., Horsch, A., and Tolxdorff, T., editors, *Proceedings of the Workshop Bildverarbeitung für die Medizin (BVM),*, pages 217–221. Springer, Berlin, Heidelberg, New York.

Vogt, F., Krüger, S., Zinßer, T., Maier, T., Niemann, H., Hohenberger, W., , and Schick, C. H. (2004). Fusion von Lichtfeldern und CT-Daten für minimal-invasive Operationen. In Tolxdorff, T., Braun, J., Handels, H., Horsch, A., and Meinzer, H.-P., editors, *Proceedings of the Workshop Bildverarbeitung für die Medizin (BVM),*, pages 309–313. Springer, Berlin, Heidelberg, New York.