Photorealistic 3-D Surface Reconstructions Using TOF Cameras

Jochen Penne[1], Kurt Höller[1], Dirk Wilhelm[2], Hubertus Feußner[2], and Joachim Hornegger[1]

[1] Chair for Pattern Regocnition, Friedrich-Alexander-University Erlangen-Nuremberg
[2] MITI Group, Klinikum rechts der Isar, Munich

Abstract. TOF (Time-of-Flight) technology enables the real-time 3-D surface reconstruction of the current field of view of a TOF camera with a constant lateral resolution. We propose algorithms for the calibration of a TOF camera which in combination with rigidly mounted standard cameras enables the photorealistic 3-D surface reconstruction of the current field of view by registrating the optical centres of the involved cameras and projecting the 3-D surface reconstruction into the standard camera image plane. For various applications not only the surface in the current field of view has to be reconstructed: We propose an algorithm to registrate consecutively acquired 3-D surface reconstructions of static scenes.

1 Introduction

In many applications, such as medical imaging, robotics, biometrics, automobile security and navigation and surveillance a three-dimensional representation of the environment is required. The acquisition of 3-D data has become one of the most important challenges in the design of modern electronic image sensors as the acquisition of 3-D data is favoured over alternative methods based on ultra-sonic or radar. Furthermore, by the direct acquisition of 3-D information time-consuming processing steps to compute the 3-D information from 2-D information are obsolete. The main advantage of *optical* systems is the capability of a very fast 3-D data acquisition, an eye-safe system setup and a high lateral resolution.
Standard acquisition techniques incorporate scanning systems or stereo systems with complex filtering and correlation processing units. But the real-time requirement which is done for most applications implies that neither the complex calculation of a distance map nor a scanning component should lead to an additional processing time-delay. Optical cost-efficient real-time 3-D cameras thus are hardly realizable with these techniques. Consequently, for 3-D imaging of the whole field of view pixel matrices are required with each pixel capable of delivering the distance information of one point in the field of view. Thus, these pixels are termed *smart pixels*.
The time-of-flight (TOF) technique is one possible method for the fast optical acquisition of distance information. The time that light needs to travel from the measurement system to the object and back again corresponds directly to the distance R, and is called time-of-flight (TOF):

$$TOF = \frac{2R}{c}, \tag{1}$$

where c is the light velocity ($c \approx 3 \cdot 10^8 \frac{m}{s}$). The method is very suitable for ranges starting from some centimetres to several hundreds of meters with relative accuracies of 0,1%. Nevertheless, the relative accuracies depend strongly on the illumination conditions and the effective supression of multiple reflections. Based on smart demodulation lock-in pixels the parallel measurement of the distance can be realized for some thousand points in a scene [1]. Two methods for measuring the TOF are distinguished:

– A light pulse is sent out and its turn-around time is measured directly.
– A continuously-modulated, e.g. sinusoidally modulated, light wave is emitted and the phase delay between the original and received light signal is estimated to compute the time or distance information.

The latter method is used in available range imaging sensors.

2 Model description

TOF technology enables the direct acquisition of the distance information about a world point which is projected on a sensor element [2]. Currently, framerates of 12...50 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 144×176 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 amplitude and distance data as one *frame* acquired by a TOF camera.
The registration of this data with simultaneously acquired color data of a standard 2-D camera will be described in the following section. Furthermore, the registration of consecutively acquired 3-D surface reconstructions is described. 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 2-D camera at position (i, j) is denoted with $p_{i,j}$. TOF camera and 2-D 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 2-D camera and denoted with a $'$ for the TOF camera.
Using homogenous coordinates (indicated by $\underline{\ }$) a 3-D point q is projected to the 2-D camera pixel p, according to

$$p = K[R|t]\underline{q} \tag{2}$$

and to the TOF camera pixel p'

$$p' = K'[R'|t']\underline{q}, \tag{3}$$

where K is the calibration matrix containing the intrinsic camera parameters [3].
Let $(p_x,\ p_y)$ and $(p'_x,\ p'_y)$ denote the physical dimensions of a sensor element of the 2-D 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 $q = (q_x, q_y, q_z)^T$ is computed by

$$q_x = \frac{d_x d_{i,j}}{d_z},\ q_y = \frac{d_y d_{i,j}}{d_z},\ q_z = \frac{f' d_{i,j}}{d_z}. \tag{4}$$

As the 3-D coordinates are associated with pixels which are arranged in a regular 2-D grid in the image plane an 3-D triangulation of the 3-D point cloud can easily be achieved. The 2-D pixel grid is triangulated and than the triangulation is transferred to the 3-D points leading to a regular triangulation. Thus, for each frame acquired with a TOF camera a consistent triangulation of the reconstructed surface can be provided. Assuming the 2-D and TOF camera being rigidly mounted

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

$$R_r = RR'^{-1},\ t_r = t - R_r t', \tag{5}$$

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

3 Calibration

Two cameras are to be calibrated: the 2-D and the TOF camera. Tsai's widely used algorithm [3] 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 intrinsic (K) and extrinsic $(R,\ t)$ camera parameters involving Levenberg-Marquardt non-linear optimization [4].

For the 2-D camera the algorithm needs no modification. Considering TOF cameras the low lateral resolution and the relative big size of the sensor elements (approx. 0.04 mm) lead to unstable calibration results. By involving all the capabilities of a TOF camera the results can be stabilized by not only minimizing the squared backprojection error

$$\sum_{i=1}^{N} \|c_i - proj(w_i, K', R', t')\|^2, \tag{6}$$

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'. Instead, the objective function 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 objective function, which is minimized for a TOF camera in the calibration routine, is

$$\sum_{i=1}^{N} (\|c_i - proj(w_i, K', R', t')\| + \alpha\|\hat{c_i} - w_i\| + \beta d(\hat{c_i}, \epsilon_c)), \tag{7}$$

where $d(\hat{c_i}, \epsilon_c)$ is the distance of $\hat{c_i}$ to the regression plane ϵ_c and $\alpha,\ \beta$ are scaling parameters. The term $\|\hat{c_i} - w_i\|$ penalizes wrong intrinsic and extrinsic camera parameters which lead to a wrong reconstruction of the calibration points. The term $d(\hat{c_i}, \epsilon_c)$ 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 Registration

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 (5) and (assuming a parallel acquisition of data) a computationally inexpensive assigning of 3-D points to color information: by using formula (4) a 3-D point can be reconstructed which is specified in TOF camera coordinate system and by applying formulas (5) and (2) the 3-D point is projected into the 2-D camera image plane.

5 Surface reconstruction

The registration of consecutively acquired frames of a TOF camera enables the 3-D reconstruction of surfaces greater than the current field of view. Assuming a static scene and considering the framerate of $\geq$12 fps which implies only relatively small camera movement between consecutive frames, two reconstructions r_i and r_j (with TOF camera poses (R'_i, t'_i) and (R'_j, t'_j)) are related via a rotation $R'_{i,j}$ and translation $t'_{i,j}$ of the optical centre of the TOF camera (assuming constant intrinsic parameters). This relative pose between two reconstruction can be estimated by the following steps:

- 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. The current extrinsic camera parameters are estimated by maximizing the correlation coefficient between edges detected in the current frame and edges already detected in preceding frames. Formally: Apply a Levenberg-Marquardt non-linear optimization on (R'_i, t'_i) to maximize
  $$\rho(proj(W, K', R'_i, t'_i), W_{cur}), \tag{8}$$
  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.
  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 $t'_{i-1,i} = t'_i - R'_{i-1,i} t'_{i-1}$.

In each optimization step the world description data is projected into the current image plane. Thus, regions in the current image plane which contain information about parts of the scene which have already been reconstructed can be identified and omitted for an update of the visualization.

Only parts of the scene which have not been seen before are added to the 3-D reconstruction and so the unnecessary accumulation of redundant 3-D information is circumvented.
Clearly, the world description data is growing. And for each optimization step the world description data has to be processed. Consequently, the estimation of the current camera pose gets slower. As the complexity of the estimation task is only depending on the size of the world description data, the world description data might be thinned out when its size exceeds a certain value to meet processing time requirements. This issue is currently out of the scope of this paper.
A global geometry is only available at the 3-D point level. A global triangulation of the scene is not provided by this algorithm. Nevertheless, as stated earlier each frame provides a consistent triangulation of the scene geometry in the current field of view. Each part which is added to the reconstructed scene after a successful estimation of the new extrinsic parameters is consistently triangulated. Thus, by stitching the scene together from consistently triangulated parts, at least the visual effect is that of a closed surface.

6 Results

To evaluate the calibration routine for TOF cameras in terms of its stability the following experiments were done with a SwissRanger 3100 (144×176 piyels). First, the TOF camera was not moved while acquiring images and performing the calibration routine 16 times. Due to the noise in the amplitude information the segmentation of the calibration circles and consequently the input data for the calibration routine do differ. Furthermore, the low lateral resolution and the relatively big dimensions of the TOF sensor elements do hinder a good convergence of the estimation of the camera parameters when minimizing the standard objective function given in equation 6. As the camera was not moved the standard deviation of the calculated intrinsic and extrinsic camera parameters characterize the influence of the effects described. The results are given in table 1.

Objective function		f[mm]	c_x[pix]	c_y[pix]	t_x[mm]	t_y[mm]	t_z[mm]	r_x[rad]	r_y[rad]	r_z[rad]
standard (eq. 6)	μ	8.37	87.855	69.882	-52.69	-55.71	284.40	0.128	0.280	-0.014
	σ	0.057	1.032	0.886	1.87	1.19	1.88	0.003	0.006	0.001
extended (eq. 7)	μ	8.21	86.68	69.33	-53.78	-54.93	279.80	0.126	0.277	-0.014
	σ	0.067	0.382	0.256	1.08	0.34	2.13	0.001	0.003	0.0006

Table 1: Mean and standard deviation of intrinsic and extrinsic camera parameters after 16 calibration routines with a non-moving TOF camera.

The same experiment was repeated but the camera was moved. In addition to the effects described previously, the occuring varying illumination of the calibration pattern leads to a destabilization of the segmentation of the calibration circles. 16 calibration routines were performed. The results are given in table 2. The experiments were repeated with a 2-D camera: a Logitech QuickCam Messenger USB (324× 248 pixel) was used to perform the calibration routine 16 times with non-moving camera. The results are given in table 3

The results indicate that the relative position of both cameras can be estimated with an accuracy of about 5mm in each translation component and an accuracy of 0.014 rad in each rotation component. This accuracy is sufficient to provide an reasonable overlay of 3-D information with color information. Figure 2 shows two simultaneously acquired TOF and Web camera images

Objective function		f[mm]	c_x[pix]	c_y[pix]
standard (eq. 6)	μ	7.95	87.681	70.515
	σ	0.528	0.874	0.830
extended (eq. 7)	μ	8.20	86.19	68.84
	σ	0.168	0.910	0.903

Table 2: Mean and standard deviation of intrinsic and extrinsic camera parameters after 16 calibration routines with a moving TOF camera.

	f[mm]	c_x[pix]	c_y[pix]	t_x[mm]	t_y[mm]	t_z[mm]	r_x[rad]	r_y[rad]	r_z[rad]
μ	3.93	124.43	152.23	-37.06	-72.08	246.95	0.282	-0.421	-0.264
σ	0.027	5.025	0.850	3.086	0.516	1.399	0.006	0.011	0.003

Table 3: Mean and standard deviation of intrinsic and extrinsic camera parameters after 16 calibration routines with a non-moving Web camera.

and the result of projection the 3-D surface reconstruction into the web camera image plane after registrating the optical centres of the cameras. Due to the hardware setup and camera characteristics the common field of view does not completely overlap. When no color information could be assigned to a 3-D surface point the corresponding amplitude gray value was used as texture.
To illustrate the reconstruction of scenes greater than the current field of view by estimating the current camera pose an image sequence was processed. The scene contained a chair. The inital viewing direction of the camera was to the left of the chair. Then the camera was moved to the left, making the chair move through the field ov view from right to left 1.

7 Conclusion and Outlook

Algorithms for the calibration of TOF cameras, the registration of TOF and standard cameras and the reconstruction of surfaces greater than the current field of view were proposed. By extending the standard objective function minimized during calibration a calibration routine especially suitable for TOF cameras was described. Being able to reliably calibrate a TOF camera the registration with standard cameras is enables as the extrinsic camera parameters for simultaneously acquired images are known. By reducing the data used for estimating the extrinsic parameters for consecutively acquired frames an on-the-fly registration of these frames is possible. The algorithm proposed assumes a static scene.
Considering the registration of standard and TOF cameras the overlay of color and 3-D information may be improoved by modelling the mismatch of color and 3-D information as a non-rigid 2-D/2-D multimodal registration problem: The color information has to be aligned with 2-D amplitude/distance data. Several non-linear effects (inaccuracies in distance measurement; lens distortions) motivate this approach. Furthermore, the identification of dynamic parts of the scene will lead to the reconstruction of dynamic scenes which are greater than the current field of view.

References

1. Lange, R.: 3D Time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. PhD thesis, University of Siegen (2000)

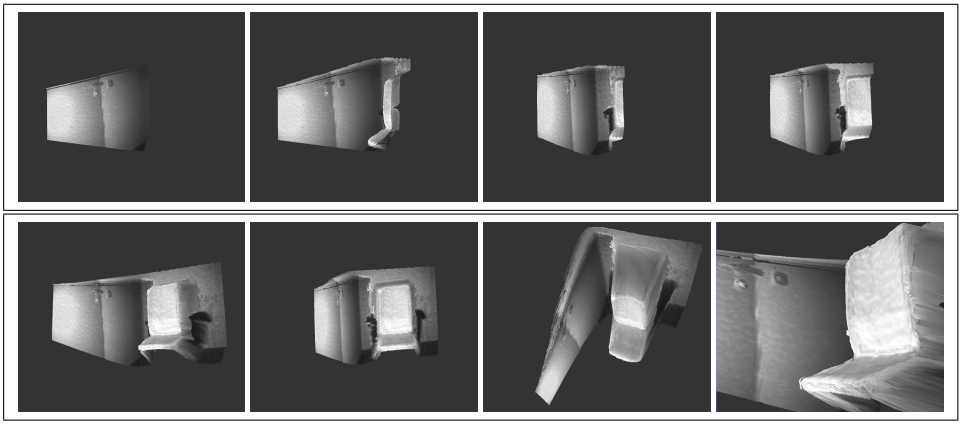

Fig. 1: Reconstructin of a scene containing a chair: the upper row of images shows the consecutive stitching of reconstructed parts of the scene by estimating the camera pose for each acquired frame. The bottom row shows different screenshots of the examination of the scenes. Note that the chair has been stitched together very well.

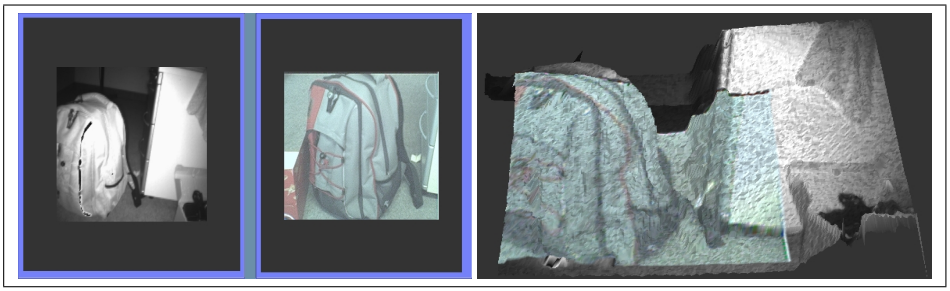

Fig. 2: Simultaneously acquired TOF and Web camera data (left image). Overlay of color information and 3-D surface reconstruction after registrating TOF and Web camera (right image). Only parts of the reconstruction can be overlayed with color information as the common field of view of both cameras is not 100% overlapping.

2. Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T., Xu, Z., Hartmann, K. In: Handbook of Computer Vision and Applications. Volume 1. The Academic Press (1999)
3. Tsai, R.Y.: 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) (1987) 323–344
4. Dennis, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice Hall, New Jersey (1983)