CALIBRATION AND REGISTRATION FRAMEWORK FOR 3D RECONSTRUCTION OF THE KIRCHE SEEFELD

T. Abmayr[1] **, F. Härtl**[1]**, G. Hirzinger**[2]**, D. Burschka**[3] **and C. Fröhlich**[1]

[1] Zoller+Fröhlich GmbH, Wangen i.A., Germany,
[2] German Aerospace Center, Oberpfaffenhofen, Germany
[3] Technical University of Munich, München, Germany
t.abmayr@zofre.de

KEY WORDS: Terrestrial Laser Scanning, Range Data, Color Scanning, Sensor Calibration, Registration, Data Fusion

ABSTRACT:

In our work we are interested in 'Modeling the Reality': Scenes are reconstructed in a combination of highly accurate geometry and real radiometric textures. Thereby, a common problem is the difficulty to gain colored panoramic 3D data which are highly accurate and have high resolution. Aligned color and range data is the basis for continuative modeling tasks, e.g. textured surface generation, 'as built' models, 3D object recognition and other virtual reality applications.
We present a multi-sensor calibration and registration framework that fuses highly accurate and robust colored panoramic 2.5D data. To achieve this, we use a panoramic 2.5D laser scanner in combination with a 2D array camera. To simplify the alignment between multiple viewpoints, we additionally make use of an electronic spirit-level, which is integrated in the scanner system. The applicability of our system is validated on a historical site called 'Kirche Seefeld'.
In summary, we present a highly accurate multi sensor system including robust methods for the calibration and registration of the data streams and its fusion to a dense 2.5D point cloud.

1 INTRODUCTION

A common problem in 'Modeling the Reality' is the difficulty to gain colored panoramic 3D data, that is both, highly accurate and has high resolution. Aligned color and range data is the basis for continuative modeling tasks, e.g. textured surface generation, 'as built' models, 3D object recognition and other virtual reality applications.

Laser scanning systems, that are currently available, often miss a panoramic option in either camera or scanner or have disadvantages in data quality and resolution of the color data stream. Contrarily, camera only based systems lack in efficiency (as much more viewpoints are needed), robustness (if the scene is "poorly" textured) and have high computational costs in the postprocessing. Furthermore, the 3D reconstruction could even fail if there is no texture. Consequently, these approaches are not applicable in some use cases.

To avoid those disadvantages, we designed a system, which consists of a panoramic laserscanner, an array camera, and an electronic spirit-level to yield highly accurate and dense 3D point clouds. To achieve the same visual field of view, the camera is attached to a vertical tilt unit and mounted on the scanner device. Such a system combines the advantages of both sensors: the direct range information of the scanner delivers a highly accurate 2.5D point cloud of the environment, whereas the color data of the camera highly improves visualization, resolution and interpretation of the data. By aligning different viewpoints, range data can be more easily combined to dense 3D models than single 2D camera data only. The electronic spirit-level additionally stabilizes the alignment between multiple viewpoints and makes the approach work robustly in a variety of different field scenarios. Finally from a measurement point of view, the combination of both sensors ideally complements each other: Camera sensors tend to blur edges less than scanners.

Figure 1: The Z+F IMAGER 5006 with color camera

Our multi sensor approach enables efficient, accurate and robust acquisition of colored panoramic 2.5D data and their alignment to dense 3D point clouds.

2 HARDWARE AND ACQUISITION PROCESS

We start with a short description of the mechanical principle of the scanner and the acquisition setup and introduce some important definitions and notations.

2.1 Hardware

The scanner we use (Z+F Imager 5006) consists of a range measurement system in combination with a mirror deflection device (see Fig. 1). The deflection system points the laser beam into the direction of measurement, the laser beam is emitted and the reflected laser light is detected. A 3D scan is acquired by two rotations: First the mirror rotates around a horizontal axis ("elevation rotation axis") and thus deflects the laser beam in a vertical direction. The second rotation is around the vertical center axis ("azimuth rotation axis") of the system (see Fig. 1 and Fig. 3). The actual direction of the laser beam is measured by two encoders: The first describes the actual horizontal rotation and is adjusted at the center axis (azimuth encoder). The second encoder describes the mirror rotation and is adjusted along the mirror rotation axis. The zero position of this second encoder is located along the negative direction of the center axis (elevation encoder). The field of view of the scanner is 360^o (azimuth) and 320^o (elevation). In

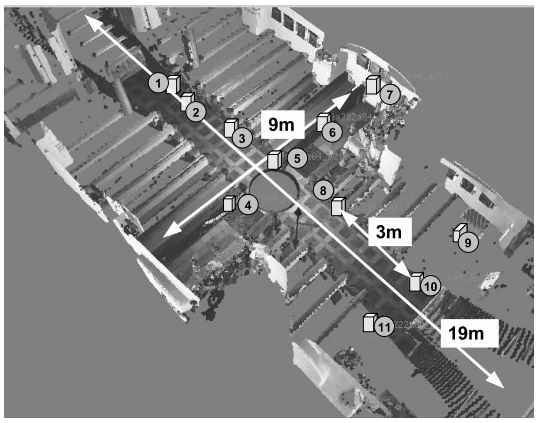

Figure 2: Viewpoints of the 'Kirche Seefeld' project

addition to range information the device also measures the reflectivity of the object-surface giving a photo like impression of the scanned area. There is a one-to-one correspondence between the reflectance and the range value of each sample with respect to the corresponding azimuth and elevation angles.

The scanning mode for our experiments has a spatial point distance of 6.4mm in 10m (1.59mrad/pixel), hence the resolution for a 360^0 scan is 10000 x 5000 pixel. The data acquisition is very fast: For a 360^0 panoramic scan, it takes an overall acquisition time of 3.22 min with a sampling rate of 250.000 Hz (see (Zoller + Fröhlich, (visited September 2008)) for more hardware specifications).

To capture color data, we use an industrial camera with a resolution of 1900 pixel horizontal and 2500 pixel vertical. By using a 4.8mm object lens, the field-of-view is 60 deg. To achieve the same visual field of view as with the scanner, a vertical tilt unit was mounted on the scanning device with the camera attached to it.

Finally, we make use of the electronic spirit-level to compensate the horizontal tilt of the system. It is mounted on the floor section of the scanner and consists of two sensors heading orthogonal to each other, indicating two angles of tilt. The domain of the level is ± 0.5 deg horizontal tilt.

2.2 Data-Acquisition

In order to acquire high resolution, well exposed pictures, data is not collected 'on the flight' in parallel with the scan, but consecutively after the scan. Overlapping color images are taken by rotating the system by predefined angle increments around the azimuth rotation axis of the scanner and vertically around the tilt unit.

In our measurement concept, the scanner as well as the intrinsic camera parameters are assumed to be stable for a long period of time, and can therefore be calibrated in the laboratory. The tilt-unit and the camera can be removed from the scanner facilitating transportation. Hence, the extrinsic camera parameters have to be calibrated after reassembly with the scanner.

Our test site was a church ('Kirche Seefeld') of approximately 9x19x10 m^3 size (see Fig. 2). In this project, the task was to acquire colored 3D data with a spatial point distance of less than 10mm in 10m (on the roof) and of less than 5mm on the horizon. To achieve this, we captured scan-data and image-data from twelve different viewpoints. The maximal distance between two viewpoints was approximately 12m. The maximal difference between neighboring viewpoints was approximately 3m. No targets were available within this site.

The church has gilded figures and some paintings and frets on the wall. Many areas exist with relatively "bad" texture. The illumination conditions differed between bright sunlight and dark regions. This resulted in several challenges, especially for the camera-calibration: The gray-value characteristics between scanner and camera sometimes differed a lot.

In order to describe the mapping between the scanner and the camera, a number of different coordinate systems are essential and are introduced in the next section.

3 COORDINATE SYSTEMS

An affine, orthogonal and right-handed coordinate system with the translation a and the basis vectors e_i is denoted throughout this paper by $K := (a, e_1, e_2, e_3)$. Based on this notation, we introduce the following coordinate systems: First, the ***scanner coordinate system*** (our reference coordinate system, see Fig. 3). Let $k_1 \in \mathbb{R}^3$, $\|k_1\| = 1$ be the elevation axis, and respectively take $k_3 \in \mathbb{R}^3$, $\|k_3\| = 1$ for the azimuth rotation axis. If we assume for now, that both axis intersect and $\langle k_1, k_3 \rangle = 0$, we can construct an affine, orthogonal and right-handed coordinate system

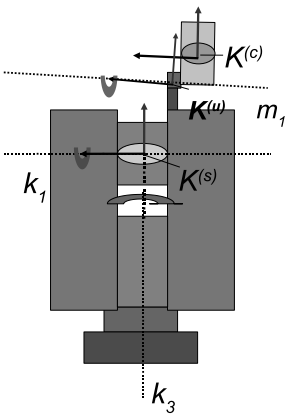

Figure 3: Coordinate systems and rotation axis

$$K^{(s)} := (0, k_1, k_2, k_3). \quad (1)$$

A second coordinate system

$$K^{(c)} := (w_0, w_1, w_2, w_3) \quad (2)$$

which describes the position of the camera with respect to the scanner is called ***camera coordinate system***. If the camera can be described as an ideal pinhole camera, then the coordinate system holds the properties:

- The origin w_0 is equal to the optical center of the camera.
- The third vector w_3 is orthogonal to the image plane.
- w_1 is parallel to the horizontal border of the image plane, and w_2 respectively to its vertical border.

Based on this we introduce a third coordinate system

$$K^{(u)} := (m_0, m_1, m_2, m_3), \tag{3}$$

which we call ***camera-tilt-unit coordinate system***. This coordinate system has the following properties:

- The origin is equal to the center of the rotation axis of the tilt-unit.
- The direction of the first vector m_1 is along the direction of the rotation axis.
- The direction of m_3 approximately corresponds to the direction of the z-axis k_3 of the scanner coordinate system $K^{(s)}$.

So far, we assumed several idealized properties for the geometry of the scanner as well as for the camera, e.g. orthogonal rotation axis. However, the real sensors deviate from these idealized models. Next, we demonstrate how these models can be transformed to apply to real sensors and how the relation between the coordinate systems can be determined.

4 MODELING THE SCANNER

Starting with the scanner, we first notice the similarity to theodolites for the mechanical set-up. In this section, we present a simple sensor model that is the basic model for approaches most commonly used in the literature (Lichti, 2007), (Deumlich and Staiger, 2002).

Let l be the azimuth encoder angle, h be the elevation encoder angle and rg the range. Then the overall transformation $\Phi(l,h,rg) : [0,\pi[^2 \times \mathbb{R} \rightarrow \mathbb{R}^3$ from un-calibrated encoder coordinate system into the scanner coordinate system $K^{(s)}$ is defined through

$$\Phi(l,h,rg) = H \circ \Upsilon(l,h,rg) \tag{4}$$

with

$$\Upsilon(l,h,rg) = \left(l + \text{sign}(h-\pi)\left(\tfrac{b}{\sin(h)} + \tfrac{a}{\tan(h)}\right), h + c, rg\right) + \eta(l,h,rg). \tag{5}$$

H describes the transformation from spherical - to Cartesian coordinates and is well known. Υ_1 corrects the non-orthogonality between the elevation rotation axis and the azimuth rotation axis as well as the non orthogonality between the laser beam and the elevation rotation axis. In photogrammetry these errors are called **trunnion axis error** and **collimation axis error** respectively. As described in Section 2, the zero position of the elevation encoder must be equal to the negative horizontal rotation axis k_1. As this is usually not the case in real sensors due to mechanical inaccuracies, it has to be corrected by a constant vertical angle offset, which is described through Υ_2. This error is called **vertical circle index error**. The unknowns $a,b,c \in \mathbb{R}$ can be determined as e.g. described in (Abmayr et al., 2005).

η is a term, which summarizes all additional calibration errors. As the focus is onto the calibration of the camera, we cannot go more into detail throughout this paper and refer to the literature (Lichti, 2007), (Rietdorf, 2005).

5 MODEL OF THE CAMERA AND THE CAMERA-TILT-UNIT

In this section we show how we combine a high resolution array camera with a panoramic laserscanner (see also (Abmayr et al., 2008a)). To achieve the same visual field of view, the camera is attached to a vertical tilt unit (see Fig. 3). Once the tilt unit and the camera are mounted on the scanner device and the relation between the sensors is calculated, both sensors can be regarded as one single system. The angle increments of the vertical tilt unit and the horizontal rotation of the scanner are highly accurate and are used as fixed input parameters. This reduces the degrees of freedom for the external camera parameters enormously.

5.1 Sensor Model

As the rotations of the scanner and the tilt-unit are highly accurate, we use these rotation angles as fixed input parameters for our model. For modeling the intrinsic parameters of the camera we use Tsai's camera model (Tsai, 1987), which is based on the pinhole camera of perspective projection, and is well known in computer vision. If we denote the horizontal rotation with the angle α around the z-axis of the scanner with Z_α and the vertical rotation with the angle β around the x-axis of the tilt-unit with X_β then the overall projection $\Xi_{\alpha,\beta} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ from a point $X := (x,y,z)$ of the scanner coordinate system $K^{(s)}$ onto the pixel (u,v) in the color image can be written as

$$\Xi_{\alpha,\beta}(X) = \varphi_{\kappa,c_x,c_y,s} \circ \pi_f \circ T_{\alpha,\beta} \tag{6}$$

with

$$T_{\alpha,\beta} := M X_\beta \widetilde{M} Z_\alpha. \tag{7}$$

$T_{\alpha,\beta}$ defines first the transformation from the *scanner coordinate system* $K^{(s)}$ into the *camera-tilt-unit coordinate system* $K^{(u)}$ and then the transformation from $K^{(u)}$ into the *camera coordinate system* $K^{(c)}$. The perspective projection onto the image plane with the *focal length* f is described through $\pi_f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and the mapping from undistorted coordinates to distorted image coordinates with the *principle point* (c_x, c_y), the parameter κ describing the *1st order radial lens distortion* and the *uncertainty scale factor* s is defined through $\varphi_{\kappa,c_x,c_y,s}\mathbb{R}^2 \rightarrow \mathbb{R}^2$. For detailed information see (Tsai, 1987).

According to the setup, the rotation angles α and β define the actual position of the azimuth encoder of the scanner and the rotation angle of the tilt unit and are assumed to be known. By considering that M and $\widetilde{M}$ are homogenous matrices and thus can be described by 6 parameters each, we get altogether 17 unknowns.

Our calibration approach for solving the external camera parameters is based on the following properties.

5.2 Properties

1.) If the vertical angle β is fixed, then the camera rotates on a circular path around the z-axis k_3 of the scanner: To show this, take any fixed $\beta_0 \in [-\pi/2, \pi/2[$ and $\alpha_0 \in [0, 2\pi[$. Then (7) holds for all $\alpha \in [0, 2\pi[$

$$T_{\alpha,\beta_0} = T_{\alpha,\beta_0} Z_{\alpha-\alpha_0} \tag{8}$$

and therefore

$$\Xi_{\alpha,\beta}(x) = \Xi_{\alpha_0,\beta}(Z_{\alpha-\alpha_0}x). \tag{9}$$

Proof With (7) we get $T_{\alpha,\beta_0} = MX_{\beta_0}\widetilde{M}Z_\alpha = MX_{\beta_0}\widetilde{M}Z_{\alpha_0}Z_{\alpha-\alpha_0} = T_{\alpha,\beta_0}Z_{\alpha-\alpha_0}$. This results with (6) in $\Xi_{\alpha,\beta_0}(x) = \Xi_{\alpha_0,\beta}(Z_{\alpha-\alpha_0}x)$.

□

If the vertical angle β_0 is fixed, then implies property (9) that corresponding points between these images and the scanner can be transformed into one single camera image. This is very convenient for the calibration as it reduces the minimal number of necessary corresponding points.

2.) If the vertical angle β_0 is fixed and one single transformation between camera and scanner is known, then any transformation along this circular path also is known. This follows directly from property (8).

3.) Let the notation be as introduced above and T_{α_0,β_0}, T_{α_0,β_1} $(\beta_0 \neq \beta_1)$ be two transformations as defined in (7). Set

$$Y_\alpha := \widetilde{M}Z_\alpha\widetilde{M}^{-1} \tag{10}$$

and

$$\begin{aligned} A &:= (X_{\beta_1}Y_{\alpha_0})(X_{\beta_0}Y_{\alpha_0})^{-1} \\ B &:= T_{\alpha_0,\beta_1}T_{\alpha_0,\beta_1}^{-1} \\ X &:= \widetilde{M}^{-1}. \end{aligned} \tag{11}$$

Then

$$T_{\alpha,\beta} = MX_\beta Y_\alpha \widetilde{M} \tag{12}$$

and

$$AX = XB. \tag{13}$$

This representation is called ***AX=XB representation of the extrinsic camera parameters***.

Proof To show (12) we use (10) and simplify (7) to

$$T_{\alpha,\beta} = MX_\beta\widetilde{M}Z_\alpha = MX_\beta\widetilde{M}Z_\alpha(\widetilde{M}^{-1}\widetilde{M}) =$$

$$MX_\beta(\widetilde{M}Z_\alpha\widetilde{M}^{-1})\widetilde{M} = MX_\beta Y_\alpha\widetilde{M}.$$

To show (13) we notice that both transformations T_{α_0,β_0} and T_{α_0,β_1} must hold

$$\begin{aligned} T_{\alpha_0,\beta_0} &= MX_{\beta_0}Y_{\alpha_0}\widetilde{M} \\ T_{\alpha_0,\beta_1} &= MX_{\beta_1}Y_{\alpha_0}\widetilde{M}. \end{aligned} \tag{14}$$

Solving (14) for $\widetilde{M}$ leads to

$$\underbrace{(X_{\beta_1}Y_{\alpha_0})(X_{\beta_0}Y_{\alpha_0})^{-1}}_{A} \quad \underbrace{M^{-1}}_{X} \quad = \quad \underbrace{M^{-1}}_{X} \quad \underbrace{T_{\alpha_0,\beta_1}T_{\alpha_0,\beta_0}^{-1}}_{B} \tag{15}$$

□

(13) shows that the extrinsic camera parameters can be transformed into a '*AX=XB*' structure. How to solve AX=XB problems, however, is well known in the literature (see e.g. (Park and Martin, 1994))

5.3 Algorithm Outline

Let us assume that we have $P := \{(P_1,p_1),..,(P_n,p_n)\}$,$P_i \in \mathbb{R}^3, p_i \in \mathbb{R}^2$ corresponding points between scan data and images taken from the same vertical encoder position β_0 and the scanner. Let us further assume that we have $Q := \{(Q_1,q_1),..,(Q_n,q_n)\}$,$Q_i \in \mathbb{R}^3, q_i \in \mathbb{R}^2$ corresponding points between scan data and images taken from a second vertical encoder position $\beta_1, (\beta_0 \neq \beta_1)$ and the scanner.

1.) *(Full Calibration - 19DoF)* If the 7 internal and 12 external calibration parameters are unknown, then all parameters can be solved for $\mid P \mid > 10$ and $\mid Q \mid > 10$ points by applying Tsai's calibration approach in combination with property 1.-3. from Section 5.2.

2.) *(Full External Calibration - 12DoF)* If the intrinsic camera parameters are known, then all 12 external registration parameters can be solved for $\mid P \mid > 7$ and $\mid Q \mid > 7$ points in closed form solution by applying property 1.-3. from Section 5.2.

3.) (*Update Procedure - 3DoF*) If the parameters were already calibrated once and the camera and the tilt-unit are reattached on the scanner as a single unit then only M has changed. Furthermore, if both devices are adjusted on the same position, then only the rotation in M has to be recalculated. This can be done for $\mid P \mid + \mid Q \mid > 2$ points by a non-linear least squares approach.

6 FEATURE EXTRACTION AND MATCHING FOR EXTRINSIC CAMERA CALIBRATION

In our approach, the scanner-calibration as well as the calibration of the intrinsic camera parameters depends on the highly accurate detection of artificial landmarks and is vision-based (Abmayr et al., 2008b). As both systems are stable for a long period of time, we perform their calibration in the laboratory. In contrast, the use of such targets in the field is time consuming and inefficient. Therefore, we apply natural landmarks for the calibration of the extrinsic camera parameters.

6.1 Feature Extraction

Let us assume that we reattached the camera on the scanner and hence let the mapping between a camera image I and the scan S be approximately known. Our registration approaches between camera and scan-data is based on corresponding points. Thereby, we only focus on those points, which are significant in their gray-value characteristics. This task is well known in computer vision (see e.g. Harris (Harris and Stephens, 1988)) and photogrammetry (see e.g. Förstner (Förstner and Gülch, 1987)). These sets of feature points extracted in each image and the scan is the input-data for our matching approach. This is introduced in the next section.

6.2 Feature Matching

Definitions Given the scan S and the set of pixel indices in the region around (m,n) through

$$M := \{(m-k,n-l),..,(m+k,n+l)\}. \tag{16}$$

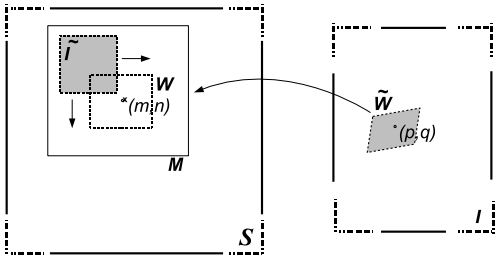

Figure 4: Scheme of the matching approach between two features in S and I: The affine linear transformation maps the image patch from $\widetilde{W}$ into the view of the scan. Then this transformed patch $\widetilde{I}$ is correlated with S in the region M around (m,n).

Then, the sub-matrix $\widetilde{S} \subset S$ of the scan is defined through

$$\widetilde{S} := (S_{ij})_{(ij)\in M} \tag{17}$$

and is called ***scan-patch around*** (m,n). Further, set

$$W := ((m-p,n-q),(m-p,n+q),(m+p,n-q),(m+p,n+q)) \tag{18}$$

with $|p| < |k|$,$|q| < |l|$ (see Fig. 4). As each pixel S_{ij} in the scan is assigned to a 3D coordinate $X_{ij} \in \mathbb{R}^3$, we use (6) to transform each scan-pixel into the image I and set

$$\widetilde{W} := (\Xi_{\alpha,\beta}(X_{W_0}), \Xi_{\alpha,\beta}(X_{W_1}), \Xi_{\alpha,\beta}(X_{W_2}), \Xi_{\alpha,\beta}(X_{W_3})). \tag{19}$$

$\widetilde{W}$ is called ***region of interest in I.***

Properties The transformation between W and $\widetilde{W}$ can be solved with a linear least squares optimization approach.

Proof Set $f(x,y) := a_0x + a_1y + a_2$ and set $g(x,y) := b_0x + b_1y + b_2$. Then

$$\sum_{k\in\{0,..,3\}} \| \begin{pmatrix} f(W_{k0},W_{k1}) - \widetilde{W}_{k0} \\ g(W_{k0},W_{k1}) - \widetilde{W}_{k1} \end{pmatrix} \| \to \min \tag{20}$$

with the unknowns $(a_0,..,a_2) \in \mathbb{R}^3$ and $(b_0,..,b_2) \in \mathbb{R}^3$ can be solved with a linear least squares approach.

□

In the region around corresponding features between scan and image the transformation between both data streams is approximately affine-linear. Therefore, for a given feature point (m,n) in the scan, we use (20) to transform image I into the view of the scan-patch around (m,n). We apply for all $(i,j) \in M$

$$(p,q) := (f(i,j), g(i,j))$$

and assign

$$\widetilde{I}_{ij} := I_{p,q}. \tag{21}$$

Then $\widetilde{I}$ is called ***into S transformed patch of I*** [1].

[1] (21) maps in general in between adjacent pixel of the image I. Therefore, interpolation (e.g. bilinear interpolation) or approximation techniques for reconstructing the image between adjacent pixel should be used to improve the quality of the transformed image.

As the mapping between camera and scan only is known approximately, the positions between $\widetilde{I}$ and $\widetilde{S}$ also only correspond approximately. To find the position of best match between $\widetilde{S}$ and $\widetilde{I}$, we apply a well known correlation-based approach (see e.g. (Hirschmüller and Scharstein, 2008 (accepted for publication))) to measure the similarity between both patches. We calculate the derivative of the input images $\widetilde{S}$ and $\widetilde{I}$. For $f \in \{\widetilde{S},\widetilde{I}\}$ set

$$\Delta^{(x)}_{i,j}(f) := \tfrac{1}{2}(f_{i-1,j} - f_{i+1,j}) \quad \text{and} \quad \Delta^{(y)}_{i,j}(f) := \tfrac{1}{2}(f_{i,j-1} - f_{i,j+1})$$

and define the discrete derivative at position (i,j) through

$$\nabla_f(i,j) := \sqrt{\Delta^{(x)}_{i,j}(f)^2 + \Delta^{(y)}_{i,j}(f)^2}. \tag{22}$$

The ***gradient-based quality measure*** is then defined through

$$\widetilde{C}_{ij} := \frac{\sum\limits_{(k,l)\in M} \nabla_{i-k,j-l}(S)\nabla_{k,l}(\widetilde{I})}{\sqrt{\sum\limits_{(k,l)\in M} \nabla_{i-k,j-l}(S)^2 \sum\limits_{(k,l)\in M} \nabla_{k,l}(\widetilde{I})^2}}. \tag{23}$$

Based on this definition for quality measure, we introduce our matching approach for camera vs. scan data.

Algorithm Outline The matching procedure can be divided in several steps:

1. *Feature Extraction:* We start with the feature extraction method introduced in Section 6.1. Denote the set of features-positions of the scan with F_S and of the image with F_I.

2. *Affine-linear Transformation:* Between any two pairs of features-positions $(m,n) \in F_S$ and $(p,q) \in F_I$ apply (21) to transform I in the region around (p,q) into the perspective view of S. We denote this transformed image with $\widetilde{I}$. Further, we apply (17) to define the scan-patch and denote this patch with $\widetilde{S}$.

3. *Feature Image Generation:* To find the best match position between $\widetilde{S}$ and $\widetilde{I}$, we slide $\widetilde{I}$ over $\widetilde{S}$ (see Fig. 4) and apply for each position (i,j) the quality measure introduced in (23). We store the result in the feature image C.

4. *Calculate Best Match Position:* The feature image C will be maximal at the index (r,s) where $\widetilde{I}$ and $\widetilde{S}$ match best (with respect to the used quality criterion) and thus the position (r,s) of best match is calculated through

   $$C_{r,s} := \max(..,C_{ij},..). \tag{24}$$

   If $C_{r,s}$ is larger than a predefined threshold, then count the feature-pair as match and add it to the set Q of matches, otherwise reject it.

5. *Back-Projection into I:* Consequently, the corresponding position in I that matches best to (r,s) is calculated through

   $$(\widetilde{m},\widetilde{n}) := \Xi_{\alpha,\beta}(X_{r,s})$$

   by using (6).

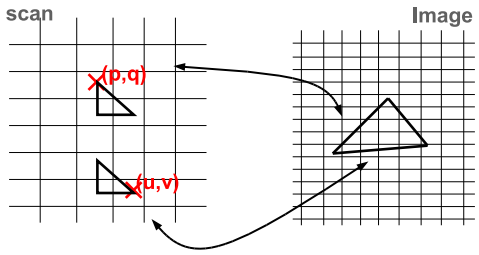

Figure 5: Two triangle-neighborhoods in the scan map to the same region in the image. Color can be assigned only to the neighborhood around the pixel which is closer to the camera origin.

6. After step 2.)-5.) was applied on all combinations of feature-pairs return the set Q of matches.

The approach we present results - depending on the parameters to adjust - in a high amount of suggested matches. On the other hand, it is obvious that these suggested matches contain a high percentage of false candidates. To detect these outliers we use the Ransac (Random Sample Consensus) method (Fischler and Bolles, 1981): Randomly, the minimal set of points needed to calculate the transformation between the images is extracted and all candidates are tested for consensus. This is repeatedly done until a break criteria is reached. The set with the largest consensus is selected for registration. Again we refer to the literature for more information.

7 DATA FUSION

Now all sensors are assumed to be calibrated and the relation among them is known. Since this paper mainly focuses on the registration framework, we only give the basic ideas of the fusion process. The basic procedure is rather straightforward, as the mapping from a scanpixel (u,v) to image coordinates (m,n) is known. As introduced in Section 2.1 each scanpixel (u,v) corresponds to encoder increments (h,l) and a range value rg. With (4) and (6) the mapping is defined through

$$(m,n) = \Xi_{\alpha,\beta} \circ \Phi(l,h,rg). \tag{25}$$

This procedure is called *Color Scan Generation*. The only problem which occurs however, is to take care of the occlusions due to parallax and the scene structure: mapping between scan and images in general is not one-to-one, and so mismatches occur, if objects are in the field of view of the scanner but not of the camera (see Fig. 5). Such z-buffer problems are well known in the literature. In (Abmayr et al., 2008a), we gave a short summary of the problem in the context of 2D camera-images and 3D-laser-scans.

Finally, to get a homogenous color crossover in overlapping image regions, a simple blending technique is used. Again, we refer to the literature (see (Abmayr et al., 2005)).

a.)

Statistics bw Targets vs. Ground Truth after Calibration				
IMAGER 5006	Nb Targets	mean proj. 10m	rms elevat.	rms azimuth
vs. Ground Truth	86	1.21 mm	0.10 mrad	0.11 mrad

b.)

Accuracy Test of Electronic Level					
roll	*pitch*	*Mean proj. 10m (4DoF)*	*roll*	*pitch*	*Mean proj. 10m (4DoF)*
≈ -0.5	$\approx +0.0$	1.67 mm	$\approx +0.0$	≈ -0.5	1.51 mm
≈ -0.3	$\approx +0.0$	1.43 mm	$\approx +0.0$	≈ -0.3	1.51 mm
≈ -0.1	$\approx +0.0$	1.46 mm	$\approx +0.0$	≈ -0.1	1.45 mm
$\approx +0.0$	$\approx +0.0$	1.63 mm	$\approx +0.0$	$\approx +0.0$	1.63 mm
$\approx +0.1$	$\approx +0.0$	1.41 mm	$\approx +0.0$	$\approx +0.1$	1.38 mm
$\approx +0.3$	$\approx +0.0$	1.41 mm	$\approx +0.0$	$\approx +0.3$	1.45 mm
$\approx +0.5$	$\approx +0.0$	1.58 mm	$\approx +0.0$	$\approx +0.5$	1.52 mm

Figure 6: Validation of the scanner accuracy: Final error between pairs of targets and ground truth data after co-registration. a.) Registration vs. ground truth with 6DoF b.) Registration after level-compensation with 4DoF

8 EXPERIMENTAL SETUP AND RESULTS

Our setup for intrinsic camera calibration allows adjustment in a laboratory. For this purpose, we assembled over 80 targets in a calibration-lab. The targets were adjusted in varying positions in relation to the system and were evenly distributed with respect to its horizontal and vertical angles. For cross-checking our results vs. a ground truth, we additionally measured the 3D positions of these targets precisely by using a highly accurate total station.

Validation of Scanner Accuracy In this test, we cross-checked the scanner calibration vs. the ground truth of the calibration lab. Our statistical measures are the remaining azimuth and elevation error (rms and mean) between pairs of points after the co-registration. This errors are shown in Fig. 6 (a). Furthermore, we calculate the remaining distance between these pairs: To norm this error, all pairs of points are projected to a 10m sphere. The small rms errors demonstrate that no parameter is over-adjusted for a certain spectrum of the field of view but that the calibration is feasible to the complete panoramic scene.

Validation of Electronic Level Accuracy Here, we cross-checked the accuracy of the electronic spirit-level. The data specification of the electronic spirit-level we used is within a domain of ± 0.5 deg (horizontal-tilt). Additional to this domain restriction, also the stability of the tripod influences the accuracy of the result. We compared results between different degrees of tilt by using the ground truth of the calibration lab: After tilting the scanner to different horizontal declines in a range of ± 0.5 deg, we extracted the 3D positions of the targets and compensated their horizontal tilt. Then, we aligned the compensated targets vs. the ground truth: However, now we restricted the registration only to a horizontal rotation and a translation and hence to 4 DoF. Fig. 6 (b) demonstrates the quality of the level-compensation: Although the registration results are worse than a full rigid motion registration with 6 DoF (compare Fig. 6 (a)), they are still highly accurate and emphasize the use of the electronic spirit-level.

a.)

Test Accuracy of Camera Calibration				
Camera	*Nb Targets*	*mean proj. 10m*	*rms elevat.*	*rms azimuth*
Calib. 19 DoF	86	3.89 mm	0.35 mrad	0.38 mrad

b.)

Test Extrinsic Camera Calibration to different Locations				
Camera	*Nb Targets*	*mean proj. 10m*	*rms elevat.*	*rms azimuth*
Calib. 3DoF	86	4.93 mm	0.46 mrad	0.57 mrad

Figure 7: Validation of camera accuracy: Final error between pairs of targets and ground truth data after co-registration. a.) Accuracy after full camera calibration b.) Accuracy after extrinsic-camera calibration in different location with 3DoF-update procedure

Kirche Seefeld	*Nb of features*	*Nb of feat. after Ransac*	*Mean Error*	*Rms Error*
Gradient	155	49	2.31 Pix	1.47 Pix

Figure 8: External camera calibration: (a) Statistical measures after outlier detection with ransac (b) Scaled error vectors of the feature extraction and matching approach

Next, we validate the results of the camera-calibration w.r.t. accuracy, robustness and generalization in field applications.

Camera Calibration Accuracy Here, we calibrated the intrinsic and extrinsic camera parameters with our full-calibration approach with 19DoF as introduced in Section 5.3 and cross-checked the result vs. the ground truth of the calibration lab. We transformed all images into the view of the scanner (see Section 7) and then extracted the target-centers (see (Abmayr et al., 2008b)).

Our statistical measures are the remaining azimuth and elevation error (rms and mean) between pairs of points after the co-registration. Furthermore, we calculate the remaining distance between these pairs: To norm the distance error, all pairs of points are projected to a 10m sphere. Fig. 7 shows the error statistics and demonstrates the accuracy of our camera-calibration: As mentioned above, the scan mode we used has a spatial point distance of 6.4 mm in 10 m and consequently the remaining error of the camera-calibration is within sub-pixel accuracy.

In summary, this test demonstrates that our camera-calibration approach is valid for the complete panoramic field of view with sub-pixel accuracy.

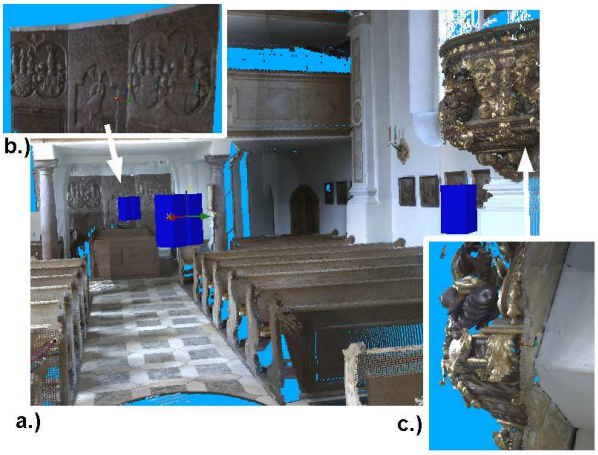

Figure 9: Kirche Seefeld: Colored 3D point cloud (i)

Generalization Test Here, we demonstrate the generalization of our extrinsic camera calibration approach to different locations: As introduced above, in our measurement concept the intrinsics are adjusted once in the laboratory, whereas the extrinsic parameters have to be updated in the field. Hence, we have to show that our calibration approach for the extrinsic parameters does not over-adjust the parameters to the environment and consequently is valid to different locations:

To show this, we first applied our method for the intrinsic camera calibration in the calibration lab as introduced above. After that, we removed the camera and the camera-tilt-unit from the scanner and changed the location. Then, we reassembled the camera and the camera-tilt-unit again in the new location and applied our automatic feature extraction and matching approach from Section 6. With this set of matches we then recalculated the extrinsic parameters with our update procedure, hence 3DoF as introduced in Section 5. Finally, we cross-checked this new calibration in the calibration lab vs. the ground truth without demounting the camera from the scanner: Equal to the camera calibration accuracy test from the last paragraph, we mapped the color images onto the reflectance data of the scan (see Section 7) and then extracted the target-centers (see (Abmayr et al., 2008b)). The results of this test is shown in Fig. 7: The scan mode we used had a spatial point distance of 6.4 mm in 10 m and consequently the remaining error is within sub-pixel accuracy. In summary, this test demonstrates the generalization of our extrinsic camera calibration approach to different locations.

The same procedure was applied in the 'Kirche Seefeld' project. Fig. 8 shows the remaining error between pairs of points after the automatic feature extraction and matching approach from Section 6 and the external camera calibration with the update procedure from Section 5.3. The remaining error between pairs of points is shown as vector. For visualization, the vector is multiplied with a scale factor. As we did not remount the camera from the scanner for the whole project, we could use this calibration result for all scans. Finally, Fig. 9 and Fig. 10 show some results of the colored 3D point cloud after the registration and data fusion

Figure 10: Kirche Seefeld: Colored 3D point cloud (ii)

of all viewpoints.

9 DISCUSSION

The angle increments of the vertical tilt unit and the horizontal rotation of the scanner are highly accurate. Using these angles as fixed input parameters reduces the unknown external camera parameters of the image sequence to only two rigid motions. Consequently, this results in large benefits for the stability of the calibration. Due to transportation conveniences the tilt-unit and the camera are designed to be removable from the scanner. Consequently, the external camera parameters must be re-calibrated in the field. After reattaching the camera on the scanner, the position and orientation between camera and scanner is approximately known: hence, we successfully applied correlation-based quality criteria for matching feature points. Although the different sensor modalities and difficult scene structures often result in a lack of uniqueness of the features, our multi sensor approach guarantees feasibility also in difficult field applications. This was validated on a historical site called 'Kirche Seefeld'.

ACKNOWLEDGEMENTS

We thank M. Mettenleiter and the r&d team of Zoller + Fröhlich GmbH for their research work on the Imager 5006. We also would like to thank H. Hirschmüller, M. Suppa and the 3D Modeling Group from the Institute of Robotics and Mechatronics at the German Aerospace Center for many useful discussions and good cooperation.

REFERENCES

Abmayr, T., Dalton, G., Härtl, F., Hines, D., Liu, R., Hirzinger, G. and Fröhlich, C., 2005. Standardisation and visualization of 2.5d scanning data and rgb color information by inverse mapping. 7th Conference on Optical 3D Measurement Techniques, Vienna, Austria.

Abmayr, T., Härtl, F., Hirzinger, G., Burschka, D. and Fröhlich, C., 2008a. Automatic registration of panoramic 2.5d scans and color images. EuroCOW 2008, International Calibration and Orientation Workshop, Castelldefels, Spain on CDROM, pp. pages 6.

Abmayr, T., Härtl, F., Hirzinger, G., Burschka, D. and Fröhlich, C., 2008b. A correlation based target finder for terrestrial laser scanning. Journal of applied Geodesy, de Gruyter 2(3), pp. pages 31–38.

Deumlich, R. and Staiger, R., 2002. Instrumentenkunde der vermessungstechnik. Herbert Wichmann Verlag; 9. Auflage, Hüthig GmbH + Co. KG, Heidelberg.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. of the ACM 24 pp. pp. 381–395.

Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centers of circular features. Proceedings of the ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data pp. 281–305.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. Proceedings, 4th Alvey Vision Conference, Manchester pp. 147–151.

Hirschmüller, H. and Scharstein, D., 2008 (accepted for publication). Evaluation of stereo matching costs on images with radiometric differences. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Lichti, D., 2007. Error modeling, calibration and analysis of an amcw terrestrial laser scanner system. ISPRS Journal of Photogrammetry and Remote Sensing 61(5), pp. 307–324.

Park, F. C. and Martin, B. J., 1994. Robot sensor calibration: Solving ax=xb on the euclidean group. IEEE Transactions on Robotics and Automation 10(5), pp. 717–721.

Rietdorf, A., 2005. Automatisierte auswertung und kalibrierung von scannenden messsystemen mit tachymetrischem prinzip, dissertation. Verlag der Bayersichen Akademie der Wissenschaften in Kommission beim Verlag C.H. Beck.

Tsai, R., 1987. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-self tv cameras and lenses. IEEE Journal of Robotics and Automation 3(4), pp. 323–344.

Zoller + Fröhlich, G., (visited September 2008). http://www.zf-laser.com.