

AUTOMATIC FEATURE MATCHING BETWEEN DIGITAL IMAGES AND 2D REPRESENTATIONS OF A 3D LASER SCANNER POINT CLOUD

N. Meierhold^{a,*}, M. Spehr^b, A. Schilling^a, S. Gumhold^b, H.-G. Maas^a

^a Technische Universität Dresden, Institute of Photogrammetry and Remote Sensing, 01062 Dresden, Germany –
(Nadine.Meierhold, Hans-Gerd.Maas)[@tu-dresden.de](mailto:tu-dresden.de), Anita.Schilling[@mailbox.tu-dresden.de](mailto:tu-dresden.de)

^b Technische Universität Dresden, SMT, Department for Computer Sciences –
(Marcel.Spehr, Stefan.Gumhold)[@tu-dresden.de](mailto:tu-dresden.de)

Commission V, WG V/3

KEY WORDS: Feature matching, Terrestrial laser scanning, Intensity image, Data fusion, SIFT, Fundamental matrix, RANSAC

ABSTRACT:

The geometric referencing of digital image data and 3D point clouds e.g. given by a terrestrial laser scanner is the prerequisite for different levels of integrated data interpretation such as point cloud or mesh model texture colourisation for visualisation purposes, interactive object modelling by monoplottting-like procedures or automatic point cloud interpretation. Therein, the characteristics of laser scanner data and camera data can be regarded as complementary, so that these data are suitable for a combined interpretation. A precondition for the geometric referencing between laser scanner data and digital images and consequently for an integrated use of both data sets is the extraction of corresponding features. With a set of corresponding features the orientation of the image to the point cloud can be obtained by spatial resection. With regard to a high automation level the paper presents an approach for finding correspondences between features extracted from laser scanner data and digital images automatically.

The basic idea of the presented approach is to use the SIFT operator to detect corresponding points in the camera image and an intensity image of the laser scanner data. Determining correspondences consists of four steps: Detection of salient characteristics, description of the features, matching of the descriptions in both images and evaluation of correct matches. RANSAC is used to find sets of consistent matches. The approach is validated with a data set taken from a baroque building in Dresden.

1. INTRODUCTION

Laser scanning is a valuable tool for the capturing of 3D object geometries and gained widespread popularity in the last years. A laser scanner delivers precise measurements in an automatic and efficient way. Cameras deliver largely complementary data, which may be valuable for integrated data processing schemes. When a camera is mounted on top of the laser scanner the orientation of the photo relative to the point cloud is directly given. But this also limits the recording of the laser scanner and the camera to one point in time. Ideally, the photos of a scanned object might be taken by a freely moving camera to allow the extensive digital documentation of the object. The usual method for such tasks is the application of signalised tie points, which can be identified in the point cloud and the photo alike. Their coordinates are then used for computing the relative orientation. The installation of the signalised tie points is costly, needs proper planning and is hardly repeatable. Therefore, the photos are limited to the moment the laser scans are taken. This denotes the main disadvantage of the procedure. A possibility to determine the relative orientation of the photos to a point cloud without signalised tie points or other supplementary information would be beneficial. In this paper, we propose a method to accomplish the computation of the relative orientation of a photo to a dense 3D point cloud from point correspondences using solely the photo and the information of the laser scanner point cloud. The point correspondences are established by matching feature points between the photo and a 2D representation of the point cloud using the intensity of the NIR laser as grey values.

The determination of point correspondences between the 3D point cloud and a 2D photo is a complex problem because of numerous possibilities which have to be evaluated. A simple but effective way is the mapping of the point cloud to a 2D image, so that the 3D coordinates to every image point are known. Hereby, the problem is reduced to finding corresponding points in the pair of 2D images. Thus, the 2D representation narrows the search space significantly and enables the usage of well researched computer vision methods. Correspondences between the image pairs are established by detecting and matching feature points. For this task, a multitude of feature detectors is available in the literature. An extensive overview is given in (Tuytelaars & Mikolajczyk, 2007). The great differences in scaling between photos and 2D representations of the point clouds require a scale-invariant feature detector. A comparison of affine region detectors is presented by Mikolajczyk & Schmid (2005) and finds that the Maximally Stable Extremal Regions (MSER, Matas et. al., 2002) show in many cases the best performance. MSER is a well established and reliable region detector but requires a monotonic transformation of the grey levels between the images. This cannot be guaranteed because of differences resulting from the distinct capturing methods of the images. Region detectors are generally dependent on clear boundaries, which are often not present in the 2D representations. Though FAST is one of the most efficient interest point detector as evaluated by Jazayeri & Fraser (2010), its performance in the context of multimodal imagery has not been tested. Traditional corner detectors like the Harris operator (Harris & Stephens,

* Corresponding author.

1988) are not appropriate, because usually they are not scale-invariant.

In (Böhm & Becker, 2007) an intensity image of a laser scanner deploying a green laser was successfully matched with a photo using the SIFT detector presented by Lowe (2004). The authors state that a portion of ca 20% good candidates for orientation computation could be determined with RANSAC from all matches found by SIFT. Becker & Haala (2007) state similar results for matching a photo with an intensity image of a green laser as well.

Inspired by these positive results SIFT was also chosen for the experiments in this paper, as it offers a suitable combination of feature detector and descriptor. Blob detectors like SIFT have the advantage over corner detectors that they perform better in feature localisation (Tuytelaars & Mikolajczyk, 2007). SIFT has already been applied successfully in quite different tasks and, according to (Lingua et al., 2009), it has the capacity to extract and match points with large geometric and photometric distortions. Furthermore, an evaluation of local feature descriptors by Mikolajczyk & Schmid (2005) has shown that the SIFT descriptor and its extension GLOH perform best in a variety of situations.

Section 2 of the paper describes two methods for calculating a 2D representation of the 3D point cloud. Based on this, an approach for an automatic feature matching is shown in Section 3. Finally, results will be presented (Section 4) and the article is closed with an outlook and a summary.

2. CALCULATION OF INTENSITY IMAGE

2.1 Imaging of polar scanner coordinates

The applied terrestrial laser scanner Riegl LMS-Z420i, measuring the time of flight of sent laser pulses, has a maximum field of view of 80° in vertical and 360° in horizontal direction. The vertical deflection is realised by a continuously rotating polygonal mirror and for the horizontal dispersion the whole optical head is rotated. In this scanning mode up to 8000 pts/sec can be obtained. By this way, the object is raster-wise scanned whereas the field of view and the scan angle width ($\Delta\theta$, $\Delta\varphi$) can be set according to the respective application.

The laser scanner software records the direction (vertical angle θ and horizontal angle φ), the covered distance (range r) and the signal amplitude (intensity i) of backscattered laser pulses. For the resulting 3D point cloud, the measured polar coordinates are typically transformed in 3D Cartesian coordinates.

A simple way for the 2-dimensional representation of the 3D point cloud is the imaging of the polar coordinates directly. Hereby, the image height is given by the obtained vertical angle range and the image width by the horizontal angle range. The number of pixels results from the defined scan resolution (Fig. 1).



Figure 1. Imaged polar coordinates with gaps (left) and filled gaps (right)

Advantageous is that in this representation, each pixel corresponds to only one laser scanner point. But like shown in Fig. 1 left, there appear gaps resulting from a deficient angle measurement. These gaps would disturb the process of feature detection so that Fig. 1 right shows the intensity image with filled gaps applying morphological operations and interpolating grey values.

Disadvantageous is the characteristic that this representation maps straight object lines as curved lines. For the acquisition of digital images the cameras are mostly equipped with central perspective lenses so that, neglecting lens distortion, straight object lines are imaged as straight lines. Thus, using the imaged polar coordinates as 2D representation would complicate the process of feature assignment, because differences of the grey values are not only effected by different illuminations but also by different geometries. To avoid additionally discrepancies to the camera image, the point cloud shall be imaged in a central perspective way.

2.2 Central perspective representation

To realise a 2D representation of the 3D point cloud, the Cartesian coordinates of each laser scanner point are projected to a virtual image plane. The coordinate system of the virtual camera is defined like shown in Fig. 2.

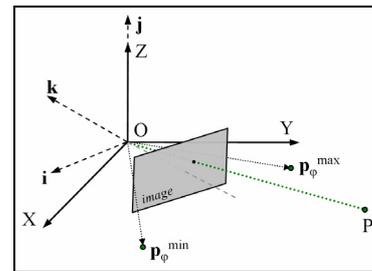


Figure 2. Definition of coordinate systems

The 3D camera coordinate system is defined by the frame $\{i, j, k, O\}$. The perspective centre relies in the origin of the laser scanner coordinate frame $\{X, Y, Z, O\}$. The virtual image plane is spanned by the vectors i and j .

The z-axis of the camera coordinate system (vector k) is calculated from the laser scanner points with smallest and biggest horizontal angle (Eq. 1) so that it is directed to the horizontal centre of the point cloud.

$$\mathbf{k} = \frac{\mathbf{p}_{\varphi}^{\min}}{|\mathbf{p}_{\varphi}^{\min}|} + \frac{\mathbf{p}_{\varphi}^{\max}}{|\mathbf{p}_{\varphi}^{\max}|} \quad k_z = 0 \quad \mathbf{k} = \frac{-\mathbf{k}}{|\mathbf{k}|} \quad (1)$$

Normally the y-axis is defined to be the up-direction of the camera coordinate system so that the approximate up-direction can be set to the Z-axis of the laser scanner coordinate system. The remaining frame vectors of the camera coordinate system result from:

$$\begin{aligned} \mathbf{i} &= (0 \ 0 \ 1)^T \times \mathbf{k} & \mathbf{i} &= \frac{\mathbf{i}}{|\mathbf{i}|} \\ \mathbf{j} &= \mathbf{k} \times \mathbf{i} & \mathbf{j} &= \frac{\mathbf{j}}{|\mathbf{j}|} \end{aligned} \quad (2)$$

The transformation of point $\mathbf{p} = (X, Y, Z)^T$ from the laser scanner coordinate system to the 3D camera coordinate system can be realised by applying the viewing matrix \mathbf{V} (Eq. 3).

$$\mathbf{V} = \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} & \mathbf{0} \\ 0 & 0 & 0 & 1 \end{bmatrix}^T \quad \hat{\mathbf{p}} = \mathbf{V}\mathbf{p}_h$$

with $\mathbf{p}_h = (w \cdot X \quad w \cdot Y \quad w \cdot Z \quad w)^T$ and $w = 1$ (3)
in homogeneous coordinates

The distance from the virtual image plane to the perspective centre is given by the principle distance c set by the user. The projection of point $\hat{\mathbf{p}}$ given in the 3D camera coordinate system to the image plane is described by the perspective projection matrix \mathbf{P} (Eq. 4).

$$\tilde{\mathbf{p}} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ \tilde{w} \end{pmatrix} = \mathbf{P}\mathbf{V}\mathbf{p}_h = \mathbf{P}\hat{\mathbf{p}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/c & 0 \end{bmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \\ \hat{w} \end{pmatrix} \quad (4)$$

The Euclidian image coordinates result from normalising the homogeneous coordinate vector $\tilde{\mathbf{p}}$ (Eq. 5).

$$\mathbf{p}' = (\tilde{x} \quad \tilde{y} \quad \tilde{z} \quad \tilde{w})^T \cdot \frac{1}{\tilde{w}} = (x' \quad y' \quad -c \quad 1)^T \quad (5)$$

The size of the resulting image can be calculated from the minimum and maximum values of the projected coordinates. With a user defined pixel size Δpx , the image height I_h and width I_w result from:

$$\begin{aligned} I_w &= (x'_{\max} - x'_{\min}) / \Delta px \\ I_h &= (y'_{\max} - y'_{\min}) / \Delta px \end{aligned} \quad (6)$$

Finally, the image coordinates related to the origin at the left up-corner of the image are calculated in pixel applying Eq. 7.

$$\begin{aligned} x_p &= (x' - x'_{\min}) / \Delta px \\ y_p &= I_h - (y' - y'_{\min}) / \Delta px \end{aligned} \quad (7)$$

Fig. 3 shows the resulting central perspective intensity image after projecting all laser scanner points and image enhancement. Originally, the measured intensity values yield an image which is poor of contrast. To enhance the image visually, it is processed with a sequence of operations: brightness and contrast enhancement, gamma correction and histogram stretching.



Figure 3. Central perspective representation of the 3D point cloud

3. FEATURE MATCHING

The process of feature matching consists of four steps: feature detection, feature description, correspondence assignment and the evaluation of the found candidate matches.

3.1 Requirements of the application

Intensity images of a laser scanner point cloud and a RGB-camera show differences concerning the image resolution, the radiometry and the viewing direction. Since its wavelength is most similar to the near-infrared wavelength of the laser scanner, just the red channel of the camera image is used to reduce illumination differences. Still the images differ largely and for the identification of point matches between both scene representations a robust feature matching scheme is needed.

In general, image feature points must meet the requirement to be clearly distinguishable from their surroundings. Such instances of interest points can include corners, blobs, edges or texture transitions. Thus, a feature does not necessarily represent an image location carrying specific semantic content. The *Scale-Invariant-Feature-Transform (SIFT)* by Lowe (2004) is designed to match corresponding points in images taken from widely different viewpoints. Linear brightness alterations, rotations and translations have no effect on their ability to find and uniquely identify point matches between images. For the experiments the SIFT implementation of the VLFeat library 0.9.4 for Matlab was used (Vedaldi & Fulkerson, 2008).

3.2 Feature detection

Achieving SIFT features starts with the construction of an image scale space pyramid by consecutively applying *Gaussian* low pass filters. A subtraction of adjacent scales generates layers of band pass filtered versions of the image which in combination are also known as a *Laplacian* pyramid. This subtraction causes the illumination invariance. Within this image pyramid local extrema are detected as keypoints and their accurate position is calculated using a Taylor expansion. Subsequently, the selected keypoints are analysed for two stability criterions. Keypoints with low contrast are discarded if the Taylor expansion at the corrected position is smaller than a *peak threshold*. Furthermore, extrema are discarded which are poorly located on edges. The indication is the principle curvature which should be smaller than a defined *edge threshold*. These thresholds were manually optimised to retrieve an optimal candidate correspondence set. Some results of the feature detection are shown in Fig. 4.

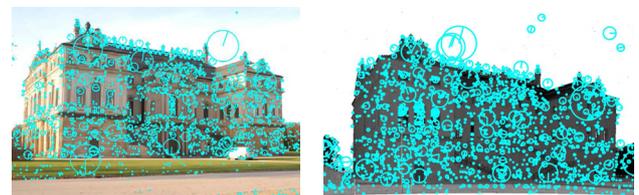


Figure 4. Interest points detected with SIFT in camera image (left) and intensity image (right)

3.3 Feature description and matching

Once a keypoint and its main orientation are detected, its feature vector is constructed using the distribution of gradient directions in its vicinity. For this purpose gradient orientations and magnitudes of a 16x16 pixel neighbourhood on the descriptor's scale are used. Gradients are normalised with

regard to the interest point's main orientation to achieve rotational invariance. For each of 4x4 subarrays around the keypoint, orientation histograms are computed and their values are summarised in a 128 dimensional vector. Candidate interest point correspondences between two images are obtained by comparing the distances of interest points within this 128 dimensional feature space. SIFT's scale invariance results from the comparison of descriptors between scales.

3.4 Evaluation of correspondences

3.4.1 Epipolar geometry: Images showing the same scene taken from different cameras and viewpoints are related by their epipolar geometry (Hartley & Zisserman, 2004). Calibration specific data is represented by the fundamental matrix \mathbf{F} . In essence, the 3x3 homogenous matrix \mathbf{F} is the algebraic representation of the mapping between corresponding points in two images.

Consider a pair of homogenous image coordinates $\mathbf{p}_1 = (x_1, y_1, 1)^T \in I_1$ and $\mathbf{p}_2 = (x_2, y_2, 1)^T \in I_2$. Both shall represent a projection of the homogenous world point $\mathbf{Q} = (X, Y, Z, 1)^T$ with:

$$\begin{aligned} (x'_1, y'_1, w'_1)^T &= \mathbf{P}_1 \mathbf{Q} & x_1 &= x'_1/w'_1 & y_1 &= y'_1/w'_1 \\ (x'_2, y'_2, w'_2)^T &= \mathbf{P}_2 \mathbf{Q} & x_2 &= x'_2/w'_2 & y_2 &= y'_2/w'_2 \end{aligned} \quad (8)$$

with $\mathbf{P}_i = \mathbf{K}_i [\mathbf{R}_i | \mathbf{t}_i]$ projection matrix of camera i
 \mathbf{K}_i interior parameters of camera i
 $\mathbf{R}_i, \mathbf{t}_i$ rotation and translation of camera i

The camera position in world coordinates is \mathbf{C}_1 and $\mathbf{t}_1 = -\mathbf{R}_1 \mathbf{C}_1$ translates points from the world coordinate system into the camera coordinate system.

If \mathbf{p}_1 and \mathbf{p}_2 really show the same scene point P , than \mathbf{p}_2 will lie on the homogenous line \mathbf{l}_2 (Eq. 9) which represents the epipolar line of \mathbf{p}_1 (Fig. 5). Mathematically this relation is described by Eq. 10.

$$\begin{aligned} \mathbf{l}_2 &= \mathbf{C}' \times \mathbf{p}'_1 \\ \text{with } \mathbf{p}'_2 &= \mathbf{P}_2 \mathbf{P}^+ \mathbf{p}_1 \text{ and } \mathbf{C}' = \mathbf{P}_2 \mathbf{C}_1 \in I_2 \end{aligned} \quad (9)$$

\mathbf{P}^+ pseudo inverse of \mathbf{P}_1

$$\mathbf{p}_2^T \mathbf{l}_2 = \mathbf{p}_2^T \mathbf{P}_2 \mathbf{C}_1 \times \mathbf{P}_2 \mathbf{P}^+ \mathbf{p}_1 = \mathbf{p}_2^T \mathbf{F} \mathbf{p}_1 = 0 \quad (10)$$

If this is not the case Eq. 10 represents the distance of \mathbf{p}_2 from \mathbf{l}_2 in units of $|\mathbf{l}_2|$. The fundamental matrix \mathbf{F} has 7 degrees of freedom since it is homogenous and $\det(\mathbf{F})=0$.

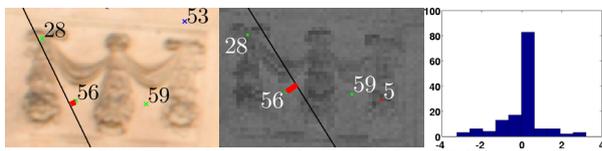


Figure 5. Close-up of camera image (left) and intensity image (middle) with SIFT interest points and epipolar lines and histogram of error values (right)

The black lines in Fig. 5 (left and middle) signifies the epipolar lines. The red orthographic lines show the error e_{56} produced under assumed \mathbf{F} .

3.4.2 Correspondence analysis: The introduced SIFT technique supplied candidate correspondence points. To identify the correct subset of matches and their associated \mathbf{F} , we employed the RANSAC procedure (Fischler & Bolles, 1981) using the entries of \mathbf{F} as model parameters to fit. RANSAC is an iterative, probabilistic optimisation algorithm which in turns randomly chooses 7 matches (consensus set), fits a model matrix \mathbf{F} using the conditions in Eq. 10 and evaluates it on the rest. The algorithm terminates either after a predefined number of iterations or if the model explains a predetermined number of matches well. The error function $E(\mathbf{F})$ that defines how well an observed matching pair $(\mathbf{p}_1, \mathbf{p}_2)$ is explained by the model \mathbf{F} , is defined by Eq. 11.

$$E(\mathbf{F}) = \mathbf{p}_2^T \mathbf{F} \mathbf{p}_1 \quad (11)$$

The histogram in Fig. 5 right shows the distribution of those error values among a candidate correspondence set. The best consensus set is chosen for the final model fit. Fig. 6 presents the result of this technique.

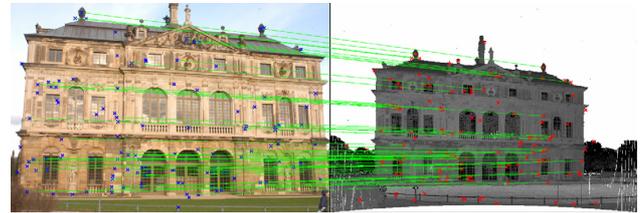


Figure 6. Correspondences detected with fundamental matrix and RANSAC

4. RESULTS AND DISCUSSION

4.1 Data set

The experiments were carried out with a data set taken from the "Palais im Großen Garten" which is a baroque building in the centre of a large park in Dresden (Fig. 7). The complete object (34 m x 45 m) was scanned from different positions using the terrestrial laser scanner Riegl LMS-Z420i. For the tests, two scan positions were picked out and the respective central perspective intensity images were calculated. Fig. 3 shows the intensity image for scan position 3.



Figure 7. Image 112 (left) and point cloud (right) acquired from scan pos 4

Additionally, images were acquired with the 6 Mpix mirror reflex camera Nikon D100 and a 28 mm lens. Before taking the object images, the camera was calibrated to determine the parameters of the interior orientation (Tab. 1).

s_0 [pixel]	c [mm]	x_0 [mm]	y_0 [mm]	A_1 [1/mm ²]	A_2 [1/mm ²]
0.04179	28.870	0.183	0.048	-8.59E-05	1.54E-07

Table 1. Parameters of interior orientation derived from camera calibration

The experiments contained the extraction and matching of feature points which afterwards were used for the relative orientation of four camera images¹ to the point cloud by spatial resection. Additionally, image orientations were calculated using points manually measured in the point cloud and in the photos.

4.2 Feature matching

For the scans, the field of view was set to about 40° in vertical and 65° in horizontal direction with a scan resolution of about 0.04° and scan distance of about 40 m. The central perspective intensity images were calculated with a principle distance of 1 mm and an empirically determined pixel size of 0.00089 mm. Hence, it results an image resolution of about 1400 x 900 pixel compared to the resolution of the camera images of 3008 x 2000 pixel which is twice larger.

parameter	camera images	scan images
peak threshold	1.5	0.0
edge threshold	3.0	3.0
pyramid level	1	0

Table 2. Defined parameters for feature detection with SIFT

The definition of the parameters for the feature point detection with SIFT (Tab. 2) was motivated by limiting the number of detected features but simultaneously cover the whole image. An increasing number of detected features affects the feature descriptor and leads to an increasing number of incorrect feature correspondences (Tuytelaars & Mikolajczyk, 2007). The parameter *pyramid level* defines the starting level of feature detection.

Fig. 8 shows the complete number of found SIFT keypoints in camera and intensity images.

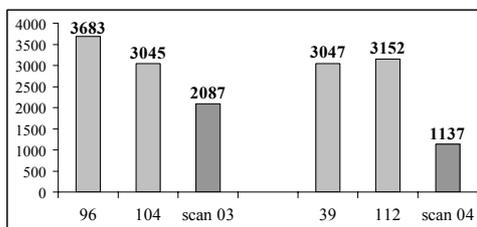


Figure 8. Number of detected keypoints with SIFT

The correspondences found by the SIFT descriptor and the remaining correspondences after fitting the fundamental matrices with RANSAC are shown in Fig. 9.

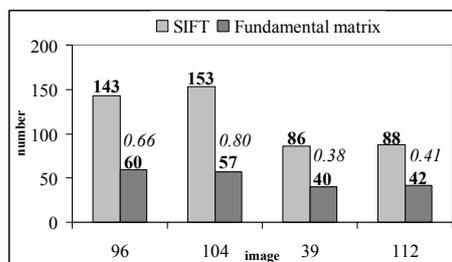


Figure 9. Feature point correspondences between camera and intensity images

¹ two photos for each scan position

Only these feature points were accepted to be correct correspondences which distances to the respective epipolar line were smaller than the hundredth of the standard deviation derived from the error values ($E(F)$) (Fig. 9 cursive values). It is obvious that only about 43% of the SIFT correspondences were detected as real correspondences. However, the remaining number of point correspondences is sufficient for the purpose of image orientation by spatial resection.

4.3 Image orientation

At first, the 3D coordinates of the 2D feature points extracted from the intensity images were identified using the known relation between intensity image and laser scanner point cloud. Afterwards, the exterior orientations ($X_0, Y_0, Z_0, \omega, \phi, \kappa$) of the four camera images as well as the interior orientation (c, x_0, y_0) and the radial lens distortion (A_1, A_2) of the camera were determined applying point-based spatial resections.

The camera parameters were compared to the values derived in course of the camera calibration. Tab. 3 shows the absolute differences as well as the empirical standard deviations of the unit weight separated for each scan position and averaged over two images. The obtained image positions and orientations were compared to the results of the spatial resections using manually measured feature points. The empirical standard deviations of the unknowns are summarised in Tab. 4.

It is obvious that the image orientation with the SIFT correspondences performs almost as good as using manually measured feature points. Due to the higher redundancy, the precision with the manual features is slightly better. Comparing the two scan positions, despite of a lower redundancy, the standard deviations and partly the differences to the calibration values show better results for the scan position 4. One possible reason could be a better spatial distribution of the object points caused by representing a more inhomogeneous part of the building facade (shown in Fig. 7).

The reached overall precision of the image orientations is about 2.4 pixel which is approximately up to 2 cm in object space at a distance of 30 m. This precision is sufficient for colouring the point cloud (Fig. 10) but for a combined interpretation of digital images and laser scanner data for modelling purposes it might be increased.



Figure 10. Point cloud of scan position 3 coloured with photo 96 oriented with SIFT keypoints

The SIFT correspondences still contain some falsely assigned feature points which result from ambiguities along the epipolar lines. Additionally, some correspondent features do not represent exactly the same part of the object. These false correspondences disturb the process of image orientation. A certain number can be discarded combining the spatial resection with a statistical test for outlier detection, but in some cases no convergence was reached. To avoid these ambiguities, the epipolar line analysis should be improved considering the range information given by the laser scanner data.

scan pos	corresp.	redundancy	s_0 [pixel]	Absolute differences of camera parameters to calibration				
				c [mm]	x_0 [mm]	y_0 [mm]	A_1 [1/mm ²]	A_2 [1/mm ⁴]
3	manual	117	2.47	0.34	0.18	0.40	2.47E-05	1.87E-07
	SIFT	102	2.72	0.44	0.34	0.03	3.32E-05	1.92E-07
4	manual	151	2.36	0.13	0.14	0.03	6.19E-05	3.04E-07
	SIFT	68	2.08	0.14	0.19	0.29	1.09E-04	6.38E-07

Table 3. Averaged absolute differences of estimated camera parameters to calibration data

scan pos	corresp.	redundancy	s_0 [pixel]	Averaged internal standard deviations of estimated parameters					
				s_{x0} [m]	s_{y0} [m]	s_{z0} [m]	s_{ω} [°]	s_{ϕ} [°]	s_{κ} [°]
3	manual	117	2.47	0.639	0.075	0.183	1.047	1.112	0.995
	SIFT	102	2.72	1.147	0.124	0.270	1.950	2.052	1.858
4	manual	151	2.36	0.358	0.133	0.086	0.665	0.631	0.683
	SIFT	68	2.08	0.580	0.209	0.124	0.929	0.970	0.966

Table 4. Averaged standard deviations of exterior orientation parameters from spatial resections

5. SUMMARY AND OUTLOOK

In this paper, an approach was presented which allows the matching between reflectance images from a terrestrial laser scanner and digital camera images independent of their specific brightness distribution.

The combination of SIFT for feature point detection and description and the fundamental matrix for the evaluation of the found SIFT correspondences is a reliable option to provide feature points which can be used to determine the orientation of camera images by spatial resection. Advantageous is, that SIFT as blob detector mainly detects interest points which are located on planar object parts and less on edges. Thus, the determination of 3D coordinates by interpolation in the point cloud can be realised more precisely.

The applied SIFT algorithm is sufficient to handle the different characteristics of digital camera images and intensity images generated from 3D laser scanner data. However, in case of larger differences between the viewing directions, the features matching with SIFT was insufficiently so that the image orientation could not be determined.

The experiments in this paper were carried out with a standard implementation of SIFT. For a better handling of perspective differences of the camera and intensity images, the algorithm will be modified. Additionally, future work will deal with an improvement of the epipolar line analysis. To realise a more reliable correspondence analysis, the length of constructed epipolar lines can be limited considering the range information of the laser scanner points. In this way, the number of ambiguities and accordingly the number of false correspondences can be reduced.

REFERENCES

Becker, S. & Haala, N., 2007. Combined Feature Extraction for Facade Reconstruction. *ISPRS Workshop on Laser Scanning and SilviLaser*, 36, pp. 44-49.

Böhm, J. & Becker, S., 2007. Automatic marker-free registration of terrestrial laser scans using reflectance features. *8th Conference on Optical 3D Measurement Techniques*, pp. 338-344.

Fischler, M. A. & Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, Vol. 24(6), pp. 381-395.

Harris, C. & Stephens, M., 1988. A combined corner and edge detector. *Proceedings of the Fourth Alvey Vision Conference*, pp. 147-151.

Hartley, R. I. & Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Second edition, Cambridge University Press, pp. 241-245

Jazayeri, I. & Fraser, C., 2010. Interest Operators for the feature-based matching in close range photogrammetry. *The Photogrammetric Record*, Vol. 25 (129), pp. 24-41.

Lingua, A., Marenchino, D. & Nex, F., 2009. Performance Analysis of the SIFT Operator for Automatic Feature Extraction and Matching in Photogrammetric Applications. *Journal of Sensors*, 9(5), pp. 3745-3766.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60 (2), pp. 91-110.

Matas, J., Chum, O., Urban, M. & Pajdla, T., 2002. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. *British Machine Vision Computing*, pp. 384-393.

Mikolajczyk, K., & Schmid, C., 2005. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 10, pp. 1615-1630

Tuytelaars, T. & Mikolajczyk, K., 2007. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision*, Vol. 3, No. 3, pp. 177-280

Vedaldi, A. & Fulkerson, B., 2008. VLFeat: An Open and Portable Library of Computer Vision Algorithms. <http://www.vlfeat.org/> (accessed 27.11.2009)