

# AN INVESTIGATION INTO THE REGISTRATION OF LIDAR INTENSITY DATA AND AERIAL IMAGES USING THE SIFT APPROACH

Abbas Abedini<sup>a\*</sup>, Michael Hahn<sup>b</sup>, Farhad Samadzadegan<sup>a</sup>

<sup>a</sup>Dept. of Surveying and Geomatics, Faculty of Engineering, University of Tehran, Tehran, Iran – (aabedeni, samadz)@ut.ac.ir

<sup>b</sup>Dept. of Geomatics, Computer Science and Mathematics, Stuttgart University of Applied Sciences, Stuttgart, Germany michael.hahn@hft-stuttgart.de

## Commission I, WG I/2 - SAR and LiDAR Systems

**KEY WORDS:** SIFT, Registration, Matching, LiDAR, MATLAB program, Transformation, Gaussian Function

### ABSTRACT:

Various methods exist for automatic registration of different kinds of image datasets. The datasets could be aerial or satellite images, maps, LiDAR data, etc. In this paper we propose a method for the registration of aerial images and LiDAR data. Our approach utilizes the so-called SIFT algorithm by which distinctive features are extracted from both aerial images and LiDAR intensity data. The extracted features are then automatically matched and the aerial image and LiDAR intensity data are registered in the subsequent stage. The invariance of the SIFT features to image scale and rotation and the robustness of the features with respect to changes in illumination, noise and to some extent changes in viewpoint may qualify those features particularly for our purpose. For SIFT feature extraction an open MATLAB implementation is used and integrated in the overall image registration process. This paper describes our approach and presents the results obtained from a LiDAR test site in Stuttgart, Germany.

## 1. INTRODUCTION

Automatic registration of airborne or remote sensing images with other images or maps has a long tradition in Photogrammetry and Remote Sensing. Generally, matching works very well if the matched image data are of the same type but is often less successful in applications using different kind of data like range and intensity data. This is the context of the research carried out in this paper. Automated registration of aerial images and LiDAR data is investigated based on a conceptual approach known as Scale Invariant Feature Transform (Lowe, 1999).

### 1.1 Related Work

Schenk and Csathó (2007) discussed fusion of imagery and 3D point clouds for reconstruction of visible surfaces of urban scenes. They represent the reconstructed surface in a 3D Cartesian reference system and introduce a feature-level fusion framework with the idea to generate a rich 3D surface description, in which surface information explicitly includes shape and boundary of surface patches and spectral properties. The purpose of the surface reconstruction is to support applications such as bold-earth determination and the generation of true orthophotos.

Habib et al. (2004) investigated registration of data captured by photogrammetric and LiDAR systems for close range applications. Their approach relies on straight lines as the features of choice upon which the LiDAR and photogrammetric dataset are co-registered. They extract LiDAR linear features and image points along the conjugate straight lines in the images interactively and estimate the parameters of a 3D

similarity transform. With experiments, they demonstrated the capabilities of the proposed method of co-registration of laser scans and photogrammetric data.

Hu et al. (2006) presented a hybrid modelling system that fuses data from three resources: a LiDAR point cloud, an aerial and ground view images. The purpose of the system is the rapid creation of accurate building models. There are three main steps to do the modelling: At first building outlines are interactively extracted from a high-resolution aerial image and mapped to the LiDAR data. Next surface information is extracted from LiDAR data and used for model reconstruction. Finally high-resolution ground view images are integrated into the model to generate fully textured urban building models.

### 1.2 Research Concept

The basic idea for registration of airborne images and LiDAR data is to use the LiDAR intensity data and the aerial images and employ the SIFT algorithm for matching these two data sets. Substantially, the LiDAR intensities represent radiation that is reflected from the scanned surface. LiDAR scanners like the one we use in our experiments mostly operate with frequencies in the NIR, therefore it is likely that a LiDAR intensity image and an aerial image do not match in all image regions similarly. On the other hand it is expected that the SIFT algorithm with its invariance to image scale and rotation and robustness to changes in illumination, noise, occlusion and to some extent changes in viewpoint should find an adequate number of suitable features for a successful matching of both data sets. By using the homologous feature points found by SIFT feature matching registration can be carried out.

---

\* Corresponding author

Image registration consists of four steps:

- Extraction of SIFT features
- Matching of SIFT features
- Removal of outliers in the matched
- Selecting the transformation model between the two images and estimating its parameters

The last two steps are often combined, in particular, if outlier detection is part of the parameter estimation procedure.

In the following section we briefly outline the SIFT algorithm and describe the matching strategy. The third section presents results of some experiments carried with data from a LiDAR test site in Stuttgart, Germany. The paper closes with some conclusions and recommendations.

## 2. IMAGE REGISTRATION BASED ON SIFT FEATURE MATCHING

### 2.1 The SIFT Algorithm

The Scale Invariant Feature Transform (SIFT) was introduced by David Lowe (Lowe, 1999). SIFT is an approach for detecting and extracting distinct features from images which can be used to perform matching between overlapping images. The features are invariant to image scale and rotation and robust with respect to changes in illumination, noise and to some extend changes in the 3D camera viewpoint.

Detection stages for the SIFT features are the following:

- Scale-space extrema detection
- Keypoint localization
- Orientation assignment
- Generation of keypoint descriptors

Key points for SIFT features correspond to scale-space extrema at different scales. Therefore, the first step towards the detection of extremas is filtering the images with Gaussian kernels at different scales, and the generation of the difference of Gaussian filtered images of adjacent scales.

Mathematically this reads as follows

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2-1)$$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2-2)$$

where  $I$  is the image,  $G$  the Gaussian kernel and  $L$  the scale-space image generated by convolution(\*).  $D$  represents the difference of Gaussian filtered images, also called DoG.

The filtered image are organised in a kind of an image pyramid in which the blurred images are grouped by octave. An octave corresponds to doubling the  $\sigma$ -value of the Gaussian kernel. On each image scale (octave), a fixed number of blurred images (sub-levels) is created. In the experiments (Section 3) we use 5 octaves with 5 differently blurred images on each octave.

In order to detect the local maxima and minima of of the DoG images across scales, each pixel in the DoG images is compared

to its eight neighbours in the current image and the nine neighbours in the neighbouring (higher and lower) scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoint.

The SIFT descriptor is a weighted and interpolated histogram of the gradient orientations and locations in a patch surrounding the keypoint. To determine the keypoint orientation, a gradient orientation histogram is computed employing the neighbourhood of the keypoint. Each neighbouring pixel contributes by its gradient magnitude to determine the keypoint orientation. Keypoints with low contrast are removed and responses along edges are eliminated. All the properties of the keypoint are measured relative to the keypoint orientation. This provides invariance to rotation. Histograms contain 8 bins each and each description contains an array of 4 histograms around the keypoint which leads to a SIFT vector of 128 elements. For more details, please refer to Lowe (1999, 2004).

### 2.2 Matching SIFT Features

In the matching stage, the descriptor matrixes of a reference and a target image have to be compared. Matching is based on the idea to find the nearest neighbours in the descriptor matrixes. For robustness reasons the ratio of the nearest neighbour to the second nearest neighbour is used. The Euclidean distance is given by

$$\|y - z\|_2 = \left( \sum_{i=1}^n (y_i - z_i)^2 \right)^{1/2} \quad (2-3)$$

where  $y$  and  $z$  are two descriptor vectors. The inner product of the vector  $(y - z)$  with itself

$$(y - z) \circ (y - z) = y \circ y + z \circ z - 2|y||z|\cos(\angle(y, z)) \quad (2-4)$$

shows that with normalized vectors  $y$  and  $z$  (length 1) the angle

$$\alpha = \arccos(\angle(y, z)) \quad (2-5)$$

can be used as a good approximation within the minimum search, in particular for small angles. For each target descriptor, the first and second nearest descriptors must be found. Then a pair of nearest descriptors gives a pair of matched keypoints if the ratio between the distances to the first and second nearest descriptors are lower than a given threshold  $t$ .

$$\text{ratio}(\text{first}, \text{second}) < t \quad (2-6)$$

In the experiments we studied the impact of  $t$  on the matching result by selecting  $t = 0.5, 0.6, 0.7, 0.8, 0.9, 1$ .

### 2.3 Removal of Outliers

Several methods exist for removing the outliers. RANSAC and Baarda's data snooping method are the methods that we have

used in this paper. In the following sections, we briefly review both methods.

### 2.3.1 RANSAC Method:

The "RANdom SAMple Consensus" RANSAC was introduced by Fischler and Bolles in 1981. If this concept is used in the context of image registration the idea is to separate outliers and inliers in a set of point pairs which are introduced to a coordinate transformation. More generally speaking, the basic assumption is that the data consist of inliers, i. e., data points which can be explained by some set of model parameters, and outliers which are data points that do not fit the model. In addition, the data points are subject to noise. An advantage of RANSAC is its ability to robustly estimate the model parameters. It finds reasonable estimates of the parameters even if a high percentage of outliers are present in the data set. A small drawback of RANSAC is that a complete search would be computationally very expensive. Therefore the number of random samples which is selected to estimate the parameters is usually limited by an upper number which may lead to a suboptimal solution (Fischler and Bolles, 1981).

### 2.3.2 Baarda's Data Snooping Method:

Baarda's method is one of the most commonly used blunder detection methods. It separates the inliers from the outliers using the estimated normalised residuals. With observation vector  $L$  and design matrix  $A$  the least squares estimate of the unknowns  $x$  is found by

$$\hat{x} = (A^T A)^{-1} A^T L \quad (2-7)$$

The residual vector  $v$  and its covariance matrix  $C_v$

$$\begin{aligned} \hat{v} &= A\hat{x} - L \\ C_{\hat{v}} &= \sigma_0^2 (I - A(A^T A)^{-1} A^T) \end{aligned} \quad (2-8)$$

are used to determine normalized residuals by computing the ratio of a residual and the square root of corresponding diagonal element of the covariance matrix  $C_v$

$$\hat{v}_{N,i} = \frac{\hat{v}_i}{\sqrt{\text{diag}(C_{\hat{v}})_{ii}}} \quad (2-9)$$

Normalized residuals follow the standard normal distribution  $N(0,1)$  if no errors are present in the data. If a normalised residual is above a critical value, the corresponding observation might be erroneous. Data snooping is the process of eliminating the observation, which produces the largest normalised residual. This process has to be repeated until no further outlier is detected anymore.

## 2.4 Procedural development

The procedural development undertaken in this research can be summarized as follow:

**Dividing the images of both datasets into patches:** Due to the large size of two datasets which causes difficulties in running the program large image data have been divided in smaller patches.

**Designing a GUI:** A GUI is developed which handles loading datasets and specifying SIFT parameters, in particular: Number of octaves, number of images in each octaves (levels), standard deviation of Gaussian Function ( $\sigma$ ), threshold of SIFT matching, and some others of minor importance. The error search with RANSAC and data snooping is visualised to simply control the impact of SIFT parameter settings onto the registration result of

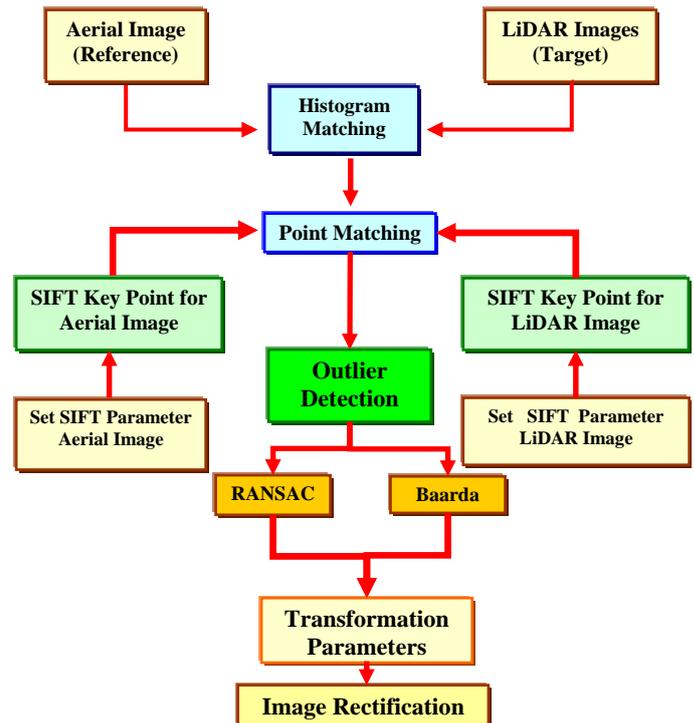


Figure 1: Flowchart of the registration program

the four types of the LiDAR data used in the experiments (i.e. LiDAR range, LiDAR intensity, both with first and second pulse).

**Overall registration procedure:** Flowchart of the overall registration process is shown in the following figure (Figure 1).

The histogram matching between LiDAR intensity and aerial images serves for convenience of visual evaluation. SIFT keypoints are extracted from the original image data. The transformation parameters are used to rectify the target image to the reference image frame.

## 3. EXPERIMENTAL INVESTIGATIONS

### 3.1 Test data set of Stuttgart city

The TopScan Company has acquired LiDAR data and aerial images in April 2006 with a sampling density of 4.8 points per square meter for the LiDAR data. Simultaneously aerial colour images with 20 cm ground resolution have been recorded using

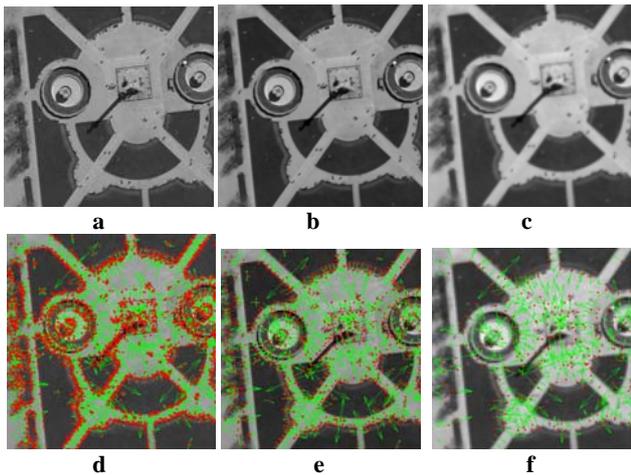


Figure 2. (a,d) Aerial image (20 cm), (b,e) reduced resolution (40 cm), (c, f) further reduced resolution (80 cm)

a Rollei AIC-modular –L-S digital metric camera with 4080 × 5440 pixels per image.

The RGB images are transformed to a gray scale images and used as input to the SIFT algorithm. The LiDAR data have been interpolated to a regular raster of 0.5 m grid size. Thus four images, two for the last and two for the first returns of LiDAR scanner are produced which in the following are called “Last Range, First Range, Last Intensity and First Intensity”. Each LiDAR image covers 2000 × 2000 grid points.

### 3.2 SIFT Algorithm applied to the LiDAR Data

In a first experiment, keypoint extraction related to LiDAR range and intensity is investigated by using all four LiDAR images. Table 3.1 shows a significant difference with respect to the number of extracted keypoints. LiDAR intensity images have 7 to 8 times more SIFT key point features than LiDAR range images which manifests a comparatively low range variation in the range images. The reflected signal

Original image 500*500 pixels	original aerial image (20 cm)	reduced image (40 cm)	reduced image (80 cm)
Number of SIFT features	3980	1326	378

Table 3.1. Key points in different LiDAR images

strength, which leads to the intensities in the range images, leads to texturing which is comparable to the greyscale aerial image. The spatial distribution of the keypoints is therefore sparse in the LiDAR range images and dense in the LiDAR intensity images. First return data tend to result in more keypoints than last return data, which was expected in advance.

### 3.3 SIFT Keypoints extracted from the Aerial Image

While LiDAR images have a resolution of 50 cm, the aerial image resolution is much higher with around 20 cm on the ground. To see the impact on the number of extracted keypoints the resolution of the aerial image is reduced to 40 cm and 80 cm

ground pixel size by Gaussian filtering and resampling. Figure 2 shows the extracted keypoints. The reduction of the resolution by a factor of 2 (linearly), i.e. 2 by 2 pixels are combined to 1 pixel, leads to a similar reduction of the number of extracted keypoints (Table 3.2).

LiDAR	Last Range	First Range	Last Intensity	First Intensity
Number of Key Points	1601	1828	11993	13393
Spatial distribution	poor	poor	good	good

Table 3.2. Dependency of SIFT keypoints from image resolution

### 3.4 Impact of Base Level Smoothing of the Datasets on SIFT Keypoint Extraction

While in the previous section smoothing and reduction of the resolution of aerial image was carried out, the base level smoothing is a filtering of the input image with the Gaussian kernel to define the base level of the scale space image according to Eqn. (2-1). Figure 3 shows the dependency of the number of extracted keypoints on base level smoothing for an aerial image, and for LiDAR First Range, Last Intensity and First Intensity. All image data sets are regions of 500 by 500 pixels. Figure 3 also shows that the curves for the aerial image and both LiDAR intensity images are close to each other at a base level smoothing with a  $\sigma$  of 1 and above. With a lower sigma value for base level filtering more keypoints are extracted from the aerial image than from the LiDAR intensities. A value of 1 for the base level  $\sigma$  was used for all further experiments. The red line indicates the very small number of extracted keypoints in the range data.

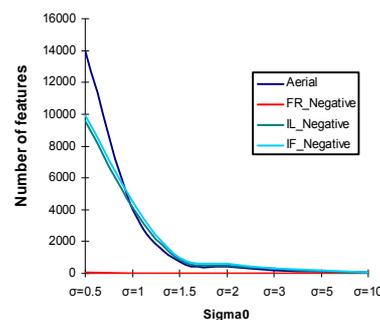


Figure 3. Impact of base level filtering on keypoint extraction

### 3.5 Impact of Octave Selection on Keypoint Extraction

Figure 4 shows that with an increasing the number of octaves an increasing number keypoints is extracted. However with more than 5 octaves the number of extracted keypoints is almost constant. Thus, 5 octaves are used for the further experiments. The size of the image was again 500 by 500 pixels. The number of extracted keypoints varies with less than 10% between aerial image and the LiDAR intensity images.

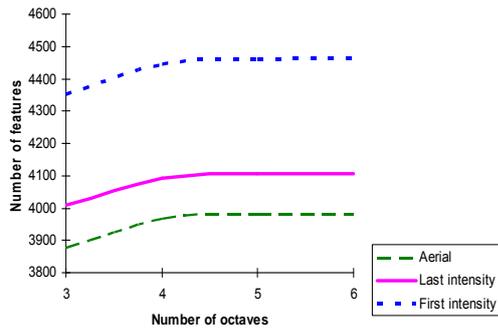


Figure 4: Dependency of matched features on the number of octaves

### 3.6 The Impact of the Number of Sub-levels per Octave on Keypoint Extraction

On each image scale represented by a certain octave a fixed number of sublevels (differently blurred images) is used in the SIFT concept to extract the keypoints as explained in Section 2.1. With only two sub-levels a fairly small number of keypoints is extracted. this increases to a maximum of features if 4 sub-levels are employed for LiDAR intensity images and 5 sub-levels for the aerial image (Figure 5). The curve for the aerial images starts from a lower number at sub-level 2 and intersects the curves for the LiDAR intensities near sub-level 4.

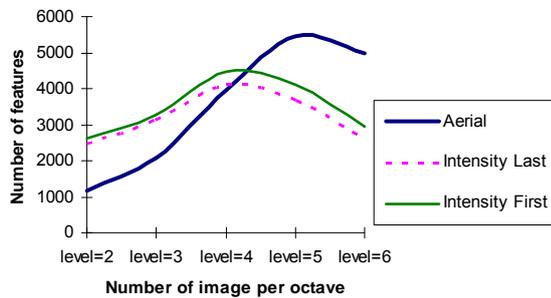


Figure 5: Impact of sub-levels on keypoint extraction

### 3.7 The Impact of the Matching Threshold on the Number of Matched keypoints

To study the dependency of the number of matched features from matching parameters again image regions of 500 by 500 pixels are used for the experiments. Matching is controlled by the threshold  $t$  (cf. Section 2.2). The lower the threshold  $t$  the more prominent should be the extracted keypoints. With values of  $t$  close to 1 a higher number of error is expected in the matched keypoints. Figure 6 shows a high sensitivity of the number of matched keypoints on the threshold  $t$ . Around a threshold of 0.85 a strong growth in the number of matched keypoints is observed. An increased base level smoothing (the red curve) produces less matched points.

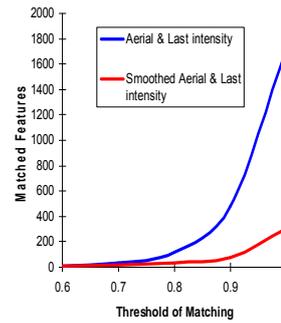


Figure 6: The number of matched features as a function of the matching threshold

### 3.8 Positive and negative LiDAR intensity images

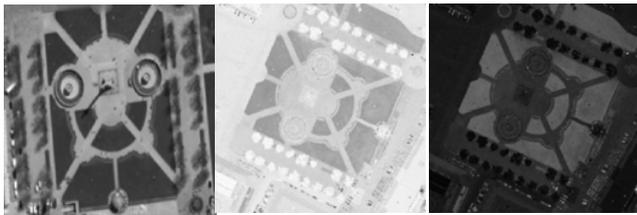
Samples of negative LiDAR intensity, positive LiDAR intensity and aerial images are depicted in Figures 7A, 7B and 7C. For the sample depicted in Figure 7A the aerial image is more similar to the positive LiDAR intensity images, whereas in Figures 7B and 7C a higher similarity of the aerial image to the negative LiDAR intensities can be observed. The reason for that is the Laserscanner operates with a frequency in the Near Infrared. Whether positive or negative LiDAR intensity images are used for keypoint extraction has almost no influence on the number of extracted keypoints. But the extracted keypoints are different which leads to a higher number of outliers if matched keypoints related to the positive LiDAR intensity images are used.

### 3.9 RANSAC and Data Snooping used for Removing Outliers

Firstly it should be emphasized that both procedures, RANSAC and Baarda's data snooping, are appropriate for removing the outliers which are among the matched keypoints. If there is a sufficient number of corresponding keypoints, which match very well both procedures eliminate the outliers efficiently. If there is a high percentage of outliers and a weak geometric distribution of the extracted points RANSAC may stuck to a local maximum of the search space. But the iterative outlier elimination also failed by eliminating some good point pairs. A simulation would probably give a more detailed insight into the pros and cons of both methods that our experiments with real image data can do.



A) Aerial image, Negative LiDAR Intensity, Positive LiDAR



B) Aerial image , Negative LiDAR Intensity, Positive LiDAR



C) Aerial image , Negative LiDAR Intensity, Positive LiDAR

Figure7: Effect of the texture

### 3.10 Visualisation of the Rectification Result

After estimating the parameters of the transformation between the images, one image can be transformed into the coordinate system of the other one. The standard method of image rectification is used for this purpose. Figure 8 shows the rectified LiDAR image, which is superimposed to the aerial image.



Figure 8 :Superimposed LiDAR and aerial images

### 4. CONCLUSIONS

Registration of aerial images and LiDAR data can be carried out successfully by using SIFT technique if the LiDAR intensity images are used for this purpose. In general, a large number of SIFT keypoints can be extracted and matched between this two data sets.

Experimental findings with our data are:

1. The number of keypoints extracted from LiDAR intensity images is 7-8 times higher than the number of extracted keypoints from LiDAR range data.
2. The grey value distribution (texturing) of a LiDAR intensity image and an aerial image might be locally very different. Sometimes the negative (inverse) LiDAR intensities fit much better to the aerial image. Those differences have to be expected as the Laser scanner operates in the near infrared while the aerial is recoding radiation of the visible electromagnetic spectrum.
3. The following parameters for extracting SIFT features are found to be most suitable: A sigma level of 1 for the base level smoothing of LIDAR and aerial images and 5 octaves and 4 sub-levels (number of images per octave) for creating the scale space representation.
4. The ratio between the distances to the first and second nearest descriptors has to be lower than a given threshold. A threshold of 0.8 produced a low outlier quota but at the expense of a small number of matched keypoints.
5. Regarding the outlier removal with RANSAC and Baarda's data snooping no clear preference could be identified.

Limitations that have been observed are:

The invariance of SIFT features to small changes in illumination does not result in an extraction of suitable features from shadow areas in an aerial image.

The polynomial transformation of second order which we used for image registration is only an approximation for a more general transformation model, e.g. a transformation based on rational functions.

### 5. REFERENCES

- W. Baarda, "A Testing Procedure for Use in Geodetic Networks," Publications on Geodesy, vol. 2, no. 5, pp. 1-97, Netherlands Geodetic Commission, 1968.
- Fischler, M.A. and R.C. Bolles, 1981. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM, 24(6):381-395.
- Habib, A., Ghanma, M. and M. Tait, 2004. Integration of LIDAR and Photogrammetry for Close Range Applications, XXth ISPRS Congress, Istanbul, Turkey.
- Hu, J., You, S. and U. Neumann, 2006. Integrating LiDAR, Aerial Image and Ground Images for Complete Urban Building Modeling. 3DPVT, Chapel Hill, USA.
- Lowe, D.G., 1999. Object recognition from local scale-invariant features. International Conference on Computer Vision, Corfu, Greece, pp. 1150 -1157.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, Vol. 60, pp. 91-110, No. 2.
- Schenk, T. and Csatho, C., 2007. Fusing Imagery and 3D Point Clouds for Reconstructing Visible Surfaces of Urban Scenes,

IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas.

SIFT: An open implementation of SIFT: <http://vision.ucla.edu/~vedaldi/code/sift/sift.html> (accessed 26. September 2007)

#### **ACKNOWLEDGEMENT**

We wish to acknowledge the support given by H. Arefi of the LiDAR research group of Stuttgart University of Applied Sciences, in particular for providing data of the Stuttgart LiDAR test site.

