Fusion of Laser Scan and Image Data for Deformation Monitoring – Concept and Perspective

Andreas WAGNER, Wolfgang WIEDEMANN, Thomas WUNDERLICH

Technical University of Munich, Chair of Geodesy
Munich, Germany
E-mail: a.wagner@tum.de, w.wiedemann@tum.de, th.wunderlich@tum.de

Abstract

Areal measurements, like laser scans or camera images, are more and more frequently used for deformation monitoring. Each single acquisition method has its advantages and disadvantages in detecting displacements, however. In dense point clouds of laser scan, distance changes in line of sight are easy to identify. In contrast, high-resolution image data is sensitive to displacements perpendicular to the viewing direction of the camera, where the analysis of scan data weakens.

A promising solution of combining the advantages and reducing the drawbacks is the fusion of laser-scan and image data in the form of RGB+D images. Point clouds are converted into a depth image, resulting in an additional D-channel to the colour (RGB) image data. In the combined RGB+D image, each pixel can be directly converted into 3D coordinates. The necessary mutual orientation of both data sets can be solved via a-priori sensor calibration or a-posteriori data registration. By identifying corresponding points in subsequent measurement epochs (RGB+D images), it is further possible to directly determine 3D displacement vectors. The results of the RGB+D method can be integrated in a rigorous geodetic deformation analysis with tests of significance.

Corresponding points in the images can be found by identifying key points and by describing them with image features, for example. The prominent representatives of image features are the SIFT algorithm, which uses floating point values, or the binary descriptor BRISK. These algorithms are used to match two images at an abstract numerical level, instead of in the original image domain.

In this analysing strategy, sensor data of any kind of acquisition system can be used, such as modern total station/multistations, mobile mapping systems, unmanned aerial vehicles or robot platforms. It will exploit the full potential of sensor data of such systems, which already provide both kind of data, for the first time.

Key words: deformation monitoring, geodetic monitoring, image matching, laser scanning, RGB+D, sensor fusion

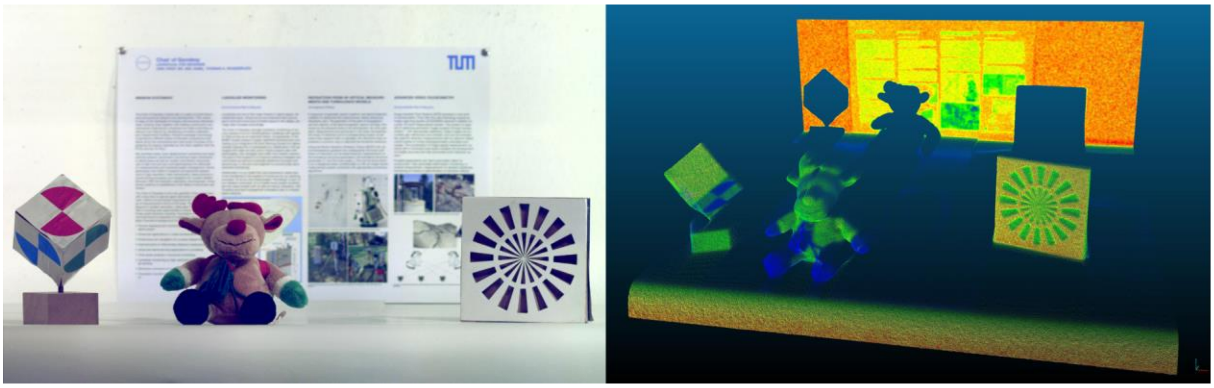
1 INTRODUCTION

A fundamental change in geodetic deformation monitoring is currently taking place: Nowadays, areal measurements are increasingly used instead of pointwise observations (of manual selected discrete points) as done in the past. Frequently, laser scanners are used, as these allow a fast, high resolution and dense acquisition of 3D information. However, it is difficult to uncover deformations and changes in multi-temporal point clouds as no discrete points are measured, and only changes in line of sight (of non-signalised areas) can be clearly detected automatically as distance variation. Further, rigorous deformation analysis and tests of significance of the results are missing (Wunderlich, Th. et al., 2016). Image analysis techniques, in contrast, are very sensible for identifying displacements in transverse direction. Hence, a combined acquisition and analysis of point cloud and image data complement each other very well.

A promising approach of such kind of fused laser scan and image data is the usage of RGB+D images (Wagner, A. 2016). In the following article, the individual task to generate such kind of images and the combined analysis are presented. For demonstration purposes, an artificial test scene is set up as described in the following section. The acquisition and method to generate a RBG+D image is presented in Section 3. If two or more epochs are measured, it is possible to derive 3D displacement vectors as shown in Section 4. A summary and outlook (Section 5) completes the article.

2 TEST SCENE

Fig. 1 shows the data of an artificial scene taken to demonstrate the procedure and result of the presented method. A high-resolution camera image (RGB) is depicted on the left, showing three different objects in front of a textured background. Fig. 1b) depicts the acquired laser scan data (XYZ) coloured according to the measured intensity values (I). Both data sets are stored within the same coordinate system, as the extrinsic parameters of the camera and the laser scan are known through an a-priori calibration.



a) b)

Fig. 1 Artificial scene: a) high-resolution camera image (RGB), b) Laser scan data (XYZ+I)

The middle foreground object (the toy) is placed at three different positions as sketched in Fig. 2. The different measurement epochs are recorded: out of the first position, the toy is shifted a couple of centimetres to the left and as third epoch moved into the direction of the camera/laser scanner.

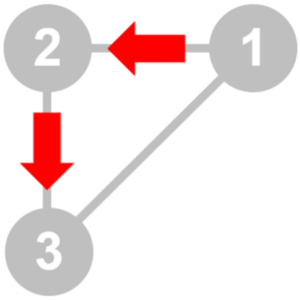


Fig. 2 Positions (top view) of the moving object in the respective epoch

3 RGB+D IMAGE

3.1 DATA ACQUISITION

To generate a RGB+D image, it is necessary to acquire image data and laser scan data of the same scene. This can be done by a single instrument, like a modern total station (multistation). These instruments are able to record high-resolution panoramic image mosaics through camera rotation, while each single image is accurately geo-referenced and oriented by high-accurate angle readings. The latest instrument releases are additionally equipped with a scanning function comparable to that of classical terrestrial laser scanners. However, by far lower scan rates. The greatest advantage of using such a (multi-sensor) instrument is the common coordinate system of all built-in sensors – an appropriate a-priori calibration provided (Wagner, A. 2016).

Another way to determine the required data is to use different instruments, possibly also from different locations. The necessary (a-posteriori) registration of both data sets must then be solved e.g. by tie points. Other approaches are image feature matching (Paar, G. et al. 2005) or the use of mutual information (Taylor, Z. et al. 2012) of the intensity values or geometrical derived information and image data.

3.2 DEPTH IMAGE

The laser scan data is further processed to use it in combination with the image data: each single point is (perspectively) projected into the camera image. In the case of a rotating camera (multistation, pan-tilt-system), it is useful to generate a spherical image panorama and consequentially also a spherical depth image. In case of different acquisition stations of the camera and the laser scan, it is necessary to filter hidden points. It must be ensured that only points are projected into the image, which could be visible in the camera image. This problem is known as z-buffer problem and different solutions are published, e.g. from Qin, R. et al. (2014).

The transformation results in an irregular grid of 2D points and a distance from the camera projection centre to each of them. A depth image can be created by interpolating the points into a regular grid with the same resolution of the image/panorama. The initial scan data is now transformed into an image in which the pixel values (in grayscale or colour) are representing the object distance. This depth image constitute the D-cannel of the RGB+D image, which is an image with four channels: red, green, blue, and depth, as shown in Fig. 3.

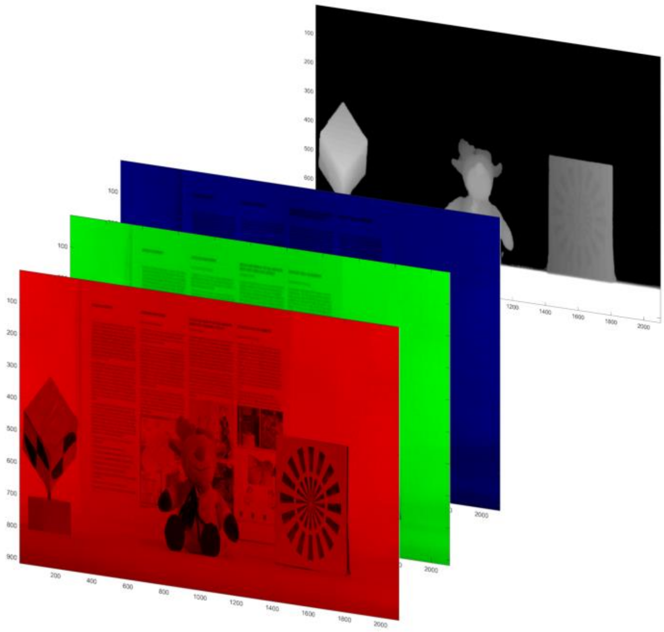


Fig. 3 RGB+D image

3.3 PROPERTIES

Main advantage of a RGB+D image is the fact that each single pixel position can directly be transformed into 3D world coordinates. With the extrinsic parameters of the camera, it is possible to determine the spatial ray of each pixel coordinate, which originates from the known camera projection centre. Together with the distance information of the D-channel, this yields spherical coordinates that can be easily converted in 3D Cartesian ones.

It should be noted that the accuracy of the depth information – and therefore the accuracy of the resulting 3D coordinates – is highly dependent on the ground sampling distance of the laser scan and the interpolation method. Assuming a high-resolution camera image or stitched panorama from a rotating platform, respectively, the image resolution is mostly higher than the distance resolution and the scan density.

4 DEFORMATION ANALYSIS

The RGB+D approach combines laser scan and image data. When comparing the sensibility for object detection of these two different data sets, it appears that they complement each other very well. Changes in line of sight can be clearly detected in consecutive laser scan data as distance variation. In image sequences, such kind of displacements appear as object scaling and are hard to detect or remain uncovered at all. On the other hand, image matching algorithms are very sensible for identifying displacements in transversal direction, where the evaluation of laser scanner data weakens (in particular with weak-structured surfaces).

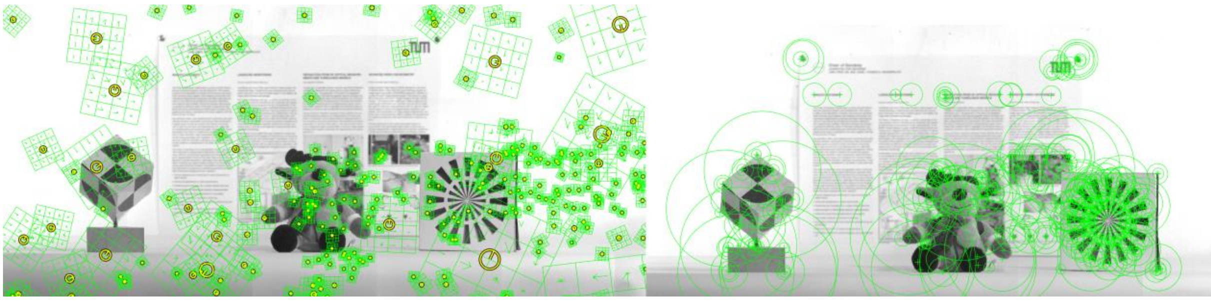
By identifying corresponding points in subsequent measurement epochs (RGB+D images), it is possible to directly determine 3D displacement vectors. As mentioned above each pixel position of a RGB+D image can be converted into 3D coordinates. If the same object point is found in both images (epochs), its pixel positions correspond to the start- and endpoint of the 3D displacement vector, which could be subject to classical testing theory.

4.1 FEATURE DESCRIPTION

To extract corresponding points or regions in images, various approaches exist in photogrammetry and/or computer vision. These methods are based on intensity values, e.g. normalised cross-correlation or least squares matching, others are using geometric or other relations between features and structures. A third technique is to describe key points by computing abstractions in the form of feature vectors and matching them against each other.

Most prominent representative of the last mentioned approach is the SIFT (Scale Invariant Feature Transform) algorithm (Lowe, D. 2004). It is based on the calculation of dominant gradient orientations around key points at a current scale. The values are weighted and aggregated into histograms of orientated gradients, forming a (normalised) vector with a length of 128 entries. Main feature of this method is its invariance to image scale and orientation changes, which provides robust matching at a substantial range of affine distortion, 3D viewpoint changes, addition of noise, and change in illumination (Lowe, D. 2004).

The newer generation of feature descriptors is focused on lower computational costs by retaining similar matching accuracy and repeatability. These descriptors are using bit strings instead of vectors of floating point values, which significantly improves the computation time. One representative of these binary descriptors is the BRISK (Binary Robust Invariant Scalable Keypoints) algorithm (Leutenegger, S. 2011). The descriptor contains three different parts: a sampling pattern, an orientation compensation and sampling pairs. The pattern contains regular spaced concentric circles, which centre points forming sampling pairs. The local gradients of long-distance pairs are used to estimate the pattern orientation. The differences in the smoothed intensity values of the short-distance pairs form the descriptor values itself.



a) b)

Fig. 4 Calculated features of the test scene using different algorithms: a) SIFT b) BRISK

Fig. 4 depicts a subset of the calculated key points and descriptors using the above-described algorithms at the test scene. The centres of the squares and circles respectively, represent the position of the keypoints. Its sizes symbolise the scale at which the individual descriptor is determined.

It should be noted that the distribution of the features is highly volatile. Depending on the proposed surveillance area, the most appropriate method must be carefully chosen.

4.2 MATCHING

To find corresponding points in two measurement epochs the above mentioned feature vectors can be used. The task to find similarities in both images is executed at an abstract numerical level, by searching for comparable vectors via distance functions. This means in detail that for each feature in the first image the one with the minimal distance in the second image is searched. One approach is to use the sum of squared differences (SSD) between the entries of the two descriptors. A matching of two binary vectors can be done by calculating their Hamming distance (Hamming, R. W. 1950) with a simple XOR operation.

Fig. 5 Result of a SIFT matching of two different epochs/toy positions

Fig. 5 depicts the result of a SIFT matching of epochs 2 & 3. The red circles indicate matched features in epoch 2, the green crosses corresponding matches of key points in epoch 3. Both markers are connected by yellow lines, representing the displacements between both epochs. In the figure, no heavy movements could be identified. The pixel positions of most correspondences coincide with each other.

A 3D view of the same scene is shown in Fig. 6. Here, it is clearly visible that these displacement vectors are de facto 3D lines. As described above each pixel coordinate of a RGB+D image can be converted into 3D coordinates and consequentially to positions of a 3D vector. The example shows the excellent suitability of RGB+D images to detect displacements. Analysing only the image data, the movement of the toy could not be identified as it happened in line of sight. Combined with the distance information of the depth image, it is possible to exactly specify the displacements.

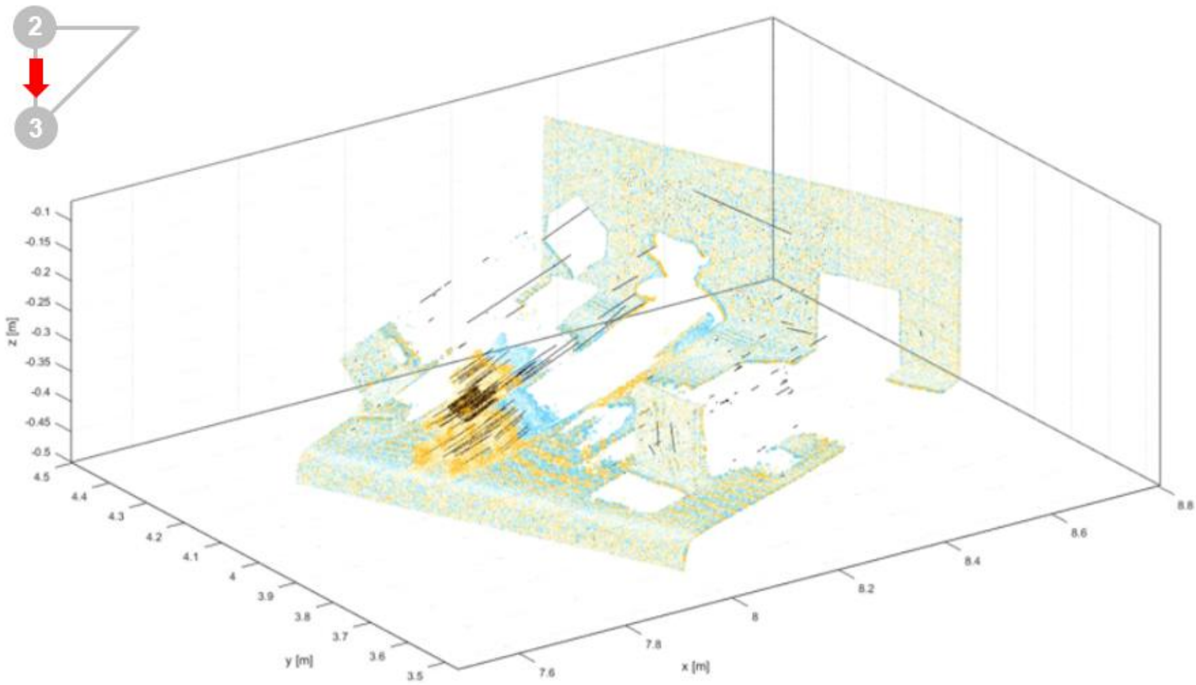


Fig. 6 3D view of the feature matching result

Some outliers are visible in Fig. 6, in particular at the borders of the foreground objects. To demonstrate the raw capacity of the method, the current result is not filtered at all. However, it would be easy to remove the outliers by comparing each vector with the orientation and length of the neighbouring elements. Additional constraints could be imposed by a forward-backward matching strategy including an appropriate threshold or by a special treatment of regions with high three-dimensional variability.

The generated data can be further processed in a classical geodetic deformation analysis, as discrete points are identified. In combination with additional measurements to stable control points, a rigorous geodetic deformation analysis with tests of significance can be calculated. Thus, it is possible to determine statistically verified results with an accuracy specification of each measurement. Concerning the inner accuracy, it could be noted that the start and end points of the displacement vectors are determined with subpixel precision. That means the image analysis is not the limiting factor of the accuracy of the result. Critical are the resolution of the image and the sampling distance of the laser scan, as described earlier.

5 SUMMARY & OUTLOOK

The presented approach for deformation monitoring – the fusion of laser scans and camera data to RGB + D images – combines the advantages of the two measurement methods. With laser scans and the resulting dense point clouds, distance changes in line of sight can be easily detected. High-resolution image data, in contrast, is most sensitive to displacements perpendicular to the viewing direction of the camera. In this way, displacements can be successfully discovered by extracting common image feature and track them over multiple measurement epochs. In combination with the depth channel, the detected image features represent 3D coordinates, which enables the calculation of 3D displacement vectors.

In this analysing strategy, sensor data of any kind of acquisition system can be used, such as modern total station/multistations, mobile mapping systems, unmanned aerial vehicles or robot platforms. It will exploit the full potential of sensor data of such systems, which already provide both kind of data, for the first time.

Future developments will be the usage of 3D point cloud features within the analysing procedure, such as e.g. 3-dimensional SIFT (Scovanner, P. 2007) or Point Feature Histograms (Rusu, R. 2009). These 3D features – similar to the images processing ones – represent the local neighbourhood of key points in the point cloud with a feature vector, which can be matched and tracked over time. Ideally, this ends up in the development of combined features of image and laser scan data.

REFERENCES

Hamming, R. W. 1950. Error detecting and error correcting codes. Bell System Technical Journal. Vol. 29, No. 2 (1950), pp. 147–160.

Leutenegger, S. - Chli, M. - Siegwart, R. Y. 2011. BRISK: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. 6-13 Nov. 2011. Barcelona, Spain, pp. 2548–2555.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*. Springer. Vol. 60, No. 2 (2004), pp. 91–110.

Paar, G. - Bauer, A. - Kontrus, H. 2005. Texture-based fusion between laser scanner and camera for tunnel surface documentation. In *Proceedings of the 7th Conference on Optical 3-D Measurement Techniques*, 3-5 Oct 2005. Vienna, Austria, pp. I-74–80.

Qin, R., - Gruen, A. 2014. 3D change detection at street level using mobile laser scanning point clouds and terrestrial images. In *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier. Vol 90 (April 2014), pp. 23–35.

Rusu, R. 2009. Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. *PhD Thesis*. Technical University of Munich. Munich, Germany. 284 pp.

Scovanner, P. - Ali, S. - Shah, M. 2007. A 3-Dimensional SIFT Descriptor and its Application to Action Recognition. In *Proceedings of the 15th ACM international conference on Multimedia*. 23-28 Sept 2007. Augsburg, Germany. pp. 357–360.

Taylor, Z. - Nieto, J. 2012. A Mutual Information Approach to Automatic Calibration of Camera and Lidar in Natural Environments. In *Proceedings of Australasian Conference on Robotics and Automation*, 3-5 Dec 2012, Wellington, New Zealand. 8 pp.

Wunderlich, Th. - Niemeier, W. - Wujanz, D. - Holst, Ch. - Neitzel, F. 2016. Areal Deformation Analysis from TLS Point Clouds – The Challenge. In *avn. allgemeine vermessungs-nachrichten*. Wichmann. Vol. 123, No. 11-12 (2016), pp. 340–351.

Wagner, A. 2016. A new approach for geo-monitoring using modern total stations and RGB+D images. In *Measurement*. Elsevier. Vol. 82 (March 2016), pp. 64–74.