PFG 2012 / 5, 0535 – 0545
Stuttgart, October 2012

Article

Identifying Correspondences in Sparse and Varying 3D Point Clouds using Distinctive Features

Daniel Muhle, Hannover, Steffen Abraham, Hildesheim, Manfred Wiggenhagen & Christian Heipke, Hannover

Keywords: photogrammetry, matching, point cloud

Summary: In a wide range of applications stereo systems are used to extract geometric information from the scene observed with the stereo cameras. One possible solution to reconstruct the motion of such a system is to establish correspondences between points of the point clouds generated from stereo matching of image features at different epochs. There exists a large variety of approaches to establish correspondences between image or 3D data. A special group of algorithms, mostly inspired by the work of Lowe (2004), is based on the notion of distinctive feature descriptions. These algorithms assume the existence of a dense neighbourhood changing not too much over time. But the prevalence of untextured regions or computational constraints hindering the use of computationally expensive dense stereo matching approaches often result in only sparse point clouds and thus these approaches cannot be used for the registration of sparse 3D data. In our work we present a new approach that uses the basic principles of distinctive feature descriptions and extends them in a way that they can be applied to identify corresponding points between sparse 3D point clouds. Furthermore, an evaluation is given investigating the advantages and limitations of our approach. The results clearly show the effectiveness of the presented distinctive features to establish point matches between sparse 3D point clouds.

Zusammenfassung: In vielen unterschiedlichen Anwendungsbereichen werden Stereosysteme verwendet, um geometrische Informationen über die aufgenommene Szene zu extrahieren. Eine dabei anfallende Teilaufgabe ist das Identifizieren von Korrespondenzen zwischen Punkten einer 3D Punktwolke, die zu unterschiedlichen Zeitpunkten durch das Stereomatching von Bildmerkmalen entstanden ist. Inspiriert durch die Arbeit von Lowe (2004) sind für die Suche nach korrespondierenden Punkten eine ganze Reihe von Ansätzen entstanden, die auf charakteristischen Beschreibungen aufsetzen. Alle diese Verfahren setzen das Vorhandensein einer dicht besetzten Nachbarschaft voraus, die sich über die Zeit hinweg nicht zu stark ändert. Allerdings führen untexturierte Bereiche oder Echtzeitanforderungen, die den Einsatz von rechenintensiven dense-matching Ansätzen verbieten, zu dünn besetzten 3D Punktwolken, so dass die bekannten Verfahren nicht unmittelbar verwendet werden können. In unserer Arbeit wird ein neuartiger Ansatz vorgestellt, der auf den Grundprinzipien der charakteristischen Beschreibungen aufbaut und diese so erweitert, dass sie für die Punktzuordnung in dünn besetzten 3D Punktwolken geeignet sind. Darüber hinaus wird eine Untersuchung vorgestellt, die die Vorteile und Grenzen des entwickelten Ansatzes aufzeigt. Die Ergebnisse zeigen deutlich die Leistungsfähigkeit der entwickelten charakteristischen Beschreibung für die Zuordnung von dünn besetzten 3D Punktwolken.

1 Introduction

One task towards reconstructing the motion of a stereo system, e.g. used as a robot's eyes as it traverses through its environment, is to establish correspondences between points of the point clouds reconstructed from stereo matching of image features at different epochs. There exists a variety of approaches to establish correspondences between image or 3D data. A special group of algorithms, mostly inspired by the work of Lowe (2004), is based on the notion of distinctive feature descriptions. Typical examples of these approaches

www.schweizerbart.de
DOI: 10.1127/1432-8364/2012/0137
1432-8364/12/0137 $ 3.00

are SIFT (Lowe 2004) and Spin Images (Johnson & Hebert 1999) for the registration of 2D and 3D data. Most of these approaches cannot be used directly for the registration of sparse 3D data, though, as they assume the existence of a dense neighbourhood that does not significantly change over time. Sparse and varying point clouds must be expected if such a system is used in a man-made environment with untextured regions on floors and walls or if real-time constraints hinder the use of computationally expensive dense stereo matching.

In our work we present an approach that uses the basic principles of distinctive feature descriptions and extends them in a way that they can be applied to identify point matches between sparse point clouds. The resulting distinctive feature vector is sparse and discrete and can be used to establish correspondences efficiently. After discussing the state-of-the-art in section 2, a detailed description of the proposed algorithm is given in section 3. The performance of the presented approach is tested on different sequences of a stereo system moving along a corridor. The established correspondences between points are used to reconstruct the motion of the system. Furthermore, in section 4 an evaluation is given investigating the overall performance and limitations of our approach. The results clearly show the effectiveness of the presented distinctive features for the matching between sparse 3D point clouds. Section 5 presents some ideas for future extensions.

2 Related Work

The point clouds that are generated at the epochs $t = i$ and $t = j$, while e.g. a stereo system mounted on a robot platform traverses down a hallway, are related by a rigid transformation ${}^{j}\mathbf{T}_{i}$, where the symbol ${}^{j}\mathbf{T}_{i}$ combines the rotation ${}^{j}\mathbf{R}_{i}$ and the translation ${}^{i}\mathbf{t}_{j}$. The rotation ${}^{j}\mathbf{R}_{i}$ rotates a point ${}^{i}\mathbf{x}$, defined in the coordinate system at $t = i$ into the coordinate system at $t = j$. The translation ${}^{i}\mathbf{t}_{j}$ between the epochs is defined w.r.t. to the coordinate system at $t = i$. Given the point clouds from two epochs one seeks to find the transformation ${}^{j}\mathbf{T}_{i}$ minimizing the error e given in (1) for all n corresponding points.

$$e = \sum_{k=1}^{n} \left\| {}^{i_k}\mathbf{x} - {}^{i}\mathbf{R}_{j}\, {}^{j_k}\mathbf{x} + {}^{i}\mathbf{t}_{j} \right\| \tag{1}$$

A prerequisite for the solution of this task is to identify point correspondences between the point clouds at $t = i$ and $t = j$. If the correspondences are known, established approaches like ICP (iterative closest point) (Besl & McKay 1992) and its variants (Rusinkiewicz & Levoy 2001) or least squares matching (Grün & Akca 2005) can be used to find the optimal solution for ${}^{j}\mathbf{T}_{i}$ that minimizes the error given in (1). As the focus of our work is an algorithm to establish point correspondences between sparse and varying point clouds we first give in section 2.1 a short overview about the generally applied matching workflow, and show in section 2.2 the general ideas behind the use of distinctive features for the task of identifying correspondence either in 2D or 3D data. In section 2.3 we give a detailed explanation of the contribution of our work. For the experiments presented in section 4 signalized targets are used that can be identified easily in the stereo image pairs to create a sparse 3D point cloud. Approaches that allow the derivation of a sparse point cloud from dense but irregularly sampled point clouds can be found e.g. in Stückler & Behnke (2011) and Novatnack & Nishino (2007).

2.1 *Common Matching Workflow*

Overviews of approaches for the matching of a large variety of input data can be found e.g. in Brown (1992), Seeger & Laboureux (2000), Zitová & Flusser (2003) or McGlone et al. (2004, chap. 6.3). In general, approaches used for the matching of a variety of input data, follow a similar scheme:

Defining the feature space

The feature space defines the input data used for matching. Typical examples for the 2D and 3D case are grey values, gradients, corners and 3D point clouds. A feature is an element taken from the set of the input data.

Defining the parameter space

The dimension of the parameter space is defined by the choice of transformation used for mapping of the input data. Typical transformations that are often used in the context of feature matching are homographies, affine and Euclidean transformations.

Establishing assignments using a similarity measure

Matched features are identified by defining a similarity measure that is computed either for all combinations or a reasonable subset of the used/extracted features. Two features can for instance be matched if a) their respective similarity is the largest and b) the similarity is discriminative, i.e. it is above and the ratio between the second-best and the best match is below a pre-defined threshold.

2.2 *Matching with Distinctive Features*

Recently, feature representations that are unique and distinctive are widely used in the area of photogrammetry and computer vision. For these representations two different terms are used in the literature: descriptor and signatures. The differentiation between these terms is not always clear and sometimes they are used ambiguously. In Lowe (2004) a descriptor is defined as a distinctive and compressed representation of the original input data. In contrast, Calonder et al. (2008) define a signature as the result of a mapping $F : \mathfrak{R}^D \rightarrow \mathfrak{R}^d$ that transforms the input data $\mathbf{x}_k \in \mathfrak{R}^D$, $\forall k = 1 \ldots n$ with the dimension D into another space with the dimension d. To avoid any ambiguities, we will use the term distinctive (feature) description in the following. The most important properties of distinctive descriptions independent of their actual realization are:

- The distinctive description is invariant w.r.t. a variety of changes of the input data. Typical changes comprise e.g. illumination changes, translation, rotation and/or scaling.
- The computation of a similarity measure between any two descriptions can be done using simple distance metrics like the Euclidean distance.

Examples for distinctive description used to match two-dimensional image data are the SIFT and SURF descriptors (Lowe 2004, Bay et al. 2008). For the identification of corresponding points in point clouds derived from either range scanner like e.g. the Microsoft Kinect system or terrestrial laser scanner, Flint et al. (2008), Wang & Brenner (2008), Lo & Siebert (2009), Barnea & Filin (2010) and Weinmann et al. (2011) directly apply variants of the SIFT algorithm that consider the special characteristics of the available 3D data. All these approaches require that the irregular 3D data must be resampled to a regular two-dimensional grid. Furthermore, the point signatures (Chua & Jarvis 1997), spin images (Johnson & Hebert 1999), the approach of Gelfand et al. (2005) and the NARF (normal aligned radial feature) developed by Steder et al. (2011) are more examples of algorithms that use distinctive feature descriptions to establish matches between points from different point clouds. All these approaches have in common, that they incorporate points or information from a densely sampled local neighbourhood to define a unique local frame of reference. The definition of the local reference frame is usually the first step to achieve invariance against rotation and translation of the input data. A well-designed computation of the distinctive description allows the usage of simple distance metrics for the matching step and makes it robust against other changes of the input data like scaling or change of illumination. Furthermore, the developed feature descriptions simplify the matching step and established approaches like clustering or binary space-partitioning trees can be used to accelerate the search for matching descriptions (e.g. Winkelbach & Wahl 2008, Nistér & Stewenius 2006).

2.3 *Contribution of Our Work*

Most of the known approaches that rely on distinctive descriptions to find matching point pairs between point clouds or surface data require the existence of a densely sampled and

regular neighbourhood that does not change too much over time. Typical examples for data fulfilling such a requirement are the results from dense stereo matching or data acquired with a laser scanner. For these types of input data it is relatively easy to establish a unique frame of reference to be robust against rotation and translation. For systems generating sparse point clouds with local neighbourhoods that change over time as new points become visible and other points move out of the field of view, the known approaches using distinctive descriptions for matching cannot be used directly. The algorithm presented in section 3 extends the existing approaches for matching 3D points and presents solutions to:

- achieve invariance against rotation and translation of the input data for sparse and varying point clouds,
- compute a distinctive description that is just as sparse and varying as the input data,
- efficiently compute a similarity measure for the sparse and varying distinctive description.

3 Matching between Sparse and Varying Point Clouds

The proposed scheme to find matching point pairs in sparse and changing 3D point clouds follows the workflow of the existing approaches that use distinctive descriptions for matching. The first step is the definition of a local frame of reference to achieve invariance against rotation and translation of the point cloud. In the second step a sparse 2D distinctive description **D** is computed from selected points in a local neighbourhood. For the third step the sparseness of the 2D distinctive description is exploited to derive a compact 1D description that allows an efficient computation of similarity between two distinctive descriptions.

3.1 *Identification of Locally Planar Neighbourhoods*

The first step for the computation of the distinctive description for a point **x** in a sparse and varying point cloud is identical to the approaches mentioned in section 2.2: the estimation of a plane normal **n** from points in a local neighbourhood. To increase the probability that the normal does not change when neighbouring points disappear or new ones enter the camera's field of view only those points are selected for the computation of **n** that lie approximately on the same plane as **x**. For the estimation of **n** and the identification of points lying in the same plane as **x** a brute-force approach using the RANSAC algorithm (Fischler & Bolles 1981) is applied. The functional model is the Hessian normal form given by (2), where d represents the distance of a point **x** to the plane with its normal **n**.

$$\mathbf{n} \cdot \mathbf{x} = d \tag{2}$$

The simple brute-force approach used here is sensible as the processed point clouds contain only few points lying mostly on the planar walls of a hallway. For these constraints the RANSAC algorithm needs only a small number of iterations to find a good solution. For the selection of the local neighbours we define the following parameters: a) maximum number n_{max} of neighbours considered, b) minimum number n_{min} of neighbours considered, and c) a maximum radius r_{max} used in the nearest neighbour search. As a result of this first step we have associated every point **x** of a point cloud with a locally planar neighbourhood, a normal vector **n** and a set $\mathbb{P} = \{x_1 \ldots x_n\}$ of its n neighbours. Those points with a non-planar neighbourhood are not considered further in the matching process.

3.2 *The 2D Distinctive Description*

The computation of the entries $\mathbf{d}_k$ for the two dimensional description **D** is done in three steps. First, for a neighbour $\mathbf{x}_i$ a local frame of reference is defined where the x-axis is given by the direction from **x** to $\mathbf{x}_i$ projected into the local plane and the z-axis is given by the local normal **n**. The y-axis is computed from the cross product of the x- and z-axis. Second, as shown in Figs. 1(a) and 1(b) all neighbours $\mathbf{x}_j : \forall j = 1 \ldots n$ are projected into the x, y-plane of the local reference frame and transformed into two dimensional polar coordi-

nates. Fig. 1(b) shows that for all neighbours $\mathbf{x}_j$ the radial distance d_j w.r.t. x and the angle θ_j w.r.t. the x-axis of the local reference frame are used as entries into **D**. Third, the steps one and two are repeated for all remaining neighbours. In contrast to e.g. the point signatures (Chua & Jarvis 1997), that use exactly one reference direction to define a local frame of reference, the approach presented here defines a local frame of reference for every neighbour. Such a strategy is advantageous if the structure of a local neighbourhood changes over time. If that is the case it is not advisable to select one of the neighbours as a reference direction, because that point might disappear from the stereo system's field of view and then a new reference direction must be selected and thus the distinctive description changes completely. The proposed computation of the entries $\mathbf{d}_k : \forall k = 1 \ldots n^2$ for the distinctive description **D** also fulfils the requirement of invariance w.r.t. rotation and translation of the point cloud. The computed entries (relative angles and distances) are not changed by a rotation around the z-axis and the use of a local frame of reference eliminates the influence of a translation. The distinctive description is not invariant w.r.t. scaling that changes the length of distances. Scale invariance can be achieved, however, when ratios of distances are used instead of distances.

3.3 *A Compact Distinctive Description*

It is clear from the algorithm in section 3.2 and Fig. 1(b) that the two dimensional distinctive description is still sparsely populated. This special structure of **D** allows, analogously to the idea given in Calonder et al. (2009), to design a more compact variant of the distinctive description. For a compact version of **D**, the non-discrete entries $\mathbf{d}_k : \forall k = 1 \ldots n^2$, $\mathbf{d}_k \in \mathbf{D}$ are mapped into an integer scalar $s_k : s_k \in \{1, 2, \ldots 2p\}$, where p is a parameter of the mapping and influences the discretisation. Finally, the s_k are pooled in an ordered result set $\mathbb{S}$ containing only unique values. As shown in section 3.4, the usage of integer values is advantageous as it allows an efficient implementation for the comparison of two distinctive descriptions. The s_k are computed by applying the mapping $\mathrm{T} : \mathbf{d}_k \rightarrow s_k$ to all entries $\mathbf{d}_k$ in **D**. For T the quad tree index (Finkel & Bentley 1974) is used that recursively divides the 2D space into discrete grids and gives an integer index for a two dimensional point. The only parameter of the quad tree index is p defining the number of quadrants used for the partitioning of the two dimensional space. This parameter controls the loss of accuracy caused by the discretization. For a chosen value of $p = 16$ the x-axis (radial distances d_j) and the y-axis (angle θ_j) will be partitioned into $2^{p/2} = 256$ bins. The value of

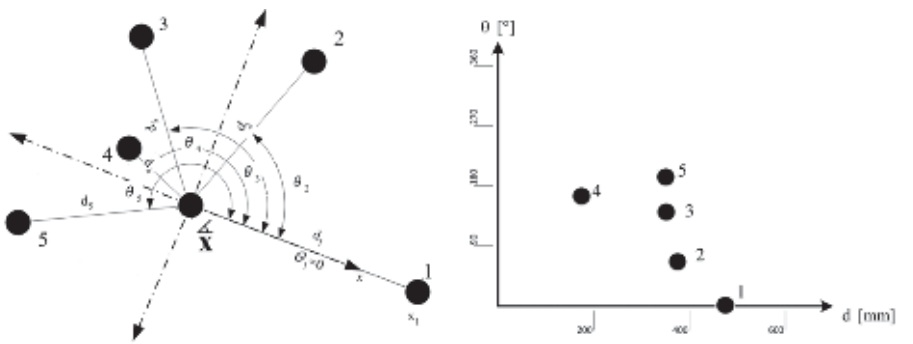


Fig. 1: Example for one iteration of the computation of the sparsely populated distinctive description **D**. (a) Definition of a local frame of reference. The x-axis is given by the direction to point 1; the z-axis points toward the reader. The n neighbours are transformed into the local frame and converted to polar coordinates (distance and direction). (b) Part of the descriptor **D** computed from the neighbourhood shown in (a). The entries into **D** are the polar coordinates of all neighbours for the local frame of reference.

p must be chosen w.r.t. to the density of the point cloud and must be large enough to avoid the event that neighbouring points fall into the same bin and are thus assigned the same index.

3.4 *Matching of 3D Points*

In order to determine the change of orientation between two epochs, it is necessary to establish correspondences between individual points of both point clouds. To find matched points a similarity measure $d_{i,j}$ is computed for all possible combinations of the n respectively m compact distinctive descriptions $\mathbb{S}_i$ and $\mathbb{S}_j$ for $t = i$ and $t = j$, where $n = |\mathbb{S}_i|$ and $m = |\mathbb{S}_j|$. The chosen similarity measure should support reliable matching of distinctive descriptions even if they match only partially and have a different number of entries. From section 3.3 it is clear that the proposed distinctive description encodes the structure of the local neighbourhood in a one-dimensional vector of unique integers: that means if two descriptions have identical entries it is very probable that they encode the structure of the same neighbourhood. Thus a possible similarity measure for the identification of matched points can be defined by the overlap of the two ordered and unique sets $\mathbb{S}_i$ and $\mathbb{S}_j$.

The first step to compute the overlap consists in determining the intersection $\mathbb{S}_\cap = \mathbb{S}_i \bigcap \mathbb{S}_j$ and the union $\mathbb{S}_\cup = \mathbb{S}_i \bigcup \mathbb{S}_j$ for the distinctive descriptions $\mathbb{S}_i$ and $\mathbb{S}_j$. Both the intersection and the union can be found efficiently as we are using sets of integer values for which a comparison of two values is very fast.

The similarity $d_{i,j} : d_{i,j} \in [0\ldots1]$ of two distinctive descriptions is then given by (3).

$$d_{i,j} = \frac{|\mathbb{S}_\cap|}{|\mathbb{S}_\cup|} = \frac{|\mathbb{S}_\cap|}{|\mathbb{S}_i| + |\mathbb{S}_j| - |\mathbb{S}_\cap|}. \tag{3}$$

Two descriptions are matched if their matching score $d_{i,j}$ is the highest (greedy approach) and is above a pre-defined threshold t_i. The threshold t_l can be computed by defining a minimum number n_{min} of neighbours that must be visible at $t = i$ and $t = j$ for a successful match. A formula for the computation of t_l is given in (4) where $n = |\mathbb{S}_i|$ and $m = |\mathbb{S}_j|$.

$$t_l = \frac{n_{min}^2}{n + m - n_{min}^2} \tag{4}$$

At the end of section 3.3 it is mentioned that it is important to select a value of p according to the point density of the observed 3D point clouds. If p is too small two neighbouring points might fall in the same bin and are assigned the same index. Such an event reduces the total number of entries in the distinctive description $\mathbb{S}$, because we allow only unique values to be present. As a result the identification of matching points might fail because the minimum number of identical entries (see (4)) is not reached.

3.5 *Robust Filtering of Matches*

Applying the matching scheme described above to point clouds generated by a stereo system at $t = i$ and $t = j$ results in a set of m candidate pairs $\{({}^{i_k}\mathbf{x}, {}^{j_l}\mathbf{x})\}$ for point correspondences. Following the insights of Sattler et al. (2009) we use a RANSAC based approach to eliminate wrong correspondences and to estimate the change of orientation between two epochs. The functional model used by the RANSAC algorithm is given in Weng et al. (1992) and needs at least three non-collinear correspondences to compute the change of orientation ${}^{j}\mathbf{T}_i$ between the two epochs $t = i$ and $t = j$.

In every iteration of the algorithm three pairs of correspondences are chosen randomly. In order to evaluate the quality of each hypothesis all points at $t = j$ are transformed into the epoch $t = i$ using ${}^{j}\mathbf{T}_i$ and a best match for every transformed point is identified with a nearest-neighbour search. The quality of the current hypothesis e_{hyp} is given by (5) where n is the number of matches identified by the nearest-neighbour search and $\|\cdot\|$ is the L2-norm of a vector.

$$e_{hyp} = \sum_{k=1}^{n} \left\| {}^{i_k}\mathbf{x} - {}^{i}\mathbf{R}_{j}\, {}^{j_k}\mathbf{x} + {}^{i}\mathbf{t}_{j} \right\| \tag{5}$$

The orientation ${}^{i}\mathbf{T}_j$ is computed from all correspondences found by the nearest-neighbour search, of the best hypothesis according to the functional model given in Weng et al. (1992).

4 Experimental Results

To evaluate the performance of the proposed matching scheme three different image sequences were recorded with a multi-stereo system traversing down a hallway. The two stereo systems A and B were mounted on a mobile platform in a way that the respective fields of view faced the opposite walls. The performed motion patterns were a pure planar motion without any rotation (E1), a planar motion where a rotation was only possible around the normal of the ground plane (E2) and a free motion with rotation around all axis (E3). An example of the planar motion of case E2 is given in Fig. 2. The sequences were recorded with a frame rate of 15 Hz and the stereo systems moved at a speed of approximately $2\,\frac{\mathrm{m}}{\mathrm{s}}$ between 2 and 4 metres down the hallway. The total number of frames captured in the different experiments is given in Tab. 1.

For the experiments signalized circular targets that can be identified reliably in the images were fixed on the walls of the hallway. The position of the centre of these targets is found using an ellipse measurement algorithm (Luhmann 1986). The result of this feature extraction step is a list of image coordinates for the targets in both images of a stereo pair. With the known epipolar geometry of the stereo system corresponding targets can be identified easily using the approach in Otepka et al. (2002), where the fact is exploited that matching points in both images of the stereo system

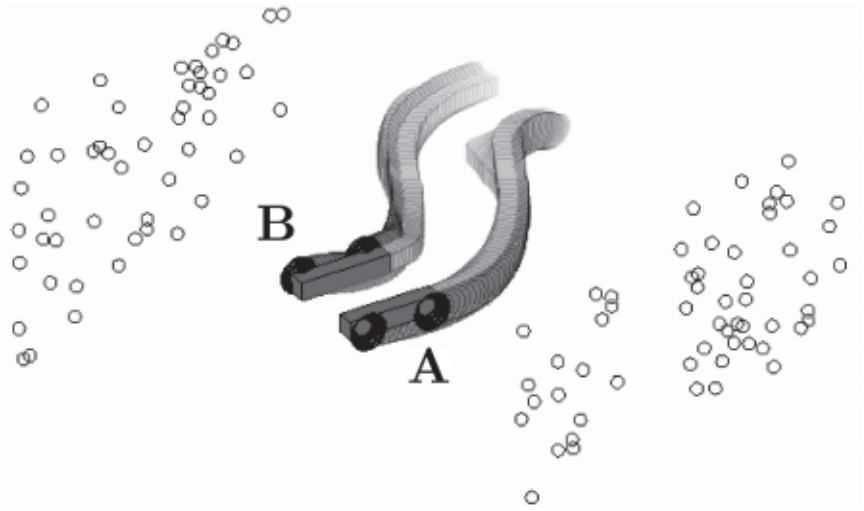


Fig. 2: Planar motion of two rigidly connected stereo systems **A** and **B** traversing down a hallway and observing signalized target on the opposite walls.

Tab. 1: Aggregated numbers used for the performance evaluation of point matching between two epochs.

	E1 A	E1 B	E2 A	E2 B	E3 A	E3 B
# stereo pairs	121	120	200	160	290	290
# connected epochs	3173	3247	5313	4964	19215	23273
# possible matches	64197	64436	91085	95359	361526	469029
# established matches	49580	50481	78121	72811	214285	314917
# wrong matches	0	0	0	674	60	5
match quality $q_{i,j}$	77 %	79 %	86 %	76 %	59 %	67 %

have identical epipolar angles. The epipolar angle for a point is given as the intersection angle of its image ray and an epipolar plane defined e.g. by the epipole and the optical axis. Finally, the 3D point cloud is given by a forward intersection for all identified stereo correspondences.

4.1 *Evaluation Criteria*

The evaluation of the performance of the proposed matching algorithm follows the scheme presented in Mikolajczyk et al. (2005) where it is used to compare different image matching approaches. While Mikolajczyk et al. (2005) analyse the performance of the detection and the matching step, we concentrate only on the matching. In our case the detection has been performed by the stereo reconstruction of the extracted image features (signalized targets) and will not be examined any further.

For the evaluation the true number of possible matches between the point clouds of any two epochs must be known. In order to provide such information a bundle adjustment for all six sequences was performed (three different motion patterns for two stereo systems **A** and **B**). Within the adjustment the change of orientations ${}^{t}\mathbf{T}_0$ w.r.t. to a reference epoch t_0 and the 3D coordinates of the point cloud, defined in the global frame of reference, were estimated. The inner orientation for all cameras, the relative orientations of the stereo systems and the lengths of their baselines were determined in advance and used as fixed parameters in the adjustment. Given the adjusted ${}^{t}\mathbf{T}_0$ the transformation ${}^{j}\mathbf{T}_i$ for any combinations of two epochs $t = i$ and $t = j$ can be computed. Then the point cloud of $t = i$ is transformed into the epoch $t = j$ and all possible matches are identified using a nearest-neighbour search. This number is used as ground truth for the evaluation.

To differentiate between different aspects influencing the performance of the matching we first compute a quality measure $q_{i,j}$ and then three values $o_{i,j}$, $v_{i,j}$ and $s_{i,j}$ for every combination of epochs characterizing three possible sources that affect the proposed algorithm:

Overall match quality

The match quality $q_{i,j}$ is computed as ratio $q_{i,j} = \frac{\left|\mathbb{P}_{m_{i,j}}\right|}{\left|\mathbb{P}_{i\cap j}\right|}$, where $\left|\mathbb{P}_{i\cap j}\right|$ is the ground truth for the number of possible matches between $t = i$ and $t = j$ and $\left|\mathbb{P}_{m_{i,j}}\right|$ is the number of matches established by the proposed algorithm.

Overlap of point clouds

The overlap $o_{i,j}$ of the point clouds at $t = i$ and $t = j$ can be computed with (6) and is identical to the computation of the similarity measure given in section 3.4.

$$o_{i,j} = \frac{\left|\mathbb{P}_{i\cap j}\right|}{\left|\mathbb{P}_i\right| + \left|\mathbb{P}_j\right| - \left|\mathbb{P}_{i\cap j}\right|} \tag{6}$$

Change of view direction

The change of the view direction $v_{i,j}$ is computed w.r.t. to the normal direction of the points in the point cloud (see section 3.1). From the set $\left|\mathbb{P}_{i\cap j}\right|$ of points visible at both epochs, those matched pairs $\{{}^{i_k}x, {}^{j_k}x: \forall k = 1\ldots n\}$ are selected for which a normal direction **n** could be computed. For all n pairs $\{{}^{i_k}n, {}^{j_k}n\}$ the angular difference v_k is given by $v_k = \arccos({}^{i_k}n \cdot {}^{j_k}n)$, where ${}^{i_k}n \cdot {}^{j_k}n$ is the dot product of two vectors and the function arccos(·) returns an angle in the interval $[0\ldots\pi]$. From all v_k the median $\bar{v}_k$ is determined and finally $v_{i,j}$ is given by $v_{i,j} = \bar{v}_k$.

Scale change

To assess the influence of scale changes, i.e. different distances between the stereo system and the point cloud, on the matching process, the number $s_{i,j}$ is computed. First, the centroids of the point clouds at $t = i$ and $t = j$ are computed according to (7).

$${}^{i}\hat{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n} {}^{i_k}\mathbf{x} : {}^{i_k}\mathbf{x} \in \left|\mathbb{P}_{i\cap j}\right|$$

$${}^{j}\hat{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n} {}^{j_k}\mathbf{x} : {}^{j_k}\mathbf{x} \in \left|\mathbb{P}_{i\cap j}\right| \tag{7}$$

The number $s_{i,j}$ is then given as a relative change $s_{i,j} = \frac{\min\left(\left\|{}^{i}\hat{\mathbf{x}}\right\|, \left\|{}^{j}\hat{\mathbf{x}}\right\|\right)}{\max\left(\left\|{}^{i}\hat{\mathbf{x}}\right\|, \left\|{}^{j}\hat{\mathbf{x}}\right\|\right)}$ of the distanc-

es at $t = i$ and $t = j$ w.r.t. to the centroids, where the functions min(·,·) and max(·,·) return the minimum and maximum values of their respective arguments.

The values $o_{i,j}$, $v_{i,j}$ and $s_{i,j}$ reflect the two major influences on the matching process. On the one hand, the overlap $o_{i,j}$ can be used to assess the influence of changing neighbourhoods on the matching process, because point clouds with a low overlap usually show large changes in the local neighbourhood of a point as well. On the other hand, uncertainties in the 3D position of points may lead to different quad tree indices during the computation of the distinctive description. The biggest influences on the point uncertainty result from large differences in the view directions and distance changes.

4.2 *Evaluation Results*

For the computation of the distinctive description for every point of every point cloud the parameters defined in section 3 must be set according to the point density of the point cloud. For the experiments the following values are chosen: maximum radius of local neighbourhood $r_{max} = 750$ mm, maximum number of selected neighbours $n_{max} = 12$, minimum number of neighbours $n_{max} = 4$ and total number of bins (2^p) for the computation of the quad tree index with $p = 64$.

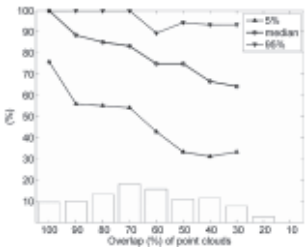
A first impression of the performance of the proposed matching scheme is given by the results presented in Tab. 1. Here the number of connected epochs are those combinations of epochs where the respective point clouds have at least three identical points, the possible matches represent the true number of matches derived from the results of the bundle adjustment (see section 4.1) and the wrong matches are the false positives before robust filtering. The differences in the number of recorded frames and connected epochs are caused by the different motion pattern. In the first experiment the stereo systems were moved in a straight and direct line down the hallway and thus only a smaller number of frames is necessary to capture the entire scene. In the last experiment, the systems were carried by hand and moved forward and backward along the corridor with changing rotations and translations. Such a motion pattern led to a larger number of frames and the continuously changing view directions resulted in more overlapping fields of view than for the translational motion and thus a larger number of connected epochs.

Tab. 1 shows that in most of the six experiments about 70–80 % of all possible matches were identified based on their respective distinctive descriptions. The most remarkable fact is that, at least w.r.t. to the large number of true matches, almost no false matches were established. None of these false positives were used to compute the change of orientation ${}^{j}\boldsymbol{T}_{i}$ between two epochs, because they were all successfully eliminated by the robust filtering process given in section 3.5.

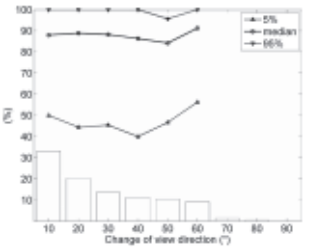
For the detailed analysis of the performance of our proposed matching algorithm in Fig. 3, the match qualities $q_{i,j}$ for any combination of epochs $t = i$ and $t = j$ are grouped w.r.t. to the different causes that possibly affect the matching process, i.e. the computed $o_{i,j}$, $v_{i,j}$ and $s_{i,j}$. For the visualization in Fig. 3 the different groups are plotted on the x-axis and the distribution of the match quality for every group is plotted on the y-axis.

For clarity of the representation the distribution of the $q_{i,j}$ for every group is represented by the 5 %- and 95 %-quantiles and the median. The graphs in the Fig. 3 show a representative subset of the results for all six image sequences. No result for the first experiment (pure translation) is shown, as that motion pattern did not give enough variety for the change of view directions and scale.

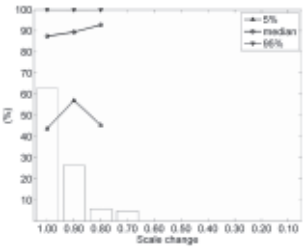
For the graphs the groups with a very low relative frequency are statistically not relevant and have been omitted. Figs. 3(a)–3(f) clearly show that the major source of influence is the change of the local neighbourhood indicated by a low overlap between the point clouds. This is supported by the observation that the median of the match quality is decreasing and the distribution is broadening with a reduced overlap. The Figs. 3(a) and 3(d) show that up to an overlap of 70 % for 90 % of all connected epochs, i.e. those between the 5 % and the 95 % quantiles, a match quality in the range of 60 %–100 % (Fig. 3a) and 55 %–95 % (Fig. 3(d)) is achieved. For an overlap of 40 % the median goes down to 70 % (Fig. 3(a)) and



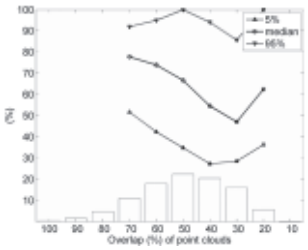
(a) Distribution of match quality $q_{i,j}$ w.r.t. overlap of point clouds (E2 **A**)



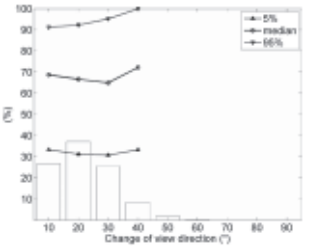
(b) Distribution of match quality $q_{i,j}$ w.r.t. change of view direction (E2 **A**)



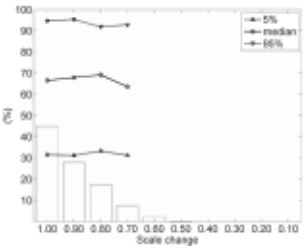
(c) Distribution of match quality $q_{i,j}$ w.r.t. change of scale (E2 **A**)



(d) Distribution of match quality $q_{i,j}$ w.r.t. overlap of point clouds (E3 **B**)



(e) Distribution of match quality $q_{i,j}$ w.r.t. change of view direction (E3 **B**)



(f) Distribution of match quality $q_{i,j}$ w.r.t. change of scale (E3 **B**)

Fig. 3: The graphs show the dependency between the achieved match qualitites $q_{i,j}$ and the different sources influencing the matching. The bars at the bottom represent the relative frequency of the respective group. The lines in the graph represent the distribution of the grouped matching quality using the 5 %-, 95 %-quantiles and the median. Groups with a relative frequency smaller than 5 % are not considered.

50 % (Fig. 3(d)) respective 70 % and 90 % of all connected epochs achieve a match quality in the interval from approximately 30 %–95 %.

The position uncertainty of a 3D point does not seem to have a strong influence on the match quality. Figs. 3(b), 3(c), 3(e) and 3(f) show that a change of the respective influence parameter does not change the median or the shape of the distribution significantly.

Larger changes in the local neighbourhood, caused by missed detections or points moving in or out of the stereo system's field of view, obviously have the effect that the similarity measure computed for two actually matching points is below the lower threshold t_i defined in section 3.4 and as a result that match is rejected incorrectly.

5 Conclusions and Future Work

A new approach to establish correspondences between points of sparsely populated and varying point clouds is presented in this paper. The identified matches can be used e.g. to estimate the change of orientation $\Delta\mathbf{T}_i$ between two epochs $t = i$ and $t = j$. The proposed algorithm is based on the known basic principles of matching using distinctive feature descriptions and extends them in a way that they can be used to identify corresponding 3D points in sparse and varying point clouds. The algorithm is an extension of the spin images (Johnson & Hebert 1999). It exploits the fact that applying the spin image algorithm to sparse point clouds gives a sparse 2D distinctive description that can be compressed further. The resulting compact 1D description is designed in a way that it allows an efficient matching of two descriptions.

The evaluation of image sequences recorded by two stereo systems shows that our approach allows an efficient and reliable matching of 3D points. The number of true positives is mostly above 70 % and the number of false positives is much smaller than 1 %. The false positives are all eliminated successfully by robust filtering of the established matches.

A limiting prerequisite of the presented matching scheme is that the point clouds at different epochs $t = i$ and $t = j$ must have the same scale as absolute distances are used in the computation of the distinctive description. For a more general variant of the description it would be possible to use ratios of distances that are invariant against scale changes. Such a variant of the compact distinctive description could be used to extend existing image based matching approaches like SIFT (e.g. Lowe 2004). The identified correspondences between two images can be used to compute 3D model coordinates. Based on these model coordinates a scale invariant distinctive description could be used differently:

Application for checking image based matching

Possible matches with further images are first established using the known image based matching algorithms and then they are additionally verified using a scale invariant distinctive description for 3D points. Only matches passing both approaches are accepted and the number of false positives might be reduced.

Application for connecting images with large perspective change

Usually the image based approaches have problems to correctly identify correspondences for larger perspective changes. Here a distinctive description for 3D points might be helpful to find additional correspondences.

It becomes apparent that it might be advantageous to combine image based distinctive descriptions with the proposed description for 3D data. Such a combination of 2D and 3D data is also presented in Wu et al. (2008).

A further extension would be the lifting of the constraint that the neighbourhood used for the computation of the distinctive description must be approximately planar. The basic principles used here can be transferred easily to the case of arbitrary 3D neighbourhoods. It remains to be investigated how much the changing 3D neighbourhood effects the results.

References

Barnea, S. & Filin, S., 2010: Geometry-Image-Intensity Combined Features for Registration of Terrestrial Laser Scans. – International Archives of Photogrammetry and Remote Sensing **38** (3a): 145–150.

Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L., 2008: Speeded-up robust features (SURF). – Computer Vision and Image Understanding **110** (3): 346–359.

Besl, P.J. & McKay, N.D., 1992: A Method for Registration of 3-D Shapes. – IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (2): 239–256.

Brown, L., 1992. A Survey of Image Registration Techniques. – ACM Computing Surveys **24:** 325–376.

Calonder, M., Lepetit, V. & Fua, P., 2008: Keypoint signatures for fast learning and recognition. – Forsyth, D., Torr, P. & Zisserman, A. (eds.): Computer Vision – ECCV 2008, LNCS **5302:** 58–71, Springer.

Calonder, M., Lepetit, V., Fua, P., Konolige, K., Bowman, J. & Mihelich, P., 2009: Compact signatures for high-speed interest point description and matching. – 2009 IEEE **12th** International Conference on Computer Vision: 357–364.

Chua, C. & Jarvis, R., 1997: Point Signatures: A New Representation for 3D Object Recognition. – International Journal of Computer Vision **25** (1): 63–85.

Finkel, R. & Bentley, J., 1974: Quad trees: a data structure for retrieval on composite keys. – Acta informatica **4** (1): 1–9.

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 **24** (6): 381–395.

Flint, A., Dick, A. & van den Hengel, A., 2008: Local 3D structure recognition in range images. – Computer Vision, IET **2** (4): 208–217.

Gelfand, N., Mitra, N.J., Guibas, L.J. & Pottmann, H., 2005: Robust global registration. – **Third** Eurographics Symposium on Geometry processing: 197–206.