18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

RoughCough - A New Image Registration Method for Radar Based Vehicle Self-Localization

Klaudius Werber*, Michael Barjenbruch†, Jens Klappstein*, Jürgen Dickmann* and Christian Waldschmidt‡

*Daimler AG, Research Center Ulm, Team Active Sensors, Ulm, Germany
†Ulm University, Institute of Measurement, Control and Microtechnology, Ulm, Germany
‡Ulm University, Institute of Microwave Techniques, Ulm, Germany

***Abstract*—Matching radar gridmap excerpts for vehicle self-localization can be regarded as an image registration task. This paper therefore presents a new, efficient image registration approach, which is suited to radar gridmaps. It does not require sharply structured input images and is applicable to all image pairs that can be aligned by a Euclidean transformation. Few line-shaped parts of the reference image are used for description. The similarity of a test image to these reference lines is calculated in a fast Hough transform based computation scheme.**
Experiments on radar gridmap excerpts derived from two different test drives demonstrate the low rate of false matches and low registration error of this algorithm. The short computation time shows the suitability of this algorithm for real-time application in a vehicle self-localization setup.

I. Introduction

The rising degree of intelligence in future vehicles presents heavy requirements on the vehicle self-localization, cf. [1]. Driver assistance functions which are increasingly autonomous must constantly know the vehicle's accurate *pose* (i.e. position and orientation). Neither the strict accuracy nor constant availability requirements are achieved by current satellite based positioning systems such as GPS, cf. [2]. Therefore, new ways of vehicle self-localization must be developed and incorporated into a fused overall self-localization system.

We pursue an approach that uses the aggregated observations of the vehicle's radar sensor for self-localization. While a vehicle is driving along a road, all observations of every radar snapshot are accumulated into one local map, taking into account the vehicle movement between two snapshots, cf. [3]. This local map (called *gridmap* in the following) represents a 2D bird's eye view radar map of the scene currently being passed through, cf. Fig. 1. Because it is based on radar data, it shares the beneficial characteristics of radar, i.e. high robustness against changing light or weather conditions.

A radar gridmap can be regarded as a gray-scale (single channel) image. Compared to camera images, it looks rather blurred. Sharp edges or corners are rare. For the transformation between two different observations of the same object in gridmaps, no expansions, shear nor nonlinear distortions are assumed. Furthermore, given a fixed cell size c_{cell} of each pixel, no scaling of the gridmap needs to be considered ($c_{\text{cell}} = 0.2\,\text{m}$ throughout this paper). Therefore, solely *Euclidean transformations* (i.e. rotation and translation) are regarded.

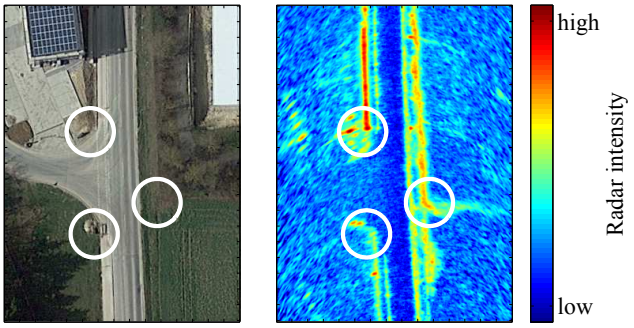

Fig. 1. Radar gridmap on the right for the setting shown on the left (examples of possible landmarks circled, aerial image taken from GoogleEarth, ©2009 GeoBasis-DE/BKG)

In this paper, small, prominent and characteristically structured excerpts of the radar gridmap along a road are regarded as landmarks for self-localization, cf. the circled portions in Fig 1. These excerpts are stored in a global database and exact information of their global pose is added. Now, a vehicle on that road can deduce its own global pose by recognizing currently observed excerpts from the gridmap that match the landmarks in the database and by calculating its relative pose with respect to that landmark through a registration of these two images.

This image registration needs to be carried out in the moving vehicle in real-time, which leads to three functional requests. First, the registration operation needs to be quick[1] ($T_{\text{reg}} \lesssim 1\,\text{s}$) to avoid continually-increasing queues of image pairs waiting to be registered. Second, for scheduling reasons the processing time should be approximately predictable. Therefore, registration algorithms with a fixed number of operations (avoiding iterative optimizations) are favorable.

[1]at least quicker than the mean expected traveling time between two consecutive landmarks along a road

Third, because the vehicle needs to either have a local copy of the landmark database on board or receive information about close landmarks via mobile data services, the description of each individual landmark should be strongly compact to keep the data load low. All three requests are met by the image registration algorithm proposed in this paper, as will be shown in the following.

A. Notation

The following notation is used throughout this paper: Scalars are designated by plain characters (e.g. c), vectors by bold lowercase characters (e.g. $\mathbf{v}$), matrices and tensors by bold capital characters (e.g. $\mathbf{A}$) and sets by blackboard bold characters (e.g. $\mathbb{T}$). Single subscript indices (x_i, x_j, or x_k) fix the corresponding variable to a specific value. Subscript words further specify the variable name without changing its mode ($\mathbf{I}_{\text{test}}$). Superscript information indicates the used coordinate system ($\mathbf{H}^{(\alpha,x,y)}$) where necessary.

II. Image Registration

There is a broad body of work on image registration. An extensive survey of different modes and approaches is given in [4]. Another more recent overview can be found e.g. in [5]. This paper follows the definitions and classifications of [4]. There are two major classes of image registration approaches: *feature- and area-based approaches*.

Feature-based approaches use manually or automatically selected prominent points (*key points*) in both images. They find corresponding key points in the different images by the specific geometric alignment of the points in each image and/or by a descriptor of each point's neighborhood, cf. [4]. Feature-based approaches usually offer very computationally efficient solutions that are comparably robust against minor changes in the image intensity. However, they require the regarded images to be strongly structured in order to find appropriate key points. This is usually not the case for the rather coarse image excerpts of radar gridmaps in our application.

Area-based approaches operate on the raw intensity values of the whole input images or certain selected excerpts optimizing some similarity measure or the mutual information of these two images, cf. [6]. On the one hand, this offers the full image information for registration. Additionally, no detection of key points is necessary, so area-based approaches usually are also applicable to less structured images. This happens by taking into account general courses of intensity values in the area rather than single outstanding points. On the other hand, these approaches require the input images to have roughly similar intensity ranges (this constraint is fulfilled for radar gridmaps), or at least some statistical dependence of different intensity levels. Furthermore, these approaches impose much stronger computational work loads than feature-based approaches, which in many cases prohibits real-time application, cf. [4].

Given increasing computation power, in recent years a slight movement from more feature-based to more area-based approaches has been observed. For example in [7] Carpin presents a method for fast registration of robotic maps of the same environment. Using Hough spectra, the difference in orientation of two maps is computed. Subsequently, the translational displacement is determined by the cross-correlation of the intensity projections onto the image axes. In Carpin's application, large maps (~ 250.000 cells) of limited environments are registered. Our application requires global databases, therefore we want to limit the gridmap excerpt size of one landmark to only a couple of thousand cells (~ 3.600).

A. Contribution

In this paper we develop an efficient algorithm for online image recognition and registration of structured gridmap excerpts. Regarding the classification of image processing algorithms in [4], it is an area-based approach operating only on a small subset of pixels of the reference image. It accepts input images that are subject solely to Euclidean transformations. Apart from radar gridmaps as in our automotive application, e.g. cameras with a fixed distance to their targets in a perpendicular plane deliver images with these characteristics. Such cameras for example are used in industrial visual quality assessment applications.

B. Concept

In radar gridmaps, many pixels contribute very little to the overall image information for two reasons. First, image areas are not equally important. There are some very relevant image parts showing the *core* of the image content and there are much less relevant image areas showing only *background*. These large connected background areas with constantly low pixel values allow neither recognition nor registration, because they are not characteristic.

Second, neighboring pixels are strongly correlated. If one pixel holds a certain value, it is very likely that its neighboring pixels hold similar values. Therefore, a gridmap image contains a lot of redundancy. For a distinct description of a gridmap excerpt, it is sufficient to regard a small subset of pixels that are spread out over the relevant image parts.

Our approach defines a mask consisting of several straight *line segments* that are pairwise parallel or orthogonal to each other. Each straight line segment has a width of one pixel and an individual length of several pixels. For the description of a radar landmark, only the image pixels masked by these line segments are regarded, cf. Fig. 2. The expression *line segment* in this context only indicates the shape of the single set elements. It does not imply that the reference image shows actual lines along these selected segments.

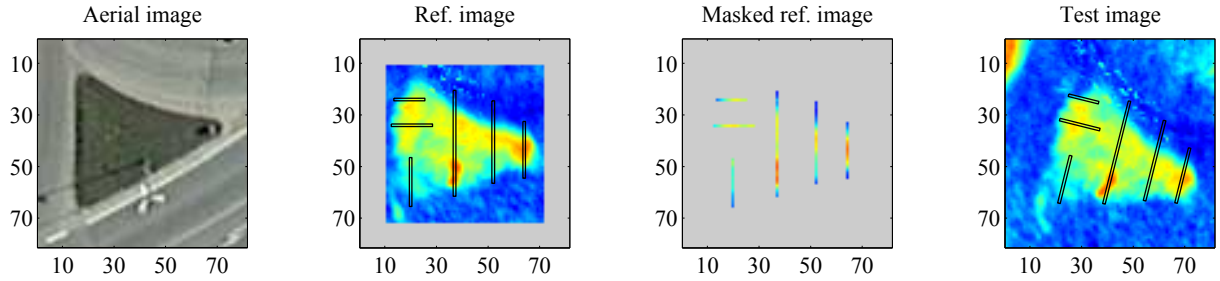

Fig. 2. Visualization of the RoughCough principle: The second to left image shows the reference image with the selected set of describing line segments framed. The second to right image shows the reference image masked by this line set. These pixels are the only information about the reference image given to our algorithm. The right-most image shows a registered test image. The closest match to the reference line set as detected by our algorithm is framed, again. The left-most aerial image shows the regarded setting (taken from GoogleEarth, ©2009 GeoBasis-DE/BKG). In all images axis values are pixel numbers.

Recognition and registration of this landmark with a test image is performed by calculating the similarity of these masked pixels with the corresponding pixels in the test image, assuming a set of all possible hypothetical Euclidean transformations. If no transformation results in a high similarity measure, the reference image is regarded as unrecognized in the test image. Otherwise, the Euclidean transformation leading to the highest similarity is regarded as the registration result for a matching pair of reference and test image, cf. Fig. 2.

In the following this registration algorithm is described in detail. It is calculated as an extension of the classical Hough transform and uses similar data fields. Therefore, identically to the Hough transform, it can be easily run in parallel and requires a fixed number of operations. Because it regards only a small subset of pixels of the reference image, its performance is fast.

III. RoughCough

RoughCough, our approach to image recognition and registration, can be regarded as an extension of the classical Hough transform for detection of straight lines, cf. [8]. The relationship between these two algorithms will be examined in Subsection III-A. Thereby, the registration of one given reference line segment in a given test image will be described. Subsequently, Subsection III-B will cover the geometric relations between multiple reference line segments and derive a simple rule to combine the results of the individual segments. These are the two core elements of RoughCough.

A. Registration of one line segment

The classical Hough transform detects straights in a binary input image $\mathbf{B}$, cf. [8]. It uses the Hesse normal form for representation of straights:

$$d_{\text{hough}} = x \cdot \cos \alpha_{\text{hough}} - y \cdot \sin \alpha_{\text{hough}}, \tag{1}$$

where α_{hough} and d_{hough} represent the angle and the distance to the origin of a straight and x and y describe the pixel coordinates along this straight. Here, the direction of α_{hough} has been chosen clockwise and the coordinate origin in the center of the regarded image.

For every hypothesis of a straight in the input image $\mathbf{B}$, the Hough transform accumulates all one-pixels $\left(\mathbf{B}\left(x_i, y_i\right) = 1\right)$ along this line. The hypothesis space is spanned by the regarded variations of d_{hough} and α_{hough}. Typically, d_{hough} and α_{hough} both are linearly sampled between a minimum and a maximum value with a given resolution. The resulting two-dimensional Hough space matrix $\mathbf{H}$ can be regarded as the similarity of every hypothesis $\left(\alpha_{\text{hough},j}, d_{\text{hough},j}\right)$ to an ideal straight in the original binary image $\mathbf{B}$.

For our image registration approach, a finite straight line segment $\mathbf{v}$ in a gray-scale test image $\mathbf{T}$ is searched for, instead of a straight in a binary input image, cf. Fig. 3. The line segment $\mathbf{v}$ is described by a given course of line values $[v_1, v_2, \ldots, v_l]$ with an individual length l, cf. Fig. 4. These line values are assumed to have the same spacing as the horizontal and vertical pixel spacing of the test image.

In addition to the two classical Hough parameters α_{hough} and d_{hough}, a third parameter is needed to fix the finite line segment on the infinite straight, in order to describe the location of each hypothetical line segment in the test image. We use the base b_{hough} to describe the index of that element of the line segment that is positioned at the perpendicular base point of this straight, cf. Fig. 3. In accordance with the two classical parameters, for the base b_{hough} a range of equally spaced hypotheses is assumed. Therefore, the resulting Hough space $\mathbf{H}$ in this case will be a three-dimensional tensor.

In the classical Hough transform, the range of α_{hough} is usually set to $[0°, 180°)$ and d_{hough} can be positive or negative, or the range of α_{hough} is set to $[0°, 360°)$ and d_{hough} can only be positive. For the case of this extension, not only the orientation, but also the direction of the straight line segment is relevant. Therefore, the range of α_{hough} is set to $[0°, 360°)$ and d_{hough} can still be positive and negative. To avoid ambiguities, we further define the vectors along the directions of d_{hough},

b_{hough} and the normal vector pointing out of the picture plane according to the *right-hand rule*, cf. Fig. 3.

The resulting value of a hypothesis $\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}\right)$ describes the similarity between the element values of the line segment $\mathbf{v}$ and the (interpolated) image values along the respective hypothetical straight line segment in the test image $\mathbf{T}$. Any similarity measure returning a scalar value for two input vectors is possible. We use a measure that accumulates all element similarities along the straight segment by a quadratic exponential function:

$$c_j \propto \sum_{i=1}^{l} \exp\left(-w \cdot \left(\mathbf{T}\left(\mathbf{p}_{j,i}\right) - v_i\right)^2\right) \tag{2}$$

with the pixel coordinates $\mathbf{p}_{j,i}$ in the test image $\mathbf{T}$

$$\begin{aligned}\mathbf{p}_{j,i} &= \begin{bmatrix} x_{j,i} \\ y_{j,i} \end{bmatrix} = \begin{bmatrix} x\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}, i\right) \\ y\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}, i\right) \end{bmatrix} \\ &= d_{\text{hough},j} \cdot \begin{bmatrix} \cos\left(\alpha_{\text{hough},j}\right) \\ -\sin\left(\alpha_{\text{hough},j}\right) \end{bmatrix} \\ &\quad + \left(i - b_{\text{hough},j}\right) \cdot \begin{bmatrix} \sin\left(\alpha_{\text{hough},j}\right) \\ \cos\left(\alpha_{\text{hough},j}\right) \end{bmatrix}\end{aligned} \tag{3}$$

and $w \in \mathbb{R}^+$ a parameter defining the strictness, with which bad agreement between the elements of the test image $\mathbf{T}$ and the line segment $\mathbf{v}$ is punished. This parameter helps to tune this measure to the specific value ranges of different applications.

Hence, the final result of this transformation is the three-dimensional Hough space volume $\mathbf{H}$ that for every assumed triple $\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}\right)$ contains the similarity of the respective line excerpt of the input image and the given line segment $\mathbf{v}$. For example Fig. 5 shows the projections onto the $\left(\alpha_{\text{hough}}, d_{\text{hough}}\right)$- and the $\left(\alpha_{\text{hough}}, b_{\text{hough}}\right)$-planes of the corresponding Hough space for the test image of Fig. 3 and the line segment of Fig. 4. This Hough space has been computed with a quantization of 0.5 pixels for d_{hough} and b_{hough} and 1° for α_{hough}.

If the line segment $\mathbf{v}$ has been derived as part of another image $\mathbf{R}$ and if applied with image $\mathbf{T}$, results in a high similarity value for some triple $\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}\right)$, it can be deduced that image $\mathbf{R}$ (or at least parts of it) can be recognized in image $\mathbf{T}$. Additionally, the translational and rotational displacement between these two images is returned. This way, the two images $\mathbf{R}$ and $\mathbf{T}$ can be recognized and registered by the extracted reference line segment $\mathbf{v}$.

B. Geometric relation between multiple line segments

In general, recognition of a reference image $\mathbf{R}$ is more distinct, when more than just one describing line segments are used (even if the combined length of the segments is kept constant). For example in Fig. 5, there are multiple strong local maxima apart from the true (encircled) maximum.

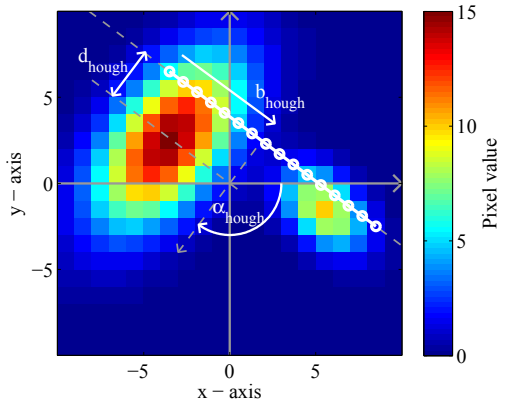

Fig. 3. Test image $\mathbf{T}$

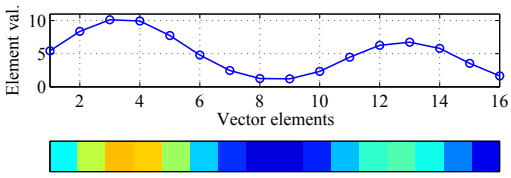

Fig. 4. Input vector $\mathbf{v}$

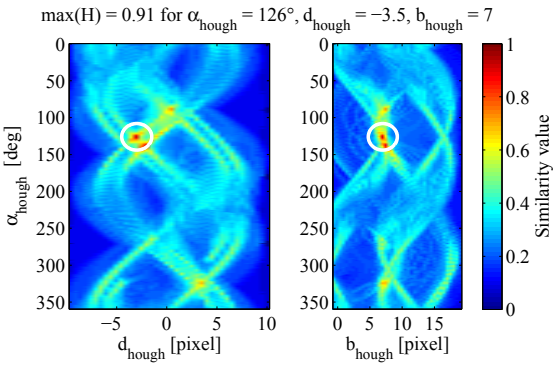

Fig. 5. Resulting Hough space $\mathbf{H}$ for the image $\mathbf{T}$ of Fig. 3 and the vector $\mathbf{v}$ of Fig. 4 projected onto the $\left(\alpha_{\text{hough}}, d_{\text{hough}}\right)$- and the $\left(\alpha_{\text{hough}}, b_{\text{hough}}\right)$-planes

Adding more describing line segments could help suppress these misleading local similarity maxima.

In the previous subsection, the resulting Hough space is represented over the coordinates α_{hough}, d_{hough} and b_{hough}: $\mathbf{H}^{\left(\alpha_{\text{hough}}, d_{\text{hough}}, b_{\text{hough}}\right)}$. For compatibility of the results of multiple line segments, we want a representation of the Hough space dependent on the rotational and translational differences α_{diff}, x_{diff} and y_{diff} between the reference and the test image: $\mathbf{H}^{\left(\alpha_{\text{diff}}, x_{\text{diff}}, y_{\text{diff}}\right)}$. α_{diff}, x_{diff} and y_{diff} are defined in the image coordinate system of the reference image. Therefore, a coordinate transformation $\left(\mathbf{H}^{\left(\alpha_{\text{hough}}, d_{\text{hough}}, b_{\text{hough}}\right)} \rightarrow \mathbf{H}^{\left(\alpha_{\text{diff}}, x_{\text{diff}}, y_{\text{diff}}\right)}\right)$ is necessary. This transformation is an angle shift between α_{hough}

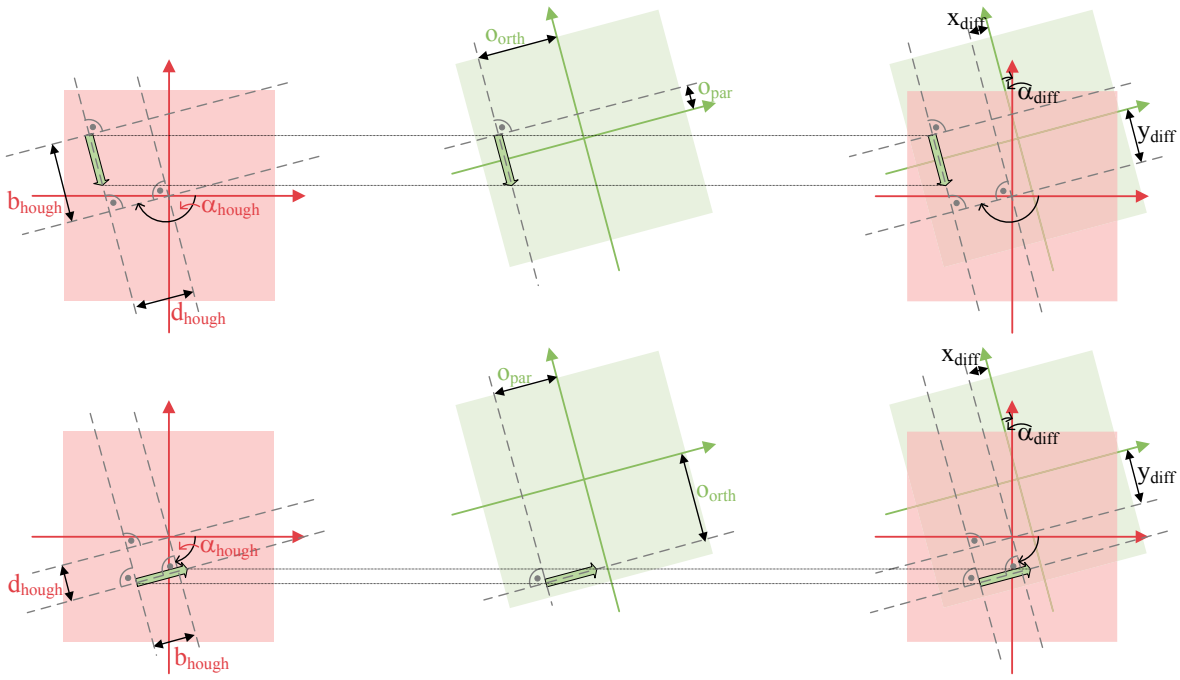

Fig. 6. Transformations for RoughCough: In the top diagrams, a column segment is regarded, in the bottom diagrams, a row segment (both indicated by the green arrows). The left hand diagrams show the Hough parameters in the test image (red), the center diagrams show the offsets in the reference image (green). The right hand diagrams show these two images aligned by the recognized common line segments and the resulting parameters of the Euclidean transformation between them.

and α_{diff} and a two-dimensional homogeneous transformation for rotation and translation between the pairs $(d_{\text{hough}}, b_{\text{hough}})$ and $(x_{\text{diff}}, y_{\text{diff}})$ depending on the position and orientation of each descriptive line segment in the reference image.

In aiming to achieve high computation speed for real-time application of this image registration method, explicit execution of this transformation should be avoided. This is possible in two special cases: when the regarded line segment is aligned with a column, or a row of the reference image.

If the line segment is aligned with a column of the reference image, this transformation becomes

$$\begin{aligned} \alpha_{\text{diff}} &= 180^\circ - \alpha_{\text{hough}}, \\ x_{\text{diff}} &= o_{\text{orth}} + d_{\text{hough}}, \\ y_{\text{diff}} &= o_{\text{par}} - b_{\text{hough}}, \end{aligned} \tag{4}$$

and if the line segment is aligned with a row of the reference image, the transformation becomes

$$\begin{aligned} \alpha_{\text{diff}} &= 90^\circ - \alpha_{\text{hough}}, \\ x_{\text{diff}} &= o_{\text{par}} + b_{\text{hough}}, \\ y_{\text{diff}} &= o_{\text{orth}} + d_{\text{hough}}. \end{aligned} \tag{5}$$

These equations are deduced from the geometry depicted in Fig. 6. o_{orth} represents the coordinate of each line segment that is orthogonal to the line orientation and o_{par} describes the position of the first line element along the coordinate that is parallel to the line orientation, both in the coordinate system of the reference image. I.e. for column segments o_{orth} equals to the x-coordinate and o_{par} to the y-coordinate of the coordinate system of the reference image. Conversely, for row segments o_{orth} equals to the y-coordinate and o_{par} to the x-coordinate of the reference image.

For these two special cases of Equations (4) and (5) this transformation becomes a trivial shift of variables. Thus, the transformation can simply be performed by storing the result $\mathbf{H}\left(\alpha_{\text{hough},j}, d_{\text{hough},j}, b_{\text{hough},j}\right)$ computed according to the previous subsection to the cell as defined by Equations (4) or (5). Furthermore, considering that the coordinate system's orientation in the reference image can be chosen arbitrarily before line selection, this constriction to line segments parallel to a coordinate axis actually only means that selected line segments need to be pairwise parallel or orthogonal to each other.

Summing up, using the extension of the Hough transform introduced in the previous subsection and the quick in-place variable transformation of this subsection, the content of a reference image $\mathbf{R}$ can now be described by an arbitrary number of row and column segments in order to recognize and register this content in a set of test images $\mathbb{T}$. This approach – *RoughCough* stands for row- and column-Hough – is our proposed new method for image registration. In the following,

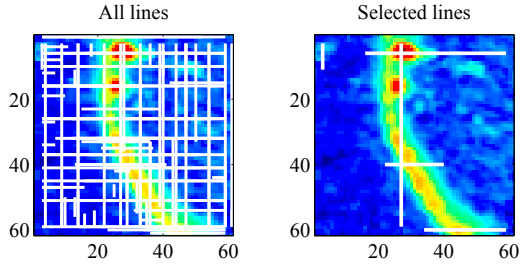

Fig. 7. Line selection results

the function

$$\mathbf{H} = \text{rghcgh}\left(\mathbf{I}, \mathbb{S}\right) \tag{6}$$

describes the application of the RoughCough algorithm to the input image $\mathbf{I}$ and the set of line segments $\mathbb{S}$. The result is the Hough space tensor $\mathbf{H}$ over the coordinates α_{diff}, x_{diff} and y_{diff}.

IV. Line Selection

The recognition and registration performance of RoughCough for a reference image is strongly dependent on an appropriate set of descriptive line segments, so care should be taken when selecting them. We regard the selection of line segments for one reference image as an offline task, but the recognition and registration of a true match using these line segments as an online task with real-time constraints. That means, that while for online image registration fast computation is a major issue, for line selection it is not a priority.

An appropriate set of line segments should:

1) contain characteristic parts of the reference image,
2) have a small number and a small combined length of all segments in order to keep computation time short,
3) be robust against slight intensity variations in a true match,
4) ideally return one single narrow mode with steep slopes in the Hough space for a true match in order to give a distinct estimate for the rotation and translation of test and reference images,
5) generally return low similarity values for all other test images to clearly reject them.

These weak constraints are partially contradictory. Especially the third constraint is in contrast to the forth and fifth ones, which makes appropriate trade-offs necessary.

We now assume to have a set of n_{ref} reference images $\mathbb{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{n_{\text{ref}}}\}$, extracted from the radar gridmap of one reference drive. The alignment to global coordinates of the reference images is assumed to be known. Furthermore, we also regard the case that additionally to the first set of reference images a second similar set $\mathbb{V} = \{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_{n_{\text{ref}}}\}$ is also available. This second set contains radar gridmap excerpts of the same image content and with the same alignment as the first one, but is taken from gridmaps generated from different radar observations than the first one. If available, this second reference set can be used for validation.

While there are various possible strategies for line selection, in this section we present one automated method. We chose a set of line segments for a reference image in two main steps. First, we quantify the image representation quality of any possible line segment in this image. Then, we optimize for a combination of these segments that best fulfills the given constraints using a fitness function.

A. Line evidence

A line segment (column or row) in an image is a connected set $\mathbb{L}$ of pixels $\mathbf{p}$. The quality, to which a line segment can represent the image it has been derived from is called *line evidence* q_{ev}, here. Generally, segments with high accumulated absolute values of the first derivative and large spreads of the image values are more characteristic than others, because they represent highly dynamic image parts. Also, we prefer shorter segments to larger ones in order to save computational power. Therefore, we have defined a quality function for column segments in an image $\mathbf{I}$ as

$$q_{\text{ev, col}} = h_{\mathbb{L}}^2 \cdot \frac{\sum_{\mathbf{p}\in\mathbb{L}} \left| \frac{\partial}{\partial y}\mathbf{I}\left(\mathbf{p}\right) \right|}{\log\left(|\mathbb{L}|\right)} \tag{7}$$

with $h_{\mathbb{L}}$ the value spread along the segment

$$h_{\mathbb{L}} = \max_{\mathbf{p}\in\mathbb{L}}\left(\mathbf{I}\left(\mathbf{p}\right)\right) - \min_{\mathbf{p}\in\mathbb{L}}\left(\mathbf{I}\left(\mathbf{p}\right)\right). \tag{8}$$

The quality function for row segments is defined accordingly.

Every possible line segment in a reference image is evaluated with regard to its line evidence. All *locally maximal* segments are stored for further processing. A line segment is called locally maximal if neither varying its line coordinate, nor its starting or ending pixel within that line by one, leads to another line segment with higher line evidence.

For a typical 60x60 reference image of our radar self-localization application, this procedure usually leads to a set of 50 to 150 locally maximal line segments $\mathbb{S}_{\text{all}}$. These segments usually are fairly well distributed in length, position and kind (row or column), cf. the example in the left plot of Fig. 7.

B. Selection of a set of lines

The goal of the second step for line selection is to find an optimal subset of line segments $\mathbb{S}_{\text{opt}} \subseteq \mathbb{S}_{\text{all}}$ that fulfills possible hard constraints (e.g. hard limits on the number or combined length of all segments) and represents an optimal compromise for the weak constraints enumerated at the beginning of Section IV.

Therefore, we use a Binary Particle Swarm Optimization (BPSO) as described and parameterized in [9] to optimize for

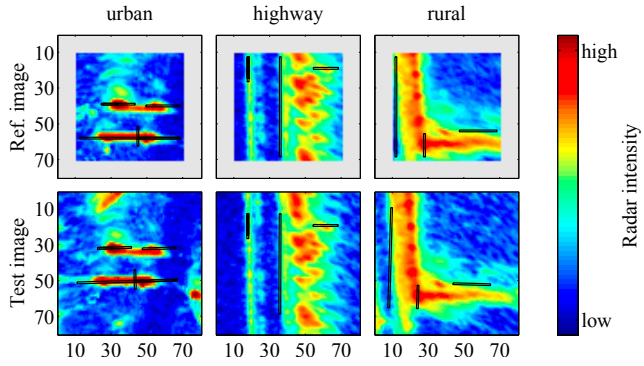

Fig. 8. Examples of registered images in all environments: Describing line segment are framed. Axis values are pixel numbers.

	total	urban	highway	rural
RoughCough for line sets selected without validation information				
selected images	**548**	186	122	240
matches	**213 (38.9%)**	94 (50.5%)	13 (10.7%)	106 (44.6%)
false matches	**11 (5.2%)**	0 (0%)	2 (15.4%)	9 (8.5%)
RoughCough for line sets selected with validation information				
selected images	**610**	328	58	224
matches	**268 (43.9%)**	166 (50.6%)	12 (20.7%)	90 (40.2%)
false matches	**16 (5.9%)**	3 (1.8%)	2 (16.7%)	11 (12.2%)

TABLE I
RECOGNITION PERFORMANCE OF ROUGHCOUGH

such a (possibly local) solution. The dimension of the particle space corresponds to the set size of all line segments $|\mathbb{S}_{\text{all}}|$. The single binary dimensions correspond to one specific line segment being chosen (*true*) or not being chosen (*false*) for a specific solution $\mathbb{S}_i$. The applied fitness function of the BPSO for the k-th reference image $\mathbf{R}_k$ and the i-th subset of line segments $\mathbb{S}_i$ is:

$$q_{\text{fit},k,i} = \frac{m_{\mathbf{V}}^2 \cdot \exp\left(-\alpha_{\text{diff},\mathbf{V}}^2 - x_{\text{diff},\mathbf{V}}^2 - y_{\text{diff},\mathbf{V}}^2\right)}{m_{\text{other}}^4 \cdot \mu_{\text{ref}} \cdot \sqrt{l_{\text{comb}} \cdot n_{\text{lines}}}} \tag{9}$$

with:

$$m_{\text{other}} = \max_{j \neq k}\left(\max_{\alpha_{\text{diff}},x_{\text{diff}},y_{\text{diff}}}\left(\text{rghcgh}\left(\mathbf{R}_j, \mathbb{S}_i\right)\right)\right) \tag{10}$$

the maximum returned value of any RoughCough operation applied to this subset $\mathbb{S}_i$ and a different reference image $\mathbf{R}_j$ (false match), and l_{comb} the combined length and n_{lines} the number of all selected line segments in $\mathbb{S}_i$. μ_{ref} is the average of $\mathbf{H}_{\text{ref}}$ with

$$\mathbf{H}_{\text{ref}} = \text{rghcgh}\left(\mathbf{R}_k, \mathbb{S}_i\right) \tag{11}$$

the result of the RoughCough operation applied to this exact subset $\mathbb{S}_i$ and reference image $\mathbf{R}_k$.

For the case that a validation set $\mathbb{V}$ is available, $m_{\mathbf{V}}, \alpha_{\text{diff},\mathbf{V}}, x_{\text{diff},\mathbf{V}}$ and $y_{\text{diff},\mathbf{V}}$ are defined as

$$m_{\mathbf{V}} = \max_{\alpha_{\text{diff}},x_{\text{diff}},y_{\text{diff}}}\left(\mathbf{H}_{\mathbf{V}}\right), \tag{12}$$

$$\left[\alpha_{\text{diff},\mathbf{V}}, x_{\text{diff},\mathbf{V}}, y_{\text{diff},\mathbf{V}}\right] = \text{argmax}_{\alpha_{\text{diff}},x_{\text{diff}},y_{\text{diff}}}\left(\mathbf{H}_{\mathbf{V}}\right), \tag{13}$$

with

$$\mathbf{H}_{\mathbf{V}} = \text{rghcgh}\left(\mathbf{V}_k, \mathbb{S}_i\right). \tag{14}$$

If no validation set is available, the numerator in Equation (9) is set to one.

This numerator in Equation (9) ensures robustness of an optimal solution $\mathbb{S}_{\text{opt}}$ against slight intensity variations in a true match due to measurement noise (weak constraint 3). If no validation set is available, this weak constraint cannot specifically be dealt with. The term m_{other} in Equation (9) leads to strong suppression of false matches (weak constraint 5). The factors l_{comb} and n_{lines} counteract high computational cost (weak constraint 2) and the factor μ_{ref} suppresses wrong or inaccurate translation and rotation results for a true match (weak constraint 4). The exponents in Equation (9) are used to weight the single terms. The result of one such optimization of all lines given in the left-hand plot of Fig. 7 can be seen in the right-hand plot of that figure.

V. EXPERIMENTS

Two measurement drives (one reference drive and one test drive) along the same route were completed, in order to evaluate the performance of our algorithm. The route is 9 km long altogether and includes urban areas (1.6 km), highways (2.2 km) and rural roads (5.2 km). The test vehicle was equipped with two experimental 77 GHz-radar sensors at both front vehicle corners and a real-time kinematics unit (RTK) for precise ground-truth measurement of the vehicle pose. The radar sensors have a bandwidth of 600 MHz, a maximum range of 42 m and an angular field of view of $\pm 60°$. The sensors' standard deviation of the range measurement is about 0.25 m and the angular standard deviation about $1°$.

For offline evaluation, the radar observations of the reference drive were accumulated to amplitude-gridmaps according to [3] in two different ways. In one way, the observations of both sensors were accumulated into the same gridmap. In the other way, the observations of each sensor were accumulated into a separate gridmap each, leading to two different, however perfectly aligned gridmaps. In this second case, a validation set $\mathbb{V}$ is available for line selection according to Equation (9).

In these reference gridmaps, characteristically structured areas were automatically searched for as gridmap excerpts of 60x60 pixels that lead to high fitness values of the final selected reference line set in an optimization according to Section IV. Structured areas in the neighborhood (closer than 40 m) of each regarded excerpt were used as excerpts of false matches $\mathbf{R}_j$ in Equation (10), necessary to optimize for good suppression of false matches. For the case without validation information, 548 reference images were selected, and 610

	total	urban	highway	rural
RoughCough for line sets selected without validation information				
true matches	**202**	94	11	97
$\mu_{\text{err},X}$	**12.4 cm**	7.9 cm	11.6 cm	16.9 cm
$\mu_{\text{err},Y}$	**15.4 cm**	4.6 cm	5.1 cm	27.0 cm
$\mu_{\text{err},\alpha}$	**0.39** °	0.37 °	0.33 °	0.42 °
RoughCough for line sets selected with validation information				
true matches	**252**	163	10	79
$\mu_{\text{err},X}$	**11.6 cm**	7.1 cm	9.0 cm	15.0 cm
$\mu_{\text{err},Y}$	**11.0 cm**	5.4 cm	6.7 cm	20.3 cm
$\mu_{\text{err},\alpha}$	**0.42** °	0.41 °	0.61 °	0.39 °
Mutual information registration				
true matches	**180**	54	51	75
$\mu_{\text{err},X}$	**11.8 cm**	12.9 cm	12.4 cm	10.7 cm
$\mu_{\text{err},Y}$	**11.4 cm**	6.1 cm	6.5 cm	18.4 cm
$\mu_{\text{err},\alpha}$	**0.59** °	0.55 °	0.45 °	0.72 °
Carpin registration				
true matches	**196**	50	38	108
$\mu_{\text{err},X}$	**37.9 cm**	36.8 cm	29.1 cm	41.5 cm
$\mu_{\text{err},Y}$	**29.3 cm**	37.6 cm	25.0 cm	26.9 cm
$\mu_{\text{err},\alpha}$	**0.62** °	0.60 °	0.53 °	0.66 °

TABLE II
REGISTRATION PERFORMANCE OF ALL METHODS

reference images for the case with validation information. The mean combined length of line segments in the first case is 68 pixels, in the second case 71 pixels, the average number of selected lines in the first case is 2.4, in the second case, 2.7.

The radar observations of the test drive were also accumulated into a gridmap, and structured test excerpts were extracted as structured image parts of these gridmaps. The size of test images was chosen as 80x80 pixels to increase the probability that the full content of a reference image is included in the test image of a true match.

The ground truth pose from the RTK was stored for both, reference and test images. The ground truth mean expected position error was 2.2 cm and the orientation error $0.03°$. This ground truth position of reference and test images was used to pre-select the test image candidates to be registered with each reference image. Only test images that are closer than 40 m to each reference image were regarded. This pre-selection simulates the effects of a rough localization given in a vehicle through GPS or a prior pose estimate. With this processing method, on average 100 test images were selected for each reference image.

Matching and registration of the reference images with their set of test images was done with the following standard parameters:

Hough space: Rotations between $-5°$ and $5°$ around the expected angle were regarded. This angular limitation is motivated by our application. Bound by the road surface, vehicles on roads in general do not have arbitrary orientations, but rather those roughly similar to the roads themselves. Translations, both in x- and y-direction, are regarded in the range of -10 to 10 pixels. That is just the displacement for which the smaller area of the reference image is still fully covered by the test image area.

Resolution: The standard resolution was chosen as $0.2°$ in rotation and 0.2 pixels in translation. These values were varied between 0.1 and 0.5 without major effects on the results.

Additional parameters: The strictness factor w in Equation (2) was chosen as $3.6 \cdot 10^{-7}$, which is appropriate for the dynamic range of our input images (unsigned 16 bit). The proportionality factor for the result of Equation (2) is chosen as $1/l_{\text{comb}}$ for each reference line set specifically. This way the RoughCough result is bound to the interval $[0; 1]$.

If the highest maximum RoughCough value returned by one of these test images is higher than $\theta_{RC} = 0.6$, we consider this test image to match the reference image. This threshold value represents an appropriate trade-off between reliability (low percentage of false matches) and number of detected matches in our application. If no test image returns a maximum value higher than θ_{RC} we regard the reference image as unrecognized within the test set. The $\alpha_{\text{diff}}, x_{\text{diff}}$ and y_{diff} coordinates of this maximum value in the Hough space are the calculated registration between this matched pair of reference and test image. If the remaining ground truth distance between this pair is greater than 1 m, this specific match is regarded as a false match.

For comparison, registration has also been done using the implementation of mutual information registration by Matthew Sochor, publicly available at [10], and with a re-implementation of Carpin's method presented in [7]. The mutual information registration applies the joint histogram approach of [6]. This code is based on [11]. For optimization Powell's direction set method is used [12]. Carpin's method originally is intended for large binary maps (pixel *free* or *occupied*), cf.[7]. It has been adapted to small excerpts of gray-scale maps. Since both implementations of the comparative algorithms require the input images to be the same size, the test images have been set to 60x60 pixels for these registrations.

These comparative methods have been applied instead of RoughCough in our processing chain. However, because this processing chain has been designed for the application with RoughCough, the *recognition performance* (i.e. the ratio of detected matches for all reference images and the ratio of true matches in all detected matches) of these methods turned out to be much worse than the performance of RoughCough. This should not be considered as a flaw of the comparative methods, but as an implication of the regarded application. Therefore, we will compare the *registration performance* (i.e. mean registration errors in translation and rotation) of RoughCough

and the comparative methods only on the detected true matches of each individual method.

An example of typical, matched pairs for each environment is shown in Fig. 8. The recognition results of RoughCough for both regarded cases are given in Table I. The ratio of only about 5% false matches in both cases shows the high reliability of the RoughCough matching result. The percentage of recognized matches of RoughCough is better for the line sets extracted with validation information than without (43.8% compared to 38.9%). However, this rate of less than 50% in both cases is still rather low. We assume that using validation information of a completely different drive could further raise this rate, as then not only disturbances due to sensor noise, but also due to the specific situation could be avoided (e.g. parked cars or field of view limitations due to other vehicles driving ahead).

The registration results of all regarded algorithms are presented in Table II. RoughCough performs well on the test data for both kinds of line segments (distance error $\sim 12\,\text{cm}$, angular error $\sim 0.4°$). The distance uncertainty is better for the line sets extracted with validation information. This shows that the nominator in Equation (9) actually does improve the robustness of the selected line set. Therefore, if available, this form of validation information should be used. Regarding the mean distance and angular errors, the mutual information registration performs equally well on the test set as RoughCough. Carpin's method also leads to low angular errors. However, the mean distance errors of Carpin's method with around 30 cm in one direction are significantly worse than the errors of RoughCough. The performance of all algorithms is worse in rural and highway environments than in urban environments. This is plausible, because in urban scenarios objects along the roads are more distinct, cf. Fig. 8.

On a computer with an Intel Core i7 X990 CPU and an Nvidia GeForce GTX 570 GPU, the RoughCough algorithm for one reference line set and one test image on average took 14 ms, using the standard parameters given above. Our implementation exploits parallel computation on a GPU. The mutual information registration on average took 980 ms and Carpin's registration 75 ms. The Hough transform used by Carpin's method also ran on a GPU. Therefore, in comparison with the two other regarded algorithms, RoughCough is the only registration method that yields both high accuracy of the registration results and low computation time.

VI. Conclusion

This paper has presented RoughCough, a new image registration algorithm, suitable for radar gridmaps. The suitability of RoughCough for radar based self-localization has been demonstrated in experiments on test data collected in test drives in real traffic scenarios. For the test data RoughCough returned more accurate results in shorter computation time than the two regarded comparative algorithms. It fulfills the functional requests of a low and approximately predictable processing time and of a compact representation of the reference image in the landmark database.

RoughCough was used with a limited range of hypothesized rotations for fast operation in this paper. However, if no rough prior estimate of the image rotation is available, it can be used with a full range of hypothesized rotations and then detect arbitrary Euclidean transformations.

Movable objects, such as parked cars, still represent a challenge for this algorithm because they look highly structured in a gridmap and therefore often are selected as bad landmark reference images. This can be avoided by multiple drives for landmark selection or by incorporating a car detection as proposed in [13] into the reference line selection process of Section IV.

In future work, new line selection methods will be investigated for raising the percentage of landmark images that are recognized in a test drive. Furthermore, RoughCough will be applied with a particle filter for self-localization. In such an application, the Hough space has to be evaluated for the particle locations, only.

References

[1] K. Werber, M. Barjenbruch, J. Klappstein, J. Dickmann, and C. Waldschmidt, "How do traffic signs look like in radar?" in *Proceedings of the 44th European Microwave Conference*. EuMA, 2014, pp. 135–138.

[2] I. Skog and P. Händel, "In-car positioning and navigation technologies - a survey." *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 4–21, 2009.

[3] K. Werber, M. Rapp, J. Klappstein, M. Hahn, J. Dickmann, K. Dietmayer, and C. Waldschmidt, "Automotive radar gridmap representations," in *2015 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM 2015)*, Heidelberg, Germany, Apr. 2015.

[4] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[5] M. V. Wyawahare, P. M. Patil, H. K. Abhyankar *et al.*, "Image registration techniques: an overview," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 2, no. 3, pp. 11–28, 2009.

[6] X. Lu, S. Zhang, H. Su, and Y. Chen, "Mutual information-based multimodal image registration using a novel joint histogram estimation," *Computerized Medical Imaging and Graphics*, vol. 32, no. 3, pp. 202–209, 2008.

[7] S. Carpin, "Fast and accurate map merging for multi-robot systems." *Auton. Robots*, vol. 25, no. 3, pp. 305–316, 2008.

[8] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[9] D. Gorse, "Binary particle swarm optimisation with improved scaling behaviour," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2013, pp. 239–244.

[10] M. Sochor. (2008) Automatic 2d rigid body image registration. http://www.mathworks.com/matlabcentral/fileexchange/19086-automatic-2d-rigid-body-image-registration, last visited Nov. 10th, 2014.

[11] J.-P. Moreau. http://jean-pierre.moreau.pagesperso-orange.fr/Cplus/tpowell_cpp.txt, last visited Nov. 10th, 2014.

[12] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes: The art of scientific computing*. Cambridge University Press, 1987, vol. 2.

[13] R. Dube, M. Hahn, M. Schutz, J. Dickmann, and D. Gingras, "Detection of parked vehicles from a radar based occupancy grid," in *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, 2014, pp. 1415–1420.