International Journal of Robotics and Automation (IJRA)
Vol. 9, No. 2, June 2020, pp. 63∼72
ISSN: 2089-4856, DOI: 10.11591/ijra.v9i2.pp63-72

General concepts of multi-sensor data-fusion based SLAM

Jan Klečka, Karel Horák, Ondřej Boštík
Department of Control and Instrumentation at Brno University of Technology, Czech Republic

Article Info

Article history:

Received Sep 30, 2019
Revised Oct 06, 2019
Accepted Feb 18, 2020

Keywords:

Data fusion
Localization
Mapping
Partially collective mapping
Simultaneous localization and mapping (SLAM)

ABSTRACT

This paper is approaching a problem of Simultaneous Localization and Mapping (SLAM) algorithms focused specifically on processing of data from a heterogeneous set of sensors concurrently. Sensors are considered to be different in a sense of measured physical quantity and so the problem of effective data-fusion is discussed. A special extension of the standard probabilistic approach to SLAM algorithms is presented. This extension is composed of two parts. Firstly is presented general perspective multiple-sensors based SLAM and then thee archetypical special cases are discuses. One archetype provisionally designated as "partially collective mapping" has been analyzed also in a practical perspective because it implies a promising options for implicit map-level data-fusion.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Jan Klečka,
Department of Control and Instrumentation,
Brno University of Technology,
Technická 12, Brno, Czech Republic
Email: klecka@feec.vutbr.cz

1. INTRODUCTION

After more than thee decades of research the Simultaneous Localization and Mapping (SLAM) algorithms provide still a variety of open topics for further development as we can see e.g. survey by C. Cadena's et al. [1] or in critique by Huang et al. [2]. These algorithms are designed to continuously process given observations of surroundings to provide observer's current position (or sometimes whole trajectory) and map of observed environment. Such information is unsubstitutable feedback for practically any navigation task e.g. trajectory planning or complex movement execution.

There can be found many application fields for SLAM algorithms. We chose to underline only thee which, as we feel, are nowadays widely discussed. Navigation of autonomous cars as discussed by Bresson et al. [3], various industry 4.0 tasks e.g. Beul presented warehouse inventory check [4] or augmented reality task as shown by Klein and Murray [5].

For several years have we been dealing with SLAM based on various sensor data-fusion and this paper aims to report some general findings we have done. Our original methodology has been originally mainly inductive process. We originally began with the concept of building map using simple geometrical entities to approximate in piece-wise manner surfaces of solids that are creating the mapped environment and during the development, we iteratively generalize this specific concept until it fits the standard probabilistic SLAM algorithms theory. However following descriptions are conducted in a more comprehensible deductive process where we start with the general and work our way to the specific.

We have been trying to use common notation customs although for maximal clarity of following descriptions we quickly state used rules. Matrices and vectors symbols are bold e.g. $\mathbf{A}, \mathbf{x}$ where uppercase

is used for matrices and lowercase for vectors. Bold uppercase symbols are also used for sets which also has lower index show range of their cardinality e.g. $\mathbf{Z}_{0:N} = \{\mathbf{z}_0, \mathbf{z}_1, \cdots, \mathbf{z}_N\}$. Scalar symbols are italics e.g. N.

Subscripts are used to express specific element of a larger collection e.g. $\mathbf{z}_n$ is a realization of $\mathbf{z}$ in time $t = n$. Superscripts in square brackets symbolize specific modality e.g. $\mathbf{z}^{[k]}$ is $\mathbf{z}$ associated with k-type sensor. For functions is used a normal font e.g. $\mathrm{h}(\cdot)$ is function named h.

2. RELATED WORKS

As we already indicate in introduction except for concept data-fusion based SLAM we also dealing with the concept of SLAM using map representation in the form of a collection of geometric entities so we split this section into respective subsections.

2.1. Data-fusion in context of SLAM

A substantial amount of papers that mention keyword fusion in the context of SLAM algorithms deals with processing observations from a single RGB-D camera (or often even specifically the Microsoft Kinect). Examples of such works are: KinectFusion algorithm presented by Newcombe et al. [6], algorithm Fusion++ by McCormac et al. [7] or ElasticFusion by Whelan et al. [8, 9].

Several teams reported also about SLAM based on observations from multiple sensors. For example with processing data from custom made sensory head equipped with two CCD cameras, two thermo-cameras and range finder has dealt Burian et al. [10] - data from rangefinder is used depth reference for camera images and therefore can be enhanced by using mathematical models of individual cameras. Fang et al. presented a SLAM capable system with CCD camera and sonar [11] which improves the reliability by utilizing feature-level data-fusion.

Let's notice that in so far listed algorithms the data-fusion is conducted always prior to SLAM iteration and so the SLAM algorithms then process already fused data. Notice moreover that various modalities are typically conceptually in mutually nonequivalent status. The dept perception modality is typically in unsubstitutable position and other modalities (like color) are used to increase the robustness of the whole solution or just for map presentation purposes.

2.2. Map as a set on non-point geometrical entities

There can be found some papers that preset solutions to SLAM problems that use representation of map in the form of a collection of geometrical entities. For example, lidar-based 2D SLAM that represents the environment by a set of lines is shown by Garulli et al. in [12] and also by Choi et al. [13]. Example of lidar-based 3D SLAM which uses plane features is presented by Ulas and Temeltas [14]. These concepts aren't specific only for Lidar. Zhou et al. [15] and Uehara et al. [16] are reported vision-based SLAM algorithms that utilize line features. Yang et al. [17] shows that utilizing planes can improve robustness of monocular SLAM against standard strictly point-based approaches .

There can be found also reports that approach only partial problems like segmentation. For example algorithm for approximation point 2D cloud by collection of lines by Jelinek et al. [18] or detection of planes in 3D point-cloud by Hulik et al. [19] and also by Pathak et al. [20].

3. PROBABILISTIC APPROACH

In this section, the mathematical background of fusion-based algorithms is presented. We present the problem from a probabilistic perspective to urge the maximal generality of given formulas. Even though some concretization had been made. We assumed strictly the static environment and from perspective of estimated trajectory, we provide solution to two variants - the "online" SLAM that aims only to estimate the most recent pose and the "full" SLAM which provide a way to estimate the whole trajectory.

3.1. Standard theory

Presented description is equivalent to thous given in standard SLAM-oriented publications e.g. survey by Durrant-White et al. [21] or book Probabilistic robotics by Thrun et al. [22]. Let's have some observer which moves in an environment given by parameterization $\mathbf{m}$ and during its movement is the observer repeatedly conducting observations $\mathbf{z}$. Observer relation to this environment, e.g. its position and orientation, is given by state $\mathbf{x}$.

Observations describe the observer surroundings and are degraded by noise. Therefore it can be defined by a conditional probability distribution that is usually called the observation model:

$$p(\mathbf{z}_n|\mathbf{x}_n, \mathbf{m}) \tag{1}$$

Because of the nature of the observer entity, the state vector will most probably be subjected to some dynamic that bounds its change between observations. This link may be dependent on some observable quantity $\mathbf{u}$ and it's also stochastic so can be defined by conditional probability distribution called motion model:

$$p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{u}_n) \tag{2}$$

Because the stochastic nature of both observation and motion model the SLAM problem lies from the general point of view in defining a probability distribution of a pose and a map conditioned by the conducted observations:

$$p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N}, \mathbf{U}_{1:N}) \tag{3}$$

This distribution has to also represent our prior belief about the state and map distribution.

Analytic solution of this problem can be found using Bayes formula as:

$$p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N}, \mathbf{U}_{1:N}) = \eta p(\mathbf{z}_N|\mathbf{x}_N, \mathbf{m}) p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N-1}, \mathbf{U}_{1:N}) \tag{4}$$

where η is an arbitrary normalization constant and second term can be defined by propagation previous believe into current time using motion model:

$$p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N-1}) = \eta \int p(\mathbf{x}_{N-1}, \mathbf{m}|\mathbf{Z}_{0:N-1}) p(\mathbf{x}_N|\mathbf{x}_{N-1}, \mathbf{u}_N) d\mathbf{x}_{N-1} \tag{5}$$

Usually the realization of equation (4) is called the update step and realization of equation (5) is called a prediction step. This recurrent form of solution is standardly referred to as an "online" SLAM and can be fairly straightforwardly seen as applicable to real-time process. The second frequently utilized form of SLAM solution is the so-called "full" SLAM that is non-recurrent and aims at the description of whole trajectory distribution.

$$p(\mathbf{X}_{0:N}, \mathbf{m}|\mathbf{Z}_{0:N}, \mathbf{U}_{1:N}) = \eta \Big[\prod_{n=0}^{N} p(\mathbf{z}_n|\mathbf{x}_n, \mathbf{m}) \Big] \Big[\prod_{n=0}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{u}_n) \Big] p(\mathbf{x}_0) \tag{6}$$

3.2. General multi-sensor based SLAM

Now, let's consider that set of observations is composed of subsets and each subset contain only observations from one particular sensor modality

$$\mathbf{Z}_{0:N} = \{\mathbf{Z}^{[1]}_{0_1:N_1}, \mathbf{Z}^{[2]}_{0_2:N_2}, \cdots, \mathbf{Z}^{[K]}_{0_K:N_K}\} \tag{7}$$

where any time indexes range $0_k : N_k \subset 0 : N$.

Then each modality has its own unique particular observation model

$$p(\mathbf{z}^{[k]}_n|\mathbf{x}_n, \mathbf{m}) \tag{8}$$

Motion model stays conceptually unchanged, we can assume the same form as in the general case.

These eventualities do not change above mentioned equations dramatically. The only change lies in the substitution of general observation models for particular ones. Specifically, the update step of the online SLAM gonna look like this

$$p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N}, \mathbf{U}_{1:N}) = \eta p(\mathbf{z}^{[k]}_N|\mathbf{x}_N, \mathbf{m}) p(\mathbf{x}_N, \mathbf{m}|\mathbf{Z}_{0:N-1}, \mathbf{U}_{1:N}) \tag{9}$$

and the probability distribution of full variant will be in the following form

$$p(\mathbf{X}_{0:N}, \mathbf{m}|\mathbf{Z}_{0:N}, \mathbf{U}_{1:N}) = \eta \Big[\prod_{n=0}^{N} p(\mathbf{z}^{[k]}_n|\mathbf{x}_n, \mathbf{m}) \Big] \Big[\prod_{n=0}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1}, \mathbf{u}_n) \Big] p(\mathbf{x}_0) \tag{10}$$

It may look like no progress at all however that because we did not take into account that with additional modalities will be changing more things than just the observation model.

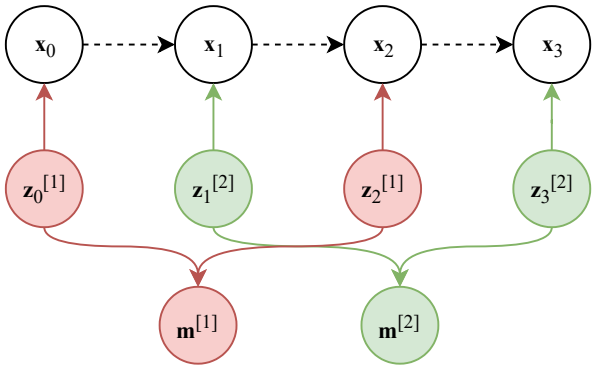

Figure 1. Conditionally independent algorithms

3.3. Special cases multi-sensor based SLAM

In this section, we specify the above-mentioned formulas by assuming specific structure derived from mutual relations of different modality observations. Specifically, we analyze thee cases that we consider to be archetypes from which the real situations can be composed of.

3.3.1. Conditionally independent algorithms

Let's consider that given modalities (or at least used style of their abstraction) does both not allow forming any cross-modality quantity that could represent a common map elements and in addition their observations are asynchronous in time of their capture - so each one belongs to different state of the observer (see Figure 1). That will leads to separation of the map parameterization $\mathbf{m}$ into a set of sensor-specific representations

$$\mathbf{m} = \mathbf{M}^{[1:K]} = \left\{\mathbf{m}^{[1]}, \mathbf{m}^{[2]}, \cdots, \mathbf{m}^{[K]}\right\} \tag{11}$$

where each particular map $\mathbf{m}^{[k]}$ is independent of any observation $\mathbf{z}^{[l]}$.

$$p(\mathbf{z}_n^{[k]}|\mathbf{x}_n, \mathbf{m}^{[l]}) = p(\mathbf{z}_n^{[k]}) \quad \forall k \neq l \tag{12}$$

If we apply these rules to the recurrent SLAM equation we can in this case, alter them into a form where the update step is separable in terms of modality.

So let's notice that only the cross-modality link is in this case established by the motion model. The weaker the motion model, the closer the uni-modal parts are to mutual independency and in an extreme case, assuming that the motion model does not exist at all, this archetype leads to completely independent parallel SLAM algorithms. Generally, we can state that particular maps can be considered conditionally independent given the state. Data-fusion is in this case scheduled to postprocessing with no benefit to runtime.

3.3.2. Super-observation

The second archetype is based on the assumption that the acquisition of the observations is conducted in a synchronized manner. So even though observer using multiple sensors their capturing times are synchronized and so all particular modalities observations always belongs to one single state realization $\mathbf{x}$ (see Figure 2). Under these assumptions, we can define the observation set as a collection of subsets that contain isochronous observations.

$$\mathbf{Z}_{0:N} = \left\{\mathbf{Z}_0^{[1:K]}, \mathbf{Z}_1^{[1:K]}, \cdots, \mathbf{Z}_N^{[1:K]}\right\} \tag{13}$$

Because from an analytical perspective it is irrelevant whether the observation is vector or set, we can define the composed observation model and then apply the single-observation theory.

$$(\mathbf{Z}_n^{[1:K]}|\mathbf{x}_n, \mathbf{m}) = \prod_{k=1}^{K} p(\mathbf{z}_n^{[k]}|\mathbf{x}_n, \mathbf{m}) \tag{14}$$

Let's notice that data-fusion, in this case, takes place in a preprocessing step.

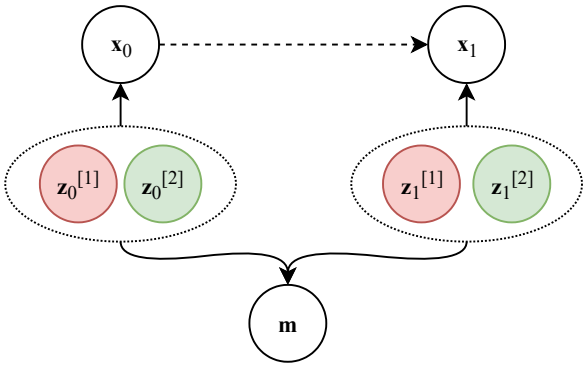

Figure 2. Super-observation

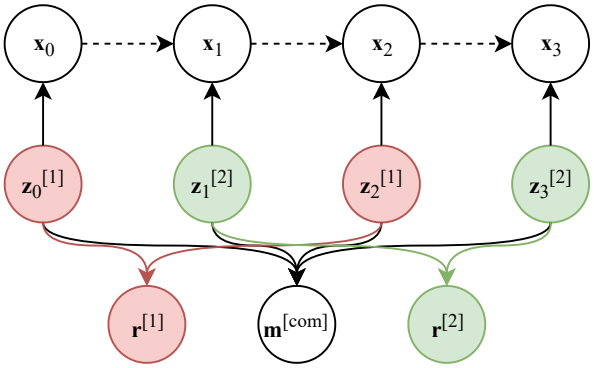

Figure 3. Partially collective mapping

3.3.3. Partially collective mapping

The third and final archetype we presenting in this section is unique in its map composition. At least part of the map representation is common to all available modalities and so on its estimation participates all sensors (see Figure 3).

Let's assume that the map representation can be defined as the following collection:

$$\mathbf{m} = \left\{\mathbf{m}^{[com]}, \mathbf{r}^{[1]}, \mathbf{r}^{[2]}, \cdots, \mathbf{r}^{[K]}\right\} \tag{15}$$

where $\mathbf{m}^{[com]}$ is a common part of map (or just a common map) and all $\mathbf{r}^{[k]}$ are modality specific remainder vectors.

Combination of common map $\mathbf{m}^{[com]}$ and a particular remainder vector $\mathbf{r}^{[k]}$ can be interpreted as a particular map $\mathbf{m}^{[k]}$. So common map $\mathbf{m}^{[com]}$ is dependent on every observation and remainder vectors $\mathbf{r}^{[k]}$ are mutually conditionally independent.

$$p(\mathbf{m}^{[com]}|\mathbf{x}_{0:N}, \mathbf{Z}^{[k]}_{0_k:N_k}) \neq p(\mathbf{m}^{[com]}|\mathbf{x}_{0:N}) \forall k \in 1:K \tag{16}$$

Data-fusion is in this case implicitly embedded into the SLAM algorithm.

4. PRACTICAL ASPECT OF COMMON MAP

By analysis of the above-mentioned archetypes, we concluded that the concept of the common map represents a promising way for the development of effective multi-sensor data-based SLAM algorithms because it implicitly enforces a high level of data fusion. However probabilistic approach to this concept is highly abstract and that's why we devoted this section to more specific and practical aspects of this concept.

There are two subsections following. In the first, we are dealing with specifics way to practically implement the concept of the common map which is composing it as parameters of a piecewise function that represent the surface of the observed environment. In the second subsection, we follow up the previous findings

into set requirements on the observation functions that lead to the categorization of real sensors accordingly to their utilizability in the context of geometrical-entities based collective map.

4.1. Geometrical-entities based collective map

Continues function that approximates the surface of obstacles is, in our opinion, an advantageous thing to utilize for the common map definition because standard SLAM capable sensors always observe this quantity in some way. For example, there is a very low probability that data from Lidar, visible spectrum (vis) camera, thermal (IR) camera would share a substantial amount of feature points in terms of belonging to the same spacial points. However, what is highly probable is that these observations would describe the same planes and curves that form the environment surfaces.

Let's have an analytical formula for an observation model, where observation is a vector that in a spatially distinguished point-wise manner describes some quantity exhibited by points of the surrounding environment.

$$\mathbf{z}_n^{[k]} = \mathrm{h}^{[k]}(\mathbf{x}_n, \mathbf{m}^{[k]}, \mathbf{v}_n^{[k]}) \tag{17}$$

where $\mathrm{h}^{[k]}$ is observation function, $\mathbf{v}_n^{[k]}$ is noise vector that models stochasticity of the process.

If we would know that some subsets of the observation elements belongs to specific geometrical-entity we can generally express this knowledge by some equality constraints

$$\mathrm{G}_i(\mathbf{m}) = \mathbf{0} \tag{18}$$

where G_i is function that define constraints specific to i-th entity.

For example, following constraint bounds the specific points to lie on the same line/plane

$$\mathrm{G}_i(\mathbf{m}) = \begin{pmatrix}\mathbf{M}_i & \mathbf{1}\end{pmatrix} \boldsymbol{\pi}_i = \mathbf{0} \tag{19}$$

where $\boldsymbol{\pi}_i$ is a vector of coefficients that defines line/plane and $\mathbf{M}_i$ matrix whose rows are spacial points that belongs to i-th entity.

Parameters that define specific form of the constraint equation (in our example $\boldsymbol{\pi}_i$) are elements that forms the common map $\mathbf{m}^{[com]}$. For practical applications, we also define a projection function g that is used in the optimization process for error evaluation.

$$\mathbf{m}^{[k]} = \mathrm{g}_i^{[k]}(\mathbf{m}^{[com]}, \mathbf{r}^{[k]}) \tag{20}$$

this function have to be from general perspective modality specifics, however, usually, it would be very similar across all modalities.

The consequence of map parametrization in this way is that dimensionality of the map is greatly reduced compared to the non-constraint case and this would very likely have positive effects on the optimization process as shown in [23, 24]. The last practical aspect we discuss in this subsection is the obvious problem that in the real-world scenarios point elements affiliation to specific geometrical entities is apriori unknown. Dividing single observations into parts where each describes the common entity is generally a segmentation problem and the probabilistic way to approach it is by statistical hypothesis testing.

$$p\big(\mathrm{G}_i(\mathbf{m}) = \mathbf{0} | \mathbf{Z}_{0:N}\big) > \alpha \tag{21}$$

where α is the significance level.

This can be practically conducted by defining statistics that evaluates whether the reprojection error can be caused by observation noise

$$t_i = d\big(\mathbf{h}(\mathbf{X}_{0:N}, \mathbf{m}, \mathbf{v} = \mathbf{0}), \mathbf{Z}_{0:N}\big) \tag{22}$$

and comparing it against given critical value $t_i < t_{crit}$.

Anyway, it is obvious that many testable hypotheses gonna be significantly higher then computational resources allow us to test, so necessary part of the segmentation algorithm has to be also a method which generates hypothesis to test. Experiment showing practical example of such algorithm can be find [25].

4.2. Sensors

In perspective of above-mentioned theory, let's analyze what properties have to the observation function meet to be compliant e.g. usable with it. Just for the formalism, we start with the obvious. Firstly, the mathematical model of the sensor has to be consistent with reality. Secondly, any sensor used as the primary source of data for the SLAM algorithm has to measure some spatially dependent quantity that is suitable to be mapped. This leads to a model's ambiguity when state or map is unknown, however, combined knowledge about both state and map forms an information gain.

$$p(\mathbf{z}|\mathbf{x}) = p(\mathbf{z}|\mathbf{m}) = p(\mathbf{z}) \tag{23}$$

$$p(\mathbf{z}|\mathbf{x}, \mathbf{m}) \neq p(\mathbf{z}) \tag{24}$$

From perspective multiple-sensor based SLAM while assuming to have limited resources, it is reasonable also to consider whether all sensors will have a perceptible contribution to overall result.

A form of the contribution is although in this context highly unclear. Generally, it can be viewed as any criterion that evaluates the result. However, we usually think about it as a noticeable improvement of a common map variance.

$$\mathrm{Var}\big[p(\mathbf{m}^{[com]}|\mathbf{Z}^{[1:K]})\big] < \mathrm{Var}\big[p(\mathbf{m}^{[com]}|\mathbf{Z}^{[1:K\setminus k]})\big] \tag{25}$$

where used probability distributions are marginalized distributions

$$p(\mathbf{m}^{[com]}|\mathbf{Z}) = \int_{\Omega} p(\mathbf{X}_{0:N}, \mathbf{m}^{[com]}, \mathbf{r}^{[k]}|\mathbf{Z}) d\mathbf{X}_{0:N}, \mathbf{r}^{[k]} \tag{26}$$

where Ω represent domain of marginalized quantities. Such criterion is however practically impossible to compute a priory and only real possibility is to evaluate it experimentally. We used this condition to classify the usage of various sensor types the overview is in Table 1 and detailed descriptions are following.

Table 1. Sensor type categorization

Category	Example	Usage
Low DOF	Thermometer	Mapping
Inertial	Accelerometer	Motion model
Modality profile	Camera	SLAM
Local structure	Lidar	SLAM
Link to ref. frame	GPS	Position reference

4.2.1. Low degrees-of-freedom

To this category belongs sensors which quite clearly cannot satisfy perceptible contribution condition because a number of degrees-of-freedom (DOF) of their observation range does not allow unambiguous enough localization in the observer's state-space. Typical members of this group are scalar sensors of local environmental quantities i.e. thermometer, light-intensity sensor, etc., but also a linear lidar can be listed here while assuming that the observer is moving in 3D space with 6 DOF. Sensors from this category can be used for unique modality map creation (assuming that pose data is provided from another source), however, direct contribution to SLAM algorithms can be considered to be none (with exception of some multi-modal localization scenarios where correct mode can be chosen only by unique environmental quantity).

4.2.2. Inertial

This is a category of sensors that provide data that brings links between subsequent observer state e.g. forms data for motion model. It is clear that these sensors do not fulfill the observing environmental quantity condition - they have no link to environment structure. This group consists of various encoders, accelerometers, gyroscopes, etc. These are the typical support sensors that have no direct way to contribute to the common map estimation but data from. Because historical reasons observations from these sensors are marked with symbol $\mathbf{u}$ rather than $\mathbf{z}$.

4.2.3. Modality profile

Sensors from this category are generally sensor that observes the properties of some ambient signal generated by the environment. From a practical perspective, these are strictly various types of cameras that measure directional characteristics of intensity of electromagnetic radiation on specific spectral interval (light). By assuming that individual parts of the obstacle surface emitting e.g. reflecting the light in such way that it is possible to identify the same spacial points in multiple images, we can use photogrammetry to reconstruct viewed structure. Characteristic property is that standard photogrammetry techniques applied on single-camera data can provide reconstruction invariant only up to unknown similarity transformation. So the scale of unknown and if needed then have to be fixed by implementing additional data into the process. Sensors of this category can be under the right conditions used for realization of SLAM as shown for example by [26] or by [27] and also can be addition to multi-sensor SLAM system.

4.2.4. Local structure

This category contains the most typical sensor used in the context of SLAM algorithms. Observations provided by these sensors represent the profile of the surrounding environment from their perspective. Typical members of this group are lidars, rangefinders, and RGB-D cameras and they have the potential to be a contribution in the sense of common map estimation.

4.2.5. Link to reference frame

As the designation probably suggests sensors of the last group provide direct information about position in some reference frame. It is sensors like global navigation satellite system (GNSS), local positioning systems (LPS) surveyed for example by [28], any similar beacon-based system or even a compass. From a formal perspective, these sensor does not observe any environmental property so primary they can not contribute to estimation common map, although they have a large potential to contribute indirectly as link to reference frame can eliminate any drift in pose estimation. The main problem is that these sensors may work poorly in urban areas or indoor (GNSS) or they require some special infrastructure (LPS), and so these data are rarely available. Let's notice that a substantial part of motivation to SLAM algorithms lies in that the pose data are directly unavailable or at least unavailable in sufficient quality.

5. CONCLUSION

We presented our theoretical analysis of fundamental aspects of multiple-sensor data-fusion based SLAM problem from probabilistic approach perspective. We concluded that the most promising way to generally approaching it is by utilizing the concept of a common map as shown by presented archetype partially collective mapping. As we see it the typical nowadays published SLAM algorithm based on data-fusion is similar to super-observation archetype, but these concepts are in our opinion suboptimal in terms of robustness. Every sensor has some limitation that determines situations where it can be used. Super observation concept will safely work in situations given by the intersection of all sensors applications fields. On the contrary, the partially collective mapping archetype can work in situations given by unification of all sensors applications fields.

From a practical perspective, we discussed options for common map implementation. As a mapped quantity we proposed to utilize the surface of obstacles and describing it as a piece-wise function composed of simple geometrical entities. After that, we find out three major problems that have to be solved before implementation. Firstly, the mathematical model of geometrical entities must be defined. That includes defining constraints equations, specific form of common map vector and sensors-specific remainder vectors and projection function. Secondly, some statistics posing as a segmentation criterion must be defined. And lastly, a strategy for selecting regions to test on the geometrical-entity hypothesis must be defined. We have confidence in the proposed method and our future work will be aimed at the creation of real implementation and conducting experiments that comparing its quality on publicly available datasets.

ACKNOWLEDGEMENT

The completion of this paper was made possible by the grant No. FEKT-S-17-4234 - "Industry 4.0 in automation and cybernetics" financially supported by the Internal Science Fund of Brno University of

Technology and by the grant No. TN01000024 by the National Competence Center-Cybernetics and Artificial Intelligence.

REFERENCES

[1] C. Cadena, *et al.*, "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, Dec 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7747236/

[2] S. Huang and G. Dissanayake, "A critique of current developments in simultaneous localization and mapping," *International Journal of Advanced Robotic Systems*, vol. 13, no. 5, pp. 1–13, 2016.

[3] G. Bresson, *et al.*, "Simultaneous Localization and Mapping: A Survey of Current Trends in Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, Sep 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8025618/

[4] M. Beul, *et al.*, "Fast Autonomous Flight in Warehouses for Inventory Applications," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3121–3128, Oct 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8392775/

[5] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, IEEE, Nov 2007, pp. 225–234. [Online]. Available: http://ieeexplore.ieee.org/document/4538852/

[6] R. A. Newcombe, *et al.*, "KinectFusion: Real-time dense surface mapping and tracking," in *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, IEEE, Oct 2011, pp. 127–136. [Online]. Available: http://ieeexplore.ieee.org/document/6162880/

[7] J. Mccormac, *et al.*, "Fusion++: Volumetric Object-Level SLAM," in *2018 International Conference on 3D Vision (3DV)*, Verona, 2018, pp. 32–41. [Online]. Available: https://ieeexplore.ieee.org/document/8490953/

[8] T. Whelan, *et al.*, "ElasticFusion: Dense SLAM Without A Pose Graph," in *Robotics: Science and Systems XI. Robotics: Science and Systems Foundation*, Jul 2015. [Online]. Available: http://www.roboticsproceedings.org/rss11/p01.pdf

[9] T. Whelan, *et al.*, "Real-time large-scale dense RGB-D SLAM with volumetric fusion," *International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 598–626, 2015.

[10] F. Burian, P. Kocmanova, and L. Zalud, "Robot mapping with range camera, CCD cameras and thermal imagers," in *2014 19th International Conference on Methods and Models in Automation and Robotics, MMAR 2014*, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 200–205.

[11] F. Fang, X. Ma, and X. Dai, "A multi-sensor fusion SLAM approach for mobile robots," in *IEEE International Conference Mechatronics and Automation*, 2005, vol. 4, no. 2002. IEEE, 2006, pp. 1837–1841. [Online]. Available: http://ieeexplore.ieee.org/document/1626840/

[12] A. Garulli, *et al.*, "Mobile robot SLAM for line-based environment representation," in *Proceedings of the 44th IEEE Conference on Decision and Control*, vol. 2005. IEEE, 2005, pp. 2041–2046. [Online]. Available: http://ieeexplore.ieee.org/document/1582461/

[13] Y.-H. Choi, T.-K. Lee, and S.-Y. Oh, "A line feature based SLAM with low grade range sensors using geometric constraints and active exploration for mobile robot," *Autonomous Robots*, vol. 24, no. 1, pp. 13–27, Jan 2008. [Online]. Available: http://link.springer.com/10.1007/s10514-007-9050-y

[14] C. Ulas and H. Temeltas, "Plane-feature based 3D outdoor SLAM with Gaussian filters," in *2012 IEEE International Conference on Vehicular Electronics and Safety, ICVES 2012*, 2012, pp. 13–18.

[15] H. Zhou, *et al.*, "StructSLAM: Visual SLAM With Building Structure Lines," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1364–1375, Apr 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7001715/

[16] K. Uehara, H. Saito, and K. Hara, "Line-based SLAM Considering Directional Distribution of Line Features in an Urban Environment," in *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, no. Visigrapp. SCITEPRESS - Science and Technology Publications, 2017, pp. 255–264.

[17] S. Yang, *et al.*, "Pop-up SLAM: Semantic monocular plane SLAM for low-texture environments," *IEEE International Conference on Intelligent Robots and Systems*, vol. 2016-Novem, pp. 1222–1229, 2016.

[18] A. Jelinek, L. Zalud, and T. Jilek, "Fast total least squares vectorization," *Journal of Real-Time Image*

Processing, pp. 1–17, Jan 2016. [Online]. Available: http://link.springer.com/10.1007/s11554-016-0562-6

[19] R. Hulik, *et al.*, "Continuous plane detection in point-cloud data based on 3D Hough Transform," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 86–97, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jvcir.2013.04.001

[20] K. Pathak, N. Vaskevicius, and A. Birk, "Uncertainty analysis for optimum plane extraction from noisy 3D range-sensor point-clouds," *Intelligent Service Robotics*, vol. 3, no. 1, pp. 37–48, 2009.

[21] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006. [Online]. Available: http://ieeexplore.ieee.org/document/1638022/

[22] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. The MIT Press, 1999.

[23] J. Klecka, *et al.*, "Non-odometry SLAM and Effect of Feature Space Parametrization on its Covariance Convergence," in *IFAC-PapersOnLine*, no. 25, Brno, 2016, pp. 139–144.

[24] J. Klecka and O. Bostik, "Effects of Environment Model Parametrization on Photogrammetry Reconstruction," in *Mendel*, Brno, 2018, pp. 151–158.

[25] J. Klecka, *et al.*, "Plane segmentation and reconstruction from stereo disparity map," in *Mendel*, Brno, 2016, pp. 199–204.

[26] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-Scale Direct Monocular SLAM," in *Lecture Notes in Computer Science*, T. T. Fleet D., Pajdla T., Schiele B., Ed. Cham: Springer, 2014, pp. 834–849. [Online]. Available: http://link.springer.com/10.1007/978-3-319-10605-2 54

[27] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, Oct 2016. [Online]. Available: http://arxiv.org/abs/1610.06475

[28] K. Curran, *et al.*, "An evaluation of indoor location determination technologies," *Journal of Location Based Services*, vol. 5, no. 2, pp. 61–78, Jun 2011. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/17489725.2011.562927