Applications of LIDAR and camera fusion

Pavel Svoboda, Pavel Zemčík
Department of Computer Graphics and Multimedia
Faculty of Information Technology, Brno University of Technology
Božetěchova 2, 612 66 Brno, Czech Republic
[isvoboda,zemcik]@fit.vutbr.cz

Abstract *— many situations exist where data fusion of different sensor provides more accurate solutions than just a single sensor measurement. Short introduction of LIDAR and camera sensors is provided in order to understand the utilization of the data fusion from those sensors. Three diametrically different fusion approaches are described and compared in this paper.*

1 INTRODUCTION

Sensory information, such as processed data from video-sensors, LIDAR, RADAR, etc. is traditionally used to obtain knowledge about the environment in many applications. Multimodal fusion can enrich such information and is used due to a number of factors. The frequently mentioned reasons are complementariness and redundancy of sensors. Better accuracy can be achieved due to complementariness. The rich input of required sensor data leads to a better and more appropriate decisions. Redundancy is used in situation when failures of one sensor can be critical.

As the progress of autonomous robots is growing, many new methods of fusion have been introduced and describe. This paper introduces different methods of fusion. The fusion is generally always advantageous. It can provide more accurate data used for the decision of autonomous robots.

This paper describes three used approaches of fusing LIDAR and camera, which were published in [1,2,3]. LIDAR and Camera sensors and their important characteristics are defined in chapter 2. Chapter 3 contains three parts describing different approaches of LIDAR and camera fusion.

First example could be understood as model verification using fusion, second method describes one of the most used applications of the LIDAR and the camera fusion – fusion of its occupancy grid maps and third example shows fusion of LIDAR and camera data using statistical methods and algorithms from field of machine learning.

2 SENSORS

Fusion of sensor deals with different sensor inputs. Methods of fusion have to produce reliable outputs – estimates and environment models which can be used by other navigation subsystems. In this sense of fusion it is integrating different sensor data for object detection, self-location, motion planning etc.

There are two different sensors described in this paper are LIDAR and camera.

2.1 LIDAR sensor

Abbreviation LIDAR means Light Detection and Ranging, is generally defined as the integration of technologies into a single system capable of acquiring data to provide knowledge about surrounding environment. These technologies are LRF – laser range finder, the Global Positioning System (GPS), and inertial navigation systems (INS). Combined, they allow on high degree of accuracy creation of model of the surrounding environment.

LRF enables accurate and instantaneous distance measurement of noncooperative targets from several tens of centimeters to several tens of kilometers. LRF is nowadays one of the most precise system for ranging. Due to its characteristics it is used in many branches of industry. From the view of application use is LIDAR often used in DMT - Digital Model Terrain gathering, 3D model of robots area reconstruction or for obstacle detection.

Because the described approaches do not use any information from GPS nor INS, LIDAR is reduced to LRF.

LRF sends out a pulse of light. The return of the pulse creates a point with an XYZ position in the real world. LRF works with principles of time-of-flight measurement. With nanosecond pulses and fast optomechanical beam scanning, LIDAR provides linear, unidirectional and parallel scan lines.

LIDARs can be divided into subcategories according to possible field of view. There are numerous types of range finder LIDAR. The most basic ones are directional fixed. These can provide range sensing only for one point. Other types have rotating mirror, which can provide more than one range sensing in some amount of time in horizontal or also in both horizontal and vertical directions.

The used wave lengths differ according to the application of LIDAR. In the environment where eye-safety is needed less powerful lasers are used, or lasers with wider beam divergence. The laser safety is significantly enhanced if the operating wavelength of the laser is shifted beyond 1400nm where it would be effectively absorbed by the ocular media or to approximately 1540nm, the so called 'eye-safe' band [4].

2.2 Camera sensor

Digital camera sensors use CMOS or CCD chips. These chips transform incident light into electrical charge. Vision can be generally divided in two categories, monocular and stereovision.

In monocular vision the data that is processed is still deeply investigated for many purposes. Filtering and transformations are done in the image processing layer, while classification, detection, etc. are part of the computer vision layer. Generally algorithms for object detection are important in monocular vision, while in stereovision the epipolar geometry or techniques based on disparity image are used. Calibration of cameras in stereovision is crucial.

Camera sensor can be described by its resolution, speed of incident light processing – frame per second and possibility of monochromatic - chromatic light receiving.

3 APPROACHES OF LIDAR AND CAMERA FUSION

The following approaches describe the data fusion of LIDAR and camera. These approaches can be divided according to the type of the data they are fusing. We consider 3 levels of processed data, the fusion of low level features – intensities of pixels, range of point in space but also specific image descriptors and feature vectors made from LIDAR data, high level features – fusion of classifiers, etc. and the main principle of fusion in robotics – the fusion of occupancy grid maps obtained from camera and LIDAR.

Exceptions exist, which can’t be classified in these categories, for example, the first described approach.

3.1 Validation of 3D model

This approach of data fusion was introduced in [1]. The authors describe the data fusion as a method of model validation. LIDAR and stereovision is used. This method is usable in the indoor conditions according to its assumptions. The approach is divided in two main steps. Creating and evaluating 3D model.

LIDAR is used for mapping the robots environment. LIDAR provides distance information in the horizontal plane. 3D points are obtained and have to be processed to create 3D model. Authors assume the simple condition, that every obstacle can be described as vertical plane with infinite height. The obtained 3D points are grouped into the line segments. These line segments are afterwards evaluated as obstacles. Then the 3D model of environment is created.

The phase of 3D model evaluation is processed thanks to the stereovision. Pixels obtained by one of the cameras are ray-traced into the 3D model. This evaluation provides 3D position of every pixel. Now image of the second camera can be used for validation. All projected pixels from the first camera are re-projected to the second camera. If the model is correct, then the re-projected and real image of the second camera has to be the same. Comparison is done by local crosscorelation of intensities or color values.

If the model is not correct, the depth information is extracted using the stereovision – epipolar geometry.

There are many difficulties in using this approach. Poor texture information, the lightening, etc. are often the problem.

This approach of fusion can be used for obstacle avoidance and mapping the robots surrounding environment. As it was mentioned, this approach is better for indoor.

3.2 Fusing grid cost maps of LIDAR and camera

In robotics and especially in branches such as mapping, localization and path planning the concept of occupancy grid map is often used. Grid occupancy map contains cells which define presence of an obstacle, free place, or unknown area. There is plethora of approaches how to obtain grid occupancy map. The method always depends on a type of a sensor.

Authors Moghadam et al. used fusion of two grid occupancy maps in [2]. Obtaining the LIDAR occupancy grid map is quite straightforward. The authors use LIDAR which scans environment of the robot in 360° degreese. Occupancy grid map is constructed according to the distance of returned point.

To obtain occupancy grid map from stereovision, there are different approaches mentioned according to place of use the robot. Generally spoken, the dual-lens camera provides complementary 3D information in a narrow field of view (66 degree) over a relatively short range (5 - 8m) in the environment. These observations provide measurements of objects of various heights and positions that permit the classification of the terrain into ‘traversable’ and ‘nontraversable’ regions taking into consideration the dimensions of the robot. The obtained 3D structures and obstacle information are transformed into the 2D occupancy grid map.

The fusion process of these 2D occupancy grid maps can be done by many less or more sophisticated methods. Authors choose the method of minimizing the probability of any failure due to omitted obstacle.

They used the maximum of the two obtained occupancy grid maps to get the resultant value (empty – traversable, obstacle – nontraversable and unknown).

The results of this method were successful. The principle of fusion was used for improvement of mapping and robot path planning. The examples showed that sometimes the 3D information derived from stereovision can detect obstacles which were not detected by the LIDAR. On the other hand, the distance of the measurement of the stereovision is quite short, while LIDAR can mitigates this weakness.

3.3 Feature vectors fusion of LIDAR and Image data

In 2009, the comparison of methods of machine learning for LIDAR and camera data fusion was published [3]. The Authors, Premebida et al. investigated the robustness of two possible ways of what and how to fuse. They evaluated the measurement on the system of pedestrian detection using LIDAR and monocular camera. The two tested ways of fusion are the centralized and decentralized method. The centralized method works with low level features while the decentralized works with high level features.

15-dimensional feature vector was defined for LIDAR data. The features are for example:

$$\sqrt{\Delta X^2 + \Delta Y^2}$$

— Normalized Cartesian dimension: corresponds to the root mean square of the segment width d_x and segment height d_y.

$$\frac{1}{np}\sum_{n=1}^{np}\left(x_n - \hat{x}_{l,n}\right)^2$$

— Linearity: this feature measures the straightness of the segment and corresponds to the residual sum of squares to a line $x_{l,n}$ fitted into the segment in the least-squares sense, where np is number of points.

The full definition of LIDAR feature vector can be found in [3]. Data from LIDAR are first projected into the 2D occupancy grid map in Cartesian or polar space. Then feature vectors are extracted from the defined segments S_k – often some kind of ROI. If vertical information is available, it has to be projected into the 2D map. This feature vector is understood as a low level feature in terms of fusion.

HOG - histogram of gradients and COV - covariance matrix descriptor are used for image data. HOG descriptor's parameters were adapted and finally 81-dimensional feature vector was used.

COV was also utilized according to data. More details can be found in [3]. In the end 180-dimensional feature vector was used.

3.3.1 Centralized fusion

Feature vectors of LIDAR data and vision data were combined into one 276-dimensional feature vector, which was used as input for classification algorithms (Fishers LDA, RBF–SVM and MCI–NN). According to the authors, the best classifier - Fishers LDA provide 0.8 true positive by 0.1 false positive on the ROC curve.

3.3.2 Decentralized fusion

Feature vectors from LIDAR and vision were passed as unique inputs into different classifiers. The most accurate methods were chosen just for the one type of feature vector and their results were fused by classification method. In another words, results of single classification are inputs in to the main classifier which makes the decision. Fishers LDA, Naive Bayes, GMM, SVM and NN were used for single classification. The most accurate single classifiers were naive-Bayes for LIDAR data and Fishers LDA for visual data. It seems to be logical to fuse results of these most accurate single classifiers, but authors confirmed, according to experiments, that for fusion is better so called "RMR - minimal-redundancy-maximal-relevance" strategy. According to the MRMR strategy the GMM for LIDAR data and the Fishers LDA for visual data were chosen. Results were fused by GMM with accuracy 0.912 true positive by 0.1 false positive on the ROC curve.

4 CONCLUSIONS

The fusion of LIDAR data and vision data can be generally divided into:

— Fusion of low level features
— Fusion of high level features
— Fusion of occupancy grid maps
  Often used for path planning and mapping
— Fusion as model verification
  Rare described as first example of fusion

Nowadays the autonomous robots and vehicles are expanding from first prototypes into industry and common life. It is challenge to move the technology step ahead. Fusion is expectable aspect of providing more accurate input data for deciding, avoiding of critical sensor failures etc.

This text describes three different methods of the LIDAR and the vision data fusion. Fusion as verification is the first usable approach.

Current and common approach is shown in the second example, where the fusion is built up on getting information from occupancy grid maps together. These are often used for defining robot surrounding environment.

The final approach uses feature extraction of both, LIDAR and visual data, to fuse them by trainable or nontrainable methods of machine learning.

The goal of the summary is to provide a general overview of today's possibilities of data fusion, especially of LIDAR and vision data.

It is not so common to use the data fusion for path planning, as it was described in the first example. On the other side, the approach of construction of the simple 3D indoor model seems to be appropriate for tasks where main basic model of surrounding walls is important and needed.

The LIDAR and camera data fusion of the occupancy grid maps is nowadays the most common approach of robot environment mapping and path planning. The application of the approaches built on combining occupancy maps seems to be effective and it is appropriate for tasks of path planning and mapping.

The third example of the LIDAR and camera data fusion shows measurable results of combining the information. The benefit is also in the use of the LIDAR features then the raw data from LIDAR, as it was described in previous examples.

Acknowledgments

This work has been supported by the project of the EU Artemis project R3COP: Robust & Safe Mobile Cooperative Autonomous Systems grant no. 100233 and by the project of BUT – FIT: Recognition and presentation of multimedia data, grant no. FIT-S-10-2.

References

[1] Baltzakis et al., *Fusion of laser and visual data for robot motion planning and collision avoidance*, Machine Vision and Applications, v.15 n.2, p.92-100, December 2003

[2] Moghadam et al., *Improving Path Planning and Mapping Based on Stereo Vision and Lidar*, Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Confe-rence on, 2008

[3] Premebida et al, *LIDAR and Vision-Based Pedestrian Detection System*, Journal of Field Robotics 26(9), p. 697-711, Wiley Periodicals Inc., 2009

[4] Bahuguna, K.C. et al., *Laser Range Sensors*, Defence Science Journal, Vol. 57, No. 6, November 2007, pp. 881-890