Heterogeneous Feature Based Correspondence Estimation

Levente Tamas and Andras Majdik

Abstract— **This paper gives an insight in the preliminary results of an ongoing work about heterogeneous point feature estimation acquired from different type of sensors including structured light camera, stereo camera and a custom 3D laser range finder. The main goal of the paper is to compare the performance of the different type of local descriptors for indoor office environment. Several type of 3D features were evaluated on different datasets including the output of an enhanced stereo image processing algorithm too. From the extracted features the correspondences were determined between two different recording positions for each type of sensor. These correspondences were filtered and the final benchmarking of the extracted feature correspondences were compared for the different data sets. Further on, there is proposed an open access dataset for public evaluation of the proposed algorithms.**

I. INTRODUCTION

The perception of the surrounding 3D environment is still a significant research field both for the industrial and researcher community. There are several applications in which the 3D representation of complex environments is essential including: architecture, automotive, mining and piping systems inside factories. Off-line 3D data is required for applications such as architecture or factory design, scenarios where the data sets can be conveniently acquired by using several post processed 3D scans. Other type of applications, e.g. autonomous navigation may require real-time data from the environment [1].

In order to capture 3D information passive sensing techniques can be used as a low cost alternative, which includes a wide variety of devices starting from cheap web-cameras to sophisticated multi-camera systems. The main disadvantage of these approaches are related to the captured scene dependency, including illumination and texture for the stereo imaging [2].

In contrast to the passive imaging methods, the active sensing techniques including laser range finder or structured light cameras emit energy towards the scene and measure the reflected values. The laser range finders capture a predefined order of point sequence on a regular spherical grid, in a time-dependent manner, hence they can be used only for static scenarios. They still provide accurate, robust and long range measurements from the environment [3]. Planar 2D laser scanners are popular in the field of robotics due to their high speed and precision. They are mainly used for 2D mapping and navigation purposes, although by augmenting the two-axial motion with an additional degree of freedom, an accurate 3D scanner is obtained [4].

On the other hand, the need for capturing complex dynamic scenes is essential for a wide range of applications. Recent research outcomes show that for close-range dynamic scenarios, the structured light cameras provide a high-update intensity and colour information. The idea of using an known projected pattern on a scene and viewing the results of the illumination does not require an expensive hardware. Practically, the 3D information can be retrieved by simple triangulation methods applied to the reflected data [5].

In this paper we are particularly interested in the evaluation of robustness regarding keypoints and local descriptors for the active laser range finder providing range data and the Kinect commercial structured light sensor mounted on the same platform with an off-the-shelf Bumblebee stereo camera. These local descriptors can be used for different applications, including object detection and classification as well as localization or mapping, which are common topics within the robotics research. Starting from this idea, we briefly present three perception devices widely used in this domain. First, a custom built 3D laser range finder based on a 2D planar laser for range data acquisition is presented. For the dense coloured depth point cloud we examine different type of stereo image processing algorithms as well as the fused infrared depth and RGB data from the Kinect. Further on, for the different data sets specific features and the correspondences among them are evaluated. Finally, the paper presents the results including the comparison between the tested approaches.

A. Related Work

The idea of fusing several type of 3D perception sensors [6] is applied in several applications including face detection [7], scene analyses [8], simultaneous localization and mapping [9], architectural industry [10], etc.

For these applications the relation between different recorded datasets represents a widely encountered problem. This problem can be summarized as the matching of similar properties between two or more registration for object recognition purposes, localization or tracking.

The recorded points are considered to represent 3D information of the surrounding, recorded from different sensors. Comparing two points recorded in different positions requires a common metric. Using some simple metric like the Euclidean one may not lead to sufficiently robust results. Instead of this local regions around the evaluated points are selected which ensure a better representation of a local zone in the space. These local patches are introduced as local descriptors in the domain literature [11].

The computation of the local feature can be time consuming, thus a common approach is to determine the descriptors only for a limited number interest points, i.e. keypoints. It is also important to select regions which are rotation

invariant and down sampling immune in order to ensure the robustness of the local descriptor. There are several possibilities for extracting keypoints and descriptors from 2D images including the popular SIFT (Scale Invariant Feature Transform) [12] Unfortunately, these rely on local gradients from a unique orientation and therefore are not directly applicable for our approach with 3D data, however some concepts may be inherited from the 2D domain. Compared to the camera images, the 3D sensors provide also the depth information, which introduce an additional robustness in the local feature estimation.

There are several approaches for the 3D range point cloud local descriptors, a part of them being rotation invariant around the normal, like in the case of the NARF descriptor [13] or even complete 3D orientation invariant as the fast point feature histogram (FPFH) [14]. Other approaches include normal based keypoint descriptors or radius based ones such as the Radius-based Surface Descriptor (RSD) introduced by [15]. For the RGB-D point cloud there are recent keypoint variants adopted such as the SIFT3D [16].

There are wide range of methods for the correspondence estimation based on different techniques for the correspondence computation like one-to-one, back and forth or sample consensus based one [17], [18], which can be used for evaluating the performance of descriptor.

B. Contributions

In this paper we propose a common framework for feature based correspondence evaluation of different 3D data including range data from laser ranger, RGB-D data from a structured light camera and the coloured point cloud from a stereo camera, calibrated in a same mobile robot platform. For the range data a custom hardware setup is presented, while for the feature estimation from stereo images different algorithms were considered. The main idea of the paper is to use these heterogeneous keypoint-descriptor pairs in a common coordinate frame and evaluate the different pair combinations of sensors, based on the performance of the estimated correspondence.

Also an open access code and recorded database will published on the internet by the time of publication in order to encourage the further investigation and evaluation of the different feature correspondence evaluation researches.

II. PROPOSED METHOD

For evaluating the different ranging devices, the data captured needs to be transformed in a common coordinate frame. In our case the mobile robot platform on which the sensors were mounted was considered to be the origin of this common coordinate frame. With this assumption, the change in the orientation and position of the platform needs to be estimated from the different correspondences.

For the correspondence evaluation metric first a rigid transformation based on the correspondences between two different point clouds is determined. This transformation is compared against the reference transformation between the two measurement coordinates. The reference point cloud is determined via the fine tuning of the 3D point cloud from the laser range finder, as this offers a precision under $1cm$ in an indoor environment. The fine tuning is done by applying a iterative closest point (ICP) algorithm to the initially aligned laser point cloud based on the information from the correspondence estimation [19].

By taking the normalized point cloud distance between the different rigid transformation estimation, the evaluation of the different feature correspondence estimation is performed. According to this idea, evaluation of the heterogeneous feature sets were tested, by means of fusing the range feature set with the kinect and the ones extracted from the stereo imaging. Also other combinations of the feature set are proposed within this framework. In order to reduce the false correspondences, a sample consensus based filtering is applied to the heterogeneous feature set [18].

The reference transformation is computed based on a FPFH type descriptor for the laser data recorded at two different positions as this is listed in the Algorithm 1. First the F_s and F_t descriptors are computed for the source and target datasets, which are used for the initial alignment guess for the ICP. The ICP based registration with a sufficiently good initial guess serves a high precision registration for a laser ranger 3D data set. The final ICP transformation t_{ref} is returned as the reference transformation for the feature based correspondence rigid transformation estimation [20].

Algorithm 1 ICP with initial alignment for reference

Require: L_s, L_t
1: F_s = ComputeFPFH(L_s);
2: F_t = ComputeFPFH(L_t);
3: $(t_{init}, C_f) = InitialAligment(F_s, F_t)$;
4: **while** $(error_{diff} < \epsilon) or (iteration < iteration_{max})$ **do**
5: $\quad A_d = getClosestPoints(t_{ref}, L_s, L_t)$;
6: $\quad t_{ref} = argmin\left(\frac{1}{|A_d|}\sum_{j\in A_d} w_j \left|t(l_s) - l_t\right|^2\right)$;
7: **end while**
8: **return** t_{ref}

The proposed feature evaluation algorithm is presented in Algorithm 2. In the first part of this algorithm the 3D features descriptor D_s and D_t are computed for both input point clouds. In the next step the correspondences are evaluated and filtered out based on a sample consensus framework. Finally, the normalized distance E_d for the estimated transformation t_c and the reference t_{ref} is determined.

Algorithm 2 Correspondence estimation

Require: P_s, P_t, t_{ref}
1: D_s = Compute3DFeatures(P_s);
2: D_t = Compute3DFeatures(P_t);
3: (t_c, C_f) = SACFilterCorrespondences(D_s, D_t);
4: $E_d = \frac{1}{|C_f|}\sum_{j\in C_f} w_j \left|t_{ref}(p_s) - t_c(p_s)\right|^2$;
5: **return** E_d

III. EXPERIMENTS

The datasets were recorded indoor at normal light conditions for spaces ranges between a few cm up till $10m$. The scanning of the environment with the *P3* type mobile robot was performed in a stop-scan-go fashion, a single scan taking up to 60 seconds depending on the used custom laser range finder configuration. All the measured data was integrated in the ROS [1] environment where each logging was timestamped for an easier off-line processing.

A. Range Image Features from Custom 3D Laser Ranger

1) Data Acquisition from Custom 3D Laser Ranger: For a scanner with pitching actuator the 3^{rd} dimension of a point is given from the pitch angle. The coordinates of one 3D point result from the distance to the surface, the yaw angle of the beam, and the pitch angle of the actuated mechanical part. Thus a scanned point can be represented as a tuple of the form $(\rho_i; \theta_i, \gamma_i)$ where ρ_i represents the depth information from the laser scanner and θ_i, γ_i the yaw and pitch measurements. The forward kinematic transformation taken as original coordinate system to the laser base link is given by:

$$p = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \begin{pmatrix} \cos\gamma & 0 & \sin\gamma \\ 0 & 1 & 0 \\ -\sin\gamma & 0 & \cos\gamma \end{pmatrix} \times \begin{pmatrix} \rho\cos\theta \\ \rho\sin\theta \\ 0 \end{pmatrix} \quad (1)$$

where p is a point in the Cartesian space with the coordinates x_n, y_n and z_n.

The key component of the 3D sensor is the 2D commercial laser scanner for which a custom rotary platform was designed. There are several possibilities to rotate the laser scanner, i.e. around the yaw, pitch or roll axis, thus achieving a yawing, pitching or rolling 3D sensor [4]. Each of these three setups has its own advantage and disadvantage. As for mobile robots the most common approach is the pitching scan, which was adopted for the current system. The mechanical design and prototype are presented in Figure 1. The design shown has two parts: one fixed containing the driving servo motor (left) and the rotation encoder (right); and the mobile rotary on which the Sick LMS200 is placed. The prototype was built using an iron frame both for the fixed and mobile part [21].

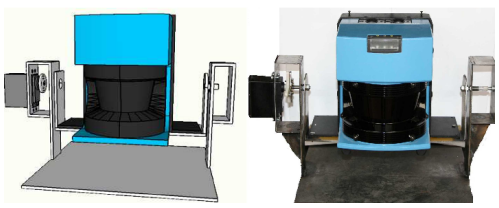

Fig. 1. The design and prototype of the actuated 3D sensor.

[1] http://www.ros.org/

2) Range Data Keypoints and Features Analysis: In this paper the Normal Aligned Radial Feature (NARF) keypoints introduced by [22] were adopted for the range image interest points extraction. This type of keypoint takes into account the information about the boarders and surfaces, ensures the detection from different perspectives and the stability for the descriptor computation phase. The most important parameter for the NARF keypoint extraction is the support size, i.e. the diameter of the sphere in which the interest point characteristics are determined [13]. In our case several values for this parameter were tested in order to gain a sufficient robust and enough number of keypoints for different type of datasets.

After the selection of the keypoints, the specific properties, i.e. descriptors for the set of extracted keypoints are determined. The role of the descriptors is the efficient comparison for discrimination between two selected points. There are several approaches for the descriptors, a part of them being rotation invariant around the normal, like in the case of the NARF descriptor [13] or even complete 3D orientation invariant as the fast point feature histogram (FPFH) [14].

For our approach we used the optimized version of the FPFH in order to augment the three dimensional space with pose-invariant local features and we also tested the NARF descriptors with Manhattan metrics for the same set of keypoints. In order to compare the two set of descriptors, the runtime (T) and the initial alignment fitness score (S) was computed for indoor (Id) cluttered and uncluttered data sets. The result of the comparison is summarized in the Table I.

TABLE I

3D RANGE FEATURE DESCRIPTOR PERFORMANCE COMPARISON

$Dataset$	T_{NARF}	T_{FPFH}	S_{NARF}	S_{FPFH}
$Id_{cluttered}$	0.17	23	0.70	0.29
Id_{plane}	0.10	6	0.91	0.55

The tests were performed on datasets containing around $10K$ points for which the extracted number of keypoints was in the magnitude of $0.1K$. For computing the runtime the average values were considered for 10 consecutive runs on an Intel Pentium 4 single core laptop running Ubuntu Linux.

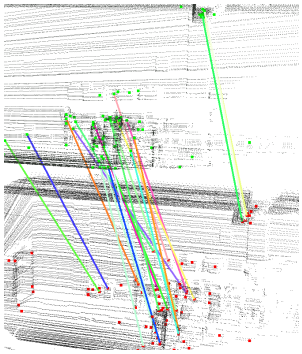

Fig. 2. Feature matching for laser range data for a typical rotated scenario

As observed, NARF descriptors are computed with several orders of magnitude faster than FPFH descriptors, but

the latter approach is more robust in terms of estimating correspondences. This would be also the case for scenes which present less clutter or variation, thus having less discriminative features, where the FPFH features ensured a better correspondence between points. A typical feature correspondence estimation applied to a rotated scene is presented in Figure 2. As it can be observed, the the majority of feature correspondences are valid ones, giving a reliable initial transformation estimation between the two scenes.

B. 3D Features for Stereo Images

1) Principle of Stereo Image Depth Extraction: The stereo cameras are popular tools for creating 3D coloured data based on two or more images recorded from different camera positions. In order to recover depth information from 2D images correspondences have to be established between pixels of the left and right image of the stereo pair, representing the same object in the scene.

Numerous algorithms are focusing on finding correspondences [2]. Generally speaking to obtain better precision more complex algorithms are applied resulting in a higher computational effort.

In many applications there is no need of a dense reconstruction for all the pixels in the image. In these cases is sufficient to use only some points, namely to apply image feature detectors [23]. Image features capture relevant distinctive information in images that can be repeatable detected under different lighting conditions, angle of view and scale increasing thus the robustness of the algorithm.

One of the fastest methods to compute dense disparity maps form stereo images is based on correlation analyses. Using this method the depth is computed at each pixel, a grey level around the pixel in the left image is correlated with the corresponding pixel in the right image. The disparity of the best match from the correspondence is determined using the Sum of Absolute Differences (SAD).

Image feature descriptors can be also used in order to generate dense depth maps. In such a scenario for every pixel in the image a complete descriptor is computed. Afterwards the corresponding pixel is searched in the other image of the pair by comparing the feature descriptors.

2) Enhanced Stereo Image Algorithm: From the numerous image feature descriptors presented in the literature Daisy [24] was chosen in order to compute the 3D point clouds presented in Figure 4. The Daisy descriptor has several advantages over other popular methods such as: SIFT [12] or SURF [25]. It can be computed very rapidly for all the pixel of an image. In our experiments for $640x480$ pixels the computational time was around $3.2s$ using a standard PC. Also the Daisy descriptor has better precision than other image feature descriptors according to the literature [26].

The processing pipeline consist of several phases as follows: (1) capture the stereo image pair; (2) rectification and aligning of the images; (3) compute the Daisy descriptor for every pixel of the left, respectively right image; (4) for every pixel of the right image search for the corresponding one in the left image, the search is performed on the same line from the minimum disparity (10 in our experiments) to the maximum disparity (respectively 40); (5) select the best match by comparing the descriptors, resulting the disparity (for some pixels the disparity could not be found); (6) using the stereo geometry of the vision system and triangulation compute the real world XYZ coordinate for the pixels; (7) get the RGB colour and save the point cloud shown in Figure 4.

3) Robust Coloured Point Cloud Features Correspondence: For the RGB-D datasets SIFT3D type keypoints were considered, while for the local descriptors both the NARF and FPFH variants were tested. As in the case of the laser range data, the FPFH proved to be more robust to the rotation type transformations. For the translations the performance of the two type of descriptors were similar.

The output of the feature estimation and the correspondence estimation for the SAD stereo algorithm is shown in the Figure 3. As it can be seen, the correspondences are valid for this type of algorithm, although the quality of the 3D point cloud is rather noisy.

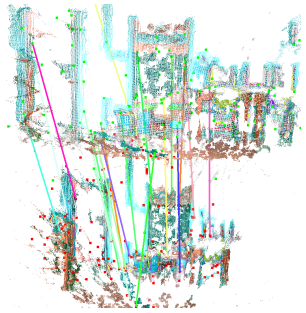

Fig. 3. Feature matching for coloured point cloud from stereo camera with SAD algorithm

The results of the feature estimation for the Daisy stereo image processing is shown in Figure 4. For this case, the generated 3D coloured point cloud has less distortions, although it has depth discontinuousness for ranges above $2m$. Also in this case the keypoint-descriptor pairs have less false correspondences as in the case of the SAD stereo imaging algorithm.

C. 3D Features from Structured Light Camera Images

1) Principle of Depth Measurement via Triangulation: The Kinect RGB-D camera captures infrared and RGB data from the same scenario. It has also an infrared emitter, which is used for structural light projection on the scene as this is described by the inventors in [27].

The projected pattern is captured by the infrared camera and is spatial correlated with the reference one. The effective distance measurement is done on the principle of computing the distance between the reference points from the emitter and the ones received at the infrared sensor via a triangulation using the known baseline distance between the emitter and the receiver.

[2]Two-frame stereo algorithms: http://vision.middlebury.edu/stereo/

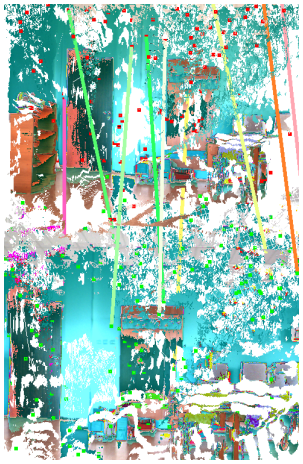

Fig. 4. Feature matching for coloured point cloud from stereo camera with Daisy algorithm

Further on the depth image and the RGB data can be fused by using a calibration based on the fixed parameters of the camera. The main error sources for this device are related to the lighting conditions of the measured scene, as for strong lighting (e.g. sun) the projected infrared pattern have low contrast. This is also the main reason for the indoor usage limitation of this type of sensor.

2) Feature Correspondence Filtering for Kinect RGB-D Data: We tested different keypoint-descriptors for this type of RGB-D data including NARF, SIFT3D and FPFH ones. The best computational effort and descriptor performance was obtained for the SIFT3D keypoint and FPFH feature descriptor pair. For this type of setup the result of the algorithm is shown in Figure 5, which shows two recordings which were taken from the same scenario with the second scene rotated with $30\,\mathrm{deg}$ to the base one.

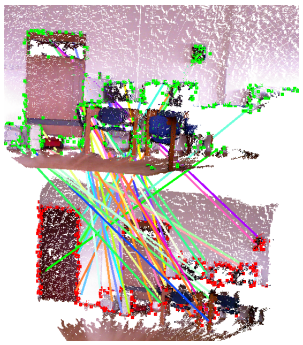

Fig. 5. Feature matching for coloured point cloud from structured light camera, unfiltered correspondences.

Further on, the SAC based filtering of the correspondences gives a reduced set of feature pairs as this is visible in Figure 6. The original correspondence pairs were reduced from 56 to 11. This was done by computing both the back and forth correspondences between the source and target datasets.

For both cases the original RGB-D point cloud was filtered by using a voxel-grid filter in order to reduce the measurement noise in the data and to have a more compact dataset [28].

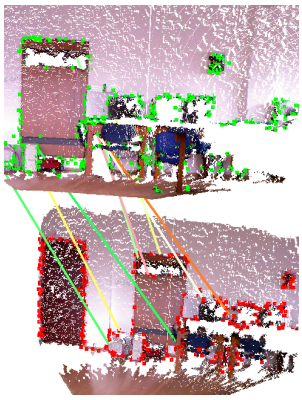

Fig. 6. Feature matching for coloured point cloud from structured light camera, filtered correspondences

D. Heterogeneous Feature Evaluation

The main idea of the heterogeneous feature evaluation relies in the simultaneous use of different type of descriptors from different type of data sets, i.e. stereo camera, Kinect or laser. The motivation behind the use of different feature descriptors is the enhancement of the robustness for the initial alignment of the rigid transformation estimation. Also in the case of the stereo imaging - Kinect combination, ensures a larger range of operation, beyond the limits of the Kinect sensor.

In our test scenario, we used the different stereo imaging algorithms and the kinect data for comparing them against the ground truth laser dataset. Four type of descriptor sets were considered: stereo from SAD, stereo from Daisy, the combination of SAD stereo with the Kinect and the combination of Daisy stereo with kinect. In the case of the RGB-D data, SIFT3D keypoints were considered. These showed to be more powerful than the NARF ones for the colored 3D data.

In Table II the results of the heterogeneous keypoint-descriptor pairs are shown. The initial estimate for the SAD stereo algorithm is compared to the Daisy one and to the combined SAD-Kinect, Daisy-Kinect sets. The Daisy correspondences returned a better transformation than the SAD, while the SAD-Kinect combination over performs both stereo algorithms. The best performance was obtained for the Daisy-Kinect pair in terms of differences from the reference point cloud transformation obtained from the 3D laser scanner considered as reference in our test cases.

TABLE II

HETEROGENEOUS 3D FEATURE DESCRIPTOR PERFORMANCE COMPARISON

$Name$	SAD	$Daisy$	$SAD-Kinect$	$Daisy-Kinect$
$Error$	0.13	0.10	0.094	0.081

The tests were performed on datasets containing around 10^4 points for which the extracted number of keypoints was in the magnitude of 10^2. For computing the runtime the average values were considered for 10 consecutive runs on a $P4$ single core laptop running Ubuntu.

The test datasets were recorded indoor at natural lighting conditions, the transformation between the source and target sets being a rotation plus translation type transformation.

IV. CONCLUSIONS

This paper presents a common framework for evaluating different heterogeneous keypoint-descriptor performances obtained from 3D datasets. The data acquisition is done by using a custom 3D laser range scanner, a stereo camera and a commercial structured light camera. The estimated correspondences were filtered by using a sample consensus based approach.

A. Result Analysis

From the comparison of the different descriptor sets including NARF and FPFH, results that the combined keypoint-descriptor sets show more robustness in terms of the correspondence estimation between two different scenes captured with stereo camera and the Kinect device. The idea of using a heterogeneous keypoint-descriptor set proved to give better results than using them apart. The computational effort is comparable for both cases.

B. Future Work

For the future work we propose the extension of the current indoor dataset to outdoor ones, with additional variants for the keypoint and descriptors and their combinations. The future applications of the proposed algorithm would include the scene recognition within a SLAM framework in an urban environment and providing an open access database for this type of evaluation algorithm both indoor and outdoor.

ACKNOWLEDGEMENT

This paper was supported by the project "Develop and support multidisciplinary postdoctoral programs in primordial technical areas of national strategy of the research - development - innovation" 4D-POSTDOC, contract nr. POSDRU/89/1.5/S/52603, project co-funded from European Social Fund through Sectoral Operational Program Human Resources 2007-2013.

REFERENCES

[1] F. Maurelli, D. Droeschel, T. Wisspeintner, S. May, and H. Surmann, "A 3D Laser Scanner System for Autonomous Vehicle Navigation," in *Proceedings of the International Conference on Advanced Robotics (ICAR)*, 2009, pp. 1–6.

[2] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, pp. 7–42, 2002.

[3] A. Zhang, S. Hu, Y. Chen, H. Liu, F. Yang, and J. Liu, "Fast Continuous 360 Degree Color 3D Laser Scanner," vol. 36, pp. 409–415, 2008.

[4] O. Wulf and B. Wagner, "Fast 3D-Scanning Methods for Laser Measurement Systems," in *Proceedings of the International Conference on Control Systems and Computer Science (CSCS)*, Bucharest, Romania, 2003.

[5] J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827–849, 2004.

[6] S. Chen and Z. Wang, "Acceleration Strategies in Generalized Belief Propagation," *Transactions on Industrial Informatics*, vol. 1, no. 1, pp. 41–48, 2012.

[7] T. Sha, M. Song, J. Bu, C. Chen, and D. Tao, "Feature level analysis for 3D facial expression recognition," *Neurocomputing*, 2011.

[8] M. Attamimi, A. Mizutani, T. Nakamura, T. Nagai, K. Funakoshi, and M. Nakano, "Real-time 3d visual sensor for robust object recognition." in *IROS*. IEEE, 2010, pp. 4560–4565.

[9] V. Castaneda, D. Mateus, and N. Navab, "Slam combining tof and high-resolution cameras," *IEEE Workshkop on Motion and Video Computing*, 2011.

[10] J. Lin, D. Cohen-Or, H. Zhang, C. Liang, A. Sharf, O. Deussen, and B. Chen, "Structure-preserving retargeting of irregular 3d architecture," *ACM Transactions on Graphics (Special Issue of SIGGRAPH Asia)*, vol. 30, no. 6, pp. 183–183, 2011.

[11] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," Ph.D. dissertation, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[13] B. Steder, R. B. Rusu, and K. Konolige, "Point Feature Extraction on 3D Range Scans Taking into Account Object Boundaries," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 2601–2608.

[14] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009.

[15] Z.-C. Marton, D. Pangercic, N. Blodow, and M. Beetz, "Combined 2D-3D categorization and classification for multimodal perception systems," *Int. J. Rob. Res.*, vol. 30, no. 11, pp. 1378–1402, 2011.

[16] M. Grand-Brochier, C. Tilmant, and M. Dhome, "Refa3d: Robust spatio-temporal analysis of video sequences," in *VISAPP (1)*, 2012, pp. 352–357.

[17] D. Chetverikov, D. Svirko, D. Stepanov, and P. Krsek, "The trimmed iterative closest point algorithm," *Pattern Recognition, International Conference on*, vol. 3, pp. 30–45, 2002.

[18] K. Pathak, A. Birk, N. Vaskevicius, and J. Poppinga, "Fast registration based on noisy planes with unknown correspondences for 3D mapping," *IEEE Transactions on Robotics*, vol. 26, pp. 424–441, 2010.

[19] P. J. Besl and H. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[20] L. Tamas and L. C. Goron, "3D map building with mobile robots," in *Control & Automation (MED), 2012 20th Mediterranean Conference on Control*. IEEE (in-press), 2012.

[21] L. C. Goron, L. Tamas, I. Reti, and G. Lazea, "3D laser scanning system and 3D segmentation of urban scenes," in *Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference*, vol. 1. IEEE, 2010, pp. 1–5.

[22] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "NARF: 3D Range Image Features for Object Recognition," in *Workshop at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, 2010.

[23] A. Majdik, M. Popa, L. Tamas, I. Szoke, and G. Lazea, "New Approach in Solving the Kidnapped Robot Problem," in *The 41st International Symposium on Robotics*, 2010.

[24] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Conference on Computer Vision and Pattern Recognition*, Alaska, USA, 2008.

[25] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Underst.*, vol. 110, pp. 346–359, June 2008.

[26] E. Tola and V. Lepetit, "Daisy: An efficient dense descriptor applied to wide baseline stereo," *PAMI*, vol. 32, no. 5, 2010.

[27] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns," Prime Sense Ltd, 2010.

[28] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1–4.