International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.12 (2016), pp.35-44
http://dx.doi.org/10.14257/ijmue.2016.11.12.04

The Research of the Key Technologies in Complex Local Navigation System Based on Binocular Stereovision

Xiao li Wang[1], *Ping Feng*[2], *Chao Zhang*[3] and *Dai-min Chen*[3]

[1]*School of Electronic and Information Engineering, Changchun University, Changchun 130022, China)*
[2]*College of Computer Science and Technology, Changchun University, Changchun 130022, China)*
[3]*Jilin Province Rehabilitation Equipment and Science andTechnology Innovation Center for The Disabled, Changchun 130022, China)*
445444726@qq.com,172046969@qq.com, 157075018@qq.com, Dai-min@sohu.com

Abstract

In this paper, binocular stereo vision platform, visual navigation for complex local range contains a lot of moving targets. The system includes a feature detection, projection reconstruction, motion analysis, clustering target four. Whereby the field of view may be obtained of the target group number, position, relative speed and direction of movement, and a comprehensive analysis of the nature of the target. The system uses real-world scenarios to test the performance of the system in most scenes to achieve the expected goal, improved method of the present system puts forward.

Keywords*: visual navigation; feature detection; projection reconstruction; motion analysis; target cluster*

1. Introduction

In the rapid development of information technology, sensor technology widely used today, automatic navigation system matures, gradually expanded from the initial military to civilian. Among them, the visual navigation system has become automatic navigation system.It is an important part for GPS, laser, inertial measurement, sonar navigation system assisted. Early single visual navigation system due to environmental factors light, viewing angle is changed, the target occlusion and exercise influence, poor real-time, long time to process and front-end analysis, applied only to locate close avoidance or indoor environment [1-2] .

But since it is too expensive to make GPS, laser and other equipment is not universal, and sonar equipment in a complex environment often brings more serious errors, so low power consumption, affordable and available to more scene information and scene within obstacle can be multi-dimensional analysis of the visual navigation system has become an important branch of research value of navigation. Visual Navigation from implementation point of view, is the process of collected visual signal is analyzed and then outputs the judgment result, and thus in addition to data acquisition front-end, the remaining process is completed by the visual computing. With hardware data processing algorithms have been continuous development and improvement in complexity and performance, part of the problem in the field of visual computing to get a better solution, certain assumptions have been realized to a certain extent, [3], so that applications visual navigation system greatly expanded.

ISSN: 1975-0080 IJMUE

2. Background

Three issues within local range of visual navigation main answer is:

(1)There may be many obstacles within the group scene;

(2) substantially spatial distribution of each obstacle;

(3) status of each obstacle relative motion.

Therefore, the algorithm functions include targets and target motion analysis and division goals, according to the information, the viewer can be adjusted to avoid the route plan in advance or on-site formulation, which also makes the problem has been greatly simplified. Its function shown in Figure 1:

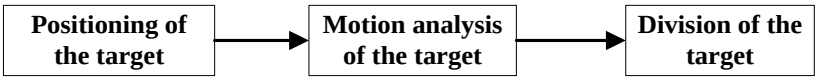

Figure 1. System Functional Diagram

3. System Design and Implementation

3.1. Feature Detection and Projection Reconstruction

Since the application for different purposes, projection is divided into dense and sparse reconstruction, since this article does not need to get all the spatial information of the scene, so the local features based sparse reconstruction. For a corrected binocular vision system, observe basic depth reconstruction formula, that formula (1):

$$z = \frac{b * f}{d} \tag{1}$$

Wherein: z is depth, b is the baseline, f is the focal length, d is the parallax.

For binocular vision system used in this system, we can learn two important points:

(1) When a fixed baseline b, To obtain the same z, f and d is proportional to, this would be reflected to the camera to select: for the same imaging device, f effective means to reduce the field of view increases, the lens tends in wide angle, but it will reduce the effective depth, effective field of view can be seen with an effective depth difficult to balance.

(2) For the correction level camera, the reaction is about parallax view pixel horizontal coordinate difference, so the feature point and a perspective view of the matching detection is the key question about the sparse reconstruction, this paper achieve the SURF feature [8,9]. SURF feature detection stage using a multi-scale non-extreme value constraint, hence, have similar SIFT scale invariant properties during this important binocular stereo matching, it will provide a strong constraint, that constraint scale. Since accuracy is not required to achieve the sub-pixel level, and therefore did not use SURF algorithm recommended way quadratic fit interpolation, but directly using the nearest neighbor interpolation, therefore, positioning accuracy error by the algorithm is a direct result of the pixel. Straightforward solution is through the very constrained to abandon the line error of point 1, and then checked by Bi-directionary Matching, If you meet the two-way matching leave this point, so this paper along the scale layer search, additional lines constraints and matching bidirectional authentication method. Due to the lowest scale and the uppermost layer of the feature points only participated in non-Extreme scale without constraints included in the test results [10], so the final match is detected through each layer on a four-level scale of 10 scale layer the feature points, and the other a corresponding view scale stage to find the corresponding points. In fact, for the image, the feature points 640 * 480 tend to be concentrated in the lower layer 5, the feature points of the top few search fast. Find the corresponding points on the ordinate available judge

sentences two points examined, because most error algorithm caused a pixel, and thus can be considered two ordinate absolute difference is less than or equal to 1 can be retained. Also, because the maximum disparity constraint is added, the same statement to determine the available two-point is determined whether the absolute difference between the horizontal axis in a predetermined threshold range, the complete left to right and then matched by matching the right to the left in FIGURE. Therefore, the matching process is added four constraints: scale constraints, constraints on the pole, the disparity constraint, two-way constraints. Photo a film about the left in Figure 2 is corrected in view of the feature points, the size of the circle indicates the level of its geographical scale level, the right is the result of stereo matching.

Figure 2. Left Image is the Response Map of Features and Right Image is the Consequence of Stereo Matching

It can be seen matching accuracy is satisfactory, but it is worth noting that, alone less robust feature points and unbound SFM technology is not directly determined by visual handicap cloud judgment, and thus can not recover all of the spatial information within the scene shown in Figure 3, above a group of pictures captured by the camera is left for 5 consecutive frames following a group of pictures is a rear perspective view of the left, as the match was almost cloud point cloud size is in accordance with the scale of the feature point extraction, the brightness high representatives from the observer closer, the farther the contrary:

Figure 3. On the Top, there are Five Continuous Frames, on the Bottom, there are Corresponding Disparity Clouds

It is worth noting that the results are sparse reconstructed spatial coordinates of the characteristic point, rather than the entire scene spatial coordinates of each point, and therefore can not directly observe the spatial distribution of the target derived only indirectly based on dense point cloud the extent and depth of change in luminance variation estimate front, thereby achieve the purpose of the obstacle prediction. Without the help of such similar SFM from point cloud point extraction method for sparse features, it can only get in front of the spatial coordinates of each feature point, can not directly estimate the spatial distribution of the target itself, Furthermore, since the low frame rate camera movement caused Fuzzy such characteristics reappear unstable, and therefore can not directly target the estimated motion situation, which require subsequent algorithm to solve.

3.2. Motion Analysis

Since moving objects, especially relative motion of objects in the local fast navigation should be an early priority to predict and avoid traditional direct collision avoidance method based stereo matching is often due to the effective range is too close (within 5-8 meters) and other reasons on the application of the large number of moving objects in the distance and calculate unsuccessful, often need to be combined with other types of sensors [11]; in addition, when using a regular camera, fuzzy target movement will produce a low texture regions, the feature point can not be stable weight now, and the moving target analysis should be more focused on the movement of health prospects were analyzed, with particular attention is the relative movement speed, so the need to implement the method of global motion analysis based on, and then the moving object individually carved out for analysis. Algorithm implementation is pyramid LK (Lucas-Kanade) optical flow [12], unlike the more popular LK Sparse Optical Flow (OpenCV has been integrated), the original pyramid LK optical flow is based on the global dense optical flow. Using dense optical flow can get view of the advantage of moving all target information, the disadvantage is the large amount of calculation. The current goal is to analyze the direction of view of the target and relative speed of movement, thus obtaining global motion information is necessary, the need for the original dense optical flow simplify or make improvements, ensuring the performance of smaller losses, in order to improve the execution speed of the algorithm .

Since LK optical flow can be simply described to some extent, as follows: luminance variation within a certain time point with respect to time to produce, respectively, the change to the point luminance in the horizontal and vertical directions x and y in proportion, thereby obtaining the formula (2), which indicates the luminance change the horizontal direction indicates the luminance change its vertical direction, indicating that the point with the brightness changes over time, u and v represent the optical flow, w represents the window size.

$$\begin{bmatrix} \sum_w I_x^2 & \sum_w I_x I_y \\ \sum_w I_x I_y & \sum_w I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_w I_t I_x \\ \sum_w I_t I_y \end{bmatrix} \tag{2}$$

Noting that the three basic assumptions of the established formula: time-continuous, continuous movement, local rigid body; wherein the continuous movement is the basis of the derivation of the formula, but because of the camera frame rate limit, the time interval is often a direct result of increased movement to reduce the continuity, Therefore, the window w tend to be smaller (for example 5) in order to detect relatively weak movement, and combined with the scale of the pyramid, the more strongly to detect motion. Combined with the actual scene of a large number of observation and algorithm purposes, based on the above, this article can make two reasonable assumptions:

(1) Although the motion scenes targets are not rigid body motion, but there is often a major sports trend that relative to other trends in terms of the target is significant, it can be detected in the upper scale is obvious optical flow;

(2) Movement is weak or relatively stationary target, there will be a scene often, movement trend of such objectives in the upper scale is difficult to detect significant optical flow.

Reflected in the code segment, on the original scale only collected three sub-scale layer, and use the size of the window 7, to collect a strong movement (no more than 56 pixels a) information, it is important in the final sub-scale setting initial iteration layer is a smaller threshold (article 0.001), no significant movement to eliminate pixels (directly to zero), reduce the computing time, the number of moving objects within the scene and shorten the extent and severity related. Figure 4, the far left is the front one, is the rightmost one, the middle are the horizontal, vertical, and the whole optical flow.

Figure 4 the more background bright areas representing the main target area movement trend from right to left or from the bottom up, on the contrary, darker than the background. The mean global level optical flow and vertical optical flow calculated respectively as the horizontal and vertical directions threshold, optical flow exceeds a threshold characteristic points individually extracted as a follow-algorithm a set of inputs below the threshold characteristic point as another set of inputs, and therefore distinguish the strong movement of the object and general object, complete motion segmentation process.

Figure 4. From Left Side to Right Side, Individually, there is Last Frame, Horizontal Optical Flow, Vertical Optical Flow, Combined Optical Flow and Next Frame

3.3. Object Segmentation

Been mentioned above, the actual decision-making, the goals are often based on their spatial properties, properties of motion are grouped, and then adjust the route, so just focus on: the number of the target group, location, size, speed, not the object of interest the exact number and categories, then we should integrate these properties on the target partition. As we focus on the algorithm is easy to implement, portability and scalability, and therefore not a priori model-based methods, the methods used in this paper is based on the clustering of target segmentation, in order to achieve the goal of full-line analysis.

Clustering method widely used in the CV is mainly based on the model and the distance are two, due to the experimental environment for this article is more complex, difficult to determine the statistical model of the target group and because of the number of unknown scene target can not be specified in advance cluster center quantity, and therefore need to use the model-independent and does not need to specify in advance the number of cluster center approach, we use AP (Affinity Propagation), the clustering algorithm based on distance, the biggest feature is not required prior to explicitly specify the number of cluster centers and convergence the degree of accuracy can be achieved based on the traditional clustering distance is difficult to achieve [13-14]. In the implementation of this article must be noted that the following two points:

(1) Because of machine learning methods belong clustering essence, the purpose is to identify the presence of the data pattern, the input data is the number of dimensions of the sample, the nature of the input data includes the type of pattern that dimension. In this paper, bound ultimately, the input feature points constituting the sample, which is a six-dimensional model dimension, namely: scale, ranks and coordinate parallax, horizontal and vertical optical flow. Wherein the degree of texture and scale are often complex area which is related to the difference between the row and column coordinates and depends directly related to the spatial location of the point, optical flow represents the movement trend that point. So the idea of this paper can be directly described as: a similar texture, similar spatial position and movement trend similar points belonging to the same set of targets.

(2) algorithm initialization phase, we must first calculate the correlation matrix between the input samples, the correlation defined by the 2-norm sample space, the correlation matrix will be updated in subsequent iterations. Since the target group for analysis, often close to the target directly into a group, so the priority value (preference)

related to the minimum value of the initial correlation matrix, we have received the least number of clusters group, thereby avoiding isolated point.

By clustering results after the division, combined with accurate three-dimensional coordinates of each point the direction of movement and projection reconstruction optical flow field was obtained, they can target a comprehensive analysis and prediction, clustering results are shown in Figure 5, the top is a different scene picture, the bottom is the corresponding clustering results shown in Figure 5.

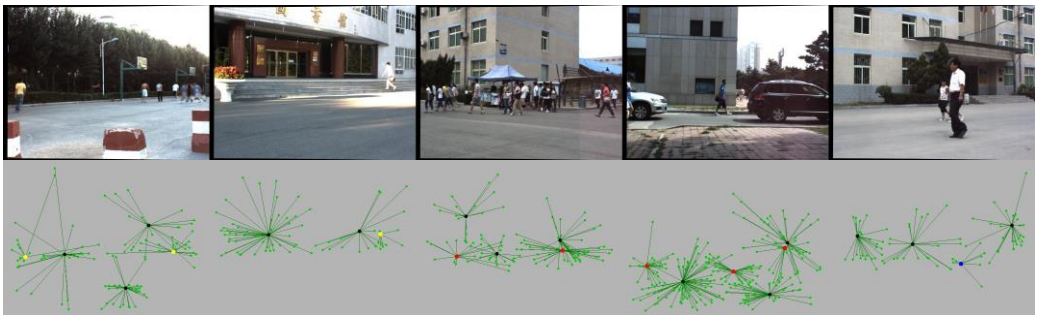

Figure 5. On the Top, there are Five Images which Represent Different Scene, on the Bottom, there are Corresponding Consequences of Clustering

4. Experimental Results and Analysis

Purpose of this paper is to seek a performance, but more dependent on the hardware platform and other types of sensors with portable, easy to implement, can be performed in the field and adjust the local visual navigation scheme, since the target medium range forecast is more practical significance, so this article using 12mm focal length lens, baseline 8cm, placed parallel to the front of the imaging platform. Due to the need for a large amount of the actual scene test object and the test should include moving foreground, background motion, occlusion, terrain changes and other factors, combined herein by relying on the project background, select the corresponding test in the campus scene range, the scene has a lot of moving target, background texture complex, foreground obscured terrain and illumination changes, and therefore has a strong representation.

This article extracted image 7 groups, and groups them analyzed to illustrate the practical application of the existing problems in accordance with the principle of similar scenes in nature. In principle, the result of a combination of optical flow analysis to judge the direction and perspective projection reconstruction after the match can determine whether the target of the current trends and possible movement of the target location, completed a rough estimate of the nature of the target, and then adjust the route . Pictures of the order left to right are characteristic response graph, stereo matching, clustering segmentation map, the global optical flow diagram.

Figure 6 contains three sets of scenarios, the observer and the moving target Figure from the moderate (approximately 8-15 meters), the outlook contains both fast-moving target and a stationary or slow-moving targets, moving targets located in the middle of the field, and with the objective between goals, objectives and background between significant occlusion, complex background texture and uniform illumination; the difference is the presence of the scene close to the top obstacle, intermediate scene is flat, the bottom of the scene due to the presence of stairs, causing changes in the terrain.

The first line of the scenario, there are three groups of central movement from the target, which is close and far away in the middle of the right of relative movement very quickly, a red dot, left relative motion is slow, a dot, and distinguish the close obstacles and distal exist complex texture background. The second row, in the scene movement

target group seriously block the background of its height, and speed faster, so the central field of clustering in red dot, and differentiate the two high background region. The third line moving objects in the scene group obviously but dispersed, and the presence of shade to the background, so in addition to distinguish moving target group also divided and movement target group at the same level as part of the background (the stairs), and relative velocity with yellow (fast), blue (ordinary), black (slower) division said movement target group. Combined with the result of matching and optical flow, can be further experience judgment according to the target properties, known algorithm has good performance in such a scene.

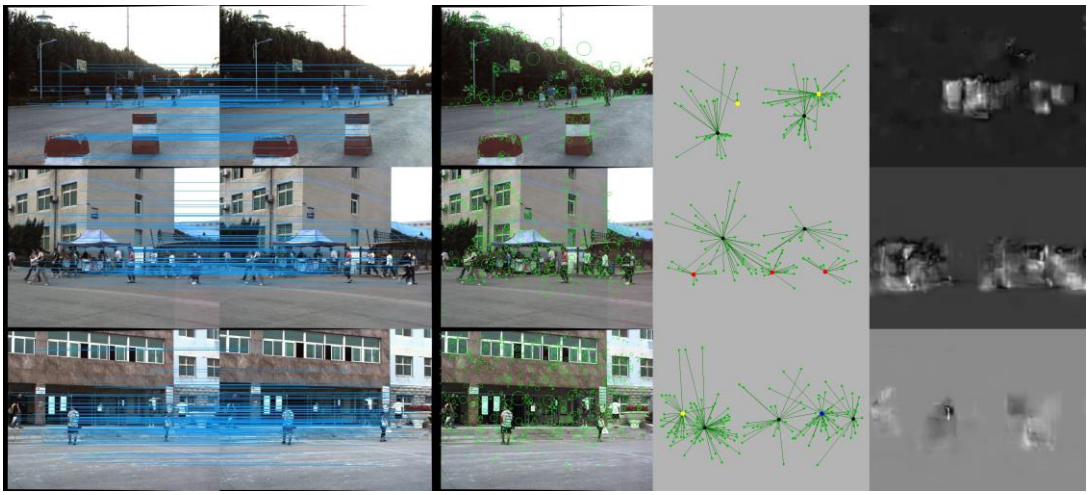

Figure 6. The Scene Have Lots of Moving Objects and the Distance Between them and Observer is Modest

Figure 7 contains two sets of scenarios and scenarios represent relatively close distance between the observer and the moving targets (5 to 8 meters) and the presence of more movement of the target group, the target of background shade is more, and the smaller the distance between the target and shade, the difference is the top in the scene because of kerbstone, cause the topography, the bottom of the scene without the topography

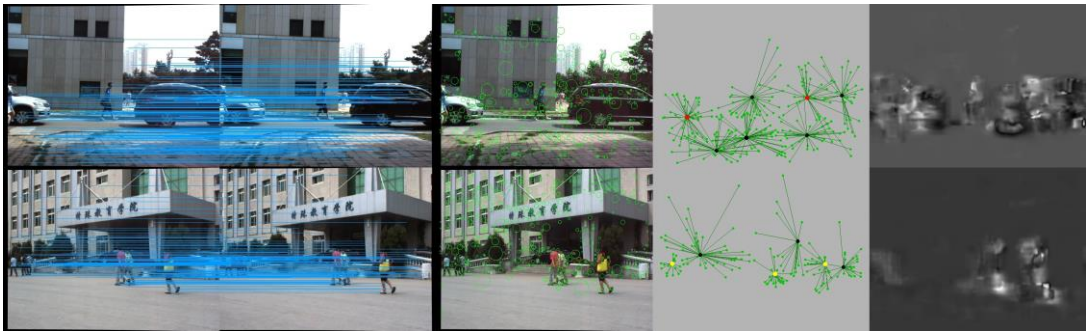

Figure 7. The Scene which Moving Objects' Speed is Fast and Objects are Relatively Close to the Observer

At the top of the scenario, the detected movement very fast and there is a certain distance between two vehicles, said red target group, and between the detected vehicle pedestrian, but due to the car pedestrian speed slower, so represented as black target group, in addition detected a complex texture background (trees), and a kerbstone and an illumination change area (error).The bottom of the scenario, the fast pedestrians moving target detected three groups and three background region, background region, but failed to

detect the entrance because the area of feature points and feature points depth and other background texture complexity, there is no obvious difference between but can be the result of comprehensive matching positioning to the area, so the algorithm in such a scenario, especially for moving target classification performance is better.

Figure 8 contains the same before and after the two consecutive frames in the scene, the scene on behalf of the observer and the moving objects very close (3 m), different target motion property (this scenario is in a different direction) and overlap.

Can be found for the two unstable frame for moving targets detection and grouping, on a frame in the failure to detect pedestrians from left to right, the next frame are two categories of moving targets are detected, but the possibility of moving target location estimate appear large deviation. Too close to the reason because of the moving targets and keep out each other cannot produce stable feature points, and due to the division of feature points based on the three dimensional space at the same time, the scale 1 d and 2 d motion, also is a sport is not only a reference standard, lead to stable feature points cannot divide, under this scenario is based on the analysis of the result of the match and optical flow is also difficult to target properties, so the algorithm should not be too close to application in moving target and the mutual keep out scenarios.

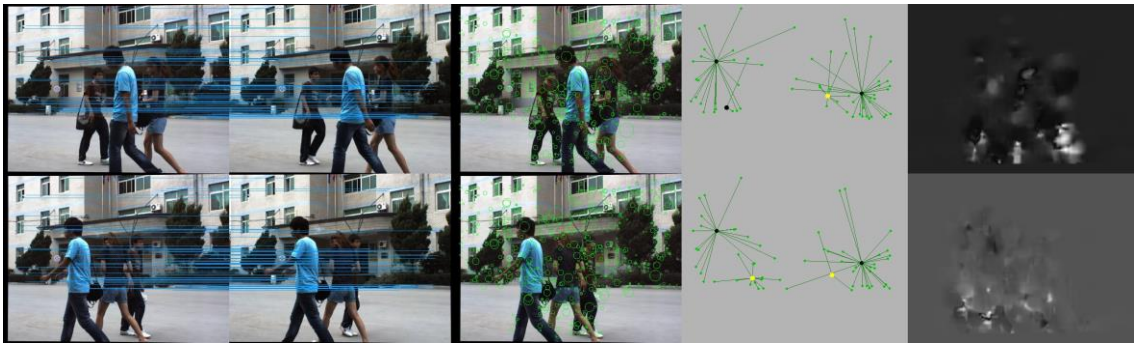

Figure 8. The Scene which Moving Objects are much Close to the Observer and Exist Interactive Occlusion

5. Conclusion

Comprehensive description of the experimental section, for the purpose of this article put forward the local visual navigation scheme can think: within 8 to 15 meters of the target group is roughly divided and estimate of the target motion properties is relatively accurate, can better detect a large number of moving targets and according to the nature of groups, the results point to the center after a group to the average distance to the target group, the author estimated the rough shape of combining projection reconstruction and optical flow can be on different target groups to locate and estimate its motion trend, then make the adjustment. The greatest characteristic of the scheme is based on the traditional positioning and motion detection function is trying to combine the space of the target, texture and motion properties of target reasonable classification and not rely on prior model, greatly reduce the algorithm complexity, and makes it easy to implement and less dependent on the performance of hardware, expand the application range, especially suitable for don't need to accurately describe the behavior of the target space.

However, this scheme has the typical 2D - 2.5D - 3D structure, so the obvious flaws is its excessive based on local feature detection, when the scene change is the local characteristics of the same target are often not stable, thus leading to the instability of the algorithm. So the appropriate combination of product data management (PDM) to join prior knowledge or intensive algorithm is a possible solution. In addition, due to the purpose of this scheme is to estimate or predict local area within the scope of the

obstacles, to consider emphatically medium distance detection and location, so at close range than on the detection performance of using wide Angle lens, wide baseline, the typical screw placement within close visual platform of obstacle avoidance, and because the close movement may produce strong motion blur, so its analysis results are not accurate, feasible solution is to combine the traditional close obstacle avoidance collision, in order to achieve the complete local area within the scope of application.

Acknowledgment

The research is carried out in the Create center of disabled recovery and equipment and is supported by Jilin Province Department of Science and Technology (No. 201215112). Second,The research is carried out in The Innovation Center of Disabled Rehabilitation in Equipment and Technolog is supported by Jilin Provincial Science and Technology Department (No.201215111)and supported by Jilin Provincial Education Department (No. 2012243).

References

[1] J. Gutmann and K. Konolige, "Incremental mapping of large cyclic environments", In IEEE International Conference on Robotics and Automation (ICRA), **(1999)**.
[2] G. Bianco and P. Fiorini, "Visual avoidance of moving obstacles based on vector field disturbances", In IEEE International Conference on Robotics and Automation (ICRA), **(2001)**.
[3] J. Leonard nad P. Newman, "Consistent, convergent, and constant-time SLAM", In International Joint Conference on Artificial Intelligence (IJCAI), **(2003)**.
[4] L. D. Whyte, "Mobile Robot Localization by Tracking Geometric Beacons", IEEE Transactions on Robotics and Automation, **(1991)**, pp. 376-382.
[5] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision(IJCV), **(2004)**:
[6] M. Agrawal and K. Konolige, "Real-time Localization in Outdoor Environments using Stereo Vision and Inexpensive GPS", In Proceedings International Conference on Pattern Recognition(ICPR), **(2006)**.
[7] R. Newcombe and J. Andrew, "Live Dense Reconstruction with a Single Moving Camera", In IEEE Conference on Computer Vision and Pattern Recognition(CVPR), **(2010)**, pp. 1498-1505.
[8] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded Up Robust Features", In European Conference on Computer Vision (ECCV), **(2006)**.
[9] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, "Speeded-up Robust Features (SURF)", Computer Vision and Image Understanding (CVIU), no. 110, **(2008)**, pp. 346-359.
[10] M. Brown and D. Lowe, "Invariant Features from Interest Point Groups", In European Conference on Computer Vision (ECCV), **(2002)**.
[11] M. Agrawal, K. Konolige and R. C. Bolles, "Localization and Mapping for Autonomous Navigation in Outdoor Terrains: A Stereo Vision Approach", IEEE Workshop on Application of Computer Vision (WACV), **(2007)**.
[12] J. Y. Bouguet, "Pyramidal implementation of the Lucas Kanade feature tracker: Description of the algorithm", Technical report, Intel Research Laboratory, **(1999)**.
[13] B. J. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points", Science, no. 315, **(2007)**, pp. 972-976.
[14] N. Lazic, I. E. Givoni, P. Aarabi and B. J. Frey, "FLoSS: Facility Location for Subspace Segmentation", In IEEE International Conference on Computer Vision (ICCV), **(2009)**.

