Visual Assistance to an Advanced Mechatronic Platform for Pick and Place Tasks

Rigas Kouskouridas and Antonios Gasteratos

Democritus University of Thrace,
Department of Production and Management Engineering,
Vas. Sofias 12, 67100 Xanthi, Greece
{rkouskou,agaster}@pme.duth.gr

Abstract. Recent advances in mechanical and electronic engineering led to the building of more sophisticated mechatronic systems excelling in simplicity, reliability and versatility. On the contrary, the complexity of their parts necessitate integrated control systems along with advanced visual feedback. Generally, a vision system aims at bridging the gap between humans and machines in terms of providing to the latter information about what is perceived visually. This paper shows how the vision system of an advanced mechatronic framework named ACROBOTER is used for the localization of objects. ACROBOTER develops a new locomotion technology that can effectively be utilized in a workplace environment for manipulating small objects simultaneously. Its vision system is based on a multi-camera framework that is responsible for both finding patterns and providing their location in the 3D working space. Moreover, this work presents a novel method for recognizing objects in a scene and providing their spatial information.

Keywords: Intelligent Systems, Mutli-camera Systems, Object Recognition, 3D Position Estimation.

1 Introduction

In the last few years, robotics researchers aimed at building autonomous vehicles in order to substitute humans in time demanding tasks. Furthermore, industries address all their efforts to developing machines capable of assisting people to everyday life. The need of robotic platforms working closely to humans necessitates the usage of intelligent mechatronics and vision systems. Towards this end, ACROBOTER (Autonomous Collaborative Robots to Swing and Work in Everyday Environment) constitutes one of the most advanced integrated sensorial system. The main concept of the ACROBOTER project is the elaboration of a new locomotion technique that extends the workspace of existing robots in the vertical direction. As a result, it is able to work on the top of tables, wardrobes and can be used for manipulating objects placed on shelves, tables, work surfaces or on the floor. It is apparent that such a robotic platform is capable of moving fast in any 3D direction in an interior working space.

H. Liu et al. (Eds.): ICIRA 2010, Part II, LNAI 6425, pp. 705–716, 2010.

This concept is highly innovative, in the sense that it combines the planar stepping motion of an arm on the ceiling and a 3D controlled pendulum like motion of the working unit. In addition, the ACROBOTER is designed to work autonomously or in close cooperation with humans and to be capable of collaborating with other robotic devices. The ACROBOTER platform consists of mechanical, navigational and control of several subsystems, including a grid of anchor points, a climber unit that moves on the grid of anchor points, and a swinging unit, which is fixed onto the climber unit. The ACROBOTER platform through the swinging unit might carry other service robot or a tool to perform the assigned tasks. The embedded controllers of these main components are integrated with the global control system, while the cameras built onto the four corners of the room provide up to date, navigational information on a dynamically changing working environment. In addition, as part of the global controller, a human machine interface enables the user to interact with and/or to control the ACROBOTER.

The vision system of the ACROBOTER adopts a multi-camera framework in order to provide the visual information needed for the accurate navigation of the robot. Moreover, it includes four individual cameras, installed in the four corners of the room that collaborate to achieve optimum results and cover the 97% of the working volume. In detail, the vision system consists of four Grasshopper cameras manufactured by Point Grey Research that are able to capture images up to 1280 X 960 pixels resolution and are connected to the PC via a firewire port, using the IEEE 1394b transfer protocol. Moreover, the ultimate goal of the vision system is to recognize objects found in the scene and estimating their location in the 3D working space. The aforementioned demanding task is adequately fulfilled by developing and implementing techniques beyond the state-of-the-art as presented in here. The exhibited methodology and results comprise the physical extension of the works presented in the last ICIRA conference [1], [2]. The remainder of this paper is structured as follows: In Section 2, we give an overview about the related work in the areas of multi-camera systems, integrated mechatronic systems and object recognition. The overall system architecture, both hardware and software, of the ACROBOTER is exhibited in Section 3, where every part of the project is presented. Furthermore, in Section 4, a detailed description of the proposed object recognition and position estimation technique is given along with the respective experimental results. Finally, the work concludes with some final notes and discussion in Section 5.

2 Related Work

A major challenge in the development of a robot for indoors applications is to equip it with the ability of moving freely and quickly in any direction. In addition, the obstacles found in buildings, specifically in a domestic environment where objects may not be static over time, is a drawback. Examples include stairs, doorsteps, chairs, tables and various surface changes including carpets. Past developments that address the general need for a robot to climb walls and ceiling inside a building are the MATS robot [3], biped walking robots like the Honda Asimo robot [4] and mobile robot devices such as the Care-o-Bot from IPA [5]. The task to move freely inside a building includes many issues, which are not easily solved. The closest solution to the ACROBOTER concept found in the literature so far is the Flora ceiling based service [6] robot. Flora is suspended onto the ceiling and uses electromagnetic force to keep the mobile cart on the ceiling and telescopic arms to navigate the working unit into the 3D workspace. The permanent magnet traction method and the rigid telescopic structure make this service robot concept less affordable for multipurpose domestic use.

In turn, the ACROBOTER is distinguished by its innovative locomotion concept, which allows to run the entire volume of a room. Additionally it is designed to be able to work autonomously or in close cooperation with humans, and to be capable of collaborating with other robotic devices. In particular, there are two main issues to work with, which are in favor for the ACROBOTER system: autonomous operation / navigation and power. The working principles of biped robots mean that they walk on the floor and thus the navigation in a domestic environment is crucial for successful operation. The power of the system is a limitation compared with the ACROBOTER that can be fed with power in a practically unlimited way.

In the field of multi-camera systems, the Stanford Multi-Camera Array [7] dominated for several years as the ultimate multi-view framework. It consists of 128 cameras that can be arranged and used in a variety of ways. The final camera architecture corresponds to an arc of separate panels aiming at a point in the middle of the room. In turn, the multi-camera system, developed by the UD Graphics Lab [8], consists of 10 Flea2 1394b cameras, where a subset of them are mounted on movable plates attached to the ceiling racks and the others are placed at different positions using tripods. The target applications of the aforementioned framework are multi-view surveillance, capturing of dynamic fluid surfaces or appearance modeling. In addition, the Mitsubishi Electric Research Laboratories (MERL) has designed a framework where the object-wise semantics are extracted from a non-overlapping field-of-view multi-camera system [9]. Its main goal is the construction of an automatic object tracking and video summarization method via background subtraction and mean-shift analysis.

In turn, with a view to the efficient accomplishment of navigation or manipulation tasks, robots must enjoy the perception of depth, for both static and dynamic working environments. As a result, over the past decade, significant research efforts were devoted to the building of autonomous vision systems, capable of providing a sense of location and direction to robots. In manipulation applications the task is to determine objects' position and orientation relative to a specific co-ordinate system. The procedures of obstacle avoidance or objects' manipulation assistance can be accomplished by integrating vital visual information derived from pose estimation techniques. Furthermore, algorithms that were recently proposed are completed by using either visual sensors [10], [11] or the later combined with inertial ones [12]. Another open research topic in computer vision is the depth estimation. In [13] the discrete Fourier Transform (FT) is utilized to estimate both local and global transformation of the spatial information. On the other hand, one of the most efficient ways to calculate scene's depth is the adoption of a disparity estimation method. Stereo vision frameworks invoke correspondences derived from two slightly different images to extract sufficient spatial information. A recent survey of existing disparity estimation methods is presented in [14].

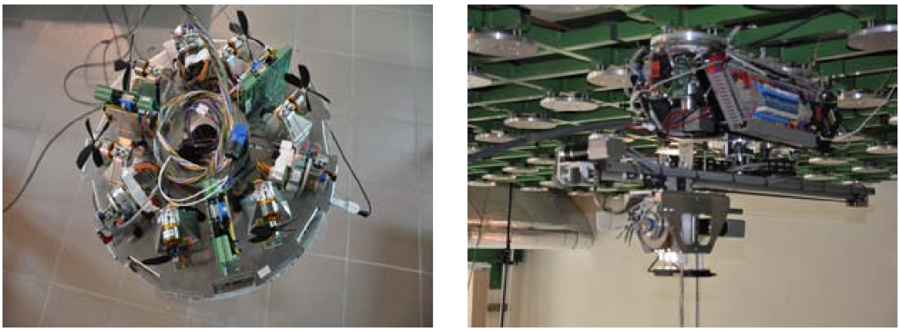
During the past few years, remarkable efforts were made to build new vision frameworks for robust object recognition in cluttered environments. To this end, researchers emphasized in creating recognition schemes based on local range features [15], [16]. Algorithms of this field extract features with local extent that are invariant to possible illumination, viewpoint, rotation and scale changes. During the past decade, several techniques that enforce the essential role of local features in demanding pattern recognition tasks were presented [17], [18]. The indispensable visual distinctiveness of an object in a scene is ensured by locally sampled descriptions. Furthermore, the two main sub-mechanisms of such frameworks are the detectors and descriptors of areas of interest. The efficiency of the two sub-mechanisms is investigated in [18], were detectors and descriptors are evaluated for object recognition purposes. Currently, the two most widely used object recognition frameworks based on local appearance features, are the Scale Invariant Feature Transform (SIFT) [19] and Speeded-Up Robust Features (SURF) [20]. Both of them implement a detector and a descriptor for fast feature extraction and, as a result, they are adopted in almost any newly appeared object recognition technique. The common issue in both of them is the fact that the their detector depicts significant efficiency [18], [21], [22] since its not affected by possible image alterations. On the contrary, the most important drawback of SIFT's and SURF's descriptor, constitutes the fact that its performance alters significantly under possible rotation, scale, viewpoint and illumination changes. In [23] and [24], two different techniques for object recognition and relative pose coherence are proposed. Both of them utilize subroutines to minimize mismatches between testing images (unknown scenes) and known sets (objects' images contained in databases).

3 System Architecture

The ACROBOTER consists of several sub-systems namely the swinging unit, the climber unit, the anchor points and the human-machine interface (HMI). In the following section we briefly describe the main components of the aforementioned frameworks since in this work emphasis is given to ACROBOTER's vision system.

3.1 Swinging Unit

The swinging unit, shown in Figure 1(a) integrates the functions of 3D motion of ACROBOTER (including positioning, orienting and stabilization), tool changing and interfacing. It can serve as a platform for moving other (existing) service robots or it can be equipped with different kinds of end-effectors in order to accomplish specific tasks. Furthermore, the swinging unit incorporates additional mechanical and control related principles. Three windable cables are used to keep the unit in horizontal position. These cables are also used to position the swinging unit vertically, which result in a redundant actuation with improved manoeuvrability and obstacle avoidance. Moreover, the required thrust is provided by ducted funs placed on the top surface of the platform. In addition, the vectoring and reversing of the thrust are provided by two ducted fans with adjustable blade pitch impellers.



(a) (b)

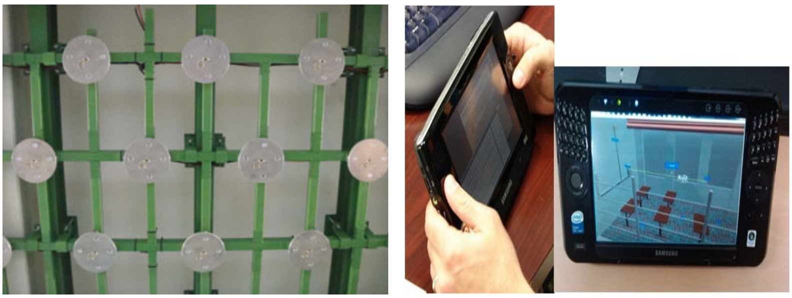
Fig. 1. (a) The swinging unit of the ACROBOTER serving as a platform. (b) The climber unit which is responsible for the overall movement of the robot.

3.2 Climber Unit

The climber unit (see Figure 1(b)) consists of the following sub-systems: Anchor arm with anchor points, rotation arm and winding mechanism. These components give the climbing unit the ability to position the swinging unit, both horizontally and vertically in the environment. The climber unit is also responsible to lead electric power from the energy source in the ceiling to the swinging unit. The climber unit itself is also powered by the same energy source. The main control tasks performed by the climber unit involve the active tracking control of the 3D trajectories along with the contribution to the compensation of real-time errors caused by simplified idealized motion planning.

3.3 Anchor Points

The anchor points system is the first system in a chain of mechanically interconnected systems and it is illustrated in Figure 2(a). The tasks of this subsystem are to support, rotate and swing the bulk of all subsystems on the anchor points that are attached on the roof of a specified structure. The subsystem is structurally divided into two parts: static anchors attached to the ceiling that provide the necessary support to the system (Anchor Points), and a mobile part (Anchor Arm) that enables rotation around the anchor points and displacement over the ceiling. The Anchor points are passive units made of structural steel. They provide structural sup-port and electrical power to the ACROBOTER system. The Anchor Arm consists of a static beam that provides the necessary support from the anchor points and a driving part that enables rotation about the anchor point.



(a) (b)

Fig. 2. (a) The anchor points of the ACROBOTER. (b) (Left) User interacting with Samsung Q1 tablet device, (Right) HMI interface computer showing the simulator running to validate a user selected task.

3.4 Human Machine Interface

The HMI acts as a gateway between the user and the task scheduling system, necessitating close integration between the two systems. Whilst the interface implementation will be specific to the candidate scenario of the ACROBOTER robot, several candidate HMI concepts have been developed which are applicable to a range of scenarios. A mobile interface device (Samsung Q1 ultra-portable PC), which is depicted in Figure 2(b), is utilized for controlling the robot since it allows the user to freely move around the robot's environment whilst issuing commands. This device is equipped with a 1GHz Intel processor, 1GB of RAM, and an Intel integrated graphics processor. This specification is capable of running the WPF based user interface, as well as the simulator. Communication with the rest of the control system is performed by an integrated WiFi link.

4 3D Object Recognition and Position Estimation

The overall goal of ACROBOTER's vision system is the simultaneously recognition of possible objects found in the working space of the robot and the 3D location estimation of the sought objects. It comprises of four cameras installed in the respective corners of the room. The proposed position estimation process is motivated by the idea that the detected features are located on given geometric positions, thus they can be considered as the corners of a polyhedron, the centre of gravity of which is computed and it is associated to the actual centre of mass of the sought object. Once the features' center of mass is known and the object is recognized, the distance of the object from the camera is trivial, given at least one recorded position of the object. Practically, the proposed technique tries to estimate the geometrical proportion between an image of a known object (contained in the database) and another one of a scene containing the same object in different arrangements. By taking into account the distribution of object's features around their centre in the image space we can estimate spatial information about the object. Especially, we are able to calculate object's distance from the camera along with its location relative to the camera's coordinate system. Afterwards, since there is a pre-knowledge about the cameras' position in space (calculated through the calibration matrix - extrinsic parameters of the camera) we estimate objects' location relative to the global coordinate system of the robot. This is accomplished by multiplying a transformation matrix to the one extracted previously. As a result, by comparing object's image to that of a scene's containing it we are able to transform information from image space to the real world. The main idea underlying our algorithm is the maintenance of the properties of SIFT, whilst making an attempt to further exploit them in order to assign the distance of an object from the camera's frame. The main stages of the proposed algorithm are as follows:

Stage I *Apply SIFT to the scene's and object's image, in order to estimate the features position in each of them.*

Stage II *Obtain the N features that match in the two images by applying the matching sub-procedure of SIFT.*

Stage III *Find the features' centers of mass for both images. This is accomplished by estimating the mean values of the features positions in the two images.*

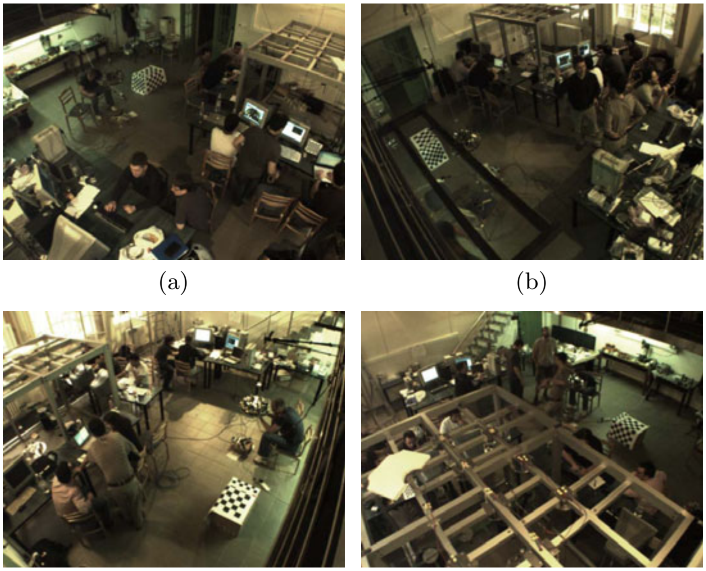
Stage IV *Calculate the mean Euclidian distance (in pixels) of each feature from the corresponding center of mass that is extracted in the previous stage. Set as* E_S *and* E_0 *the mean Euclidian distances in the scene and object image, respectively.*

Stage V *Estimate* d_S *which corresponds to the ratio of the two mean distances* E_S *and* E_O*. Furthermore, we introduce the pre-computed depth* d_O*, which is obtained, during the training session and while the object is captured alone.*

Stage VI *The previously extracted spatial information relative to the camera's plane are transformed to the global X,Y and Z of the object.*

Stage 1 could be apprehended as the training session of our algorithm. In this phase, for each image in the database keypoint features are extracted using SIFT. Each object is captured at different distances from the camera and the pre-computed depth d_O is stored for further use. This process is performed while the system is offline, thus, executable time is not taken into account. The results are stored for further use at the next phases. In Stage 2, the matching sub-procedure of SIFT is performed. Especially, descriptors that are common in both images (scene and object) are extracted. It is apparent that, one image representing the scene is compared with several others, representing the object from different viewpoints. Furthermore, the locations of the common features are stored for further exploitation. In Stage 3, the position estimation sub-procedure takes place till the end of the algorithm. Moreover, at this phase, the features' centers of mass in both images are calculated. The last is obtained by estimating the mean values of features locations in both representations. In Stage 4, the distance of each keypoint from the center of mass is calculated. This is measured in pixels with the use of Euclidian Distance. By the end of this sub-routine, we are able to collect significant spatial information of an object in a scene. This is accomplished by simply estimating the distribution of trained features around their center of mass. Finally, in Stage 5, the object's distance from the camera is computed. The pre-computed depth d_O measured during the training session (Stage 1), is taken into account. The ratio d_S is used to measure the proportion of object's features to those found in the scene. After the necessary training session and the database construction at the initial stages of the method, an on-line search engine follows. This is responsible for querying in the scene for objects contained in the trained database. When an object is found, the scene's image is compared to this object, which provides the majority of common matches. Finally, features' information from both images is interpolated with a view to object's position allocation.

The proposed method was exhaustively tested at the final integration meeting of the project. The scene captured corresponds to the real working space conditions under which the ACROBOTER will operate. The cameras have been successfully installed in the four corners of the room and calibrated in order to calculate the extrinsic and intrinsic parameters. Furthermore, the multi-camera network has been adequately established through the respective IEEE 1394b firewire ports and the necessary data related bandwidth adjustment has been accomplished. The dimensions of the cameras' array are: length 6.2 m, width 4.4 m and the height of each camera from the ground is 3.3 m. In Figure 3 the four viewpoints of the multi-camera system are illustrated. The tests were executed on a typical PC with a core2duo 2.2 GHz processor, 2 GB RAM and Windows XP operating system. Furthermore, the procedure executes in approximately 4 sec per camera. It is apparent that the overall computational cost of the position estimation process is mainly depended on the trained images per object. An over-enhanced database results into high execution times. In detail, the proposed method requires approximately 0.8 sec per training image in order to estimate the necessary features extracted through the recognition process.



(a) (b)

(c) (d)

Fig. 3. The same scene captured from four different viewpoints. The 3(a), 3(b), 3(c) and 3(d) correspond to camera no. 1 viewpoint, camera no. 2 etc.

The proposed method initially aims at estimating object's distances from the cameras. This estimation provide to the robot raw information concerning the relative position of the object to the camera. The results of this first phase are shown in Figure 4. The efficiency of the proposed technique remains at very high standards and especially the mean value of the distance estimation is 93.4 % and the standard deviation oscillates around 5.2. In turn, as a next step the initial spatial information derived is taken into account for the relative and absolute location assignment of the recognized object. Figure 5 depicts the estimations derived by the second phase of the proposed method. The X,Y and Z positions of the object relative to the camera are estimated through the calculated distance of the first phase. In turn, the absolute object's position are extracted by multiplying the transformation matrix T of each camera with the relative coordinates. Once more the efficiency concerning the relative to camera coordinates remains high with mean value 86.7% and with standard deviation around 9.34. As far as the absolute coordinates are concerned, the proposed method's yield has a mean value 84.4 % and a standard deviation 10.7. It is obvious that concerning the fourth camera a decrease of the algorithms efficiency is noted. This is due to the fact that scene as captured from camera 4, contains several textured objects that are incorrectly matched to the sought object.

	camera 1			camera 2		
	measured	groundtruth	efficiency	measured	groundtruth	efficiency
Distance	618	635	97.3%	485	510	95.1%
	camera 3			camera 4		
	measured	groundtruth	efficiency	measured	groundtruth	efficiency
Distance	365	350	95.7%	455	530	85.8%

Fig. 4. The proposed method is able to estimate object's distance from the camera with remarkable efficiency

Coordinates relative to camera 1	measured	groundtruth	efficiency
X	360	380	94.7%
Y	445	480	92.7%
Z	248	260	95.4%
Coordinates relative to robot			
X	355	380	93.4%
Y	175	200	87.5%
Z	52	70	74.3%

(a)

Coordinates relative to camera 2	measured	groundtruth	efficiency
X	67	60	88.3%
Y	425	480	88.5%
Z	299	260	85.0%
Coordinates relative to robot			
X	410	380	92.1%
Y	190	200	95.0%
Z	80	70	85.7%

(b)

Coordinates relative to camera 3	measured	groundtruth	efficiency
X	70	60	83.3%
Y	150	140	92.9%
Z	259	260	99.6%
Coordinates relative to robot			
X	396	380	95.8%
Y	176	200	88.0%
Z	74	70	94.3%

(c)

Coordinates relative to camera 4	measured	groundtruth	efficiency
X	341	380	89.7%
Y	205	140	53.6%
Z	200	260	76.9%
Coordinates relative to robot			
X	295	380	77.6%
Y	255	200	72.5%
Z	100	70	57.1%

(d)

Fig. 5. The absolute and relative positions of the found object along with the proposed method's attributes

5 Conclusions

In this paper an intelligent multi-camera agent responsible for providing the necessary visual information to an advanced mechatronic system is presented. The vision system of the ACROBOTER project comprises of four cameras installed in the four corners of a room and it is responsible for two tasks that affect the overall efficiency of the system. Initially, it aims at the adequate recognition of objects found in the scene and are already contained in a database. The ultimate goal is to provide to the robot the position of the sought objects in its 3D working space. Experimental results representing those obtained during the final integration meeting prove our claim that once the features' center of mass is known and the object is recognized, the distance of the object from the camera is trivial, given at least one recorded position of the object. This in turn yields to an accurate location assignment of the object and allows the swinging unit to fly over the object and manipulate it. Moreover, the proposed multi-camera framework can be exploited in several other indoor applications where simultaneous recognition and localization of objects is prerequisite. Finally, we firmly believe that due to the fact that all the sub-systems utilize methods, which are beyond the state-of-the-art, and its anthropocentric concept, ACROBOTER will make a great breakthrough at the field of autonomous mobile assistant robots.

Acknowledgements. This work is supported by the E.C. under the FP6 research project for Autonomous Collaborative Robots to Swing and Work in Everyday EnviRonment ACROBOTER, FP6-IST-2006-045530.
http://www.acroboter-project.org

References

1. Kouskouridas, R., Kyriakoulis, N., Chrysostomou, D., Belagiannis, V., Mouroutsos, S., Gasteratos, A.: The Vision System of the ACROBOTER Project. In: International Confernce on Intelligent Robotics and Applications, pp. 957–966 (2009)
2. Kouskouridas, R., Badekas, E., Gasteratos, A.: Simultaneous Visual Object Recognition and Position Estimation Using SIFT. In: Xie, M., Xiong, Y., Xiong, C., Liu, H., Hu, Z. (eds.) ICIRA 2009. LNCS, vol. 5928, pp. 866–875. Springer, Heidelberg (2009)
3. Balaguer, C., Gimenez, A., Huete, A., Sabatini, A., Topping, M., Bolmsjo, G.: The MATS robot: service climbing robot for personal assistance. IEEE Robotics & Automation Magazine 13(1), 51–58 (2006)
4. `http://www.honda.co.jp/ASIMO/`
5. `http://www.care-o-bot.de/english/`
6. Sato, T., Fukui, R., Morishita, H., Mori, T.: Construction of ceiling adsorbed mobile robots platform utilizing permanent magnet inductive traction method. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2004, vol. 1 (2004)
7. Vaish, V., Levoy, M., Szeliski, R., Zitnick, C., Kang, S.: Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2 (2006)
8. Zhang, J., McMillan, L., Yu, J., Hill, U.: Robust tracking and stereo matching under variable illumination. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1 (2006)
9. Zwicker, M., Vetro, A., Yea, S., Matusik, W., Pfister, H., Durand, F.: Resampling, antialiasing, and compression in multiview 3-D displays. IEEE Signal Processing Magazine 24(6), 88–96 (2007)
10. Schweighofer, G.: Robust pose estimation from a planar target. IEEE Trans. Pattern Anal. Mach. Intell. 28(12), 2024–2030 (2006)
11. Chandraker, M., Stock, C., Pinz, A.: Real-time camera pose in a room. In: Crowley, J.L., et al. (eds.) ICVS 2003. LNCS, vol. 2626, pp. 98–110. Springer, Heidelberg (2003)

12. Xu, D., Li, Y.F.: A new pose estimation method based on inertial and visual sensors for autonomous robots. In: IEEE International Conference on Robotics and Biomimetics, Sanya, China, pp. 405–410 (2007)
13. Torralba, A., Oliva, A.: Depth estimation from image structure. IEEE Trans. Pattern Anal. Mach. Intell., 1226–1238 (2002)
14. Naplapntidis, L., Sirakoulis, G., Gasteratos, A.: Review of stereo vision algorithms: from software to hardware. International Journal of Optomechatronics 2(4), 435–462 (2008)
15. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA, pp. 2161–2168 (2006)
16. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, pp. 1470–1477 (2003)
17. Liao, M., Wei, L., Chen, W.: A novel affine invariant feature extraction for optical recognition, vol. 3, pp. 1769–1773 (August 2007)
18. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)
19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
20. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. 110(3), 346–359 (2008)
21. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. Int. J. Comput. Vision 65(1-2), 43–72 (2005)
22. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. Int. J. Comput. Vision 60(1), 63–86 (2004)
23. Ferrari, V., Tuytelaars, T., Van Gool, L.: Simultaneous object recognition and segmentation by image exploration. In: Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, pp. 40–54 (2004)
24. Moreels, P., Perona, P.: A probabilistic cascade of detectors for individual object recognition. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 426–439. Springer, Heidelberg (2008)