The Vision System of the ACROBOTER Project

Rigas Kouskouridas, Nikolaos Kyriakoulis, Dimitrios Chrysostomou, Vasileios Belagiannis, Spyridon G. Mouroutsos, and Antonios Gasteratos

Democritus University of Thrace, School of Engineering,
Vas. Sofias 12, 67100 Xanthi, Greece
{rkouskou,nkyriako,dchrisos,vm6872,agaster}@pme.duth.gr,
sgmour@ee.duth.gr

Abstract. The ACROBOTER project aims to develop a radically new locomotion technology that can effectively be used in a home and/or in a workplace environment for manipulating small objects autonomously. It extends the workspace of existing indoor service robots in the vertical direction, whilst its novel structure allows covering the whole volume of a room. For the adequate accomplishment of demanding manipulating tasks, its vision system must provide vital visual information concerning the position of the robot in the 3D working space and the topology of possible objects/obstacles in robot's trajectory. Thus, the proposed system is capable of: estimating robot's pose in the room; reconstruct the 3D working space and; recognize objects with remarkable efficiency. In this work, initially, we present the basic structure of ACROBOTER and its vision system and, we also evaluate the aforementioned functions.

Keywords: Pose Estimation, 3D Reconstruction, Object Recognition, Vision Systems.

1 Introduction

In the last few years, a remarkable increase of autonomous robots' usage in domestic workplace environments has been discerned. A wealth of research is devoted in building new frameworks capable of assisting people in everyday life. Moreover, industries address all their efforts to developing machines for substituting humans in house chores such as tidying child's bedroom after a party or collecting clothes before they enter the washing machine. The need of robots working closely to human beings makes a necessity the usage of intelligent systems. One of the most interesting evolutioned system is the ACROBOTER (Autonomous Collaborative Robot to Swing and Work in Everyday EnviRonment) whose vision system is presented in this work.

The main idea underlying the ACROBOTER project is the development of new locomotion technology for manipulating small objects and/or assisting humans in their movements or exercises. This new type of mobile robot is able to move fast and in any 3D direction in an interior environment. Due to the fact that it extends the workspace of existing robots in the vertical direction, it is

M. Xie et al. (Eds.): ICIRA 2009, LNAI 5928, pp. 957–966, 2009.

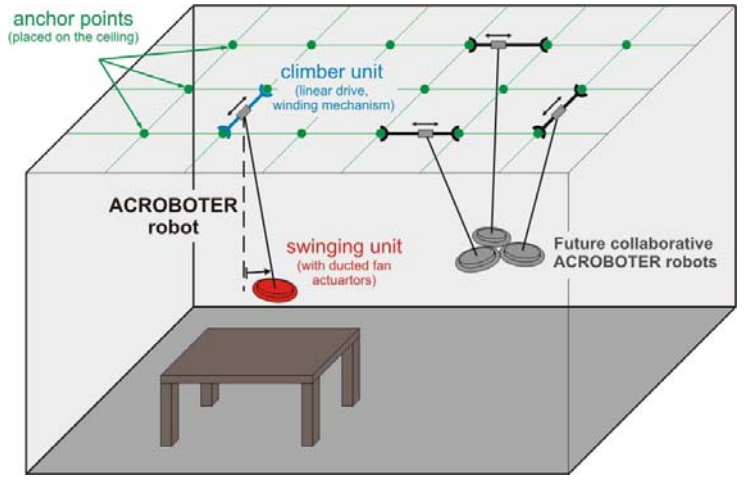


Fig. 1. Concept of the ACROBOTER project

able to operate on the top of tables, wardrobes and can be used for manipulating objects placed on shelves, tables, work surfaces or on the floor.

The conceptual idea of the ACROBOTER project is the development of a completely new technology of a wire suspended robot as it is shown in Fig. 1. Specifically, the whole system is divided into several sub-systems whose architecture, apart from the vision's one, is not presented in this work. The moving platform depends on the anchor points-units placed in a raster fixed to the ceiling of the room. The pendulum-like structure shown in the figure corresponds to the swinging unit (SU) that hangs on a wire, whilst the necessary vertical movements are provided by a winding mechanism (WM) placed on the climber unit (CU). The horizontal transportation of the robot are fulfilled by the CU combined with a linear drive for the fine adjustment of SU's position between the anchoring points. In addition, the SW is equipped with ducted fan actuators that provide a free motion inside a conic volume. Therefore, ACROBOTER's ducted fan system is also used for the posing and as well as stabilization of the robot's motion, and as a result, the unit can fly freely around a suspension point.

As far as the vision system is concerned, it is responsible for three vital tasks that affect directly the overall efficiency of the project. Especially, the vision part emphasizes in the estimation of the SU's pose, the reconstruction of the 3D working environment of the robot and the recognition of objects found in the scene. It consists of five Grasshopper cameras dispersed in different positions. The Grasshopper camera manufactured by Point Grey Research is able to capture images up to 1280 X 960 pixels resolution and it is connected to the PC via a firewire port, using the IEEE 1394b transfer protocol. As far as 3D reconstruction and object recognition are concerned, they are based on four cameras installed in the four corners of the room shown as shown in Fig. 1. The last camera, which provides vital data to compute the SU's pose, is mounted onto

the CU and is fixated towards the ground. The aforementioned demanding tasks are adequately fulfilled by developing and implementing techniques beyond the state-of-the-art.

The remainder of this paper is structured as follows: In Section 2 we give an overview about the related work in the areas of pose estimation, 3D reconstruction and object recognition. ACROBOTER's pose estimation method is presented in Section 3, where both the hardware architecture and the first experimental results are shown. In addition, the 3D reconstruction technique adopted along with the corresponding results are demonstrated in Section 4. Furthermore, essential information concerning the object recognition scheme and how it is trained for the purposes of the ACROBOTER, is presented in Section 5. The last Section 6, concludes with some final notes and an outlook to the future work.

2 Related Work

The essential idea underlying the ACROBOTER project is this new type of autonomous mobile platform, which is designed to move fast and in any direction in an indoor environment. Furthermore, the main challenge is to navigate around any kind of obstacles such as stairs, doorsteps and various other everyday objects that can be found in a room. The ACROBOTER concept outperforms existing developments that require a robot to climb walls and ceiling, due to the fact that it extends significantly the vertical working space of any state-of-the-art similar robotics application. The MATS robot [1], walking robots like Honda's ASIMO or even the Care-o-Bot from IPA proved to have many drawbacks compared to ACROBOTER. The only system or mobile platform that approximates the design or the concept of the presented project is the Flora ceiling based service robot [2]. The latter uses some kind of telescopic arms for the navigation of the working unit into the 3D working space and electromagnetic force for the stabilization of the moving cart on the ceiling.

As far as pose estimation is concerned, it is achieved by using two sensors, a visual and an inertial one. The visual sensor is mounted on the CU and is fixated towards the ground, where the SU operates, and computes the extrinsic parameters of the SU in real-time. An IMU senor lies on the top of the SU and its measurements are fused with the camera ones. This visuo-inertial fusion has been used to many robotics applications as their complimentarity provides efficiency and robustness to the system. The visuo-inertial applications can be divided into three categories, namely the correction, the colligation and the fusion one [3]. The correction deals the readings of the camera as the desired values, while the IMU readings are the estimated ones and vice versa. Concerning the colligation type of integration, a combination of all the measurements takes place. Usually there is a control loop to verify, the camera measurements with the IMU ones and vice versa [4]. Finally, the fusion category takes both measurements into account. Gemeiner *et al.* have estimated the egomotion and the environmental structure by assessing a binocular vision system integrated with an inertial sensor

[5]. Shademan *et al.* [6] have demontrated a comparative study for fusing inertial sensors with visual ones via EKF and iterated extended Kalman filter (IEKF). The main difference of the pose estimation subtask (PES) in the ACROBOTER project with the aforementioned work is that the visual sensor has a different reference frame from the IMU. Furthermore the fact that the visual sensor is constantly moving increases the subsystem's complexity to a high degree. The PES architecture used in the ACROBOTER project is selected after examination described in [7].

The key framework where the 3D Reconstruction task of the ACROBOTER is based on is the computation of the representation of every object in an everyday room using wide baseline stereo techniques. Previous approaches on this field encompass the use of small correlations windows among the views [8] or optimization techniques like graph-cuts [9]. The drawback of using such small image windows and optimization techniques is that are very sensitive to illumination changes and textureless surfaces. Recently, methods that rely on the simplicity and discriminative power of feature detectors and descriptors have been presented [10] and have been used mostly for depth estimation applications.

Recognizing objects in a scene is fundamental task in image understanding and still constitute one of the most challenging tasks. Every pattern recognition technique is directly related with the decryption of vital visual information contained in the natural environment. During the past few years, researchers emphasized in building new recognition frameworks based on appearance features with local estate. Algorithms of this field extract features with local extent that are invariant to possible illumination, viewpoint, rotation and scale changes. In addition, several techniques that enforce the crucial role of local features in demanding pattern recognition application were presented [11]. One of the most efficient object recognition scheme, that is adopted for the purposes of ACROBOTER, is the Scale Invariant Feature Transform (SIFT) [12]. The latter was selected among a set of high-level algorithms to describe patterns and objects. Furthermore, by using information derived from SIFT, we were able to estimate the distance between camera and objects found in a scene.

3 ACROBOTER's Pose Estimation Task

In the ACROBOTER project the pose estimation task provides the orientation and the position vectors of the SU. The PES, is composed of two senors and two computers. A camera is mounted onto the CU and is fixated towards the ground, while a IMU sensor is placed on the top surface of the SU. Both sensors are connected to the respective subsystem computers, while their communication is built on top of sockets. The ACROBOTER system operates in real-time, demanding so, the PES to be able to operate also in high frequencies. In order to meet the requirements of low processing times the PES utilizes the C# OpenCV (Open Source Computer Vision) libraries [13] for the vision tasks, which exhibit low computational load and, thus, they achieve real-time operation. The IMU sensor readings, are acquired by a program code which is implemented in the

C# environment, as well. The measurements of the sensors are fused by the extended Kalman filter (EFK) as it is capable to deal with non-linear problems, and its recursive nature eliminates the errors [14].

The ACROBOTER[1] pose estimation subsystem is depicted in Fig. 2(a). The two computers are connected with each other, through a high speed network protocol as it is illustrated in Fig. 2(b). The visual pose estimation measurements are fused at the SU PC. The EKF is utilized in a function which has two inputs, one for each sensor. The inputs to the EFK are vector states, which includes the positions and the rotations of the three axis. The filter's error covariances, R and Q, concerning the measurement and the process noises, respectively, are tuned to have a ratio of 10 (R/Q=10). These tuning variables determine whether the output follows more the IMU's measurements or the camera's ones. Depending on the reliability of the sensors, the value of Q can be increased, to give more weight to the visual pose, or on the contrary can be decreased, in order the IMU's pose estimates to gain more weight. However, there is a trade off between the cumulative drift error of the inertial unit sensors and the possible occlusions to the field of view regarding the camera. As a result, the tuning process of the filter is highly application dependent and it is mostly a trial and error one.

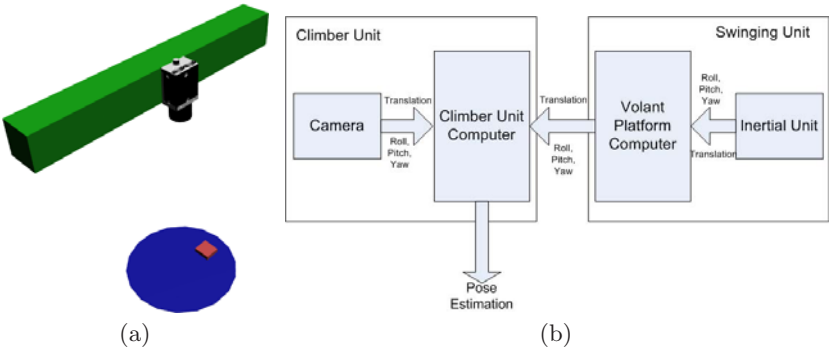


Fig. 2. (a) The ACROBOTER system hardware architecture. The orange box represent the inertial unit for SU, blue cylinder. (b) The proposed system architecture.

The first prototype for the assessment of the visual pose estimation system was performed by identifying a chess board with known geometry. The algorithm identifies the SU by recognizing all the features, i.e. the corners of the chess board squares. The final rotation and translation matrices are given by the OpenCV libraries. Although there is not going to be used a chess board as a feature for the ACROBOTER, the same procedure applies. The landmarks are going to be an arrangement of markers with known geometry. Thus, instead of having as

[1] The ACROBOTER project is still under development. There are copyright restrictions for demonstrating the proposed parts into images. The only parts that can be illustrated are the camera and the inertial sensors.

features the square corners, the markers will be the new features. In order to test the efficiency of our system we mounted the camera on a pan-tilt mechanism and rotated the camera to known angles. The translation measurement where tested with a laser sensor. The accuracy the system is demonstrated in Table 1(a).

Table 1. Overall results for the rotations (left) and for the translations (right). The pan-tilt and the laser sensor measurements are used as the ground truth.

	Pan-Tilt	Camera	IMU	EFK
Roll	14.66	13.99	14.83	14.62
Pitch	4.47	4.58	4.36	4.42
Yaw	9.08	8.84	8.93	8.89
Accuracy	100%	96.7%	98.2%	98.8%
Error	0	0.34	0.15	0.1

	Laser	Camera	IMU	EFK
x	56.21	55.87	57.60	56.47
y	34.82	33.31	33.26	33.39
z	76.63	76.07	77.33	76.51
Accuracy	100%	98.11%	97.4%	98.33%
Error	0	0.80	1.22	0.64

4 ACROBOTER's 3D Reconstruction Task

Reconstruction of surfaces from multiple images has been a central research problem in Computer Vision for a long time. Early previous work in this area focused on developing stereo algorithms mostly for binocular camera configurations. More recently, however, due to significant advances in computational power, vision systems using multiple cameras are becoming increasingly feasible and practical. Multi camera systems like the *Vi Room* which is a low cost synchronized multi camera system developed by [15] and the *3D Room* developed by [16] are systems able to capture multiple synchronized images of indoor scenes. They were developed mainly for tracking applications without providing any knowledge for the structure of the 3D environment.

Due to the wide baseline nature of the 3D reconstruction subtask of the ACROBOTER project, where only four cameras mounted on the four respective corners of the ceiling must provide the coordinates of every object inside the scene, voxel-based algorithms can not deliver accurate results and robust measurements. Thus, a system similar to [17], where point-wise similarity measures for two consecutive views are used, has been expanded to trifocal plane in order to meet the requirements of the task.

Our system gradually combines the strengths of a point-wise similarity measure and a discriminant feature descriptor to deal with the huge amount of information and acquire the desirable result. First of all, we extract features from four views. Due to the unique position of the cameras we need an algorithm that can deal with extreme image tranformations and extract robust feature points. Thus, the use of an algorithm which can provide both quality features and robust matching among them is considered. The ASIFT [18] descriptor can provide the great results of the SIFT algorithm even when affine transformations will occur. The features of different images are then compared using the similarity function of SIFT and lists of potential matches are established. Based on these matches

the relations between the views are computed. We first compute the relationship among the two images which can give us imformation about the exact position of the points among those two views. The next step is the application of trilinear constraints in order to process additional information from another camera. The trifocal tensor is used because of its unique characteristic to transfer corresponding points of the two views to the corresponding point in a third view. Thus, we attain robust point correspodences between the three views of the four cameras in the room. Finally, we compute trifocal tensors over the four triplets of images and a dense point cloud is gradually produced, providing exact information about the position of the objects inside the scene.

In Fig 3 an example of three views from our tests is illustrated. Three wide baseline views are shown at top left side which are the original images from the cameras. Below them the corresponding points between them are depicted and the final 3D point cloud after the application of the trifocal tensor is shown on the right part of the figure.

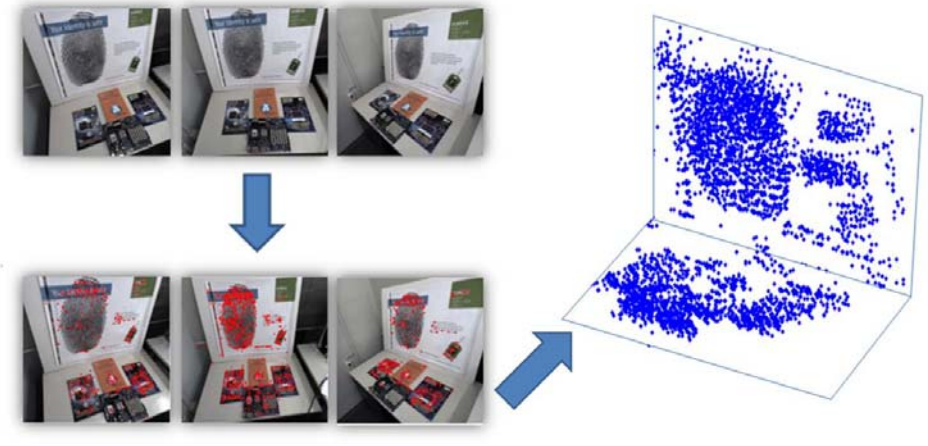


Fig. 3. Three views of the cameras on the top left, three views showing the correspodences between them on the bottom left and the final 3D point cloud after the trifocal tensor is applied

5 ACROBOTER's Object Recognition Task

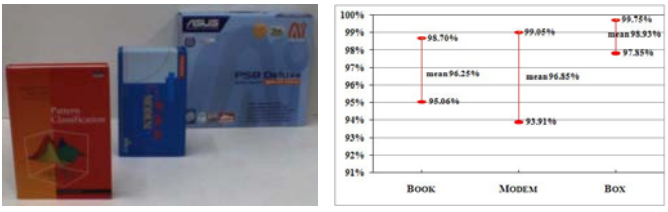
The main idea of local appearance-based recognition methods is the decryption of locally visual information. Thus, the indispensable visual distinctiveness of an object in a scene is ensured by locally sampled descriptions. The efficiency of algorithms based on features is directly related to the maintenance of this regional-based data. Furthermore, the two main submechanisms of such frameworks are the detectors and descriptors of areas of interest. The latter provide special attributes such as, insensitivity against rotation, scale, illumination or

viewpoint changes. Generally, the main idea behind interest point detector is the pursuit of points or regions with unique information in a scene. These spots or areas contain data that distinguish them from others in their local neighborhood. The most important methods of this field are Harris Corner detector [19], Scale Invariant Feature Transform (SIFT) [12] and Speeded Up Robust Features (SURF) [20]. On the other hand, the descriptor organizes the information collected from the detector in a discriminating manner. Thus, locally sampled feature descriptions are transformed into high dimensional feature vectors. From time to time, several approaches that implement a descriptor have been proposed. The most important are the Moment Invariants [21], Gradient Location and Orientation Histogram (GLOH) [22] and SURF [20]. For the purposes of ACROBOTER SIFT was adopted and expanded in order to estimate the distance between camera and objects found in a scene.

ACROBOTER's object recognition task is based on the four cameras installed in the four corners of the room. The main idea behind the proposed method is to maintain SIFT's properties whilst we make an attempt to enhance them. Thus, we have constructed a large database containing images from several objects. With a view to database's enrichment, these objects were photographed from different viewpoints and distances from the camera. Moreover, we used SIFT's matching sub-procedure to build an on-line scene search engine. Estimations derived from this engine are taken into account for the position estimation task. Initially, for each image in the database keypoint features are extracted, using SIFT. Then, the training session, where each object is photographed at different distances from the camera that are stored for further exploitation, takes place. Afterwards, the matching sub-procedure of SIFT is performed, where one image representing the scene is compared with several others (one per sample), representing the object from different viewpoints. Moreover,the features' centers of mass in both images are calculated. Next, the distance, measured in pixels with the use of Euclidian Distance, of each keypoint from the center of mass is calculated. Finally, by taking into account camera's distance from the object during the training session we are able to estimate objects' distance from the camera.

In Fig. 4(a) a scene that contains three different objects (e.g. a book, a modem's box and a motherboard's box), is illustrated[2]. With a view to reader's better understanding, objects found in the scene are referred as book, modem and box, respectively. During the training session, where the system remained offline, each object was captured separately from different viewpoints under varying illumination and geometrical (distance from the camera) conditions. Afterwards, and when the system started, the on-line scene search engine came into operation. The proposed method was evaluated through exhaustive tests containing several scenes and objects. The results of the recognition process were remarkable (approximately 98%), whilst the ones concerning the position estimation procedure for the three aforementioned objects are shown in Fig. 4(b).

[2] The ACROBOTER project is still under development. Due to the limited space only a part of the experimental results is demonstrated.



(a) (b)

Fig. 4. (a) The three objects (book, box, modem) used for evaluation purposes. (b) Accuracy of ACROBOTER's position estimation algorithm.

6 Conclusion

The basic structure and the vision system of the ACROBOTER project has been presented in this paper. The ACROBOTER's system is capable of: estimating robot's pose in the room; reconstruct the 3D working space and; recognize objects with remarkable efficiency. For each aforementioned process we implemented methods that are beyond the current state-off-the-art. Moreover, although the project is ongoing, the first experimental results prove that the overall efficiency of the developing vision system ranges at very high standards. As a result, and with a view to the integration with other systems of the project, we believe that the ACROBOTER will make a great breakthrough at the field of autonomous mobile assistant robots.

References

1. 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)
2. Sato, T., Fukui, R., Morishita, H., Mori, T.: Construction of ceiling adsorbed mobile robots platform utilizing permanent magnet inductive traction method. In: Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 1 (2004)
3. 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, 2007. ROBIO 2007, pp. 405–410 (2007)
4. Zufferey, J.C., Floreano, D.: Fly-inspired visual steering of an ultralight indoor aircraft. IEEE Transactions on Robotics 22(1), 137–146 (2006)
5. Gemeiner, P., Einramhof, P., Vincze, M.: Simultaneous motion and structure estimation by fusion of inertial and vision data. Int. J. Rob. Res. 26(6), 591–605 (2007)
6. Shademan, A., Janabi-Sharifi, F.: Sensitivity analysis of EKF and iterated EKF pose estimation for position-based visual servoing. In: Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005, pp. 755–760 (August 2005)
7. Kyriakoulis, N., Karakasis, E., Gasteratos, A., Amanatiadis, A.: Pose estimation of a volant platform with a monocular visuo-inertial system. In: IEEE International Workshop on Imaging Systems and Techniques, IST, pp. 1–6 (2009)

8. Strecha, C., Fransens, R., Gool, L.V.: Combined depth and outlier estimation in multi-view stereo. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2394–2401 (2006)
9. Kolmogorov, V., Zabih, R.: Multi-camera scene reconstruction via graph cuts, pp. 82–96 (2002)
10. Tola, E., Lepetit, V., Fua, P.: A fast local descriptor for dense matching, Alaska, USA (2008)
11. Liao, M., Wei, L., Chen, W.: A novel affine invariant feature extraction for optical recognition, vol. 3, pp. 1769–1773 (August 2007)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
13. Open Source Computer Vision (OpenCV) home page, `http://sourceforge.net/projects/opencvlibrary`
14. Welch, G., Bishop, G.: An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA (1995) (Revised, July 2006)
15. Svoboda, T., Hug, H., Gool, L.V.: Viroom - low cost synchronised multicamera system and its self-calibration. LNCS, pp. 512–522. Springer, Heidelberg (2002)
16. Saito, H., Baba, S., Kimura, M., Vedula, S., Kanade, T.: Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3d room, pp. 516–525 (October 1999)
17. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R.: Visual modeling with a hand-held camera. International Journal of Computer Vision V59(3), 207–232 (2004)
18. Morel, J., Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. SIAM Journal on Imaging Sciences 2(2), 438–469 (2009)
19. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3d object modeling and recognition from photographs and image sequences, pp. 105–126 (2006)
20. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding 110(3), 346–359 (2008)
21. Van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns, pp. 642–651. Springer, London (1996)
22. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)