

Under Vehicle Perception for High Level Safety Measures Using A Catadioptric Camera System

Caner Sahin and Mustafa Unel
Faculty of Engineering and Natural Sciences
Sabanci University
Istanbul, Turkey
{canersahin, munel}@sabanciuniv.edu

Abstract—In recent years, under vehicle surveillance and the classification of the vehicles become an indispensable task that must be achieved for security measures in certain areas such as shopping centers, government buildings, army camps etc. The main challenge to achieve this task is to monitor the under frames of the means of transportations. In this paper, we present a novel solution to achieve this aim. Our solution consists of three main parts: monitoring, detection and classification. In the first part we design a new catadioptric camera system in which the perspective camera points downwards to the catadioptric mirror mounted to the body of a mobile robot. Thanks to the catadioptric mirror the scenes against the camera optical axis direction can be viewed. In the second part we use speeded up robust features (SURF) in an object recognition algorithm. Fast appearance based mapping algorithm (FAB-MAP) is exploited for the classification of the means of transportations in the third part. Proposed technique is implemented in a laboratory environment.

I. INTRODUCTION

To prevent the dangerous situations that will be resulted from vehicles in certain areas such as shopping centers, government buildings, customs stations or army camps under vehicle surveillance and classification of the vehicles are indispensable missions. The main challenge to achieve this task is to monitor the under frames of the means of transportations for detecting the hidden objects and classifying the vehicles. Since the conventional camera systems have limited field of view, realization of this task becomes formidable. In such a scenario, conventional systems require high number of cameras that give rise to extraordinary computational cost. Moreover, displaying the under frames of the vehicles by typical perspective cameras that have different orientations or a single rotating camera requires wide installation space and extensive calibration. On the other hand, because of the fact that catadioptric camera systems are able to capture omnidirectional images of the environments, i.e. providing 360 degree field of view, one can monitor the under frames of the vehicles, detect the undercovered materials and classify the vehicles just using a single catadioptric camera. This unique feature of the catadioptric cameras eliminates disadvantages of perspective cameras. Moreover, increase in the number of extracted features from panoramic images maintains stability for object detection and classification.

A catadioptric camera system consists of a convex mirror such as a parabolic, a spherical, an elliptical or a hyperbolic

mirror and a single conventional perspective camera. They are also called as omnidirectional vision systems and have been studied extensively [1, 2]. Catadioptric camera systems can be categorized into central and noncentral catadioptric systems. In a central catadioptric camera system convex mirror is aligned with a central camera where it has a single projection center. For more details, interested readers may refer to [3, 4]. Nevertheless, in practice, the real catadioptric cameras have to be treated as noncentral cameras since they have multiple effective viewpoints. Misalignment between the perspective lens and convex mirror, structural imperfection in the convex mirror types, inexact positioning of the perspective camera in one of the focal points of the convex mirror should cause the noncentrality [5]. Regarding the utilization of the multiple catadioptric cameras, different omnidirectional vision systems are designed for different tasks. Schönbein et al. propose two different catadioptric stereo camera systems in [6] that are the combination of the catadioptric-perspective and catadioptric-catadioptric systems mounted on a car. In [7] Lui and Jarvis present vertically aligned stereo catadioptric system that has a variable vertical baseline. Gandhi and Trivedi design an omnidirectional stereo system for visualizing the nearby environment of a vehicle [8]. Schönbein et al. combine three catadioptric cameras and align them horizontally in [9] to increase the robustness of the ego motion estimation and localization by 3D features all around the autonomous vehicles.

From the point of under vehicle surveillance, various monitoring systems are proposed. In [10] a vehicle inspection system is proposed that uses an image mosaic generation technique for different perspective views. A mobile robot equipped with a 3D range sensor to inspect the under frames of the vehicles is offered by Sukumar et al. [11]. A combination of the vehicle recognition and the inspection system is proposed in [12] to improve safety precautions. In [13] an automatic under vehicle inspection system is utilized to monitor the under frames of the vehicles. Regarding the under vehicle surveillance in most of the proposed solutions, different computer vision and image processing algorithms are utilized with perspective cameras.

In this study we propose a new catadioptric camera system consists of a perspective camera pointing downwards to the convex mirror mounted to the body of a mobile robot to monitor the under frames of the vehicles. We show how to solve one of the most common safety measure problems in structures where extra safety precautions must be taken by

displaying under frames of the vehicles that cannot be dealt with conventional perspective cameras easily. While mobile robot navigates under the means of transportations, it starts to detect the hidden materials attached to the under vehicles and classify the vehicles utilizing the fast appearance based mapping (FAB-MAP) algorithm [14]. If the robot detects a peculiar material such as a bomb it warns the detection of the material by drawing a line between the object image in the database and the object that is seen in the video frame.

The rest of the paper is organized as follows: In section II, the imaging model and the construction of the catadioptric camera system are introduced. In section III, object recognition algorithm is presented. FAB-MAP algorithm is described in section IV. Experimental results are provided in section V, and finally, the paper is concluded in section VI.

II. CATADIOPTRIC CAMERA SYSTEM

A. Catadioptric Camera Model

In the design of the catadioptric camera systems one important property that must be considered is determining the shapes of the mirrors in such a way that the single effective viewpoint condition is ensured. The reason why a single effective viewpoint is desirable is that it allows the derivation of the epipolar geometry of two omnidirectional images and it is a requirement for the generation of pure perspective images from the sensed images. Regarding our omnidirectional vision system, we used hyperbolic convex mirrors and the projection model that Mei et al. propose in [15]. In the following steps we summarize the imaging model (Fig. 1):

1) The projective ray \mathbf{x} coming from \mathbf{X} intersects the unit spherical surface in M ,

$$(M)_o = \frac{\mathbf{X}}{\|\mathbf{X}\|} = \left(\frac{x}{\|\mathbf{X}\|}, \frac{y}{\|\mathbf{X}\|}, \frac{z}{\|\mathbf{X}\|} \right)^t \quad (1)$$

where $\|\mathbf{X}\| = \sqrt{X^2 + Y^2 + Z^2}$.

2) Once the world points are projected onto the unit sphere, the points are changed to a new reference frame centered in $O_c = (0, 0, -\xi)$,

$$(M)_{O_c} = \left(\frac{x}{\|\mathbf{X}\|}, \frac{y}{\|\mathbf{X}\|}, \frac{z}{\|\mathbf{X}\|} + \xi \right)^t \quad (2)$$

where ξ is the difference between O_c and sphere center O and is a mirror parameter.

3) These points are projected onto the normalized image plane. A second projective ray \mathbf{y} is defined that intersects the points O_c and M . The intersection of \mathbf{y} with the plane $Z = \psi - 2\xi$ is the catadioptric image of \mathbf{X} ,

$$\mathbf{x}_i = \left(\frac{x}{z + \xi\|\mathbf{X}\|}, -\frac{y}{z + \xi\|\mathbf{X}\|}, 1 \right)^t = f_i(\mathbf{X}) \quad (3)$$

4) The final projection matrix includes a camera projection matrix K with γ the generalized focal length, (u_0, v_0) the principal point, s the skew and r the aspect ratio [15].

$$\mathbf{p} = k(\mathbf{x}_i) = \begin{bmatrix} \gamma & \gamma s & u_0 \\ 0 & \gamma r & v_0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_i = K \mathbf{x}_i \quad (4)$$

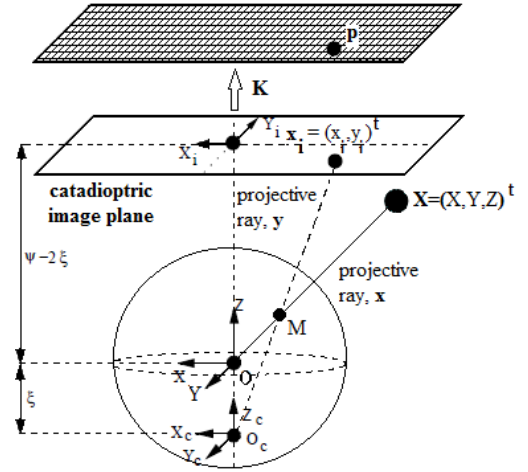


Fig. 1. Modeling central catadioptric image formation [15]

B. Catadioptric Camera System

The catadioptric camera system proposed in this paper is a combination of a hyperbolic mirror and a perspective camera. The hyperbolic mirror is attached to a plexiglass plate and it is passed through a four sided transparent plexiglass tube in such a way that the mirror is settled down in the base. Top side of the tube is covered with a hole centered transparent plate that the camera lens is able to point down to the hyperbolic mirror. Some example photos taken using the catadioptric system are depicted in Fig. 2. Since the perspective camera points downwards we can see the ceiling of the laboratory in these images. Once we designed this system, we mounted it to the body of a nonholonomic mobile robot. The main advantage of such a system is to monitor the vehicle under frames that cannot be achieved easily using conventional camera systems. Other benefits obtained from this system can be listed as: It increases the field of view and as a result not only the frontal direction but also the right, left and back sides of the mobile robot are displayed. The number of extracted features from single catadioptric image is higher than a perspective image and so matching between two consecutive images taken from catadioptric cameras gives rise to much more consistent results in terms of object recognition and classification, localization and mapping. In this study we use just a single catadioptric camera for object recognition and vehicle classification that is able to monitor upper side of the camera mounting area. One can also design a catadioptric stereo system to utilize for 3D reconstruction, visual simultaneous localization and mapping, structure from motion and pose estimation etc.

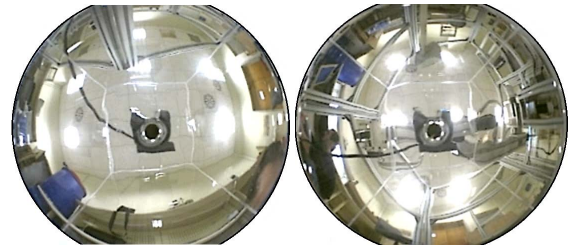


Fig. 2. Catadioptric images

III. OBJECT RECOGNITION

In a typical object recognition system, extracted features from a test object are matched against the features of the object model database to determine the identity of an object as shown in Fig. 3. There are two main approaches in object recognition: model-based recognition and appearance-based recognition. In model-based recognition problem, an object model is being used and it is subjected to geometric transformation that maps the model in 3D world into the camera sensor coordinate frame. In such a recognition approach, efficient algorithms for estimating geometric transformations are central to many model-based recognition systems. In contrast, appearance-based approach does not require any prior knowledge of an object. The latter approach is suitable for the algorithms such as simultaneous localization and mapping which deals with unknown environments [16].

Several approaches are proposed for appearance based object recognition. Santos et al. present the support vector machine (SVM) learning technique as an option to perform appearance-based object recognition [17]. Itti et al. propose the saliency based region selection strategy that extracts multi-scale image features to find salient objects in a cluttered natural scene [18]. Lowe's Scale Invariant Feature Transform (SIFT) features [19] provide invariance to change in rotation, scale and viewpoint and are successfully used in object recognition.

A. Speeded Up Robust Feature (SURF) Extraction and Matching

Bay et al. propose well-known SURF features in [20] and SURF features are exploited in various object recognition algorithms. SURF descriptor represents a distribution of Haar wavelet responses within interest point neighborhood. It is based on the Hessian matrix and relies on integral images to reduce the computation time. In [20] three different versions of the descriptors have been examined and compared with the SIFT descriptor: the standard SURF descriptor, which has a dimension of 64, the extended SURF which has a dimension of 128 and U-SURF version that is not invariant to rotation and has a length of 64 elements. According to the results of the performances for 3 different versions it is indicated that while SURF, extended SURF and upright SURF (U-SURF) extraction processes take 354ms, 391ms, 255ms computational time respectively, SIFT feature extraction method takes 1036ms. In a comparison between the performance of SURF and SIFT feature extraction methods, it is shown that for a scene requiring about 1000ms with SIFT, the extraction of the SURF features takes about 250ms, meaning that the time is reduced by a factor of 4 [21]. Because of the fact that SURF features are not only scale and rotation invariant but also offer the advantage of being computed very efficiently compared other feature extraction methods, in this work we utilize SURF features in our object recognition algorithm. Extracted SURF features in a catadioptric image are shown in Fig. 4 and 252

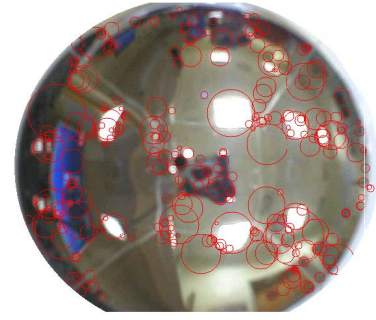


Fig. 4. Extracted SURF features

SURF keypoints are extracted.

Once the SURF keypoints are detected in both database object image and video frames the nearest neighbor matching algorithm is implemented between SURF keypoints. A keypoint in the test image is compared to a keypoint in the database object image by calculating the Euclidean distance between their descriptor vectors. In our work we use SURF descriptor vectors that have lengths of 64 elements. After detection of a matching pair it is examined that the distance is closer than 0.7 times the distance of second nearest neighbor.

IV. VEHICLE CLASSIFICATION VIA PLACE RECOGNITION

In this section, we describe the place recognition algorithm to classify the under frames of the vehicles and utilized Cummins and Newman's appearance based mapping algorithm in [14]. This important work proposes an appearance based probabilistic solution for many problems such as loop closure and perceptual aliasing in SLAM that cannot be solved easily using standard Extended Kalman Filter (EKF).

To recognize places, the world is modeled as a set of discrete locations and each location is described by a probability distribution over appearance words. Extracted features from images are converted into a bag-of-words representation and a vocabulary is generated. Also, for each location, observation probability of coming from a place in the map or not is examined.

A. Bag-of-Words Model

In bag-of-words model, an image is represented as a sort of document, and it contains a set of local descriptors. In order to obtain visual words from images the feature space of the descriptors must be quantized. Thus, a new descriptor vector can be held in terms of the discretized region of feature space to which it belongs. Then, the vocabulary that includes collection of words is generated collecting a large sample of features from a representative corpus of images and quantizing the feature space according to their statistics. In [22], Sivic and Zisserman propose quantizing local image descriptors for the sake of rapidly indexing video frames with an inverted file. They show that local descriptors extracted from features can be mapped to visual words by computing prototypical descriptors with k-means clustering, and that having these tokens enabled faster retrieval of frames containing the same words. Once the descriptor vectors are quantized into visual words, weighting and indexing processes are applied to the vector model as

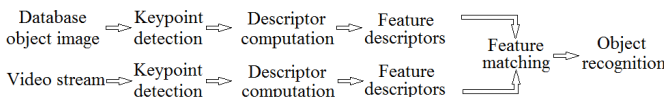


Fig. 3. Object recognition

follows: In a vocabulary which includes k words, each document is represented by a k -vector $V_d = (t_1, \dots, t_i, \dots, t_k)^T$ of weighted frequencies with components

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i} \quad (5)$$

where n_{id} is the number of occurrences of word i in the document d , n_d is the total number of words in the document d , n_i is the number of occurrences of term i in the database and N is the number of documents in the whole database. The weighting is the multiplication of the word frequency and the inverse document frequency. The word frequency weighs words occurring often in a particular document, while the inverse document frequency downweighs words that appear often in database [22]. All of these steps are applied before actual retrieval, and the set of vectors representing all the documents in a corpus are organized as an inverted file. An inverted file index is almost the same as an index in a book, where the keywords are mapped to the page numbers where those words are used. In the visual word case we have, we have a table that points from the word number to the indices of the database images in which that word occurs. Retrieval via the inverted file is faster than searching every image, assuming that not all images include every word. In this work, we utilize SURF features and descriptors to have a bag-of-words representation.

B. Loop Closure Detection

Detection of loop closure requires the capability of recognizing a previously visited place from current visual sensor measurements. To make it clear one can consider the following illustrative example: Suppose that you are making camp on an island you have not visited before. You would like to discover the camping environment. At the beginning you mentally keep track of the path you travelled but after some time it would be a challenging work for you to remember in what point you are with respect to the camping area. Instead, if you follow a circular path you pass the places that you have visited before. Thus, recognizing previously visited places will allow you to estimate your trajectory and where you are with respect to the camping area.

In this paper, we address the problem of loop closure detection as an image retrieval task. To classify the vehicles, a newly visited place is examined if it is a new under frame or an old one that is seen before. To achieve this aim, the probability of an observation at a particular sample time k is calculated utilizing the given observations till sample time $k - 1$. For the theory of loop closure detection interested readers may refer to [14].

V. EXPERIMENTAL RESULTS

Our proposed solution is implemented using a nonholonomic mobile robot in a laboratory environment. In our implementations the bottom of the tables in the laboratory are considered as the under vehicles. A database that includes eight different under vehicle images is used in this experimental work. They are attached to the bottom of the tables and the mobile robot is made navigate under the tables. All the

algorithms are implemented in Microsoft Visual C++ and OpenCV 2.4.4 that are installed to the on-board computer mounted to the body of the mobile robot.

A. Experimental Setup

The mobile robot that is used in our experimental work includes a processor, an on-board computer, a catadioptric camera system and a rechargeable lithium polymer battery (Fig. 5). Working principle of the experimental setup is depicted in Fig. 5. The power of both of the processor and the computer are supplied via 14.8 V lithium polymer batteries. The processor is inserted to the mobile robot body for sending the control commands to the wheels of the mobile robot. Philips USB camera with a catadioptric mirror is connected to the on-board computer. The communication between the processor of the mobile robot and computer is provided using RS-232 communication protocol. A network is established between the on-board computer and a laptop and shown with a dashed line in Fig. 6. The laptop is used as an external device to display the camera results on the screen.

B. Results for Object Recognition

While the mobile robot navigates under the tables it starts to monitor under vehicle images that are attached to the bottom of the tables. In a certain image, we attach a test object which is one of the database images in the mobile robot and make the mobile robot detect the object. In this implementation we use the SURF features for extraction and matching between the object and the video frames and Random Sample Consensus (RANSAC) algorithm to neglect the matches that are found as

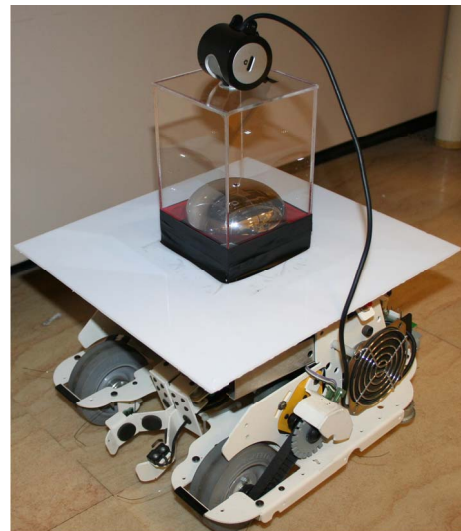


Fig. 5. Experimental setup

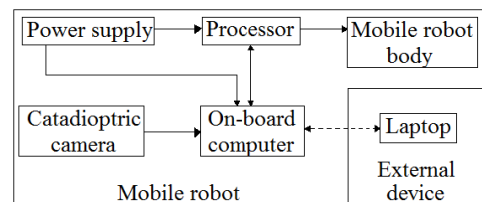


Fig. 6. Working principle of the experimental setup

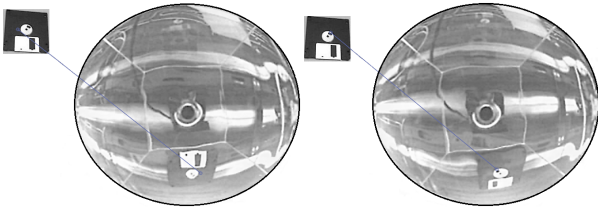


Fig. 7. Detected objects

outliers. If the mobile robot detects the object in a catadioptric image it is shown using a line as in Fig. 7.

C. Results for Vehicle Classification

In the first experiment, to show the accuracy of the FAB-MAP algorithm we use a hand-held perspective camera with taking under frame images of the vehicles. Seven different under frame images of the vehicles are used to calculate the resultant confusion matrix that is shown in Fig. 8. When a new place is seen the relevant diagonal element of the matrix is assigned with a high probability distribution and this element is depicted bright in the matrix. Regarding the loop closure detection off-diagonal elements of the matrix are used and indicated bright on the off-diagonal region.

In Fig. 8, it is seen that all of the diagonal elements of the matrix are bright and it is understood that all of the visited places are new and there is no loop closure detection. Namely each of these images belongs to different under vehicles and they can be classified in seven groups. The probability of being a new vehicle under frame for the third one is 0.995 whilst the fifth one is 0.996968.

Once we obtain this resultant confusion matrix we try a different set of images to show the loop closure detection. The relevant loop closure detections are shown in Fig. 9. In Fig. 9 (a) and (b), two different images of the same vehicle under frame for the first and ninth discrete places are shown whilst in Fig. 9 (c) and (d) the same under vehicle images are depicted for the third and tenth places (see Fig. 10). While the loop closure probability for the ninth and first images is 0.961524, the probability of being a new place for the ninth image is 0.0150896. Similarly, loop closure probability for the third and tenth images is 0.954927 and assigned probability for being a new place for the tenth image is 0.00117. These ten vehicles can be classified under eight groups because of the two loop closure detections in the ninth and tenth steps.

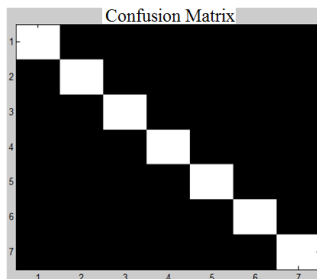
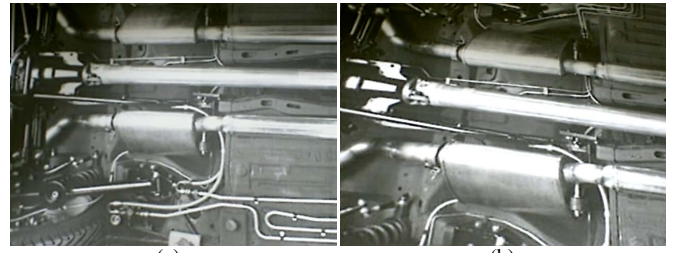


Fig. 8. Confusion matrix: all visited places are seen first



(a)

(b)



(c)

(d)

Fig. 9. Loop closure detections: between (a) and (b) for the ninth place and between (c) and (d) for the tenth place

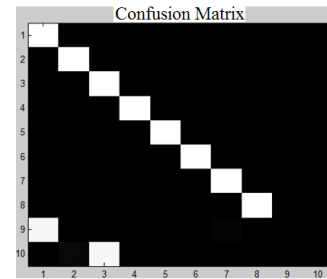


Fig. 10. Confusion matrix: loop closures in the ninth and tenth places

These loop closure results allow us to classify the vehicles merely using their under frames. In the experiment we use Open FAB-MAP software released by Glover et al. [23].

In the second experiment, we take six different under frame images of the vehicles using our proposed catadioptric camera system mounted to the body of the mobile robot. Some example images are shown in Figure 11. As it is seen from Fig. 11, a different place is assigned for each different vehicle under frame. Firstly, we capture the omnidirectional images of the six consecutive different under vehicles and related confusion matrix is depicted in Fig. 12 (a). Because of the fact that all images are different, the diagonal elements of the matrix are indicated bright with high probability that explains the related visited place is newly seen. For example, the probability of being a new under frame for the third place is 0.996 and for the sixth place is 0.997. Then, we deliberately enlarge the database by two additional images that belong to the same under vehicles in the database. This time, the resultant confusion matrix is shown in Fig. 12 (b) that explains the loop closures between the fourth and seventh places and between first and eighth places. While the loop closure probability for the fourth and seventh images is 0.9742, the probability of being a new place for the seventh image is 0.01296. Similarly, loop closure probability for the first and eighth images is 0.96289 and

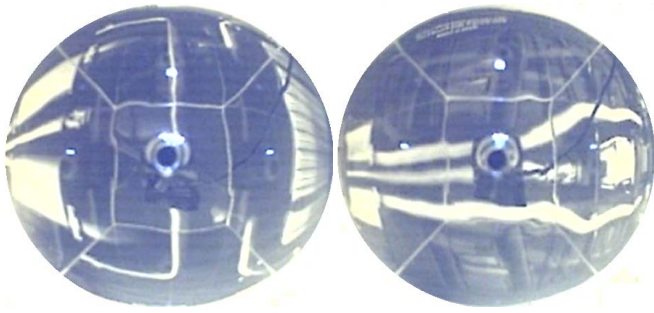


Fig. 11. Omnidirectional images of under vehicles

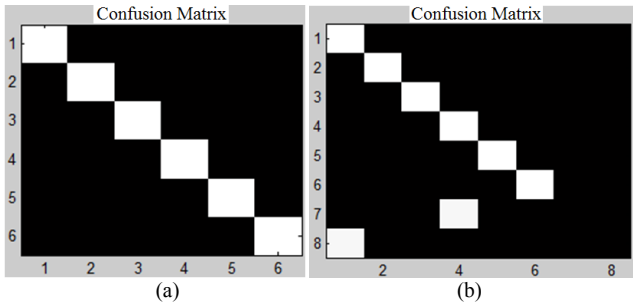


Fig. 12. Confusion matrices for omnidirectional images: (a) all visited places are seen first, (b) loop closures in the seventh and eighth places

assigned probability for being a new place for the eighth image is 0.0023. These eight vehicles can be classified under six groups because of detecting two loop closures in the seventh and eighth steps.

VI. CONCLUSION AND FUTURE WORKS

In this paper, a new solution for under vehicle surveillance and the classification of the vehicles is proposed. A catadioptric camera system where the perspective camera points downwards to the convex mirror is used to monitor the under vehicles. The vehicles are classified utilizing the FAB-MAP algorithm and hidden objects are recognized under vehicles using SURF features. Experimental works verify the feasibility of the proposed solution. In the future, we will expand our work towards the vision based control of the mobile robot using catadioptric cameras.

REFERENCES

[1] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int. Journal of Computer Vision*, vol. 35, no. 2, pp. 1-22, 1999.

[2] R. Benosman and S. E. Kang, *Panoramic Vision: Sensors, Theory, and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, 2001.

[3] C. Geyer and K. Daniilidis, "Structure and motion from uncalibrated catadioptric views," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 279-286, 2001.

[4] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *Int. Journal of Computer Vision*, vol. 49, no. 1, pp. 23-37, 2002.

[5] B. Micusik, and T. Pajdla, "Autocalibration and 3D reconstruction with noncentral catadioptric cameras," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol.1, pp. 58-65, 2004.

[6] M. Schönbein, B. Kitt, and M. Lauer, "Environmental perception for intelligent vehicles using catadioptric stereo vision systems," *Proc. of the European Conference on Mobile Robots (ECMR)*, Sweden, pp.1-6, 2011.

[7] W. L. D. Lui and R. Jarvis, "Eye-Full Tower: A GPU based variable multibaseline omnidirectional stereovision with automatic baseline selection for outdoor mobile robot navigation," *Robotics and Autonomous Systems*, vol. 58, pp. 747-761, 2010.

[8] T. Gandhi and M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps," *IEEE Trans. on Intelligent Transportation Systems*, vol. 7, no. 3, pp. 293-308, September 2006.

[9] M. Schönbein, H. Rapp and M. Lauer, "Panoramic 3D reconstruction with three catadioptric cameras," *Advances in Intelligent Systems and Computing*, vol. 193, pp.345-353, 2013.

[10] P. Dickson et al., "Mosaic generation for under vehicle inspection," *Proc. of Sixth IEEE Workshop on Applications of Computer Vision*, pp. 251-256, 2002.

[11] S. R. Sukumar, D. L. Page, A. V. Gribok, A. F. Koschan and M. A. Abidi, "Robotic three dimensional imaging system for under vehicle inspection," *Journal of Electronic Imaging*, vol. 15, 2006.

[12] C.N. Anagnostopoulos, I. Giannoukos, T. Alexandropoulos, A. Psyllos, V. Loumos and E. Kayafas, "Integrated vehicle recognition and inspection system to improve security in restricted access areas," *IEEE Annual Conference on Intelligent Transportation Systems*, pp. 1893-1898, 2010.

[13] E. E. Ruiz and K. L. Head, "Use of an automatic under vehicle inspection system as a tool to streamline vehicle screening at ports of entry and security checkpoints," *IEEE European Intelligence and Security Informatics Conference*, pp. 329-333, 2012.

[14] M. Cummins and P. Newman, "FABMAP: Probabilistic localization and Mapping in the space of appearance," *The Int. Journal of Robotics Research*, vol. 27, no. 6, pp. 647-665, 2008.

[15] C. Mei, S. Benhimane, E.Malis and P. Rives, "Homography-based tracking for central catadioptric cameras," *IEEE Int. Conf. on Intelligent Robots and Systems*, pp. 669-674, 2006.

[16] Y. J. Lee and J. B. Song, "Visual SLAM in indoor environments using autonomous detection and registration of objects," *IEEE Int. Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 671-676, 2008.

[17] E. M. D. Santos and H. M. Gomes, "Appearance-based object recognition using support vector machines," *IEEE Computer Graphics and Image Processing*, pp. 399, 2001.

[18] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.

[19] D. G. Lowe, "Distinctive image features from scale invariant keypoints," *IJCV*, vol. 60, no.2, pp. 91-110, 2004.

[20] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "SURF: Speeded up robust features," *Computer Vision and Image Understanding*, vol.110, no.3, pp. 346-359, 2008.

[21] V. Höglman, "Building a 3D map from RGB-D sensors," Master thesis, 2012.

[22] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," *IEEE Int. Conf. on Computer Vision*, vol.2, pp. 1470-1477, 2003.

[23] A. Glover, W. Maddern, M. Warren, S. Reid, M. Milford and G. Wyeth, "OpenFABMAP: an open source toolbox for appearance based loop closure detection," *IEEE International Conference on Robotics and Automation*, pp.4730-4735, 2012.