ELSEVIER

Pattern Recognition Letters 20 (1999) 1381–1387

Pattern Recognition Letters

www.elsevier.nl/locate/patrec

Target recognition via 3D object reconstruction from image sequence and contour matching

Mostafa G. Mostafa *, Elsayed E. Hemayed, Aly A. Farag

Department of Electrical Engineering, Computer Vision and Image Processing Laboratory, University of Louisville, Louisville, KY 40292, USA

Abstract

This paper proposes an automatic target recognition (ATR) system based on the three-dimensional (3D) reconstruction of the target from an image sequence. The main contribution of this work is twofold: (1) we present a modified voxel coloring reconstruction algorithm and (2) we employ the 3D reconstructed target model to generate the front and side target templates at zero depression angle to be used in the target recognition process. Target recognition is performed by matching the generated templates to a library using a subpixel contour matching algorithm. Experimental results on simulated scenes show the accuracy of the approach presented in this paper.

Keywords: Automatic target recognition; Three-dimensional object reconstruction; Voxel coloring reconstruction; Template matching

1. Introduction

Automatic target recognition (ATR) has been an active area of research in the past two decades (Bhanu et al., 1997; Ratches et al., 1997). It has attracted many researchers, due to its theoretical challenge and its application in image exploitation. An ATR is a system that is expected to detect, recognize and track a target or multiple targets without human intervention. In the recent years, it has been shown that an ATR based on range imaging is more robust than other ATR systems. This is due to the fact that range images provide information on the 3D structure of the sensed object/scene which efficiently is used for object localization and recognition (Bhatnagar et al., 1997; Miller et al., 1997). However, three-dimensional (3D) imagers, e.g., laser radars (LADARs), are less likely to be used in some scenarios because it is an active sensor. This makes the 3D scene reconstruction from intensity images an important candidate to the problem of object recognition in the ATR systems.

In the past two decades, various algorithms were proposed for performing ATR based on varieties of sensing modalities (Bhanu et al., 1997). Many approaches have been proposed to solve the object recognition problem by using 2D and 3D matching algorithms. Stevens and Beveridge (1997) developed a 3D model-based ATR system which is based on matching edge features of the target's 3D CAD model to those extracted from

* Corresponding author.

E-mail addresses: mostafa@cvip.louisville.edu (M.G. Mostafa), sayed@cvip.louisville.edu (E.E. Hemayed), farag@cvip.louisville.edu (A.A. Farag)

0167-8655/99/$ - see front matter
PII: S0167-8655(99)00109-9

combined multisensory data including LADAR, intensity and FLIR imaging sensors. Serra and Berthod (1997) developed an ATR algorithm to localize and recognize a 3D model in a sequence of monocular images, based on matching the model with the 3D edge map extracted from the reconstructed scene.

In this paper, we present a multisensor ATR system, described in Section 2, that is based on the target's 3D model. A modified voxel coloring technique, which is described in Section 3, is used to reconstruct the target's 3D model from an image sequence. Section 4 describes the target recognition process which is based on template matching. Section 5 presents the results and discussions. The paper is finally concluded in Section 6.

2. The proposed ATR system

The proposed ATR system, schematically outlined in Fig. 1, includes a hyperspectral imager and a video camera of known motion. The operation scenario is as follows. The hyperspectral images are used to find the region of interest (ROI) in the images by globally detecting and localizing the targets in the scene, using a fuzzy classifier (see Yamany et al., 1999a). After detecting the ROI, a monocular camera of known motion can be zoomed on the target and acquire a sequence of images. These images are used to reconstruct the 3D structure of the observed target.

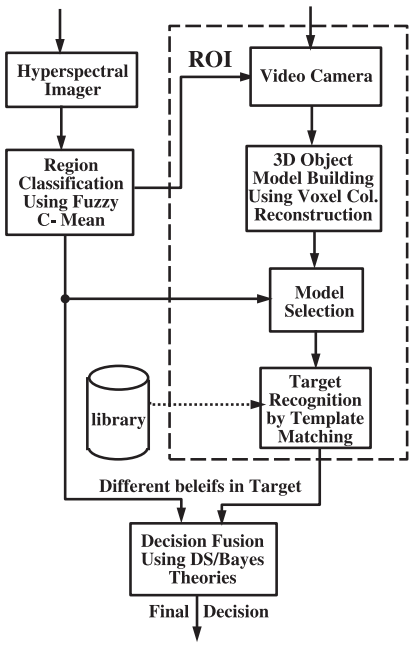

Fig. 1. Schematic representation of the proposed ATR system.

The 3D model reconstruction is carried out using a modified voxel coloring algorithm. The reconstructed 3D target is then used to generate front and side contours of the target on which the target recognition process is based. Classification of the detected target is performed using the results of both the hyperspectral classification and the 3D reconstructed model. Results of both the recognition process from the template matching and the hyperspectral classification may provide two different beliefs about the target which can be combined into one belief using Dempster–Shafer or Bayes decision theories. This paper focuses on the target's 3D model reconstruction and the recognition processes.

3. 3D object reconstruction from a sequence of images

One of the goals of machine vision is to understand the visible world by inferring 3D properties from 2D images. Making such an inference requires modeling of the relationship between the 2D images and the 3D world. Camera calibration is a process which models this relationship.

The camera is calibrated only once if it is stationary. A moving camera has to be recalibrated at each position. However, in the following we show that for a translationally moving camera, the camera calibration can be done automatically. Considering the pinhole camera model, the 3D coordinates of a point $\boldsymbol{M} = [X\ Y\ Z]^{\mathrm{T}}$ in the world coordinate system and its 2D retinal image coordinates $\boldsymbol{m} = [x\ y]^{\mathrm{T}}$ are related by $s\tilde{\boldsymbol{m}} = \boldsymbol{P}\tilde{\boldsymbol{M}}$, where s is an arbitrary scalar, $\tilde{\boldsymbol{m}} = [\boldsymbol{m}^{\mathrm{T}}\ 1]^{\mathrm{T}}$ and $\tilde{\boldsymbol{M}} = [\boldsymbol{M}^{\mathrm{T}}\ 1]^{\mathrm{T}}$ are the homogeneous coordinates of the points $\boldsymbol{m}$ and $\boldsymbol{M}$, respectively, and $\boldsymbol{P} = [\boldsymbol{B}\boldsymbol{b}]$ is the camera perspective projection matrix (Faugeras, 1993), where $\boldsymbol{B}$ is a 3×3 matrix and $\boldsymbol{b}$ is a

3×1 vector. The location of the camera optical center $\boldsymbol{M}_c = -\boldsymbol{B}^{-1}\boldsymbol{b}$ can be tracked as the camera moves, by a GPS system to obtain the translation vector $\boldsymbol{T}_c$ between the new and original optical center. Thus $\tilde{\boldsymbol{M}}_c^{\text{new}} = \boldsymbol{T}_c\tilde{\boldsymbol{M}}_c^{\text{original}}$. Using this transformation, we can write $\boldsymbol{P}^{\text{new}} = \boldsymbol{P}^{\text{original}}\boldsymbol{T}_c^{-1}$ which is used to recalculate the perspective projection matrix after each translation of the camera in terms of its previous values and the translation vector.

3.1. 3D Reconstruction by voxel coloring

Recently, image-based reconstruction has gained much interest. Our modified voxel (volume element) coloring reconstruction technique is inspired by the voxel coloring reconstruction algorithm of Seitz and Dyer (1997) that avoids the image correspondence problem by discretizing the scene space into a set of fixed voxels that is traversed and colored in a fixed visibility ordering. The voxel coloring problem is to assign colors (radiances) to voxels in a 3D volume so as to maximize photo integrity with a set of input images.

Instead of using fixed voxel size and discretizing the whole space, we propose another approach for the reconstruction using a sequence of images that do not cover the whole object's surface. Our approach, as schematically illustrated in Fig. 2, reconstructs the 3D object with respect to a reference image in the image sequence. For a given sequence of N images I_i, $i = 0, 1, 2, \ldots, N-1$, the reference image is taken to be I_0. Note that the reference image can be any image of the sequence. The camera optical center of each image is denoted by

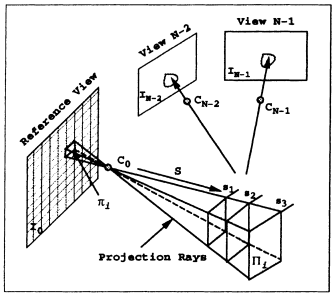

Fig. 2. Illustration of the voxel coloring technique.

C_i, $i = 0, 1, \ldots, N-1$. First we discretize the reference image I_0 into K non-overlapping segments such that $I_0 = \bigcup_{i=1}^{K} \pi_i$, where π_i denotes the segment i. Each segment is represented by its four corners, i.e., $\pi = \{\boldsymbol{m}_i;\ i = 1, \ldots, 4\}$. From the geometry of the perspective projection in a pinhole camera model, we derive an equation for the projection ray of each point $\boldsymbol{m}$. The projection ray is the line that passes through the optical center C and $\boldsymbol{m}$. The projection ray equation is written in terms of the scale factor s as

$$\boldsymbol{M}(s) = \boldsymbol{B}^{-1}(-\boldsymbol{b} + s\tilde{\boldsymbol{m}}),$$

which indicates that for a calibrated camera, the 3D point $\boldsymbol{M}$ of a projection $\boldsymbol{m}$ is uniquely determined by s. Therefore, the space of s values that bounds the scene is uniformly discretized into L values such that $S = \{s_i \,|\, i = 1, \ldots, L\}$. The value of s_i is positive with a minimum value equal to 0, which is the optical center (i.e., $M(0) = C$). The reconstruction algorithm searches for the value $s_m \in S$ that maximizes the photo integrity over the entire sequence (i.e., N images). The photo integrity is measured by the similarity between the projection of $M(s)$ in the N images. In order to speed up the search process, we assume that each segment $\pi \in I_0$ is approximated by a planar patch Π in the 3D space. So, we compute s for segments not points. In this case, the similarity measure λ is computed between N regions as follows:

$$\lambda = \frac{1}{A}\sqrt{\frac{\sum_{i=0}^{N-1}(a_i - A)^2}{N}},$$

where a_i denotes the mean of region π_i and A_i the mean of the N regions $\bigcup_{i=0}^{N-1} \pi_i$.

The voxel coloring algorithm is outlined as follows:

1. Tessellate the reference image, I_0 into K regions each of area π.
2. Initialize the object space $\mathcal{O} = \Phi$, $i = 1$.
3. Let $\pi = \pi_i$.
4. Discretize the ray projecting to π into a set S.
5. **For** every $s \in S$ **do**.
   - Compute the corresponding 3D segment Π_s computed at s.
   - Project Π_s into the other images $I_1, \ldots, I_{N-1}$, then compute the photo-integrity measure λ_s.

- Select s_m that maximizes the photo-integrity λ_s.
- Update the object space $\mathcal{O} = \mathcal{O} \cup \Pi_{s_m}$.

6. $i \leftarrow i + 1$ and **repeat** from step 3 **until** $i = K$.

4. Target recognition using template matching

Since we only have a *partial* 3D data representation for the target from the reconstruction process (i.e., the visible surface), we cannot use 3D–3D matching algorithms for the recognition task. Partial surface matching techniques will result in erroneous measurements, due to the presence of outliers and noise in the reconstructed data. Therefore, we propose to generate the intrinsic target's 2D templates, front and side views at zero depression angle and use them in the recognition process. The templates are produced by orthogonal projection of the 3D reconstructed data onto the x–z and y–z planes, after aligning the data such that its longest side is parallel to the x-axis. Then we match the target's contours generated from those templates with those of the selected 3D models from the library.

Template matching has been used extensively in object recognition. It is used to recognize patterns by matching extracted features to library models. In this work we used a subpixel contour matching algorithm (Yamany et al., 1999b) that allows translation and rotation in the matching process. We use the algorithm to obtain a matching measure between the templates generated from the reconstruction process and the templates selected from the library. The matching measure is used as the decision criterion.

It should be noted that the proposed ATR system in this paper is robust to scaling and rotation, since we use the 3D reconstruction model to generate the 2D template images. Therefore, we generate the images to be of the same scale and orientation as the model images in the library.

5. Results and discussion

We used the 3D Studio to generate sequences of two scenes of a 3D CAD model of a Panther tank. The two sequences are created for the Panther tank without and with occlusion. The camera location is chosen such that it looks at the target at 10° depression angle. In order to create a proper sequence of images, the camera is translated parallel to the ground while fixating to the tank model and the images are taken at different camera positions of known translation. We also chose a dim lighting condition to simulate a real environment. From this setup, we obtained a sequence of 10 images for the tank, some of which are shown in Figs. 3 and 4, with the 3D reconstructed data for the full Panther tank and the occluded tank, respectively. The metric measurements of the target's height, length and width from the reconstructed data are 323.62, 866.61 and 364.17 units, while the model values are 305.83, 861.36 and 336.32. These results show the accuracy of the reconstruction algorithm.

The side and front views obtained by projecting the reconstructed tank and its CAD model onto the y–z and x–z planes are shown in Fig. 5 with their outer contours. We used the same algorithm to generate a library of the side and front views of the other CAD models. These contours are shown in Fig. 6. Fig. 7 shows the matching results of the

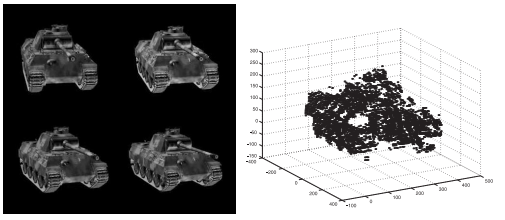

Fig. 3. (left) Four images from a sequence of ten images of a Panther tank and the 3D data of the reconstructed tank (right).

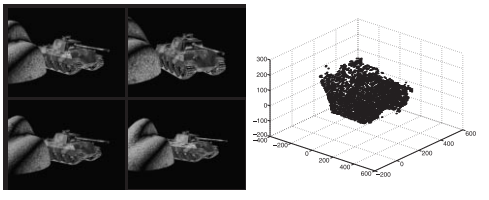

Fig. 4. (left) Four images from a sequence of ten images of a partially occluded Panther tank and the 3D data of the reconstructed tank (right).

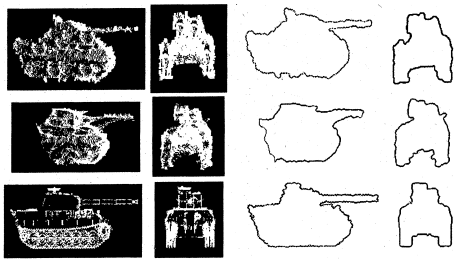

Fig. 5. The side and front projections (left) and their outer contours (right) of the reconstructed Panther tank without occlusion (top), without occlusion (middle) and its CAD model (bottom).

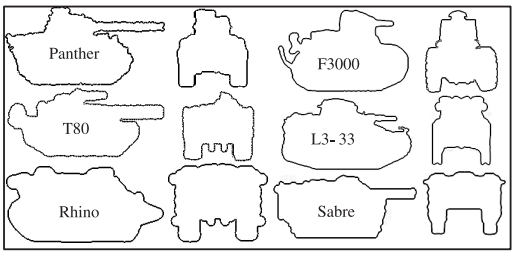

Fig. 6. The contours of some CAD models obtained from their projection on the side and front views.

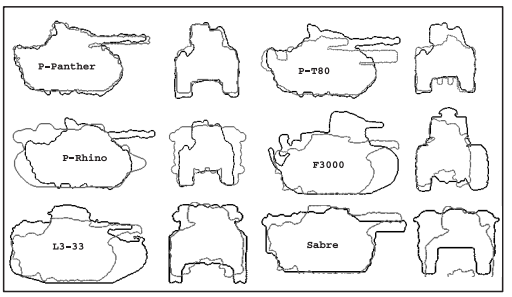

Fig. 7. The side and front contour matching of the reconstructed (black) Panther (P) tank with the contours of the CAD models (gray).

contours of the reconstructed tank with the contours of the CAD models. Fig. 8 shows the matching results of the reconstructed occluded tank with the contours of the CAD models.

To quantify the contour matching results, we calculated the best match between the reconstructed data and the models in terms of the percentage of the number of pixels of the two contours that are close to each other within a threshold distance. Table 1 shows these percentage measurements. The percentage shown in the table is the average value of those obtained by matching the model to the image and matching the image to the model. As seen from the table, the best match is between the reconstructed Panther and its model. The matching results also show that filtering the reconstructed 3D data with a median filter slightly enhances the recognition. The matching results of the partial contours of the occluded tank to those of the CAD models are shown in Fig. 8. The percentage of matched pixels between the partial contours and the model contours are also given in Table 1. The table shows that our matching algorithm is also robust in partial contour matching, where the best matches are obtained between the partial side and front contours with the contours of the Panther CAD model.

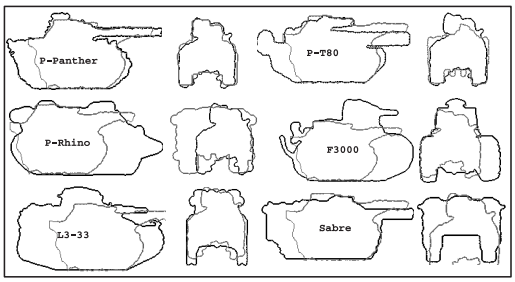

Fig. 8. The side and front contour matching of the occluded Panther (P) tank (black) with the contours of the CAD models (gray).

Table 1
Matching percentage between the contours of the CAD Models and the reconstructed Panther (P) tank with and without occlusion[a]

Target model	No occlusion		Occlusion	
	SV	FV	SV	FV
P-Panther	86.2	74.4	61.4	65.7
P-T80	52.9	57.4	46.4	53.1
P-Rhino	34.2	42.8	39.6	40.9
P-Sabre	24.9	21.7	44.4	33.3
P-F3000	17.8	26.3	34.9	40.3
P-L3	20.3	19.3	36.2	56.2

[a] SV and FV refer to side- and front-views, respectively.

6. Conclusion and future work

We presented in this paper an ATR system which is based on the 3D reconstruction of the target using a sequence of intensity images. The reconstruction process is performed using a modified voxel coloring algorithm. Instead of using a 3D–3D matching algorithm, which produces erroneous measurements in case of partial surface matching, we base the target recognition process on matching the target's front and side contours generated from the 3D reconstructed model to the selected contours from a model-library. A subpixel contour matching algorithm that supports rotation and translation is used in the matching process. Since we generate the contours from the 3D data, this approach is also invariant to target scaling. The results show the precision of the reconstruction and matching algorithms. Quantitative measurements are performed by measuring the percentage of matched pixels between the reconstructed contours and the model contours. Partial contours due to occlusion in the scene are also considered. Despite the incompleteness of the contours, the matching results show the best match is between the partial contour and the model of the same tank.

Future work will focus on the application of this technique to real data. In particular, we will examine the fusion of stereo-based reconstruction with the voxel coloring technique in order to speed up the search. In addition, we will investigate the fusion of data from other imaging modalities that may be available with certain ATR applications.

Discussion

Gimel'farb: My question is not about tanks, although tanks are very interesting things. My question is about the 3D reconstruction. How do you take possible occlusions into account in your voxel colouring technique?

Farag: This is a very nice problem and, quite frankly, it is not solved yet, especially if the occlusion is in all images (views).

Gimel'farb: In 1997, the work by Robert Haralick and myself was published in the Proceedings of the CAI conference in Kiel. Of course, it is a very hard problem, I do agree with you, but some approximate solutions based on so-called confidence maps exist. If you want, I can give you the exact reference to this publication. (*Note of the editors*: *G.L. Gimel'farb*, *R.M. Haralick*. *Terrain Reconstruction from Multiple Views*. *Proceedings of the Seventh International Conference on Computer Analysis of Images and Patterns* (*CAIP97*), *September 1997*, *Kiel*, *Germany*. *Lecture Notes in Computer Science*, *Vol. 1296*, *Springer*, *Berlin, 1997*, *pp. 694–701*). Our experiments were not with tanks, but with RADIUS imagery of a model-based scene.

Farag: Very good, I will be more than happy to get your papers.

Gimel'farb: One more question: why do you think that all the stereo approaches are too slow?

Farag: Well, they are too slow if you have so many images to do correspondence, and if the image size is big. And in a military situation, when you need almost real-time interactivity, then computation time becomes important. Of course, if there is no camouflage, and the target is small, then you could expedite the process a little bit. I should also say that here I have chosen very simple cases. Usually, a target can be an entire landscape that has tanks, cities, etc. So the word target here is also fuzzy. That is why the image can be extremely big and the correspondence problem can definitely be a nightmare. Incidentally, in our laboratory, we have a supercomputer, so we have no problem with computation time. But when you have an unmanned air vehicle and a closed loop military environment, then computation time becomes a problem.

Gimel'farb: I can show you this stereo algorithm on my laptop. To accelerate the reconstruction, it could be implemented in hardware, while the 3D reconstruction technique by voxel colouring is time-consuming and has only a software implementation. So, for real-time implementations, some stereo techniques could be better than multi-view reconstruction.

Farag: I agree, if you have some components micro-programmed in a chip, then this would expedite the process. I should also add that stereo would not be able to provide a dense depth map as achieved by the technique reported in this paper.

Kanal: I would like to refer you to some work done by my student Stockman, who is now a professor at Michigan State University. Many years ago, I think it was 1978, we developed a system for recognition of occluded objects in a bin of parts. He had a paper published on this technique.

Farag: I would be more than happy to read your earlier work. One final thought: there are two publications related to this, both in 1997. One is by Ratches et al. (*Note of the editors*: *see* (*Ratches et al., 1997*) *in this paper*), giving a survey of the advances in ATR systems from the US Department of Army point of view. The other is a special issue on ATR, which is in IEEE Transactions on Image Processing, also in 1997. (*Note of the editors*: *see* (*Bhanu et al., 1997*) *in this paper*). These are very good sources to give you more updates on ATR.

Acknowledgements

This work has been partially supported by grants from the DoD under contract No. USNV N00014-97-11076. M.G. Mostafa is on leave from the Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt.

References

Bhanu, B., Dudgeon, D.E., Zelnio, E.G., Rosenfeld, A., Casasent, D., Reed, I.S., 1997. Special issue on automatic target recognition. IEEE Trans. Image Processing 6 (1).

Bhatnagar, R., Horvitz, R., Williams, R., 1997. A hybrid system for target classification. Pattern Recognition Letters 18 (11–13), 1399–1403.

Faugeras, O., 1993. Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press, Cambridge, MA.

Miller, M.I., Grenader, U., O'Sullivan, J.A., Snyder, D.L., 1997. Automatic target recognition organized via jump-diffusion algorithm. IEEE Trans. Image Processing 6 (1), 157.

Ratches, J.A., Walters, C.P., Buser, R.G., Guenther, B.D., 1997. Aided and automatic target recognition based upon sensory inputs from image forming systems. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (9), 1004.

Seitz, S.M., Dyer, C.R., 1997. Photorealistic scene reconstruction by voxel coloring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Santa Barbara, Califonia, June 1997, IEEE Computer Society, p. 1067.

Serra, B., Berthod, M., 1997. 3D model localization using high-resolution reconstruction of monocular image sequences. IEEE Trans. Image Processing 6 (1), 175.

Stevens, M.R., Beveridge, J.R., 1997. Precise matching of 3D target models to multisensor data. IEEE Trans. Image Processing 6 (1), 126.

Yamany, S.M., Farag, A.A., Shin-Ye, H. 1999a. A fuzzy hyperspectral classifier for automatic target recognition (ATR) systems. Pattern Recognition Letters, 20 (11–13), 1431–1438.

Yamany, S.M., Ahmed, M.N., Farag, A.A., 1999b. A new genetic-based technique for matching 3D curves and surfaces. Pattern Recognition 32 (10), 1817.