Proceedings of the 3rd Asian Conf. on Computer Vision, ACCV'98,
LNCS 1351, Vol. 1, pp. 648-655, Hong-Kong, January 1998.

A Bayes Nets-Based Prediction/Verification Scheme for Active Visual Reconstruction

Éric Marchand, François Chaumette

IRISA / INRIA Rennes
Campus de Beaulieu
35042 Rennes-cedex, France

Abstract. *We propose in this paper an active vision approach for performing the 3D reconstruction of polyhedral scenes. To perform the reconstruction we use a structure from controlled motion method which allows an accurate estimation of primitive parameters. As this method is based on particular camera motions, perceptual strategies able to appropriately perform a succession of such individual primitive reconstructions are proposed in order to recover the complete spatial structure of complex scenes. The algorithm described in this paper is based on the use of a prediction/verification scheme managed using decision theory and Bayes nets. It allows the visual system to get a more complete high level description of the scene.*

1 Overview

Our goal is to obtain a complete and precise description of a scene using the visual data provided by a controlled camera mounted on the end effector of a robot arm. The idea of using active schemes to address vision issues has been introduced a few years ago. Since the major shortcomings which limit the performance of vision systems are their sensitivity to noise, their low accuracy and their lack of reactivity, the aim of active vision is generally to elaborate control strategies for adaptively setting camera parameters (position, velocity,...) in order to improve the knowledge of the environment. In our application, the purpose of active vision is handled at two levels: a **local aspect** where active vision is used to constrain the camera motion in order to improve the quality of the reconstruction results, and a **global aspect** which is used to ensure full scene reconstruction.

The method we have used to estimate the 3D structure of the primitives assumed to be present in the scene is fully described in [3]. It is based on the measure of the camera velocity and the corresponding motion of the primitive in the image. More precisely, we use a "*structure from controlled motion*" method which consists of constraining the camera motion in order to obtain a precise and robust estimation of 3D geometrical primitives. When no particular strategy concerning camera motion is defined, important errors in the 3D structure estimation can be observed. This is due to the fact that the quality of the estimation is very sensitive to the nature of the successive camera motions. An

active vision paradigm is thus necessary to improve the accuracy of the estimation results by generating adequate camera motions. Indeed, it has been shown that two vision-based tasks (called *fixation* and *gazing* tasks) have to be realized in order to obtain a robust and non-biased estimation. In this paper, we restrict ourselves to polyhedral objects. The only considered primitives are thus 3D segments, which must appear centered and vertical (or horizontal) in the image during the camera motion.

Since the proposed structure estimation method involves fixating at and gazing on the different primitives in the scene, this can be done on only one primitive at a time, and reconstructions have to be performed in sequence for each primitive of the scene [7]. Our incremental strategy leads to an exploration process which is handled at two levels:

- When a new primitive appears in the camera field of view, it is estimated. In that case, we do not need to compute explicitly new viewpoints. This level is called **local exploration**. It allows to split the observed areas into free-space and reconstructed objects and to bridge the gap between a local model of the scene in terms of isolated 3D segments and a global model in terms of objects (segments, junctions, and polygons).
- When a local exploration ends, a more complex strategy is used in order to gaze on parts of the 3D space which have not been already observed. This level is called **global exploration** and is described in [7].

This paper deals with the local exploration strategies. They are based on a prediction/verification scheme which is described in Section 2. Finally, we present real-time experimental results carried out on a robotic cell in Section 3.

2 A Bayes Nets-Based Prediction / Verification Scheme

Let us consider the scene depicted on Figure 1.a. The model obtained using a simple incremental reconstruction algorithm is given in Figure 1.b. We can notice that the 3D model is composed of five segment which *a priori* came apart, a few segments have not been taken into account because of their small size, and a long segment has not been estimated (because it was always occulted).

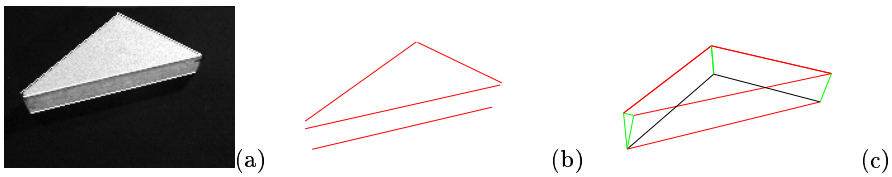


Fig. 1. Polyhedral scene: (a) view of the scene, (b) model of the "polyhedron" scene acquired using an incremental algorithm, (c) model of the same scene acquired using the prediction/verification scheme

To cope with these problems, we propose a prediction/verification scheme based on the use of Bayesian networks which allows us to obtain a high level

description of the scene. As an example, the method proposed afterwards allows us to complete the model of the object depicted in Figure 1.a as shown on Figure 1.c.

2.1 Prediction/Verification Scheme and Bayes Nets

Prediction/verification approaches intend to solve the problem of the agreement between models and data. In our application, uncertainty appears either in the 3D data acquired using the structure from motion approach or in the extraction of segments in images. Mainly, the consequence of this uncertainty is the confrontation of different possible alternatives for guiding the reconstruction and the exploration of the scene. The goal of decision theory is to provide well defined and mathematical approaches for making a decision in presence of uncertainty. Bayes nets [8] seem to be well adapted to our problem. They allow us to model "expert" reasoning. They are adapted to the automatic generation of action while performing this reasoning. Thus we can directly introduce perception strategies within the scene interpretation process. Bayes nets have been already used in computer vision (*e.g.* [1], [10]). Using Bayes nets in active vision is more recent. Most discriminant works have been proposed by Rimey and Brown [9] with the TEA-1 system (selective perception for visual search), Buxton and Gong [2] (traffic analysis), and Djian and Rives [4] (for object recognition). The goal of these systems, including ours, are obviously different ; however, the realization of the task always requires the execution of perception actions such as sensor motions.

Bayes nets allows us to represent join probabilities distributions of a set of variables using a set of *a priori* knowledge on the relations between these variables. A Bayes net is a directed acyclic graph where nodes represent the discrete random variables and where links between nodes represent the causality between the variables. Such a net can be used to represent the knowledge available on a particular domain. The graph structure and the *a priori* knowledge introduced in the graph (as conditional probability tables) must be defined in function of the application. The advantages of Bayes nets lies in the ability to reflect the *a priori* knowledge available on the application. The knowledge is reflected at two levels: first in the structure of the net through the nature and the number of nodes (variables), the different states of these variables and the relations (links) between these variables; second in the conditional probability tables associated with the variables of the net and which reflect the expert reasoning. These tables also model the uncertainty associated with the observations. Finally, the propagation allows us to take each new observation into account. The influence of an observation is propagated to the other variables of the net according to the causality relations.

2.2 Our approach

The information available on the scene is composed by a set $\mathcal{S}(\mathcal{T}_0^{t-1})$ of 3D segments. $\mathcal{S}(\mathcal{T}_0^{t-1})$ is a subset of $\mathcal{O}(\mathcal{T}_0^{t-1})$ which represents all the known objects of the scene (*i.e.* 3D segments but also junctions, polygons, etc.). The goal is to

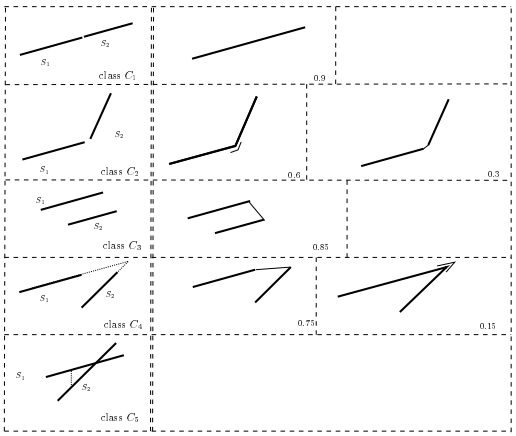
determine the relations between segments and to infer either the presence of new segments either the existence of more complex objects. As our reconstruction is incremental, we have to determine the consequence of the introduction of a new segment S_t in $\mathcal{S}$. Therefore, this module is used each time that a new segment is introduced in $\mathcal{S}$.

Our approach can be decomposed into three steps. For each couple of segments $(S_{t'}, S_t), t' \in [0, t-1]$, we propose hypotheses on the relation between these two segments. Then, we verify if these hypotheses match the observations. Finally, the system propose a new model of the scene resulting from the integration of the new segment.

Prediction Dealing with two segments $S_{t'}$ and S_t, the possible actions are the followings: fuse the segments, create a junction, or add a link (a new segment) between $S_{t'}$ and S_t. Therefore the aim of the prediction step is to create some hypotheses leading to the realization of one (or more) of these actions. The hypotheses are directly linked to the actions:

- H_1: there is a junction between $S_{t'}$ and S_t ;
- H_2: there are one or two segments between $S_{t'}$ and S_t.
- H_3: $S_{t'}$ and S_t are identical ;
- H_4: there are no (or some other) relation between $S_{t'}$ and S_t.

We have a multi-step strategy. First, we compute the belief we have in simple topological relations between $S_{t'}$ and S_t (proximity ($p(N)$), coplanarity ($p(C)$), and collinearity ($p(P)$)). Then, according to these beliefs, it is possible to classify the pair of segments into five classes (see the first raw of the Table in the Fig. 2). Classes are $\mathcal{C}_1$: CNP (coplanar, neighbor and parallel) , $\mathcal{C}_2$: $CN\neg P$, $\mathcal{C}_3$: $C\neg NP$, $\mathcal{C}_4$: $C\neg N\neg P$, and $\mathcal{C}_5$: $\neg C\neg N\neg P$.



	Classes				
hypotheses	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
H_1 fusion	0.90	0.025	0.05	0.05	0.025
H_2 junction	0.025	0.60	0.05	0.15	0.025
H_3 string	0.025	0.30	0.85	0.75	0.025
H_4 other	0.05	0.075	0.05	0.05	0.925

closest extremities

	Class				
hypotheses	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_5$
H_1 fusion	0.025	0.025	0.025	0.025	0.025
H_2 junction	0.025	0.025	0.025	0.025	0.025
H_3 string	0.025	0.5	0.7	0.5	0.025
H_4 other	0.925	0.45	0.25	0.45	0.925

distant extremities

Fig. 2. Elementary classes and associated hypothesis (closest extremities), and conditional probabilities table P(hypotheses | classes) for the closest and distant extremities

Using the belief we have in the belonging of the couple of segments to each class, the system can infer the belief in each possible hypothesis. We have defined decision strategies which are able to determine the best hypothesis according to the available knowledge. These strategies are coded in conditional probability tables $P(\mathcal{H}|C)$ where $\mathcal{H}$ is the hypothesis and C the class (see Fig. 2). These tables are defined in an empirical way from a set of elementary considerations about topological relationship that we usually find in a group of segments. These considerations often reflect the truth, though they provide no guarantee. However, extreme precision is not required. Rather, they must reflect the knowledge we want to transmit to the system.

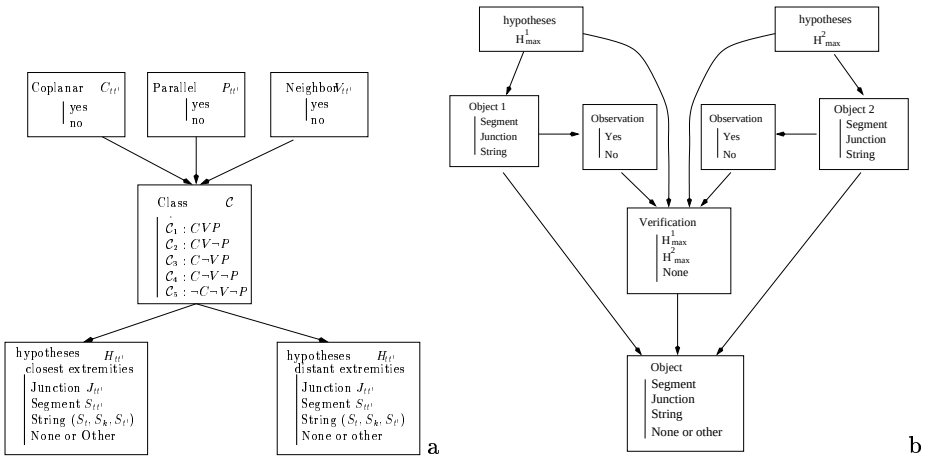


Fig. 3. (a) Prediction net, (b) Verification net

The prediction step reasoning can be encoded in a simple Bayes net (see Fig. 3.a). It is composed of six nodes. Links between these nodes depict the causality relations between the different steps of reasoning and thus its progression. One node is associated to each topological relation, another to the class, and one node is associated to each set of hypotheses. Indeed, two sets of hypotheses are emitted. The first concerns the relation between the closest extremities of the segments (see Fig. 2) and the second concerns the relation between their distant extremities. In both cases, the same hypotheses can be emitted, though the associated conditional probabilities can be very different. As already stated, the hypothesis with the higher belief is not always the correct one, and this is the reason why we always consider for each case (closest and distant extremities) the two hypotheses with the highest belief (H^1_{max} and H^2_{max}). These two hypotheses are then verified or invalidated.

Verification In order to verify the two selected hypotheses, we use the reasoning encoded in the Bayes net depicted in Fig. 3.b. We use two similar nets, each associated with one of the two sets of hypotheses (*i.e* close and distant extremities). Considering the two hypotheses, we first define the nature (segment,

junction, string) and the position of the created object associated with each hypothesis. Then, we compute the belief in the existence of this object using the observation node. Finally, knowing the belief in each hypothesis and the belief in the related observation, it is possible to determine the most probable hypothesis (or to reject both).

The most important node in the verification net is the observation node. Sometimes, the hypotheses can be verified (or invalidated) using direct observation in the images previously acquired. In such cases, the validation is performed using the 3D information associated with the hypotheses and the 2D observation. We perform a back-projection of the 3D objects in each image previously acquired by the camera and we try to associate this projection to the observed data in more than one image (to avoid false matching). For each possible matching, we compute the belief granted to this matching. The case of a single segment or of a junction is simple. If this junction exists, it has already been observed (because the presence of the two segments has been already verified). Thus, the verification is performed as described above. In the case of a string, with three segments, the presence of two of them is certain (they have been used to predict the presence of the third). However, the last one has not been yet reconstructed (most of the time), and its presence is not validated. When no matching is found in the images previously acquired, it is necessary to know why. The first possibility is that the segment under consideration does not exist, the second is that it is occluded by another object. In the latter case, it is necessary to move the camera to a new viewpoint from which the segment can be observed. Rather than computing explicitly a viewpoint (*e.g.* [11]) and researching *off-line* the considered segment, we prefer to turn the camera around a segment which belongs either to the occluding polygon or to a plane to which the considered segment belongs. During this motion, automatically generated by visual servoing [5], an image processing is performed *on-line* to detect the appearance of the researched segment.

Modeling. At this step of the reconstruction process, we have a model of the scene composed of 3D segments, 3D junctions, or even a coplanar string of segments. It is finally quite easy to use this information in order to get 3D polygons. To this end, we use the junction information and the coplanarity information already used in the hypotheses generation (see [6] for further details). This three-step approach allows us to get a high level and more complete representation of the scene.

3 Experimental results

We present in this section the reconstruction results obtained for a polyhedral object (see Fig. 4.a). This scene allows us to illustrate the interests of the proposed method. As already stated, as they are too small, some of the vertices of the polyhedron may be not reconstructed using a simple incremental reconstruction process. Furthermore, due to the local approach used in that process, others remain occluded and thus non reconstructed. We now focus on two aspects of the Bayes nets prediction verification scheme.

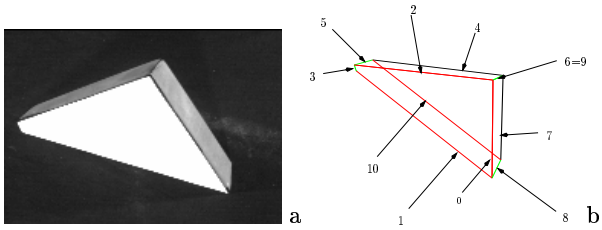


Fig. 4. Polyhedral scene: (a) view of the scene, (b) model of the same scene acquired using the prediction/verification scheme and numbering of the reconstructed segments in the order of their introduction in the 3D map

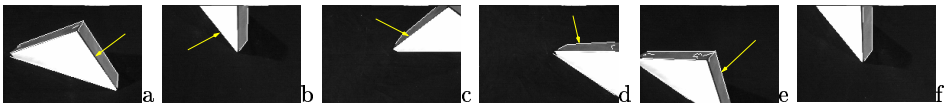


Fig. 5. Polyhedral scene: arrows point at the next primitive to be estimated

Consider that segments S_0 and S_1 have been already estimated and that S_2 has just been reconstructed (see Fig. 5.abc), the system considers the relation between S_2 and S_0 and between S_2 and S_1. Dealing with S_2 and S_0, the system concludes easily to the presence of a junction between them. Dealing with the couple (S_1,S_2), there is around 1cm between their closest extremities. The belief for S_2 and S_1 to be neighbor is 61% and to be coplanar is 99% ; thus they are likely to belong to the class $\mathcal{C}_2$. According to the strategies encoded in the hypotheses Bayes net, it is likely that there exists a junction with a 46% belief and a a segment between them with a 41% belief. The remaining 13% are shared between the two other hypotheses. After the verification process, and according to the observations, the former hypothesis (junction) is verified with a 60% belief. This high value (even if this hypothesis is false, see Fig. 4.a) results from the fact that these two segments are very close in the different images (around 5 pixels). Thus the observations reinforce this hypothesis. However, the latter hypothesis is verified with a 95% belief. Indeed, a 2D segment is observed at the predicted position in many images. Finally, according to the belief in each hypothesis, to the belief in the observations, a new segment S_3 is added to the model of the scene (with a confidence of 53%, while the confidence in a junction creation is only 37%). This underlines the interest to consider a multi-hypotheses approach. A classical approach might have chosen the first (and wrong) hypothesis.

Let us now consider a second interesting case. When segment S_7 is reconstructed, within relations with other segments, the system proposes the creation of a junction with S_4 and the creation of a segment between their two distant extremities. Such a segment has never been observed (and could not have been observed according to the current knowledge on the scene and on the camera trajectory). Therefore, as described in the previous section, the camera gazes on S_7, and turns around it (see Fig. 6). During this motion, automatically generated by visual servoing, observers are looking for a moving segment located at its expected position in the images. The discovered segment is then reconstructed and introduced in the scene model (see Fig. 6.c).

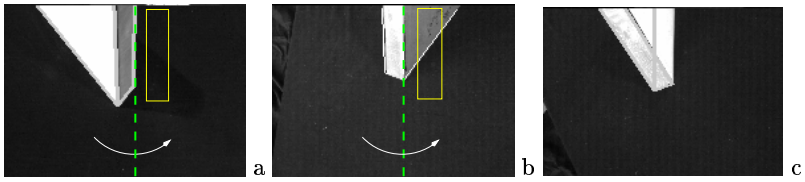


Fig. 6. Verification of a hypothesis: (a) rotation around S_7, (b) S_{10} is discovered, (c) and reconstructed

4 Conclusion

The main goal of the prediction/verification approach was to bridge the gap between a representation in terms of isolated primitives and a representation in terms of objects. Sub-goals were to deal with small segments and to propose a partial solution to the occlusion problem. Although the considered scene and images are quite simple, the resulting system is able to perform a complete and accurate reconstruction of these scenes using images acquired and processed at nearly video rate. Finally, experiments carried out on a robotic cell have proved the validity of our approach.

References

1. J.M. Agosta. The structure of Bayes nets for vision recognition. In *Proc. of 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, Aug. 1988.
2. H. Buxton, S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78(1-2):431–459, Oct. 1995.
3. F. Chaumette, S. Boukir, P. Bouthemy, D. Juvin. Structure from controlled motion. *IEEE PAMI*, 18(5):492–504, May 1996.
4. D. Djian, P. Probert, P. Rives. Active sensing using Bayes nets. In *Proc. of Int. Conf. on Advanced Robotics, ICAR'95*, pages 895–902, Sant Feliu de Guixols, Spain, Sep. 1995.
5. B. Espiau, F. Chaumette, P. Rives. A new approach to visual servoing in robotics. *IEEE Trans. on Robotics and Automation*, 8(3):313–326, June 1992.
6. E. Marchand. *Stratégies de perception par vision active pour la reconstruction et l'exploration de scènes statiques.* PhD thesis, Univ. of Rennes 1, IRISA, June 1996.
7. E. Marchand, F. Chaumette. Controlled camera motions for scene reconstruction and exploration. In *CVPR'96*, pages 169–176, San Francisco, June 1996.
8. J. Pearl. *Probabilistic reasoning in intelligent systems : Networks of plausible inference.* Morgan Kaufmann Publisher Inc., 1988.
9. R.D. Rimey, C. Brown. Control of selective perception using Bayes nets and decision theory. *IJCV*, 12(2/3):173–207, Apr. 1994.
10. S. Sarkar, K. Boyer. Integration, inference, and management of spatial information using Bayesian networks: perceptual organization. *IEEE PAMI*, 15(3):256–274, Mar. 1993.
11. K. Tarabanis, P.K. Allen, R. Tsai. A survey of sensor planning in computer vision. *IEEE Trans. on Robotics and Automation*, 11(1):86–104, Feb. 1995.