Towards a coarse real-time stereo fusion based on triple junctions

M. GONZALO-TASIS[1], J. FINAT[2]

[1]Dept. of Computer Science, E.T.I.T Campus Miguel Delibes,
47011 Valladolid, University of Valladolid, Spain
marga@infor.uva.es
[2]Faculty of Sciences, University of Valladolid, Spain
jfinat@infor.uva.es

Abstract. In this paper we develop a hybrid method to establish the correspondence between common triple junctions appearing in two views of the same scene taken by an uncalibrated bicameral heading with known disparity between both cameras and under constant illumination conditions. After a segmentation process, we get a set of edges. Beginning from theses edges, we extract different types of junctions (↑, T, L or Y). These geometric features are robust even when the noise arising from the real scenes appears. We provide a very fast process of comparison between points placed at a similar depth in real scenes. Next, we apply a labeling relaxation procedure to determine possible homologues between two images. Decision criteria are based on probabilistic methods, allowing us an easy extension under uncertainty conditions or under incomplete information due to partial occlusions or noise.

Keywords: Computer graphics, Image processing, Stereo Vision

1 Introduction

Geometric data are present as elementary primitive almost in every image. Our approach is based on locally linear representation of the scene which is captured from a binocular stereo system, and our efforts are concentrated in provide a fusion procedure between both views as soon as possible.

Even in absence of mobile objects, experimental psychophysics studies show that saccadic movement of the eyes fix their attention in vertices or corners of objects appearing in real scenes. In the human case, this selective attention is due to the neurons of the visual system must rapidly adapt to stationary stimuli, allowing to fix reference points. These reference points are usually some kind of vertices. We propose to use triple junctions as candidates because they provide visual cues which are stable in real scenes, but also for oculomotor perception; hence, their 3D reconstruction from data contained in two views give still stable points, too. In addition, in a experimental way, it is possible to verify how these points are still standing after applying different kinds of filtering.

This approach based on triple junctions provides an artificial basis for simulating the stereopsis (the appearance of three dimensional image) based on retinal disparity (see [Tov94]). Damages associated to perception appear as related with failures in depth perception, spatial orientation or grouping tasks (holes in scenes, e.g.). All of them are located in different parts of the cerebral cortex. Thus, it seems reasonable to try of separating all tasks related to spatial attention and motion when we propose to develop the artificial models of visual perception and justifies our emphasis in to isolate and improve methods based in the ”mise en correspondence”. From the knowledge of disparity, we introduce the notion of related regions in different views as those appearing overimposed (up to a tolerance factor) after a translational motion in one of the images. This approach is initially based on metrical properties relative to proximity relations between possible homologues located at related regions (in [SB92],[SLH91]). This euclidean viewpoint is complemented with a probabilistic one, where we use our own method like maximum likelihood method to discriminate possible homologues. The neck-bottle is still linked to the right evaluation of depth.

We have followed a classical approach for the segment extraction ([Can86]), based on the method of gradient (by using recursive Deriche’s filters[Der90]). Intersection between segments give us information about the localization of corners. Junctions are given by these corners where two or three segments coalesce. Double junctions are labelled as T or L , whereas triple junctions are labelled as Y or ↑, depending on discontinuities for two or three segments, respectively. Next, one selects the triple junctions appearing in the left view and we match them together with those appearing in the right image. Even under deficient illumination and reflectance conditions, one obtains a sufficient number of triple junctions allowing us to make a coarse comparison between pairs of images. After choosing the meaningful double or triple junctions, we connect these junctions between them, by using those segments (previously stored in a file) arising or arriving to these vertices. In this way, we obtain grouping criteria for related regions in both views which are robust ones for changes in orientation, elongation and scale. Problems related to the identification of 2D regions appearing in both views under rotations and scale changes are considered in [Per92].

This constitutes the core of our work because of its efficiency and accuracy. This selection involves the rejection of those vertices which are not paired due to noise, mismatching of partial occlusion.

Main advantages of our approach is to achieve a very fast localization of triple junctions (instead of using a more complete information about the scene), to increase the efficiency in lowering the execution time of the comparison phase between both views and the accuracy of the ”mise en correspondence” under complete information conditions for the disparity of stereo system and the depth of scene. In the way, we are able of evaluating how small variations in the depth give errors in pairing procedure. Inversely, if we have previous information about the scene, these errors provide information about scale factor.

About this paper, in the following sections, we review the general basis of the matching process and the labeling relaxation method . This way, we introduce our scheme and its methods associates.

2 Searching for correspondences

In stereo vision, the recognition problem is one of searching for a match, that is, how to associate some features of one image with corresponding features in the other image. Matching (in [Gri90]) implies finding correspondences between two different images taken by the same scene.

We have developed a matching algorithm which is based on the selection of a little set of matchings through topological and metrical constraints. Furthermore, we have added new constraints in relation to the image’s characteristics chosen. The purpose of these restrictions or constraints is to obtain a set (as reduced as possible) of possible homologous for each given vertex. Novelty of our approach consists of relaxing by means probabilistic criteria some of these classical constraints; this is in the same vein as it appears recently in [Kan97]. If 3D scene is ”simple”, the process based in triple junctions analysis provides fundamental information for the segmentation process and consequently, for pose evaluation of objects in the scene. Moreover, we need to define criteria to choose the correct homologous element in the search region, so that we do not match wrong junctions. Selected criteria are: similarity in the length and slope characteristic, in the type of junction and in the spatial characteristic.

The search process consists of looking over all triple junctions in the left image, looking for those junctions (triple or not) in the right image. In every case, we must discount the disparity effect (horizontal disparity). Disparity effect comes from the distance between optic axes of heading’s cameras.

2.1 2D Constraints

In our approach, main problem is to provide sufficient criteria for an efficient search process. Difficulty arises from the amount of elements where we must develop the searching process. Then, we need to implement constraints to reduce this search, by avoiding wrong elements to be compared. Another said, our goal is to reduce the number of possible homologous of a given vertex. First, each triple junction in left image has as possible homologous all the list of vertices (double or triple) in the right image S^D.

After applying filters that we will describe in next section, we obtain a set of matching hypotheses to each triple junction Vt_i^I. Then:

$$H_{Vt_i^I} = V_{l,i}^D, ..., V_{u,i}^D \qquad (1)$$

2.1.1 Spatial constraint

To match triple junctions we delimitate which region in the right image is restricted to search to. This region is bounded with two ranges: a horizontal range and a vertical range, because we suppose that optics axes of cameras are approximately aligned in a parallel way. Considering vertex in the right image V_j^D as a possible homologous in relation of triple junctions in left image Vt_i^I

$$iff\ V_j^D\ \cap Region\,(Vt_i^I)\ \neq\ 0 \qquad (2)$$

This restriction is operative to compare those triple junctions locating in a similar relative depth in relation of camera position. Experimentally, we can show scenes where it is rejected as possible homologous, junctions placed in virtual planes moved away from reference plane that, nevertheless, it corresponds to images acquired simultaneously with a stereo heading; then, it was necessary determine an threshold for relative depth.

In analysed scenes, we have opted for a compromised solution based in comparing triple junctions placed in next region from a fixed plane orthogonal to the optic axis.

2.1.2 Slope and Length constraint

First , slope and length constraint eliminates those junctions whose edges have not a similar slope and length in relation of ones in the reference image. But we have detected that slope segments in triple junctions can have important modifications. For this reason we need additional information about the type of junction.

2.1.3 Type of Junction constraint

This restriction is about if character double or triple of a junction is preserved. We have proved , experimentally, that if ratio between camera disparity and depth is minimum, the junction character have not variation, that is, double junctions like T-junctions or L-junctions are applied in T-junctions or L-junctions respectively, and the same happens with triple junctions. But, if this ratio is greater , it can happen that Y-junctions evolution-ate to ↑-junctions going through to T-junctions.

3 Classification Criteria for Labeling Relaxation

Features as junctions (common or not) are codified as labels to help us in verification and comparison process. Relaxation (as in [HS92]) is any computational mechanism in parallel processes which can act locally (each one on each image), and can be actualized iteratively with a labeling process, until obtaining consistent interpretation of data which appeared in the image. Usually, labeling schemes are probabilistics, where each feature are labeled with a weight or probability.

In general, there are two methods to relaxation : maximum likelihood (M.L.) and minimum distance (M.D.) classification. Maximum likelihood classification is the most commonly method for relaxation labeling. It is based in Bayesian approaches for Computer Vision (or for Stereo vision as in Belhumeur[Bel94]). Using the Bayes paradigm, we seek to extract scene information from an image, or sequence of images, by balancing the content of the observed image with prior expectations about the content of the observed scene. Maximum likelihood efficiency depends on the estimation of mean and variance of a set of samples for each class. The process consists in comparing the value of each point with the estimation of mean and variance of each class to know if this point belong to this class.

If it was no possible to have enough samples, it will be better to choose another classification which not bear in point of variance and depends on the mean, like minimum distance classification.

Minimum distance classification is based in discriminant function of samples'means. It is quicker than Maximum likelihood but it need also samples to compute the value of the mean for each class. The process is similar than maximum likelihood .

We have developed another approach for the relaxation labeling. Our interest is in distinguish between a set of elements of a class which have similar characteristics and we need select only one of these elements. In the class, there will be only possible homologous of a triple junction.

Thus we have a set of classes, ω_i , $i = 1,..,M$ where M is the total number of classes. Each class has a generator which represents i-esim triple vertex in the left image χ_i , $i = 1,..,M$ Therefore, let us m_{ij} , $j = 1,..,N$, where $m_{ij} \in \omega_i$, is the possible homologous (i.e. double or triple junctions in the right image) of the class ω_i.

Then, we said that m_{ij} is the homologous of χ_i iff

$$d(\chi_i, m_{ij})^2 < d(\chi_i, m_{ik})^2 \;\; \forall j \, \forall k \; with \; j \neq k \qquad (3)$$

The decision to develop a new method of labeling relaxation is based in the next reasons. First, as we said, we have taken real images with deficient illumination (non Lambertian illumination), this fact implies that some points of important objects was not in the image for partial occlusions or some inexistent points was added because they belonged to object's shadows. These data is not useful to compute the mean and the variance for the M.V. or the M.D. methods.

Second, we don't need samples for each class (as M.V. and M.D. need) to begin this relaxation method.

Third, mainly M.V. and M.D. process points or edges. One of our images have between 150-200 edges. Our relaxation method is based in junctions. Junctions appear in less quantity in the same image than edges (i.e. between 40-50 junctions). The difference in these amounts of information implies a notable difference in the time of process.

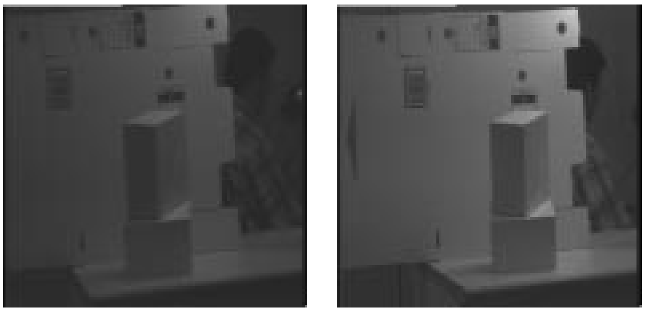

Figure 1: Left and Right Image of a Scene

4 Experiments on real scenes

We have two images of the same scene, i.e. right and left image, taken with a stereo bicameral heading (see *Figure*1). Then, we analyse these images with a HP-workstation. While extracting junctions in both images, we identify their type and character analysing the intersection between edges.

Only triple junctions are interesting in left image (taken as reference's image) , while in the right image we are interested in double and triple junctions. The reason is that we do not know if triple junctions in the left image preserve their character (double or triple) in the right image.

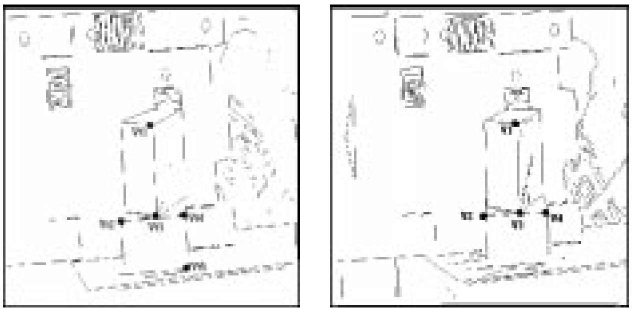

Figure 2: The triple junctions in the left Image and their Homologues in the right Image

Then, we apply the matching algorithm to select a little set of matchings for each triple junction in left image . Those junctions (double or triple) in the right image which have similarities (constraints) in length, slope, type of junction and spatial with one in the left image, are selected as homologous. After applying the match algorithm, a set of possible homologous of each triple junction in the left image is obtained. In order to know which is the best homologous of each set of possible homologous, it is necessary to do a labeling relaxation process.

Thus, in the left image (see *Figure*2), there is a set of triple junctions $\{$Vt1,...,Vt5$\}$ and their homologous in the right image are $\{$V1,...,V4$\}$.In [XZ96] is explained why is so important to recognize four elements in stereo vision. In *Figure*2 we can see that Vt5 vertex has no homologous in the right image. The reason is that there is a shadow area that prevent the matching process.

Experimentally, we have verified in the same workstation and with the same couple of images that our approach for a real-time stereo fusion gets homologous in seconds while the process with maximum likelihood takes minutes.

5 Conclusions

In this paper, we have described another approach for the matching and relaxation process in stereo vision using junctions extracted beginning from edges and classified by type of junction (i.e. arrow, T, L and Y).

The main result is that these new algorithms process junctions in images with noise, partial occlusions and deficient illumination and establish the correspondence between junctions that appear in two views of the same scene.

The major advantage of these algorithms is their simplicity and speediness even though they need a previous process to extract junctions.

It should make them attractive in a variety of image processing applications as in [FGT98].

References

[Bel94] Peter N. Belhumeur. Bayesian models for reconstructing the scene geometry in a pair of stereo images. Technical report, Harvard University, 1994.

[Can86] J. Canny. A computational approach to edge detection. *IEEE PAMI*, 8(6):679–689, 1986.

[Der90] R. Deriche. Fast algorithms for low level vision. *IEEE PAMI*, 12(1):78–81, 1990.

[FGT98] J. Finat and M. Gonzalo-Tasis. Fast recognition of postures for a simplified three-fingered artificial hand. In Proceedings IEEE-SMC San Diego, 1998.

[Gri90] W. Grimson. *Object Recognition by computer*. The MIT Press, 1990.

[HS92] R. Haralick and L. Shapiro. *Computer and Robotic vision*. Addison-Wesley, 1992.

[Kan97] K. Kanatani. Statistical optimization and geometry visual inference. In G. Sommer and J.J. Koenderick, editors, *Algebraic Frames for the perception-action cycle*, pages 306–322. Springer-Verlag, 1997.

[Per92] P. Perona. Steerable-scalable kernels for edge detection and junctions analysis. In G. Sandini, editor, *Computer Vision, ECCV'92, Santa Margherita Ligure, Italy*, pages 4–18. Springer-Verlag, 1992.

[SB92] L. Shapiro and J.M. Brady. Feature based correspondence and eigenvector approach. *Image and Vision Computing*, 10:283–288, 1992.

[SLH91] G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two images. In *Proc. Royal S. London*, volume 244, pages 21–26, 1991.

[Tov94] M. Tovee. *An introduction to the Visual System*. Cambridge University Press, 1994.

[XZ96] Gang Xu and Zhengyou Zhang. *Epipolar Geometry in Stereo motion and Object recognition, A Unified Approach*. Kluwer Academic Publisher, 1996.