A PYRAMIDAL STEREOVISION ALGORITHM BASED ON CONTOUR CHAIN POINTS

Aimé Meygret, Monique Thonnat, Marc Berthod
INRIA Sophia Antipolis, 2004 route des Lucioles
06565 Valbonne cedex, France

Abstract

We are interested in matching stereoscopic images involving both natural objects (vegetation, sky, reliefs,...) and man made objects (buildings, roads, vehicles,...). In this context we have developed a pyramidal stereovision algorithm based on "contour chain points." The matching process is performed at different steps corresponding to the different resolutions. The nature of the primitives allows the algorithm to deal with rich and complex scenes. Goods results are obtained for extremely fast computing time.

Introduction

The fundamental problem of stereovision is matching homologous visual characteristics extracted in several images of the same scene observed from different view points. These visual characteristics are also called images primitives. Marr and Poggio [Grim81] have noted that the difficulty of the correspondence problem and the subproblem of eliminating false targets is directly proportional to the range and resolution of disparities considered and to the density of matchable features in an image. It is also crucial for the subsequent triangulation process to localize these primitives very accurately. Thus, many stereo matching algorithms have been developed. Multi-resolution approaches provided an efficient way to limit the complexity of the matching process [Marr79], [Hann84]. In these algorithms ambiguous matches are solved by enforcing the continuity of disparity in the neighborhood of ambiguous points. However, the continuity constraint is no more available when the neighborhood used for resolving ambiguities, crosses an occluded contour. Mayhew and Frisby [Mayh81] partially solved this problem by enforcing continuity of disparity along edges in the image. This constraint which they called "figural continuity" is more realistic than the surface smoothness assumption since, while disparity varies discontinuously accross surface boundaries, it rarely varies discontinuously along such a boundary. Grimson [Grim85] implemented a new version of his earlier algorithm, incorporating this constraint to eliminate random matches.
Kim and Bovik [Kim86] have also used the continuity of disparity along contours for both disambiguation and matching control. In a first step they match extremal points (high

curvature edge points, terminations and junctions) by enforcing continuity of disparity along contours. They check the consistency of the disparity at a matched extremal point by examining disparities at its neighbours which have been matched and propagate the disparities along the contours in order to match other edge points. Their algorithm gives good results but it has apparently only been tested on indoor stereo images with few objects.

We think that this constraint is an important continuity law stereo consequence and we have tried to include it in a pyramidal stereovision algorithm.

We can also facilitate the matching process by using richer primitives such as edge segments [Medi85], [Long86], [Ayac85], regions [Wrob88],...

Within Prometheus European Project we are interested in three-dimensional localization of obstacles in road scenes. The nature of images we deal with (diversity and complexity of shapes) made us choose contour chain points as primitives. For a pedestrian which is a typical obstacle, it is very difficult to extract linear features (as edge segments) because of the smoothness of the surface of human body. On the other hand it is difficult to match chains of contours for two reasons at least: the chain geometry description difficulty and problems involved by chains cutting out management. Moreover the chain as an entity is not localizable very accurately (same problem with regions). As against this, contours chains points are primitives with rich information, so easy to match, are very accurately localizable and are suitable to describe any type of scene, man made, natural or mixt. Moreover, contour chain points provide a studied three-dimensional world more complete description.

The hierarchical structure of Prometheus images (important objects, details), the existence of strong disparities and time processing constraint (for obstacles detection we must be able to take a decision very quickely) made us choose a coarse-to-fine approach.

Presentation of the method

Data structure: We use a pyramidal image data structure in which the search for objects starts at a low resolution and is refined at ever increasing resolutions until one reaches the highest resolution of interest. We consider a pyramid at the four highest resolution levels. The consolidation is made in a 2x2 neighborhood. We have chosen a pyramidal approach rather than a classical multi-resolution approach essentially for time computing reasons. In a pyramid of resolution the image size is reduced by the consolidation process.

Primitives: Contour chain points extraction is performed in three steps: gradient computation [Deri87], hysteresis thresholding to eliminate noise and contour points chaining [Gira87]. The contour chain is used on the one hand to eliminate false targets and on the other hand to propagate the disparity. The three-dimensional world description provided by contour chains is richer and facilitates the further recognition process.

Matching: Matching of the two stereo images is performed by optimizating a similarity function. For two contour points, (x_l , y_l) in the left image and (x_r , y_r) in the right image, we define the similarity function as:

$$f(x_l, y_l, x_r, y_r) = \frac{[G(x_l, y_l) - G(x_r, y_r)]^2}{S_G^2} + \frac{[\theta(x_l, y_l) - \theta(x_r, y_r)]^2}{S_\theta^2}$$ where

$G(x, y)$ and $\theta(x, y)$ respectively design the gradient norm and orientation at (x, y) point.
S_G and S_θ are respectively thresholds on the gradient norm and orientation difference.
For all (x_i, y_i) in the left image (x_j, y_j) is a potential matching in the right image if:

- $|G(x_{li}, y_{li}) - G(x_{rj}, y_{rj})| \leq S_G$
- $|\theta(x_{li}, y_{li}) - \theta(x_{rj}, y_{rj})| \leq S_\theta$
- $f(x_{li}, y_{li}, x_{rj}, y_{rj})$ minimum with regard to j

(x_{rjo}, y_{rjo}) is a potential matching;
The pair $((x_{li}, y_{li}), (x_{rjo}, y_{rjo}))$ is validated if $f(x_{li}, y_{li}, x_{rjo}, y_{rjo})$ is minimal with regard to i. The matching process is then symmetric, and uniqueness is guaranteed. Matched points should have similar properties, of course, because they are both projections of the same surface point , but in many cases there will be ambiguous. So we have to determine criterions allowing to decide which matches are correct. The three-dimensional spatial continuity of real world surfaces constrains the two-dimensional spatial distribution of disparity in the image plane. This is the second stereo law. The continuity of disparity over most of the image can be used to avoid false matches based on similarity alone, by suppressing matches in the absence of supporting local evidence. We have chosen contour chains as local support to check the matches consistency. Edge points along the chain belong to the same surface, except when the edge crosses an occluding contour: we have then to stop the chain; we avoid this problem by using Giraudon chaining algorithm which cuts chains when it finds triple points. It may happen that the occluded contour gradient is too weak to be detected or is eliminated as noise. The disparity continuity assumption along the chain does not hold, but we can limit this occurence with a coarse to fine algorithm. As the search of a potential matching is hierarchically governed, we can choose very low thresholds at the step of hysteresis thresholding and so increase the number of primitives without too much complicating the matching process. We can generally consider that the disparity varies smoothly along the chain. Then we avoid the problem of the region overlapping an occluding contour.
For each pair of matched points we determine the disparity vector (its norm L, and its orientation β) and we check the vectors continuity along contour chains. A couple i of matched points is validated if:

$$\sum_{j \in V_i} \frac{\frac{|L_i - L_j|}{(L_i + L_j)} \frac{1}{|i - j|^n}}{\sum_{j \in V_i} \frac{1}{|i - j|^n}} \leq S_L \quad \text{and} \quad \sum_{j \in V_i} \frac{\frac{|\beta_i - \beta_j|}{(\beta_i + \beta_j)} \frac{1}{|i - j|^n}}{\sum_{j \in V_i} \frac{1}{|i - j|^n}} \leq S_\beta$$

V_i : i neighborhood along the chain.
S_L et S_β are thresholds, n define the neighborhood size.

The hierarchy in the pyramid: After consolidation, contours (local maximas of the first derivate) are extracted on each stereo image and then, matched at finer and finer resolutions. The disparity information obtained at a given resolution is used to specify the search space for finding a matching point at a finer resolution. So we limit the matching algorithm complexity by controlling at each step, the number of primitives and the search window size. At the first step (coarsest resolution) the search window size is defined by the minimum and maximum depths estimated in the image. At the following steps it is defined by the size of the filter for the extraction of contours. The use of epipolar geometry permits us to reduce strongly the correspondence problem dimension: the search windows are reduced to strips along epipolar lines. At each step we interpolate the disparity along the chain. The disparity at a chain point is given by the average of the neighbours disparities weighted by the inverse of the distance between the current point and its neighbours. The disparity interpolation along the chain gives us a richer disparity map to specify the search space of a potential matching at finer resolution. At a given resolution, the search space of a potential matching is obtained by searching at the nearer coarse resolution the nearer neighbour which has been matched.

Results

We have treated two very different Prometheus scenes in order to test the algorithm robustness. The images were taken using two CCD cameras mounted parallel on top of a car with a height of about 1.60 m and a distance of 40 cm between the cameras. The first stereo pair (figure 1) is a countryside scene and is characterized by a lot of discontinuities. The figure 2 represents a part of the reconstructed scene in a three-dimensional space. The car profile, the dividing line and the post on the right appear clearly. There are two important sources of errors in the matching process:

- the lack of precision in cameras calibration which has constrained us to use epipolar bands and not epipolar lines
- matching errors which arise at horizontal lines due to the cameras relative geometry; we expect to limit this problem by swinging the cameras support in an appropriate direction.

For this scene, though we have chains, we have only represented points in three-dimensional space because few false matches (which often arise at horizontal lines) are sufficient to blur the reconstructed scene. The program was coded in C and implemented in SUN4 110. For 512x144 stereo images, the consolidation process, the edges detection and the chaining take less than 1 minute for the two images. At the last step 1445 points have been matched and after the disparity interpolation along chains we had 2522 points. The matching process (including interpolation) takes less than 30 secondes.

fig. 1: The stereo images (countryside scene).

fig. 2: The reconstructed scene, in a 3D space

The second scene takes place in town (many discontinuities in the scene) and shows a typical obstacle: a cyclist. We present in figure 3 the stereo images.The figures 4a, 4b and 4c represent the reconstructed chains in a three-dimensional space viewed by an observer turning around the scene. The software displays thready chains, so according to the view point some chains are seen by transparency; a surface modelisation could cope with this problem. In the figure 4a we clearly distinguish the file of cars on the left, the middle road line mark and the cyclist. As the observer moves towards the left, in the figure 4b, the cyclist appears very clearly. In the figure 4c we still distinguish the cyclist; the road right boundary and a parking car appear.

fig. 3: The stereo images (town scene).

fig. 4a: The reconstructed chains in a 3D space; the observer has moved slightly towards top right.

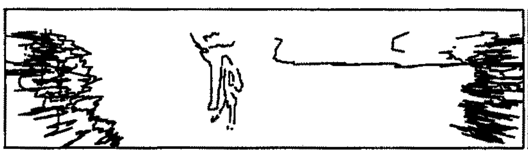

fig. 4b: The reconstructed chains in a 3D space, front view

fig. 4c: The reconstructed chains in a 3D space; the observer has moved slightly towards top left.

The presented multi-resolution approach gives very interesting results as well for the fullness of the reconstructed information as for the quality of this information. We have now to deal with three-dimensional data to provide a three-dimensional description of the environment seen from the car.

Conclusion

We have presented a matching stereovision algorithm reliable enough to provide a three-dimensional description of the environment seen from the car. The hierarchy permits to avoid aberrant matches by matching main structures before details.
Using a symetric similarity function garantees non ambiguity according to the stereo unicity law. The second stereo law, the continuity, is checked when validating the matches. Using a chain contour as a local support for consistency makes the method indifferent to the types of handled scenes (man made, natural) since the contours density and nature (occluded or not) does not affect the method and so makes it more general than methods adapted to robotic scenes [Ayac85] or to natural scenes [Grim81] .
Finally, a pyramidal approach reduces greatly the computing time (by a factor between 3 and 4) with regard to classical multi-resolution approach. Although this algorithm permits to work with rich scenes (involving man made or natural objects), its computing time is nevertheless extremely fast.

Acknowledgements
We would like to thank Gerard Giraudon and John Fairfield for many valuable discussions and Professor Nagel who provided us the stereo images.

References

[Ayac85] N. Ayache and B. Faverjon, "Fast Stereo Matching of Edges Segments Using Prediction and Verification of Hypotheses", CVPR, 1985.

[Deri87] R. Deriche, Optimal Edge Detection Using Recursive Filtering, Proceedings of of the first International Conference on Computer Vision (ICCV), pp. 501-505, Londres, 1987.

[Gira87] G. Giraudon, "Chainage rapide sur des images de contour," Rapport de recherche INRIA, 1987.

[Grim81] W. E. L. Grimson, "From Images to surfaces," MIT Press, Cambridge, 1981.

[Grim85] W. E. L. Grimson, "Computational Experiments with a Feature Based Stereo Algorithm," IEEE Trans. PAMI, vol. PAMI-7, pp. 17-34, January 1985.

[Hann84] M. J. Hannah, "Decription of SRI's Baseline Stereo System," Technical Note No.342, Artificial Intelligence Center, SRI International, Oct 1984.

[Kim86] N. H. Kim and A. C. Bovik, "A solution to the Stereo Correspondence Problem Using Disparity Smoothness Constraint," Proc. of IEEE Conf. Systems, Man, and Cybernetics, Atlanta, October 1986.

[Long86] P.Limozin-Long, "Stereo Matching Using Contextual Line Region Primitives," ICPR Paris, Oct 1986.

[Marr79] D. Marr and T. Poggio, "A theory of Human Stereo Vision", Proc. R. Soc. Lond.' vol. B 204, pp. 301-328, 1979.

[Mayh81] J.E.W. Mayhew and J.P. Frisby, "Psychophysical and computationnal studies towards a theory of human stereopsis," Artif.Intell., 17, pp. 349-385, 1981.

[Medi85] G. Medioni and R. Nevatia, "Segment-Based Stereo Matching", Computer Vision, Graphics, ang Image Processing, vol. 31, pp. 2-18, July 1985.

[Wrob88] B. Wrobel-Dautcourt, "Coopération entre l'extraction et la mise en correspondance symbolique de régions," MARI Paris, pp 181-188, 1987