Cooperative Segmentation and Stereo Matching

Laurent Vinet Peter Sander Laurent Cohen

André Gagalowicz

INRIA
Domaine de Voluceau-Rocquencourt, B.P.105
78153 Le Chesnay Cedex, France
tel: 39 63 54 08

Stereo matching based on points and lines as the entities has become a well developed industry. In this paper, we investigate region based matching as we feel that many of the shortcomings inherent in other approaches can be overcome by taking more developed entities (to cite but two examples: mismatches over pairs of line elements are to be expected frequently due to the lack of distinguishing features; and occlusion can effects points or segments more than regions). We present a novel approach to the combined problem of image segmentation and object matching, based on interaction between a segmentation component and a stereo component. We believe the combination to be better than the parts taken individually. This paper is limited to segmentation/matching results—we are not concerned with the explicit depth computations here (but see [1]).

The basic idea developed in this paper is that, since objects in the world give rise to events in both stereo images , segmentation in each image should be carried out in conjunction with segmentation in the other, thus, hopefully, producing a more reliable segmentation in both. Of course, a 'vicious circle' arises in that cooperative segmentation presupposes matching, and matching is dependent on a prior segmentation. We propose breaking the circle by iteratively using partial segmentation results to suggest tentative matches, which then feed back into the segmentation procedure, and so on.

Thus, segmentation (by merging and splitting regions) in one image depends on what matches have been found with the other. Some of the computation can be done independently, however, prior to any matching: a number of (candidate) segmentations of the images are computed for a range of parameters and organized in a tree structure so that merging/splitting regions just amounts to moving up/down in the tree. The complete procedure consists of two steps (see [2]for more details):

1. computing, independently for each image, fine to coarse hierarchical candidate segmentations;
2. determining a 'final' segmentation from among the candidates, cooperatively with region based stereo matching between the images.

Segmentation

Following [3], let a segmentation $S = \{R_1, R_2, \ldots\}$ of a set E be defined by a predicate P: S is a partition of E; $P(R_i)$ is true for all i; if $i \neq j$ then $P(R_i \cup R_j)$ is false. A hierarchical segmentation is a sequence $S_0, S_1, \ldots, S_n$, where each level S_i is a segmentation defined by a predicate P_i and which contains the previous S_{i-1}, i.e., $\forall R \in S_{i-1},\ \exists \bar{R} \in S_i$ such that $R \subset \bar{R}$. Note that each segmentation level may result from the successive application of several predicates, P_i^j, $j = 1, 2, \ldots, n_i$, say (see also [4]).

Both images are first segmented independently into a hierarchy of candidate, or potential, segments $S_0, S_1, \ldots, S_n$, but with no commitment to any particular one. Segmentation proceeds 'upwards' (fine to coarse) from an initial level by merging neighbouring regions satisfying homogeneity conditions.

To effect a substantial reduction in the number of initial regions, we begin with regions created by standard quadtree operation [5]. Contour information [6] is already incorporated at this level, so that blocks are not merged when they are separated by sufficiently strong edge elements.

We build the hierarchy by successive use of the region-growing algorithm. Each new level S_i is obtained from the previous one S_{i-1} by merging adjacent regions which satisfy a condition derived from the predicate P_i. The parameter of the predicate is then progressively relaxed, permitting more permissive merges, and the resolution of the segmentations of the hierarchy moves from fine to coarse.

A segmentation depends on the order of the merges. To avoid having the order depend on an arbitrary image traversal strategy, we carry out the merges in order of increasing cost, according to the appropriate predicate. Adjacent regions R_i^l, R_j^l are merged into one region whenever the predicate P_k^l is true, where, given a cost-of-merging function C_k^l and threshold t_k^l,

$$P_k^l(R_i^l, R_j^l) \equiv \left(C_k^l(R_i^l, R_j^l) < t_k^l \right).$$

Cooperative matching

We stress the distinction between computing the graphs representing multiple levels of segmentation of the images (step 1 above), which is done independently in each, and commitment to a particular set of regions as the resulting segmentation (step 2) , which is done *while* stereo matching.

The region based stereo matching associates regions in the left graph with regions in the right which are likely to be images of the same physical object. Matching may occur across levels of segmentation, since image formation parameters can differ, and the same segmentation parameter is not guaranteed to give similar results in both images. Beginning at the the top (coarsest) level of segmentation, region L of the left image matches region R of the right whenever $\max_{\hat{R} \in \Lambda} s(L, \hat{R}) = R$, where Λ is the set of regions of the right image eligible to match L , and $s(L, R)$ a measure of similarity between regions (a weighted average of resemblance functions using region intensity and geometry attributes). The set Λ could, in principle, be all regions of the right image, but when we are given the geometry of the cameras, we can restrict Λ to regions whose centre of gravity is 'close' to the epipolar of the centre of gravity of L. If L at some level fails to find its match in the other image, then its descendents in the segmentation hierarchy are added to the list of regions to be matched. Whenever a match is found, then both matched regions and all their descendents are no longer open for matching. Pairs are considered in order and removed form the list until the measure of similarity between the next pair falls below a given threshold. Matching stops when there is no remaining pair of sufficiently similar admissible matches—it is also only at termination of the segmentation/matching that we consider a definitive segmentation of the both images to have taken place (see Figs. 1–3).

Figure 1: Original stereo pair.

Figure 2: Levels of segmentation of the left image(fine to coarse from upper right to lower left).

Conclusion

Our approach to stereo image analysis is based on three tenets, which address the basic problem of how to make use of as much image information as possible. First, image segmentation and matching should not be independent sequential processes. There is information

Figure 3: Matching regions.

in each image relevant to the analysis of the other, and this should be incorporated into the segmentation as well as the matching. Second, regions possess more structural information which is stable to small changes of viewpoint than do edges or points. Hence, we expect to make more stable matches by taking regions as the primitive elements. Third, and related to the previous point, edge- and region-based methods are naturally complementary, and should be used together for segmentation; neither should be considered as an end in itself. We feel that our results demonstrate the promise of this approach.

References

[1] Michel Peyret and André Gagalowicz. Reconstruction 3D å‘ partir de couples de régions-image appariées. In *Proceedings of Pixim*, Paris, October 1989. In preparation.

[2] Laurent Cohen, Laurent Vinet, Peter T. Sander, and André Gagalowicz. Hierarchical region based stereo matching. In *Proceedings of Computer Vision and Pattern Recognition*, San Diego, June 1989. To be presented.

[3] S.L. Horowitz and T. Pavlidis. Picture segmentation by a directed split-and-merge procedure. In *Proceedings of the Second International Joint Conference on Pattern Recognition*, pages 424–433, 1974.

[4] André Gagalowicz and Olivier Monga. A new approach to image segmentation. In *Proceedings of the Eighth International Conference on Pattern Recognition*, Paris, October 1986.

[5] Hanan Samet. The quadtree and related hierarchical data structures. *Computing Surveys*, 16(2):187–260, June 1984.

[6] Rachid Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 167–187, 1987.