

Robust Region-Based Stereo Vision to Build Environment Maps for Robotics Applications*

M. Angeles López and Filiberto Pla

Departament d'Informàtica. Universitat Jaume I
Campus Penyeta Roja, s/n
E-12071 Castelló, Spain
e-mail: {lopeza, pla}@inf.uji.es

Abstract. Stereoscopic vision is an appropriate tool for building maps of the environment of a robot. When matching regions of the images, segmentation errors should be avoided. In this paper an algorithm to deal with errors in region matching is proposed, and the results in the presence of noise are analyzed. The selection of an appropriate similarity criterion to create the initial nodes in the graph-based matching process is very important for reducing the time of computation considerably. The experimental results show that the method is robust in the presence of noise.

1 Introduction

Many applications in robotics need a map of the robot environment. Stereo vision can be used to construct such a map, given that it provides three dimensional information about the scene: the distances from the point of view to the viewed objects.

One of the most difficult problems in stereo vision is to find the corresponding point for a given feature point. It is called the *correspondence problem*. Several matching techniques have been developed, which differ in the imaging geometry, the matching primitives, or the matching strategy. For a survey on stereo vision techniques see [2] [4] [5].

Feature-based approaches establish correspondences between features extracted from the images. Most of the existing approaches use edge-based features as primitives for matching [7] [11]. There are also some works on the use of regions [9] which have a higher semantic content compared with edge segments or points.

The use of regions make some of the matching constraints implicit or easier to introduce, and there are typically a lower number of regions than edges. Therefore, establishing image correspondences would be more efficient and the

* This work was partially supported by grants TAP95-0710 (CICYT, Ministerio de Educación y Ciencia) and GV-2110/94 (Consellería de Educació y Ciència, Generalitat Valenciana).

number of mismatches would be reduced. However, region-based matching provides a very sparse disparity map. So, region matching may be the first step in a hierarchical stereo system [9].

A graph-based method for dealing with segmentation errors in region matching was developed in [8]. The method proposes a preprocessing stage to the classical association graph method where appropriate merged regions are constructed. Here, we present an improvement to the method, where a new similarity criterion is proposed in order to obtain a measurement that reduces the number of false matchings, and thus, the computational cost (section 2.2). The experimental results in presence of noise are presented and the resulting disparity maps are compared in order to analyze the robustness of the method.

1.1 Previous Work

Region-based techniques have to be able to handle segmentation errors:

- *Partly occluded regions*: regions that appear in both images, but they have different shapes due to partial occlusion, noise, etc.
- *Occluded regions*: regions that appear only in one image.
- *Splitted regions*: some adjacent regions in one image correspond to one or more regions in the other image. This problem is also known as oversegmentation and undersegmentation in the area of object recognition.

Some papers try to solve these problems using graph-based methods [7] [1]. Many of these methods have been also applied to object recognition [12] [3], where the matching is performed between image regions and parts of an object model.

The usual technique [12] is to construct an *association graph* where a node represents a mapping between a region in the left image and a region in the right image, and an arc between two nodes represents compatibility between the two mappings.

Each node may be assigned a weight to indicate the similarity of the regions to be matched, and each arc may be assigned a weight to indicate the compatibility between two mappings. Therefore, the problem of finding the best match can be achieved by finding the best maximal clique (the maximal clique whose node weights sum is maximal). Several improvements to the basic method have been developed, like using a relaxation algorithm [11] [12] which incorporates contextual information into each node in order to avoid erroneous weights.

2 Solving segmentation errors

2.1 The incompatibilities association graph

Let the left image consist of N regions (L_1, L_2, \dots, L_N) and the right image consist of M regions (R_1, R_2, \dots, R_M) . The first step consists of finding all the possible mappings between regions in both images, using the *epipolar constraint*

to reduce the search space and some *similarity criterion* to measure node weights and reject pairs of non-similar regions.

The second step consist of finding all the incompatibilities between possible mappings. Two nodes are incompatible if they try to match the same region. One splitted region will have at least one incompatibility with each splitted region corresponding to the same region. However, not all the incompatibilities in the graph lead to pairs of regions that should be merged.

The third step consist of rejecting all the incompatibilites that will not provide adecuate merged regions and stablsh arcs in the graph for the remaining incompatibilities. Let (L_k, R_p) and (L_k, R_q) be two incompatible nodes and $R_{(p,q)}$ a region resulting from joining R_p and R_q ,

$$\begin{aligned} \textit{if } S(L_k, R_{(p,q)}) > S(L_k, R_p) \quad \textit{and} \quad S(L_k, R_{(p,q)}) > S(L_k, R_q) \\ \textit{then } \textit{establish arc between node } (L_k, R_p) \textit{ and node } (L_k, R_q) \end{aligned} \quad (1)$$

where $S()$ is the criterion of similarity between two regions. In other words, we test every pair of incompatible mappings to see if they generate a merged region which matches to the common region better than each splitted region. If both regions together do not match to L_k better than separately, they should not be merged, and then the arc is not created.

The fourth step consist of finding all the maximal cliques in the incompatibilities association graph. Each maximal clique represents a group of regions in one image that match to a common region in the other image. For each maximal clique, a new region is constructed, as the union of all the single regions. The new regions will provide new possible mappings that have to be added to the graph. All the possible matches will compete in the matching process in order to give the best solution.

This preprocessing provides all the regions to be taken into account in the matching process, which can be for example the association graph technique. The experiments carried out so far show that the method provides satisfactory results with respect to the matching of splitted regions in one image to a whole region in the other one [8].

However, there was an important drawback. Although we used for the similarity criterion some region attributes (as mean intensity, first-order moments and second order moments) which are proved to be significant region features [10], some pairs of regions to be matched could have very different sizes and shapes, and the similarity criterion was not able to match for example one very little splitted region with one big region in the other image.

In order to avoid these regions left unmatched, we could consider a lower threshold in the creation of graph nodes. In this case, the number of nodes is incremented, which makes the cost of the matching process to be increased exponentially, due to the algorithm for finding the best maximal clique. In the other hand, we must say that also many false mappings were included in the graph and the matching process proved to be robust and selected always the good ones.

2.2 A new similarity criterion

The similarity measure should be able to

1. match partly occluded regions
2. match one splitted region with the whole region in the other image
3. match one merged region with its corresponding region in the other image better than each simple region separately, and
4. reject non matching regions.

The similarity criterion used so far, provided good measures in the first and third cases, but not in the second one, which is not well separated from the last case. We propose here a correlation-based criterion in order to achieve all the objectives.

Let us consider L_i a region in the left and R_j a region in the right. Both regions rarely match exactly, so that we have to slide one region over the other in order to localize the point of maximum correlation. The pixels to be correlated are those defined by the intersection of both regions at a given disparity d . Thus, the similarity measurement can be defined as the maximum *correlation coefficient* given a set of disparities:

$$S(L_i, R_j) = \max_{d \in [d_{min}..d_{max}]} (C_{ij}(d)) \quad (2)$$

where $C_{ij}(d)$ is the correlation between the intersecting areas of L_i and R_j at disparity d , and $[d_{min}..d_{max}]$ depends on the position and shape of both regions, and the disparity limits, if known.

Let us call $L_{i,j,d}$ and $R_{i,j,d}$ the intersecting areas of regions L_i and R_j at disparity d in each image. Let $\tilde{L}_{i,j,d}$ and $\tilde{R}_{i,j,d}$ be the mean intensity in each area and $\sigma_{L_{i,j,d}}$, $\sigma_{R_{i,j,d}}$ the standard deviations. The correlation measurement used in the experiments is the *Zero-Mean Normalized Cross Correlation* (ZNCC):

$$C_{ij}(d) = \frac{1}{2} \left(1 + \frac{1}{K} \sum_{\forall (x,y) \in L_{i,j,d}} (I_1(x,y) - \tilde{L}_{i,j,d}) (I_2(x+d,y) - \tilde{R}_{i,j,d}) \right) \quad (3)$$

where $I_p(x,y)$ is the intensity value of pixel (x,y) in image p , $K = N\sigma_{L_{i,j,d}}\sigma_{R_{i,j,d}}$, and N is the number of points in the area of intersection.

Other correlation techniques were tested (Sum of Squared Differences (SSD), Zero-Mean SSD) that resulted in poor results. Only the Zero-Mean Normalized SSD (ZMSSD) measurement resulted in results similar to ZNCC. The normalization given by the standard deviation of both intersected regions seems to be very important to achieve a significative correlation result, given that the number of correlated points is very different in each calculation.

However, the similarity criterion in (2) does not satisfy the last objective of the list. Some information about the size of the matched area should be introduced in the criterion. Let us introduce a *matching coefficient* that calculates the number of matched pixels with respect to the size of both regions. Let us define

$N_{ij}(d)$ and $N'_{ij}(d)$ as the matching coefficients with respect to the maximum and minimum size, respectively.

$$N_{ij}(d) = \frac{A(L_{i,j,d})}{\max(A(L_i), A(R_j))}$$

$$N'_{ij}(d) = \frac{A(L_{i,j,d})}{\min(A(L_i), A(R_j))}$$

where $A(L)$ indicates the number of pixels of region L .

If we maximize the product of the first matching coefficient with the correlation coefficient, we find the disparity value of best coincidence between both regions. Let us call it the *best disparity*,

$$d_{ij} = \operatorname{argmax}_{d \in [d_{min}..d_{max}]} (N_{ij}(d)C_{ij}(d)) \quad (4)$$

We propose a combination of two criteria:

$$S_1(L_i, R_j) = N_{ij}(d_{ij})C_{ij}(d_{ij}) \quad (5)$$

$$S_2(L_i, R_j) = N'_{ij}(d_{ij})C_{ij}(d_{ij}) \quad (6)$$

1. For creation of graph nodes, it is necessary to have a similarity measurement that satisfies requirements 1,2,4 of the list. For this purpose, the appropriate criterion would be S_2 (equation 6).
2. For comparison between nodes and all the operations in the matching process, it is necessary requirements 1,3,4 of the list. For this purpose, the appropriate criterion would be S_1 (equation 5).

Note that the use of two similarity measurements does not cause an important increment in the computational cost of the algorithm, given that the more expensive task is the searching of the best disparity (equation 4), which is done once per match.

A simple improvement to the computational cost of the best disparity calculation is to avoid calculating correlation at those disparities where the size of intersected area is small enough. As the term $C_{ij}(d)$ is less or equal to 1, the matching coefficient should be at least greater than the selected threshold for similarity, T , in order to be considered in the creation of a graph node. All disparities that satisfy

$$A(L_{i,j,d}) < T \min(A(L_i)A(R_j)),$$

will not provide a graph node, so we can omit all the involved calculations.

The results of some of the experiments are shown in section 3. These experiments showed that the selected criteria have a significative meaning:

- High values for $S_1(L_i, R_j)$ and $S_2(L_i, R_j)$ means that probably L_i and R_j are matching regions without need of merging other regions.
- Low values for $S_1(L_i, R_j)$ and $S_2(L_i, R_j)$ means that they are not matching regions.

- Low value for $S_1(L_i, R_j)$ and high value for $S_2(L_i, R_j)$ means that they may be corresponding regions but either some part of the region is occluded, or some splitted regions should be merged in order to achieve a better correspondence.
- High value for $S_1(L_i, R_j)$ and low value for $S_2(L_i, R_j)$ is not possible.

Therefore, we can summarize that $S_2(L_i, R_j)$ is mainly used for rejecting false matchings and $S_1(L_i, R_j)$ is an appropriate similarity measurement when comparing different matchings.

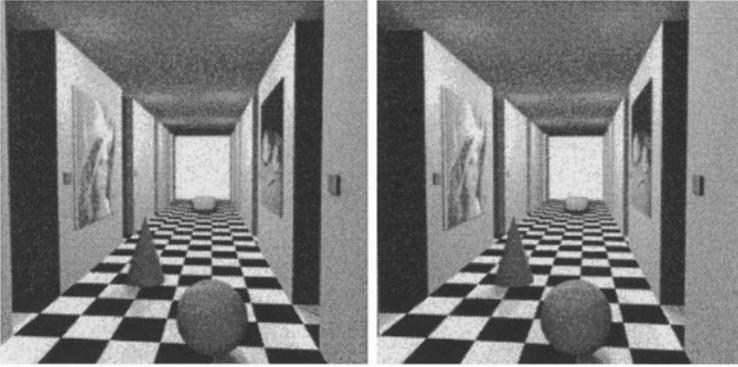


Fig. 1. Synthetic stereo image pair corrupted with noise of variance 100.

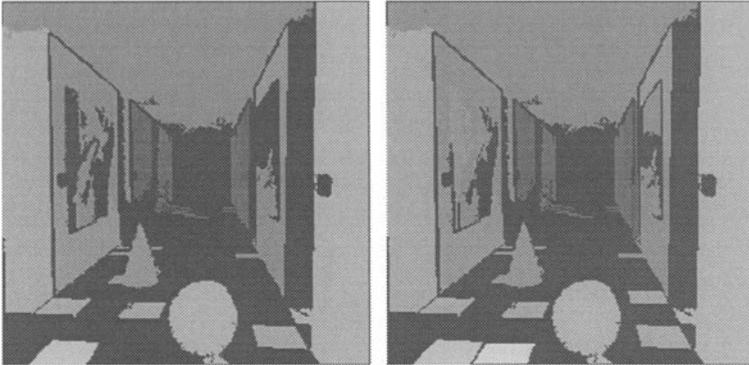
3 Experimental Results

Figure 1 shows one example of the stereo pairs that we have used in the experiments. It is a synthetic stereo pair of images corrupted with noise of variance 100 [6]. The method used to segment the images is a common region merging segmentation method [13]. The process of segmentation applied in both images is identical. However, several errors can be appreciated. Each region labelled in the left does not correspond to one region in the right. In some cases, one or more regions in the left correspond to one or more regions in the right.

Table 1. Percentage of matching and computational cost.

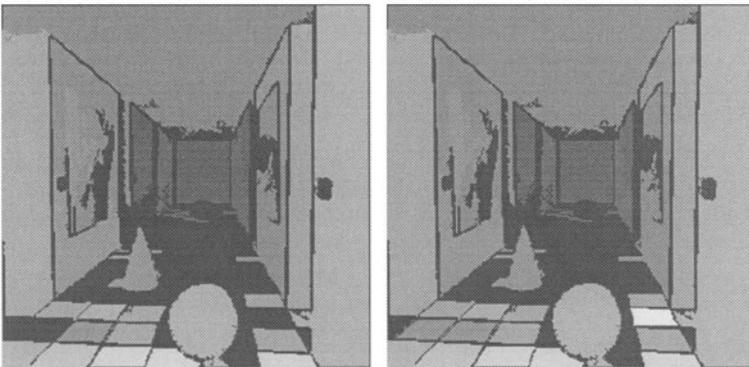
threshold	0.65	0.60	0.55	0.50	0.49
% matching	65.48	78.85	87.26	96.17	97.56
time	1m.35s.	1m.37s.	2m.12s.	4m.41s.	8m.06s.

Not all the regions are considered in the matching, given that very little regions provide a lot of false matchings, which would produce a huge graph. In figure 2(d), the black areas correspond to these not considered regions. Except for the black areas, darker areas correspond to further points in the scene.



(a) Similarity threshold 0.60

(b) Similarity threshold 0.55



(c) Similarity threshold 0.50

(d) Similarity threshold 0.49

Fig. 2. Disparity maps corresponding to the example in figure 1.

3.1 Evaluation of the results with different thresholds

Figure 2 shows the disparity maps produced by the method using different thresholds in similarity. The lower the threshold, the higher the number of possible matchings in the graph and the better the results. In the other hand, the

higher the number of graph nodes, the more expensive computationally, due to the searching of the best maximal clique.

For the example of figure 1, and using an HP 9000 SMP class K200, with 4 CPUs PA-RISK model 7200 (120 MHz) and 512M of RAM, the computational time of the different disparity maps is in shown in table 1.

Note the difference in time of computation from threshold 0.50 to 0.49. This is due to the searching of the best maximal clique, which is a problem of exponential complexity.

The percentage of matching is expressed as the rate of pixels matched with respect to the number of pixels of all the regions considered for matching. The number of matched regions grows when the threshold decreases, because more possible matchings are considered in the matching process.

The computational time of the whole process has been reduced considerably with respect to [8] given that most of the false matchings are rejected and not included in the association graph. This makes the algorithm of finding the best maximal clique much faster.

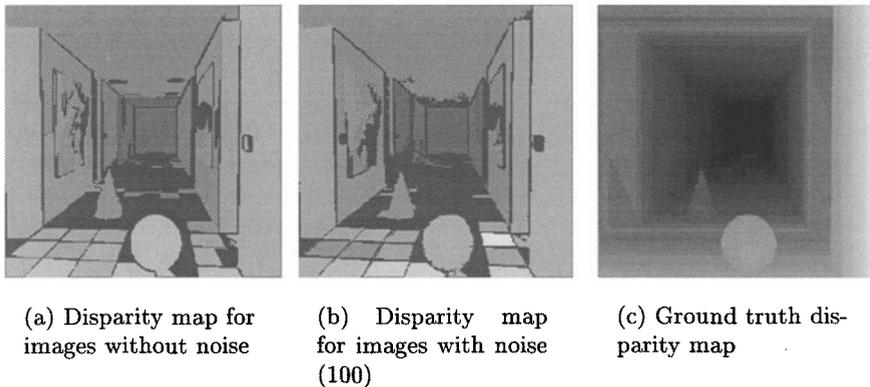


Fig. 3. Disparity maps corresponding to pairs with and without noise.

3.2 Evaluation of noise influence

In the experiments, we used stereo pairs with different variance in noise. For example, in figure 3(a) and 3(b), we show the results for a stereo image pair without noise, and another with noise of variance 100.

The resulting disparity maps are similar, in the sense that they provide an approximation to the solution, which is shown in figure 3(c). Although the results are not accurate, both maps give an idea of the position of the objects in the scene. Note that different gray levels have been selected in order to be able to distinguish the matched and unmatched area.

Table 2 shows the percentage of matching and time of each experiment. Note the values of the thresholds in similarity. Because of the noise, matching regions result in lower similarity values, and a lower threshold in similarity is needed in order to obtain similar results. That means that more nodes are included in the graph and the time of computation is greater.

Table 2. Percentage of matching and computational cost for different levels of noise.

noise variance	0	100
threshold	0.60	0.49
% matching	98.10	97.56
time	2m.5.74s.	8m.6s.

Given that we assigned constant disparity within each region, the resulting map shows only fronto-parallel planes. This constraint is also implicitly incorporated in the use of correlation. However, this approximation could serve for a number of robotics applications, where the needed map can be less accurate. This map gives a general idea of the location of objects in the scene. For applications that need accuracy, this map can be a good initialization for the matching of edges and points in the scene. Note that with this initial approach, the matching of edges should be much easier.

4 Conclusions and Further Work

A new improvement to the graph-based method for finding region correspondences has been proposed. The use of two similarity criteria, one for creating graph nodes and another for the rest of the matching process, provides less number of ambiguities, thus reducing the time of computation and producing more reliable region matches.

The experiments carried out show that the method provides satisfactory results, with respect to the segmentation problem. The construction of a graph with the incompatibilities between matches is a robust solution for merging regions in order to achieve better correspondences.

The results obtained so far can be used for a number of robotics applications which require low accuracy. They can also be used for initializing some matching process in a hierarchical stereo system in order to provide finer resolution.

Further work is directed to:

1. Perform the matching of edges and points.
2. Avoid the fronto-parallel constraint.

References

1. N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *IJCV*, 1(2), April 1987.
2. S. T. Barnard and M. A. Fishler. Computational stereo. *ACM Computing Surveys*, 14(4):553–572, December 1982.
3. Robert C. Bolles and Ronald A. Cain. Recognizing and locating partially visible objects: The local-feature-focus method. *IJRR*, 1(3):57–82, 1982.
4. Umesh R. Dhond and J.K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1489–1510, 1989.
5. Olivier Faugeras, Pascal Fua, Bernard Hotz, Ruihua Ma, Luc Robert, Monique Thonnat, and Zhengyou Zhang. Quantitative and qualitative comparison of some area and feature-based stereo algorithms. In Wolfgang Förstner and Stephan Ruwedel, editors, *Robust Computer Vision: Quality of Vision Algorithms*, pages 1–26. Wichmann, Karlsruhe, Germany, 1992.
6. Thorsten Froehlinghaus. Stereo images with ground truth disparity and occlusion. Technical Report <http://www-dbv.cs.uni-bonn.de/ft/stereo.html>, Computer Vision Group, Computer Science III, University of Bonn, 1997.
7. R. Horaud and Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. *IEEE Transactions on PAMI*, 11(11):1168–1180, 1989.
8. M. Angeles Lopez and Filiberto Pla. Solving oversegmentation errors in graph-based region matching. In *Proceedings of the 8th. Portuguese Conference on Pattern Recognition*, pages 387–394, Guimaraes, Portugal, March 1996.
9. Suresh B. Marapane and Mohan M. Trivedi. Region-based stereo analysis for robotics applications. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1447–1464, November 1989.
10. Joseph L. Mundy and Andrew Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, 1992.
11. N. M. Nasrabadi. A stereo vision technique using curve segments and relaxation matching. *IEEE Transactions on PAMI*, 14(5):566–572, May 1992.
12. Heggere S. Ranganath and Laure J. Chipman. Fuzzy relaxation approach for inexact scene matching. *IVC*, 10(9):631–640, November 1992.
13. A. Rosenfeld and A.C. Kak. *Digital Picture Processing*, volume 1. Academic Press, New York, 1982. Second Edition.