HIERARCHICAL FUSION OF COLOR AND DEPTH INFORMATION AT PARTITION LEVEL BY COOPERATIVE REGION MERGING

Felipe Calderero, Ferran Marques

Department of Signal Theory and Communications
Technical University of Catalonia (UPC), Barcelona, Spain
{felipe.calderero, ferran.marques}@upc.edu

ABSTRACT

A high level scheme for information fusion to create hierarchical region-based image representations based on a region merging process is presented. The strategy is based on an iterative evolution where the different merging criteria work independently and cooperate at the partition level to obtain a further consensus that increases the reliability of the resulting partitions. This cooperative scheme is applied to the creation of hierarchical region-based representations of the image based on color and depth information. The proposed technique is compared with approaches using only one source of information or linear combinations of both, in datasets with ground truth as well as estimated disparity information.

***Index Terms*—** Image segmentation, region merging, information fusion, median partition

1. INTRODUCTION

Image segmentation is a key step into image analysis. However, commonly, a unique solution for the image segmentation problem does not exist. To overcome this situation, a hierarchical segmentation approach can be used where, instead of a single partition, a hierarchy of partitions may be provided. An important type of hierarchical bottom-up segmentation approaches are region merging techniques. Starting from an initial partition or from all pixels, regions are iteratively merged until a stopping criterion is reached.

Classical merging criteria are based on color. Nevertheless, in general, color information is not enough to correctly segment natural images. For that reason, an increasing attention has been focused on adding new features to the merging process. Some researchers have proposed linear combinations of different region features. For instance, a linear combination of color and contour complexity is applied in [1] to obtain object-oriented hierarchies of partitions. Approaches based on a probability or belief framework have been also proposed [2]. Other segmentation approaches combine color and depth information, for instance, into a variational framework [3], or in the context of Markov random fields [4].

Conclusions extracted from the previous approaches agree that a correct criteria combination improves the segmentation results compared to a single color-based criterion. This improvement comes at the cost of a parameter estimation or a weight setting stage using a sufficiently large database.

The motivation of the current work is to propose a high level scheme for criteria combination on a region merging process. In our approach the fusion of information does not take place at the criterion level but at the partition level, after applying each criterion without being interfered or biased by the other criteria. The separated segmentations for each criterion are fused to obtain a common consensus partition representing the basic agreement between the different information sources. Using the basic consensus partition as initial partition for each independent segmentation, the process is iterated. The consensus partitions obtained at each iteration form a hierarchy of partitions, that is provided by the cooperation of the different criteria in their search for a further consensus.

In this work, this cooperative scheme is used to create object-oriented hierarchical region-based representations, that is, a hierarchy of partitions where objects are well represented. We assume that an object is formed by a single region or by the union of texture/color homogeneous regions which are placed at similar depth. For that reason, we create a hierarchy based on the combination of color and depth information. Particularly, we focus our investigation on stereo image pairs, where color and depth information correspond to the RGB image representation and the associated disparity map, respectively. In this context, we extent the validity of our proposal considering both cases: ground truth and estimated disparity maps.

The paper is structured as follows. Section 2 introduces some preliminary definitions and properties related to the mathematical concept of image partition. Section 3 presents the main components of the cooperative merging scheme for the fusion of color and depth information. In Section 4, experimental results using ground truth as well as estimated disparity maps are presented and compared with the use of a single criterion (color or depth) or a linear criteria combination. Finally, conclusions are outlined in Section 5.

2. PRELIMINARY DEFINITIONS

Following [5], let π be an *image partition*, that is, a division of the image into nonempty disjoint sets (known as *regions*) which completely cover the image. Let denote by Π the set of all possible partitions of an image. The set Π is ordered by the refinement order, that is, for $\pi, \pi' \in \Pi$, we can say that $\pi \leq \pi'$, meaning that regions in π' are obtained as the union of regions of π. Then, we say that π is *finer* than π', and that π' is *coarser* than π.

A *lattice* structure is a partially ordered set in which every pair of elements has a greatest lower bound and a least upper bound. A lattice structure is associated with the refinement order, with the meet and the join binary operations defined as follows: the *meet* (sup) $\pi \wedge \pi'$ of two partitions π and π' is their greatest lower bound, that is the coarsest of all the partitions finer than both π and π'; their *join* (inf) $\pi \vee \pi'$ is their least upper bound, that is the finest of all the partitions coarser than both π and π'.

If we have a set of partitions, or *profile*, $\underline{\pi} = \{\pi_1, \ldots, \pi_n\}$, a *median partition* of the profile $\underline{\pi}$, denote by μ, is defined as a partition minimizing the function:

This work has been partly supported by the projects CENIT-VISION 2007-1007 and TEC2007-66858/TCM PROVEC of the Spanish Government.

$$f(\pi) = \sum_{1 \leq i \leq n} \delta(\pi, \pi_i) \quad (1)$$

for $\pi_i \in \underline{\pi}$, and for $\delta(\pi, \pi_i)$ being the *symmetric difference* between the partitions π and π_i, defined as the minimum number of pixels labels that must be changed for π to become identical to π_i, or viceversa. Hence, a median partition of a profile is defined as:

$$\mu = \arg\min_{\pi} \sum_{1 \leq i \leq n} \delta(\pi, \pi_i) \quad (2)$$

Median partitions have some interesting properties [5]. One of them is the *Pareto principle*, defined as follows:

Pareto principle. *If μ is a median partition of profile $\underline{\pi} = \{\pi_1, \ldots, \pi_n\}$, then:*

$$\bigwedge_{1 \leq i \leq n} \pi_i \leq \mu \quad (3)$$

Thus, the median partition is coarser than the meet of all the partitions of the profile. This property will be fundamental for the creation of the cooperative region merging scheme (see Section 3.2).

3. COOPERATIVE REGION MERGING

The cooperative region merging strategy is presented in Figure 1. It is formed by three main steps: the region merging step, where the separated segmentation for each criterion is performed; the meet step, where the final partitions for each criterion are fused at each iteration; and the scale controller, where the scale consistency of the partitions is assured and the number of regions is modified to build the hierarchy.

The central idea is to let the system evolve by itself, starting with a basic agreement (given by the meet of the partitions) and searching for partitions with decreasing number of regions by further consensus iteration after iteration. This is done instead of finding a coarser direct consensus partition for color and disparity output partitions (risking to introduce under-segmentation errors). It can be shown that this iterative scheme provides with a partition hierarchy, i.e. $\pi^k \leq \pi^{k+1}$ (at each iteration, a partition equal or coarser than the partition at the previous iteration is obtained).

3.1. Region Merging Step

A region merging step is associated with each criteria or information source. Starting from an initial partition of the image data (or directly all pixels at the first iteration, π^0), this step performs a region merging process until it obtains an output partition with the number of specified regions (see Figure 1).

The region merging techniques used in this work are based on a modified version of the general region merging techniques based on information theory statistical measures proposed in [6]. Concretely, a merging process formed by a Bhattacharyya merging criterion and a scale-based merging order is chosen for its good compromise between under- and over-segmentation errors, in the context of both color homogeneous and texture region segmentation. In addition, it can be used in a completely unsupervised manner, automatically providing with an ordered set of the most significant partitions into the whole hierarchy [6] (this feature will be specially useful in Section 3.3.2).

The region model proposed in [6] was based on the empirical distribution of the region pixels. The pixel values were quantized into a reduced number of bins (typically 5 or 10) to improve the performance and reduce oversegmentation. However, the quantization can increase the undersegmentation error and the emergence of fake contours. To mitigate these effects, in this work we have used an estimator of the empirical distribution based on an averaged shifted histogram (ASH). The ASH estimator [7] provides a low-pass filtered version of the normalized histogram, obtained by convolving the histogram with a triangular window. This modification provides smoother probability function estimations.

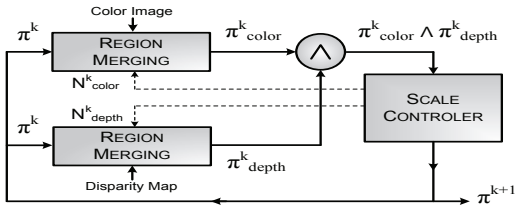

Fig. 1: Cooperative region merging scheme.

3.2. Meet Step

This step computes the meet operation (defined in Section 2) between the output partitions provided by the region merging processes, i.e. $\pi^k_{\text{color}} \wedge \pi^k_{\text{depth}}$ in the context of Figure 1.

The functionality of this step is based on the Pareto principle, presented in Equation (3). Given the output partitions of the region merging processes, their meet provides a finer (i.e. oversegmented) version of a median partition for the profile formed by the color and depth output partitions. In other words, it provides a conservative combination or basic consensus of the output partitions in terms of merging errors (undersegmentation).

3.3. Scale controler

This block is formed by two steps: the scale-based filtering (Section 3.3.1), that assures the scale consistency of all the obtained partitions; and the scale adapter (Section 3.3.2), that controls the number of regions in the color and depth partitions to build the hierarchy taking into account the particularities of the information sources.

3.3.1. Scale-Based Filtering

The meet operation between the color and disparity partitions can generate small regions formed by few pixels into the resulting partition. The scale-based filtering removes from the meet partition the set of regions that are too small to be significant at the current scale, assuring that all partitions are scale consistent. For that purpose a scale threshold is defined on the region areas, similarly to the scale-based merging order used in [6]:

$$T_{\text{scale}} = \alpha \cdot \frac{\text{Image Area}}{\text{Number of Regions}} \quad (4)$$

The α parameter controls the minimum resolution at each scale. In our experiments, we have chosen a low value for this parameter, $\alpha = 0.03$, to be sure that only clearly meaningless regions at that scale are discarded (in [6], it was typically set to 0.15).

Out-of-scale regions are merged using the same color-based merging criterion that is applied to compute the color output partitions. The reason for prioritizing color is that, in most cases, these regions are generated in the contours of the image objects due to disparity errors or inaccuracies in the estimation, even for so-called ground truth disparity maps (see Section 4).

3.3.2. Scale adapter

This block controls the number of regions into the color and depth partitions to build the partition hierarchy. The strategy is based on two points: first, the ability of the unsupervised region merging techniques in [6] to provide with a set of the most statistically significant

Fig. 2: From left to right: Aloe and Baby1 datasets (from [8]), Ballet and Breakdancers (from [9]); including one of the color views (top) and its corresponding disparity map (bottom).

partitions; second, the assumption, stated in Section 1, that objects are formed by a single region or by the union of regions at similar depths. This means that a correct color segmentation must be finer that a correct depth segmentation and, consequently, that the number of regions into the color partition must be larger than the number of regions in the depth partition.

Before starting the cooperative process, a maximum number of regions for the created color and depth partitions is defined, $N_{\text{color}}^{\text{MAX}}$ and $N_{\text{depth}}^{\text{MAX}}$, respectively. These maximum values are not crucial, as far as they are large enough to avoid any dramatic errors into both output partitions (if important errors occur only in one of the partitions, they are commonly corrected by the meet operation). In the first iteration, the system is initialized with $N_{\text{color}}^{0} = N_{\text{color}}^{\text{MAX}}$ and $N_{\text{depth}}^{0} = N_{\text{depth}}^{\text{MAX}}$ (in general, we use the same value in both cases).

At each iteration, N_{color}^{k} and N_{depth}^{k} are automatically adapted to the current problem situation. This is done by analyzing the current set of color and depth meaningful partitions provided by the region merging blocks. The number of regions of the largest significant partition, not exceeding $N_{\text{color}}^{\text{MAX}}$ and $N_{\text{depth}}^{\text{MAX}}$, is adopted.

It may happen after some iterations that $\pi^{k+1} = \pi^{k}$, for $N_{\text{color}}^{k+1} = N_{\text{color}}^{k}$ and $N_{\text{depth}}^{k+1} = N_{\text{depth}}^{k}$. This means that no further consensus is possible between the color and depth partitions at this given scale. For that reason, the scale is modified taking into account that the number of regions in the color partition should be larger than in the depth partition. Particularly, if the color partition with second largest number of regions exceeds the number of regions in the current depth partition, the color scale is decreased, setting N_{color}^{k+1} to the second largest value and $N_{\text{color}}^{\text{MAX}} = N_{\text{color}}^{k}$. Otherwise, the depth scale is decreased, setting N_{depth}^{k+1} to the second largest number of regions in the significant set, and modifying $N_{\text{depth}}^{\text{MAX}} = N_{\text{depth}}^{k}$.

When the system arrives to the coarsest color and depth significant partitions, the last part of the hierarchy is created exclusively based on depth information. The reason is that, at the lowest level of resolution, depth is more reliable than color to merge regions that represent image objects. Hence, in this situation, each time $\pi^{k+1} = \pi^{k}$ occurs, the hierarchy is built by, first, decreasing the number of regions in the color partition and, when a single region in color is reached, decreasing the number of regions in the depth partitions.

4. EXPERIMENTAL RESULTS

4.1. Ground Truth Disparity Maps

The first set of experiments was performed into a subset of the Middlebury Stereo Datasets [8] (some examples are shown in Figure 2). The datasets include color images and ground truth disparity maps obtained by structured light (see references in [8]). The examples shown in this section were computed on images with a third-size resolution (413-465×370 pixels). Disparity values are represented

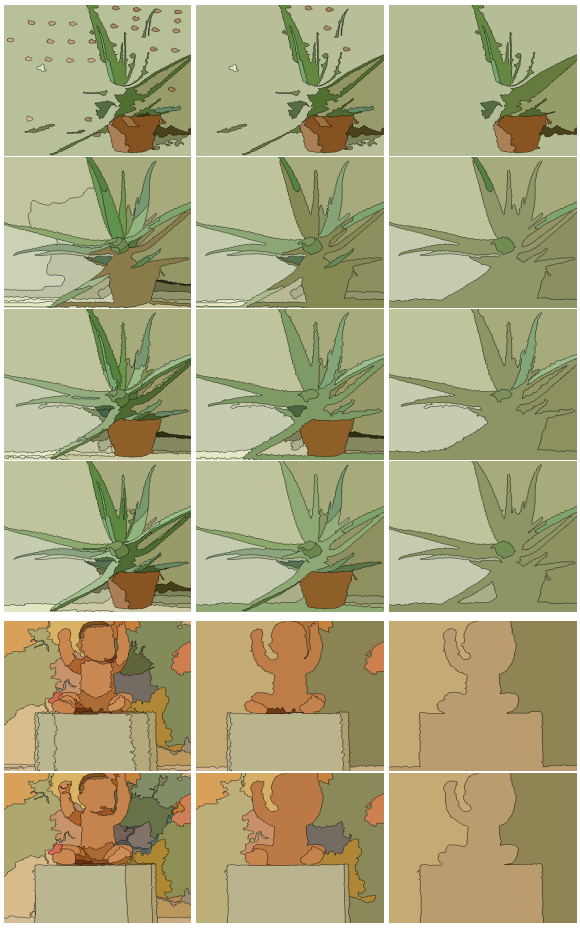

Fig. 3: Middlebury datasets. Comparison between different hierarchy creation strategies. See description in Section 4.1.

by intensity values from 0 to 60, corresponding to real disparities except for the intensity 0 (black areas) that corresponds to unknown disparity values. Thus, note that though being considered as ground truth disparity maps, they may include some unknown disparity areas that, for our purpose, can be considered as estimation errors.

Figure 3 presents a comparison between the hierarchy of partitions obtained using different information sources and combination strategies. Concretely, for the upper dataset (Aloe), partitions extracted from the hierarchy with 48, 23 and 10 regions (these are some of the values automatically obtained after the scale controller) are shown in columns from left to right, respectively. The first row corresponds to a color-based hierarchy, created using the same region merging technique introduced in Section 3.1. The second row presents the hierarchy created by the same region merging technique but exclusively using disparity information. The hierarchy of partitions into the third row was computed using a weighted combination of color and disparity information. Heuristically, we found that the best performance was obtained using equal weights for color and depth information. Finally, the last row presents the hierarchy obtained by the proposed cooperative approach. In turn, partitions with 54, 14 and 3 regions are shown for Baby1 dataset. In this case, only hierarchies obtained by the linear (upper row) and the cooperative combination (bottom row) are compared. In both examples, $N_{\text{color}}^{\text{MAX}}$ and $N_{\text{depth}}^{\text{MAX}}$ were initialized to 50 regions.

As commented in Section 1, there is a great improvement into the segmentation results when not only color or depth information is

used. Nevertheless, note that partitions from the depth-based hierarchy with a reduced number of regions provide with good segmentation results from an object-oriented point of view, while partitions with a larger number of regions have not this property. This fact confirms the assumption that a partition correctly describing the objects into the scene has to be finer than a correct depth segmentation, and that the top part of the hierarchy must be build based on depth information.

For a large number of regions, the cooperative-based partitions present, in general, less oversegmentation and more regular and stable contours compared to weighted-based partitions. This is due to the iterative process that allows the region merging process to correct possible merging errors and improve the contour accuracy thanks to their mutual cooperation through the meet partition. For a smaller number of regions, some color and depth errors are not present in the cooperative approach and the color characteristics of the objects are better preserved than in the direct combination approach. For instance, in Aloe dataset, part of the background is merged with the plant in the partition with 10 regions in the weighted-based hierarchy. In Baby1 dataset, most of the heterogeneous color regions of the background have disappeared in the partition with 14 regions, and part of the background is merged with the foreground in the bottom of the image in the partition with 3 regions.

An extension of these results for other 20 images of the same dataset can be seen at *http://gps-tsc.upc.es/imatge/_Felipe/icassp09/*.

4.2. Estimated Disparity Maps

Another set of experiments was performed using a single frame, and their corresponding estimated disparity maps, from the multiple view sequences Ballet and Breakdancers [9] (shown in the third and fourth columns of Figure 2, respectively). Images were downsampled to a fourth of their full size (256×196 pixels). The depth information was estimated using the stereo technique described in the references in [9]. As it can be seen in Figure 3, the disparity maps have a good visual quality although they include some estimation errors and disparity discontinuities.

The first two rows in Figure 4 present the results obtained from Ballet dataset. From left to right columns, the partitions with 130, 31, and 10 regions, extracted from the weighted-based hierarchy (first row) and the cooperative-based hierarchy (second row) are shown. Results in the last two rows of Figure 4 correspond to Breakdancers dataset. Partitions with 103, 43, and 4 regions from the weighted-based hierarchy (first row) and the cooperative-based hierarchy (second row) are shown. In both cases, $N_{\text{color}}^{\text{MAX}}$ and $N_{\text{depth}}^{\text{MAX}}$ were initialized to 100 regions.

In the case of partitions with a large number of regions, the cooperative partitions present slightly more detail and accuracy in terms of color regions. For instance, the black hair and the feet of the ballet dancer, and the hand and face of the man, for Ballet image; and the head and the stretched hand of the dancer in first plane, for Breakdancers dataset. For partitions with less regions, the direct criteria combination preserve the most color heterogeneous regions, while the cooperative-based partitions provide regions that correspond with main objects and depth planes of the scene (see the Ballet partitions with 10 regions, and Breakdancers partitions with 4 regions). Note that, also in both datasets, the estimated disparity for the ground presents a large number of discontinuities and errors. It is remarkable that this fact generates a large number of oversegmented regions into the weighted-based partitions, while the cooperative partitions present a better segmentation of the ground despite the estimation errors (see partitions with 31 and 43 regions for Ballet and Breakdancers, respectively).

Fig. 4: Ballet and Breakdancers datasets. Comparison between different hierarchy creation strategies. See description in Section 4.2.

5. CONCLUSIONS

The cooperative region merging scheme, applied to combining color and depth information, provides in general more stable (contour regularity), accurate (richer color description) and semantic (better object representation) hierarchical region-based image representations than a direct criteria combination, and without requiring any parameter adjustment, model estimation, or training stage. Although the computational load of the direct combination approach may be lower, the cooperative structure can be easily parallelized (for instance, each region merging block can be independently run in a different CPU). In this case, the increase in the number of region fusions computed by each block is bounded in the worst case (the number of regions at each iteration decreases only by one) by the square of the number of regions at the first iteration. Considering that the first meet partition has 150 regions (a typical value for N_{color}^{0} and N_{depth}^{0} set to 50 regions), the increase in the number of mergings is bound in the worst case to 9%, 2.25%, and 0.56% for images with 500×500, 1000×1000, and 2000×2000 pixels, respectively.

6. REFERENCES

[1] V. Vilaplana, F. Marques, and P. Salembier, "Binary partition trees for object detection," *IEEE Trans. Image Process.*, Accepted. 2008.

[2] T. Adamek and N.E. O'Connor, "Using Dempster-Shafer theory to fuse multiple information sources in region-based segmentation," *ICIP'07*, vol. 2, pp. 269–272, 2007.

[3] S.H. Lee, N.I. Cho, and J.I. Park, "Stochastic diffusion for correspondence estimation and objects segmentation," *CVPRW'04*, pp. 183–183, 2004.

[4] T. Pock, C. Zach, and H. Bischof, "Mumford-shah meets stereo: Integration of weak depth hypotheses," *CVPR'07*, vol. 0, pp. 1–8, 2007.

[5] J.P. Barthelemy and B. Leclerc, "The median procedure for partitions," *Partitioning Data Sets, DIMACS Series in Descrete Mathematics*, vol. 19, pp. 3–34, 1995.

[6] F. Calderero and F. Marques, "General region merging approaches based on information theory statistical measures," *ICIP'08*, 2008.

[7] D. W. Scott, "Averaged shifted histograms: Effective nonparametric density estimators in several dimensions," *The Annals of Statistics*, vol. 13, no. 3, pp. 1024–1040, 1985.

[8] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," *CVPR '07*, pp. 1–8, June 2007.

[9] Microsoft Research, "Ballet and breakdancers sequences," http://research.microsoft.com/ivm/3DVideoDownload/.