HIERARCHICAL REGION-BASED REPRESENTATION FOR SEGMENTATION AND FILTERING WITH DEPTH IN SINGLE IMAGES

Mariella Dimiccoli, Philippe Salembier

Universitat Politecnica de Catalunya (UPC)
Dept. of Signal Theory and Communications (TSC)
Jordi Girona 1-3, 08034, Barcelona, Spain

ABSTRACT

This paper presents an algorithm for tree-based representation of single images and its applications to segmentation and filtering with depth. In a our recent work, we have addressed the problem of segmentation with depth by incorporating depth ordering information into a region merging algorithm and by reasoning about depth relations through a graph model. In this paper, we extend this previous work giving a two-fold contribution. First, we propose to model each pixel statistically by its probability distribution instead of deterministically by its color value. Second, we propose a depth-oriented filter, which allows to remove foreground regions and to replace them with a plausible background. Experimental results are satisfactory.

***Index Terms*—** Image segmentation, filtering

1. INTRODUCTION

Recently, the work in [1] has started a new line of research in incorporating depth ordering information into a statistical region merging algorithm, allowing the merging process to exploit local depth information provided by T-junctions. Experimental results have showed the effectiveness of the proposed strategy in extracting a correct depth map. However, the results of the segmentation can be criticized in terms of accuracy and regularity of region boundaries. This weakness is attributable to a poor modeling during the early stages of the tree construction process. Indeed, the tree construction in [1] is based on the work in [2], which has recently proposed a new region model based on color histogram and a new family of merging criteria based on information theory statistical measures between the region models. This kind of modeling has demonstrated a noticeable improvement with respect to first order statistic models where mean or median color values are used as region model since they do not assume that regions are homogeneous in color. However, the tree construction process starts by considering that each pixel is a single region, which is modeled still deterministically by its color value and therefore the effect of the statistical modeling become really important only in the late stages of the merging process. In this paper, we solve this problem by modeling each pixel statistically by a probability distribution. This same statistical pixel model has been used in the seminal work of Efros [3] for texture synthesis and then in [4] for image and video denoising [5]. To the best of our knowledge it is the first time that this statistical pixel model is exploited for segmentation purpose. In addition, we propose a novel ”depth criterion” for filtering based on connected operators, which allows to automatically simplify the image by removing the closest regions to the viewpoint and to fill the hole that is left behind in a visually plausible way by using a state-of-the-art image completion technique [6]. The idea of using image completion as restitution strategy for filtering based on connected operators has been originally proposed in [7], but in the context of a depth-oriented filter it represents a natural solution and it can be applied directly to the removed regions without the need of any preprocessing. Background substitution, a problem usually addressed in the multiview scenario [8, 9], and automatic foreground object removal are typical examples of applications where this kind of filtering can be useful.

The next section describes the proposed approach for constructing a hierarchical region-based representation of images. Section 3 and section 4 are devoted to the presentation of the segmentation and filtering algorithms respectively. Finally, section 5 reports the main conclusions of this work.

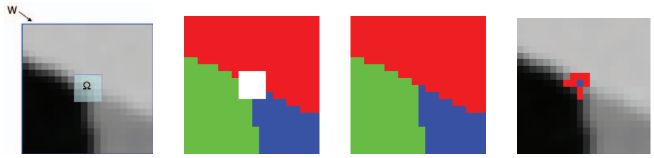
2. TREE CONSTRUCTION

Segmentation by region merging techniques and filtering by connected operators based on region-tree pruning are closely related. Both rely on a hierarchical region-based representation of the image obtained using merging techniques and act by pruning the tree structure. The difference is in that while the output of the segmentation is a label image, the output of the filtering by connected operators is the original image from which some elements have been removed. This section details the construction of a hierarchical region-based representation of the image. The tree structure which is created in this work is a Binary Partition Tree (BPT), a structured representation of a set of hierarchical partitions in which the finest level of detail is given by the initial partition of image pixels. The nodes of the tree are associated to regions that represent the union of two children regions and the root node represents the entire image support. The method used for creating this tree structure consists in first detecting T-junctions, then in iteratively merging pairs of neighboring regions following a statistical criterion and preserving the previously detected T-junctions. Due to space limitations, some details of the algorithm as well as some references have been omitted, but can be found in [1].

2.1. T-junction detection

The algorithm for T-junction detection involves three main steps. The goal of the first step is to reduce in a fast way the number of points to be processed without loosing any true T-junction. In order to do this, candidate points are detected by the SUSAN filter [10] which is applied to a simplified version of the image, obtained with a multi-scale leveling [11]. The second step consists in extracting the T-junction branches in a close surrounding of candidate points (W), omitting a 4x4 neighborhood (Ω) centered on them (Fig. 1(a)).

978-1-4244-5654-3/09/$26.00 ©2009 IEEE



(a) (b) (c) (d)

Fig. 1. (a) A window W centered at a candidate point: Ω is the neighborhood we consider unreliable. (b) Branch extraction in $W - \Omega$. (c) Branch propagation in Ω. (d) Validated points: The point having the smallest value of sum of average curvature on each branch is marked in blue

The branch extraction in $(W - \Omega)$ is performed by a region merging algorithm. While the algorithm in [2] starts from an initial partition of flat zones and models each region by its mean of color values, we starts by the partition made of all image pixels and model each pixel by its probability distribution. Assuming that the image u is a fairly general stationary random process, and therefore there exist many similar patches for all details of the image, the probability distribution of a given pixel x can be evaluated by comparing a whole patch centered at x with patches at other pixel locations y. The value of the central pixel y of these similar patches will thus be considered as another realization of the random process at x. In practice, to reduce the computation cost, the search for similar patches is restricted to a search window of size $S \times S$. The similarity $w(x, y)$ between the pixel in consideration x and a pixel $y \in S \times S$, is given by: $w(x, y) = e^{\frac{-\sum_i (u(x_i) - u(y_i))^2}{h}}$, where i indicates the position of a pixel on the patch in comparison and h is the filtering parameter. To give more importance on the patch to pixels closer to the reference pixel, the distance between two patches is also weighted by a gaussian-like kernel decaying from the center of the patch to its boundary. Each pixel is then modeled through its probability distribution, obtained by summing the normalized values of similarity of each of given set of color ranges. The normalization of all similarity values in $S \times S$ is performed by dividing them by the sum of all similarity values. In all experiments for detecting T-junctions, we have fixed the size of the patch to $5x5$ and the search window to $8x8$.

With this modeling, pairs of neighboring regions in $W - \Omega$ are iteratively merged following a statistical criterion, based on the probability of fusion of their region histograms (Bhattacharyya merging criterion) [2], until three regions are obtained (Fig. 1(b)). When all three obtained branches intercept Ω, they are propagated inside Ω according to the *good continuation principle* [12] and constrained to meet at the candidate point (Fig. 1(c)). To validate the candidate point, three different criteria based on geometrical or photometrical region properties have to be met [1]. Each validated T-junction is assigned a measure relying on the regularity of its branches (Fig. 1(d)), which is used in the third step, for reducing clusters of validated points by selecting the validated point that minimizes the sum of the branch curvature.

2.2. Region-merging preserving T-junctions

The next step consists in computing a hierarchical region-based representation of the image through region merging but preserving T-junctions. The region model we use is the same defined in the previous section, whereas, in order to preserve T-junctions, the merging order is modified by introducing the concept of *incompatibility*. Two regions are said *incompatible* if they are involved in an occlusion relation and therefore cannot be merged.

3. SEGMENTATION WITH DEPTH

The goal of segmentation with depth is to define a partition by pruning the tree-representation of the image and to construct a global and *consistent* depth interpretation based on this partition as well as on the local depth assessments provided by T-junctions. Although locally, the region delimited by the head of the T appears to be in front of the ones delimited by the leg, the interpretation of pairs of T-junctions that share an edge may give rise to an inconsistency. In Fig. 2 (b) there is an example of first order inconsistency. Region C is in front of region A for one T-junction, while the converse is true for another T-junction. Higher order inconsistencies involve more than two T-junctions. In this work, the regions involved in the final partition are the incompatible regions, that is all regions that are involved in at least one occlusion relation. The nodes of the partition corresponds to the set of the biggest incompatible regions.

To deal with possible inconsistencies and obtain a global depth ordering a Direct Graph (DG) is used. A DG is specified by $DG = (V, E_A, A)$, where V is a set of nodes which represent image regions, E is a set of directed edges,which represent the relative depth relations between regions and A is the matrix of weights attached to the edges, amounting the number of occurrences the depth relation represented by the edge has been inferred from different occlusion relationships. For instance, the weight of the edge (C, A) is 2, whereas the weight of the edge (A, C) is 1. The usefulness of this formalization is in that it allows easy detection and correction of inconsistencies. The search for inconsistent pairs of T-junctions is reduced to the search of cycles on the DG (dashed thick red arrows in Fig.2(c)), which is performed by a Depth-First Search (DFS) algorithm [13]. Inconsistencies are solved by suppressing the edge(s) on the cycle with lowest cost. Since the depth relation associated to the edge with the lowest cost is considered unreliable, the other edge (dashed thin blue arrow in Fig.2(c)) associated with the T-junction from which the unreliable depth relation arise is also removed. As a result, a Directed Acyclic Graph (DAG) is obtained Fig.2(d). The depth map is exactly the Hasse Diagram (HD) corresponding to the transitive reduction of the DAG (Fig.2(e)). Since there is no depth order between the regions forming the leg of a T-junction, they appear on the HD as leaves (A and B), without any information about their respective depth, unless of course, an order between them can be inferred by other T-junctions. In Fig.3 we show some examples of comparison between the segmentation results obtained via the algorithm in [1] and the segmentation results obtained by using the proposed approach. As can be observed, the new modeling proposed here improves both accuracy (see Fig. 3(a)) and regularity (see Fig. 3 (a) and (b)) of region boundaries. Note that in both these examples there is a case of conflict between the regions A and C: while region A is interpreted as foreground and region C as background for one T-junctions, the contrary is true for two T-junctions. The solution of this conflict leads to a correct depth interpretation.

4. DEPTH-ORIENTED FILTERING

In this section we explain how to take advantage of the tree representation of the image as well as of the depth information encoded on the transitive reduction of the DAG, to create a novel depth-oriented filter, whose goal is to remove the foreground region. The strategy used to construct this new filter is based on region-tree pruning [14, 15].

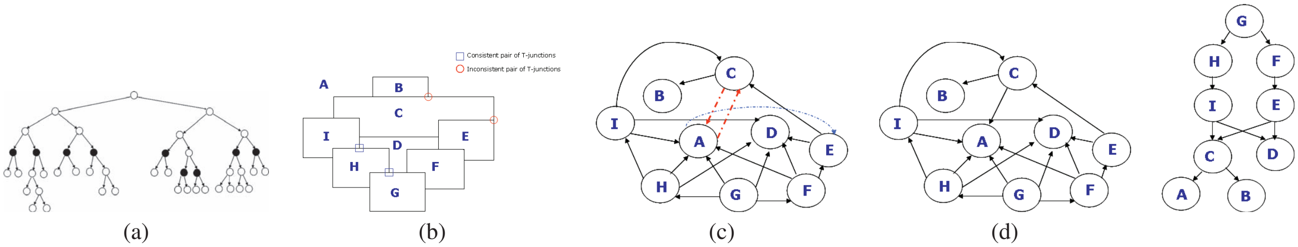


Fig. 2. (a) BPT: the black nodes represent the regions of the final partition (b) Final partition (c) Associated DG (d) Associated DAG (e) Hasse diagram resulting from the transitive reduction of the DAG

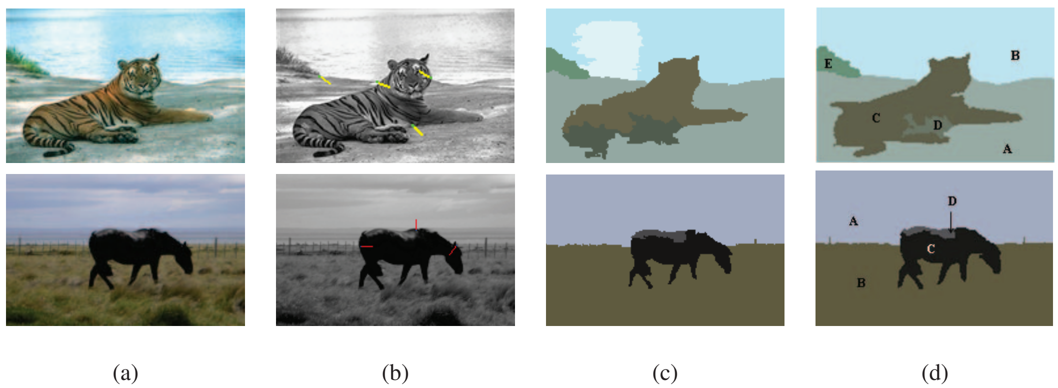


Fig. 3. Example of segmentation: (a) Original image (b) T-junction detection (c) Segmentation by the algorithm in [1] (d) Segmentation by the proposed approach

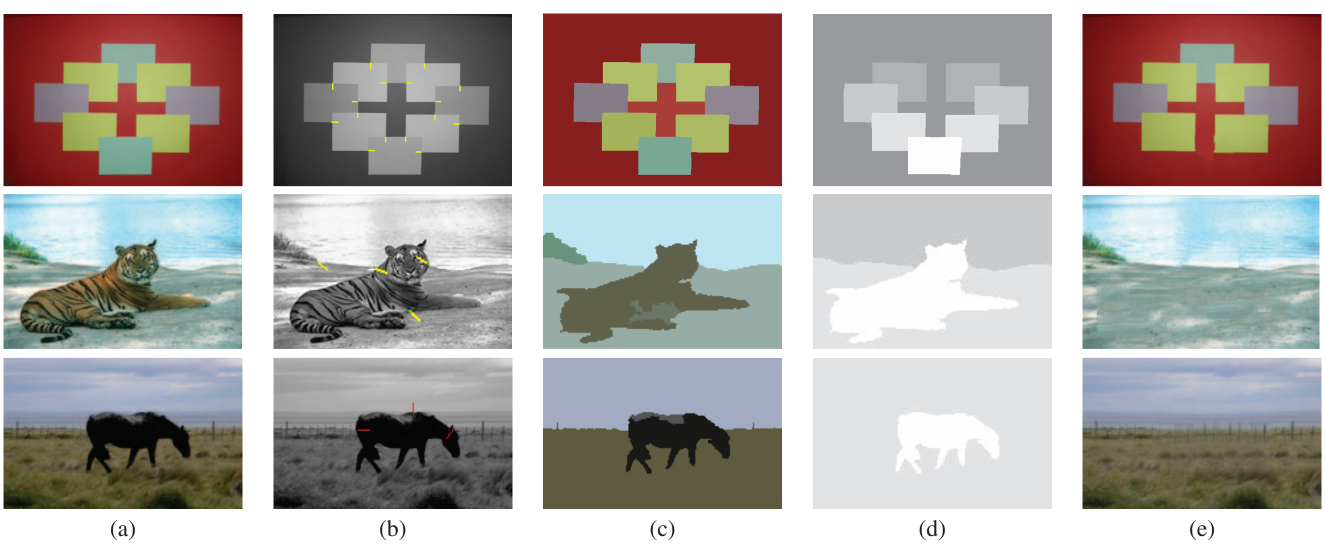


Fig. 4. Examples of segmentation and filtering with depth.(a) Original image (b) T-junction detection (c) Segmentation (d) Depth ordering (e) Filtering

The decision on which node has to be pruned is taken by looking at the transitive reduction of the DAG (Fig.2(e)). The node having only outgoing arrows corresponds to the region closer to the viewpoint and therefore it is removed from the tree. The reconstruction of the image from the pruned tree involves the use of a state-of-the-art image completion technique [6]. Based only on the observed part of an incomplete or consciously masked image, the goal of image completion is to fill the missing part so that a visually plausible outcome is obtained. In [6] this task is posed in the form of a discrete global optimization problem, whose objective function corresponds to the energy of a Markov Random Field and is optimized by using an novel optimization scheme, called priority-Belief Propagation. The use of this technique allows to restores the removed regions as if there were originally not present in the image, giving a visually plausible result. We tested our algorithm on a set of real images. For each experiment we show five images: the original image; a gray level version of the original image where detected T-junctions are represented through a vector pointing to the roof; the segmented image; the depthmap, which is rendered as a gray level image (high values indicate regions closer to the viewpoint),the filtered image, where the region closest to the viewpoint has been removed and replaced by a visually plausible background. In all experiments, we segmented the image until only regions involved in at least one occlusion relation are obtained(Fig.4(c)).

5. CONCLUSIONS

This paper has presented an algorithm for the computation of a hierarchical region-based representation of single images and its applications to segmentation and filtering with depth. Both applications act by pruning the tree representation. In the case of segmentation, the output image is a partition where each region represents a distinct depth level in the image, whereas the result of the filtering is an image in which the closest regions to the viewpoint is replaced by a visually plausible background that mimics the appearance of the surrounding regions. Our approach extends the previous work in [1], which incorporates depth ordering information provided by local occlusion into a statistical region-merging algorithm. The depth relations between the regions of the final partition are formalized through a graph representation which allows to solve possible ambiguous interpretations leading to a global depth ordering. Contrary to this previous work, which models each pixel deterministically by its color value, we model each pixel statistically by its probability distribution,improving the accuracy of region boundaries.

The filtering consists in removing the foreground region and replacing it with a visually plausible background obtained by applying a state-of-the-art image completion technique.

Currently, we are investigating extensions for obtaining a more accurate depth map by incorporating other depth cues, as for instance convexity which allows to infer a depth ordering even in absence of T-junctions.

6. REFERENCES

[1] M. Dimiccoli and P. Salembier, "Exploiting t-junctions for depth segregation in single images," in *To be presented at ICASSP*, Taipei (Taiwan), April 2009.

[2] F. Calderero and F. Marques, "General region merging approaches based on information theory statistical measures.," in *Proc. of ICIP*, San Diego (CA), October 2008.

[3] A.Efros and T.Leung, "Texture synthesis by non-parametric sampling," *Proc. of ICCV*, October 1999.

[4] A. Buades et al., "The staircasing effect in neighborhood filters and its solution," *IEEE Trans. on IP*, vol. 15, no. 6, pp. 1499–1505, 2006.

[5] A. Buades, B. Coll, and J.M. Morel, "Nonlocal image and movie denoising," *Int. Journal of Computer Vision*, vol. 76, no. 2, pp. 123 – 139, 2008.

[6] N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," *IEEE Trans. on IP*, vol. 16, no. 11, pp. 45–78, 2007.

[7] M. Dimiccoli and P. Salembier, "Perceptual filtering with connected operators and image inpainting," in *Int. Symp. on Mathematical Morphology*, Rio de Janeiro (Brazil), October 2007.

[8] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother, "Bi-layer segmentation of binocular stereo video," in *Proc. of CVPR*, San Diego, CA, June 2005.

[9] A. Criminisi, G. Cross, A. Blake, and V. Kolmogorov, "Bilayer segmentation of live video," in *Proc. of CVPR*, New York, NY, June 2006.

[10] S.M. Smith and M. Brady, "Susan - a new approach to low level image processing," *Int. journal of Computer Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[11] P. Maragos and G. Evangelopoulos, "Leveling cartoons, texture energy markers, and image decomposition," in *Proc. 8th Int. Symp. on Mathematical Morphology*, Rio de Janeiro (Brazil), October 2007.

[12] N. Montanari, "On the optimal detection of curves in noisy pictures," *Communications of the ACM archive*, vol. 14, no. 5, pp. 335 – 345, 1971.

[13] T.H. Cormen et al., *Introduction to Algorithms*, chapter Depth-first-search, pp. 540–549, MIT Press and McGraw-Hill, 2001.

[14] P.Salembier, A.Oliveras, and L.Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Trans. on IP*, vol. 7(4), pp. 555–570, 1998.

[15] P.Salembier and L.Garrido, "Binary partition tree as an efficient representation for image processing, segmentation and information retrieval," *IEEE Trans. on IP*, vol. 7(4), pp. 561–576, 2000.