EXPLOITING T-JUNCTIONS FOR DEPTH SEGREGATION IN SINGLE IMAGES

Mariella Dimiccoli and Philippe Salembier

Technical University of Catalonia (UPC), Dept. of Signal Theory and Communications, Barcelona, SPAIN

ABSTRACT

Occlusion is one of the major consequences of the physical image generation process: it occurs when an opaque object partly obscures the view of another object further away from the viewpoint. Local signatures of occlusion in the projected image plane are T-shaped junctions. They represent, in some sense, one of the most primitive depth information. In this paper, we investigate the usefulness of T-junctions for depth segregation in single images. Our strategy consists in incorporating ordering information provided by T-junctions into a region merging algorithm and then reasoning about the depth relations between the regions of the final partition using a graph model. Experimental results demonstrate the effectiveness of the proposed approach.

Index Terms— Image Segmentation, T-junctions.

1. INTRODUCTION

Although the importance of T-junctions has been emphasized by Gestalt psychologists [1] and investigated in human vision, their role is often downplayed in practical applications. Their usefulness in extracting non-trivial information about 3D scenes has been recently demonstrated in a variety of applications such as stereo vision, multiview geometry and video segmentation [2, 3, 4]. However, the potential of T-junctions in the single image scenario has been a little explored until now. This is due partly to the lack of robustness of T-junction detection without relying on redundant information in space or time, and partly to the ambiguity of their depth interpretations. In fact, whereas all instances of occlusion produce a T-junction [5] (Fig.1 (a)), the converse is not true. For instance, the T-junctions in Fig.1 (b) are likely the result of a reflectance discontinuity and not of an occlusion. Although locally, the region forming the *roof* of the T appears to be in front of the ones forming the *stem*, a complete assessment of the information carried out by a T-junction is usually not possible. In this paper however, we attempt to show that a global analysis allows to solve ambiguities and to costraint possible consistent interpretations. To this goal, we have developed a new algorithm for robust T-junction detection in single images and a new segmentation strategy based on region merging that integrates ordering information arisen from T-junctions. Occlusion relationships between the regions of the final partition are finally encoded through a Directed Graph (DG). The depthmap is directly obtained as transitive reduction of the underlying Directed Acyclic Graph (DAG). The following section reviews the literature on occlusion reasoning. Section 3 describes the proposed approach. Section 4 details experimental results while Section 5 reports the conclusions.

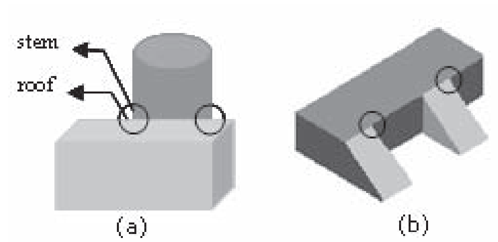


Fig. 1. (a) Occluding T-junctions (b) Non-occluding T-junctions

2. RELATED WORK

Until recently, most work on occlusion reasoning has been limited to line drawing [6, 7, 8, 9]. Thanks to the introduction of Markov Random Field (MRF) models, the last decade has seen an increase of interest in occlusion reasoning for scene description. Stella et al. [10] developed an ordered partitioning method in spectral graph theory that handles directional grouping cues. Gao et al. [11] proposed a Bayesian inference framework that uses a graph representation consisting of two types of nodes, atomic regions and the corresponding terminators of T-junctions, that makes the problem a Mixed MRF. The most recent works are learning-based approaches [12, 13, 14]. They all rely on the use of a large database of images, annotated with human-marked segmentation and depthmap, for training and qualitative evaluation. Most of these approaches have been tested only on a limited set of synthetic images [10], or on images previously segmented by interactive methods [11]. Learning-based approaches have shown often impressive results, but often they rely on strong assumptions on the image structure [14], or on the use of human segmentation [12]. In all cases, they are obtained using a ratio between the number of test images and the number of training images almost equal to 1, which may result in a lack of robustness in real applications.

Here, we propose a framework that does not rely on any assumption on the image structure neither on intensive learn-

(a) (b) (c) (d) (e) (f) (g)

Fig. 2. (a) A window W centered at a candidate point: Ω is the neighborhood we consider unreliable. (b) Partitioning of $\Omega = \Omega_1 \cup \Omega_2$. (c) Branch extraction in $W - \Omega$. (d) Branch propagation in Ω_2. (e) Branch propagation in Ω_1. (f) Validated points: The point having the smallest value of sum of average curvature on each branch is marked in blue. (g) Cluster reduction.

ing strategies and does not require any human marked segmentation. As a first example, we illustrate here the interest of this framework using only T-junctions as depth cue but the set of cues will be extended in the future.

3. PROPOSED APPROACH

The method proposed here consists in first detecting T-junctions, then in partitioning the image preserving the T-junctions previously detected and, finally, in depth ordering the regions of the partition.

3.1. T-junction detection

The algorithm for T-junction detection involves three main steps. A first selection step provides candidate points that represent potential T-junctions to be characterized and possibly validated in a second step. The characterization, namely the branch extraction, is performed on a close surrounding of candidate points (W), omitting a 4x4 neighborhood (Ω) centered on them (Fig. 2(a)). The obtained branches are then propagated inside Ω according to the *good continuation principle* and constrained to meet at the candidate point. This procedure also is supported by psychophysical experiments [15] suggesting that junctions are detected even when the center is occluded. Each validated T-junction is assigned a measure relying on the regularity of its branches, which is used in the third step, devoted to the reduction of clusters of validated points.

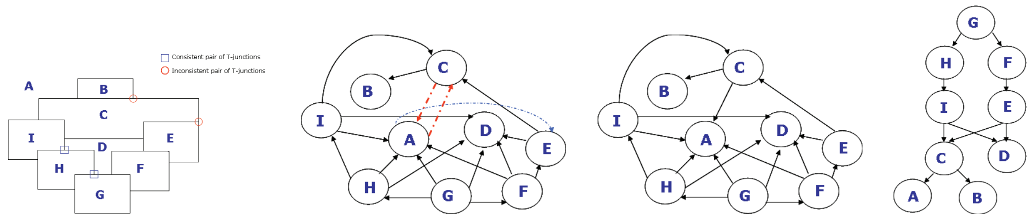
3.1.1. Candidate point selection

Since T-junctions are structural features, the search for candidate points is performed on the structural part of the image, also called *cartoon component*, obtained by a simplification of the original image with a hierarchy of leveling [16]. Candidate points are localized by the local filter SUSAN [17]. To take into account the localization inaccuracy of the local filter, coordinates of candidate points are allowed to vary on a small neighborhood. In practice, we apply a dilation on the mask of candidate points obtained by applying SUSAN. The branch extraction is applied directly on the original image at points marked by the mask.

3.1.2. Branch extraction and validation

The branch extraction in $(W - \Omega)$ is performed by a region merging algorithm. Starting from an initial partition of flat zones, pairs of neighboring regions are iteratively merged following a similarity criterion until a termination criterion is reached. Each region is modeled by its mean of color values and the similarity criterion between two regions is the color distance between the region mean color values. As T-junctions are modeled by three piecewise constant regions, called wedges, the merging is done iteratively until three regions are obtained. The region merging strategy does not guarantee that the three final regions will reach Ω. If this is the case, the candidate point is discarded. To guarantee the visibility of each branch and to distinguish from corners, we impose a threshold on the minimum color difference between the mean color of each pair of wedges. The branch extraction in $(W - \Omega)$ relies on photometric information. However, inside Ω, where the photometric profile is unreliable, it is achieved in two steps (first in Ω_1 and then in Ω_2, see Fig.2(b)) using a curvature-based criterion that minimizes the sum of the absolute curvature at the new branch points created by the hypothetic labeling, with the constraint that branches meet at the candidate point (Fig.2(c) and Fig.2(d)). The curvature at the candidate point is computed eliminating the stem of the candidate T-junction.

To validate the candidate point three different criteria have to be fitted. The first criterion is geometrical. In most cases corresponding to spurious T-junctions, one wedge is composed of a very small number of pixels or looks like a narrow band. We then use a "size criterion" that is as follows: if at least one region completely disappears after an erosion (binary) with a structuring element, then the candidate point is discarded. The structuring element is a square whose size s is related to the size w of W. To allow keeping T-junctions whose contours converging at the junction center form a small angle, s is taken as follows: $\frac{w}{4} - 1$. The second criterion deals with the branches orientation and is necessary in order to distinguish from other junction types. For each branch, we first compute the vector that represents its medium orientation in W and then the angles between each pair of vectors. We say that a candidate point represents a T-junction if there is a pair of vectors such that the angle between them is equal to π with precision $\frac{1}{4}$. The third criterion relies on the the assumption that object contours and thus T-junction branches are smoothed. A branch is considered to be smooth if the



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

Fig. 3. (a) Partitioned image (b) Associated DG (c) Associated DAG (d) Hasse diagram resulting from the transitive reduction of the DAG

value of the integral of the absolute curvature on the three branches is below to a given threshold k, which is computed as follows: $k = wc$, where c corresponds to a curvature close to zero. w addresses the *scale* issue. Its minimum value is 16.

3.1.3. Cluster reduction

As result of the validation step we often obtain clusters of validated points (Fig. 2 (f)). In fact, the wedge shape of neighbors candidate points varies smoothly and if a candidate point is validated, its neighbors have a high probability of being validated too. We solve this problem by selecting, for each cluster, the candidate point that minimizes the sum of the branch curvature.

3.2. Segmentation

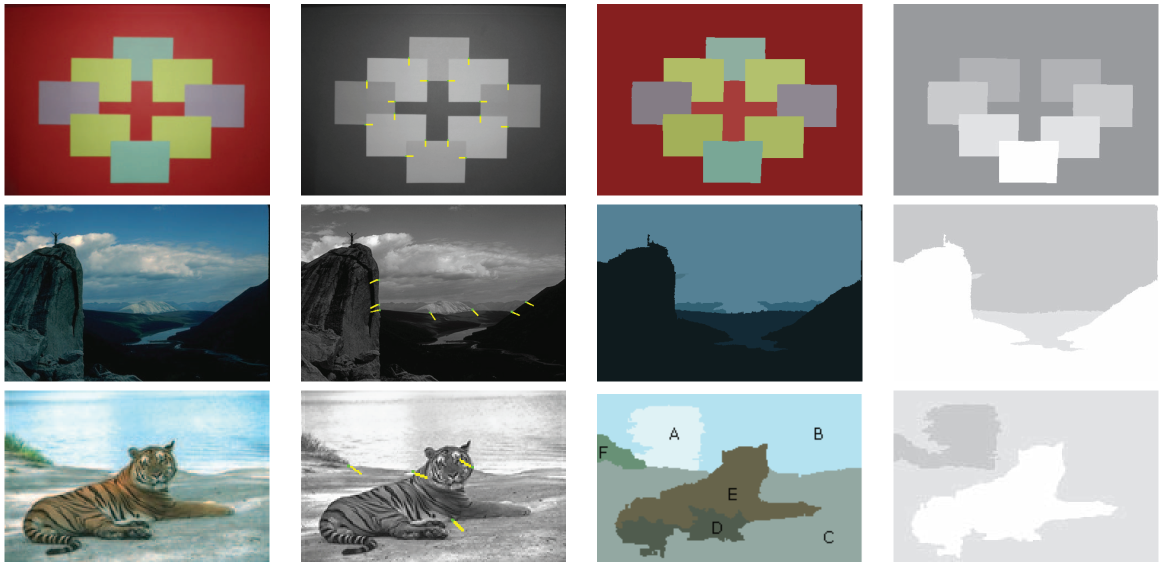
The next step consists in computing a partition of the image through region merging but preserving T-junctions. The merging order we use is mainly defined by a statistical measure on the regions [18] and each region is modeled by its color histogram. In order to preserve T-junctions, we introduce the concept of *incompatibility*. Two regions are said *incompatible* if they are wedges of a T-junction and therefore cannot be merged. As will be shown in Section 4, the incompatibility derived from T-junctions improves the segmentation quality since it preserves low-level data structure.

3.3. Depth Ordering

The goal of this last step is to construct a global and *consistent* interpretation from the local depth assessment previously obtained. In Fig. 3 (a) there is an example of first order inconsistency (involving a pair of T-junctions). Region C is in front of region A for one T-junction, while the converse is true for the other. Higher order inconsistencies involve more than two T-junctions. We formalize the problem of finding a global consistent solution through a DG. A DG is specified by $DG = (V, E_A, A)$, where V is a set of nodes, E is a set of edges and A is the matrix of weights attached to the edges. In our formalization, each node represents an image region and each directed edge represents the relative depth relation between two regions. Edges are specified as ordered pairs: an edge $e = (x, y) \in E$ is considered to be directed from x to y meaning that the region x is in front of the region y. The weight attached to each edge corresponds to 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. With this formalization, local constraint are allowed to propagate along the graph and the search for inconsistent pairs of T-junctions is reduced to the search of cycles on the DG (dashed thick red arrows in Fig.3(b)). The search for directed cycles is performed by a Depth-First Search (DFS) algorithm [19]. 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.3(b)) associated with the T-junction from which the unreliable depth relation arise is also removed. As a result, a DAG is obtained. The depth map is exactly the Hasse Diagram (HD) corresponding to the transitive reduction of the DAG. The transitive reduction of a DAG removes redundant edges while maintaining identical reachability properties. An edge $e(x, y)$ is said *redundant* if there is a path from x to y that does not contain the edge. The HD is constructed as follows: for $e(x, y)$ belonging to the DAG, x is positioned higher than y; the edge $e(x, y)$ is drawn if and only if it is not redundant. The transitive reduction of a finite DAG is unique. Since there is no depth order between the regions forming the stem of a T-junction, they appear on the HD as 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.

4. EXPERIMENTAL RESULTS

We tested our algorithm on a set of real images. For each experiment we show four 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). In all experiments, we segmented the image until only regions involved in at least one occlusion relation are obtained(Fig.4(c)). As can be observed in all examples (Fig.4(d)), the last level of depth includes two or more regions, corresponding to leave nodes of the HD. In the example on the last row, there is a case of conflict between the regions E and C: while region E 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



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

Fig. 4. Example of depth segregation.(a) Original image (b) T-junction detection (c) Segmentation (d) Depth ordering

to a correct depth interpretation. The results obtained for the last two examples would be improved by integrating into the proposed framework more monocular depth cues, such as convexity, size cues and texture gradient.

5. CONCLUSIONS

In this paper, we have proposed a new approach for depth segregation in single images. Contrary to the state of art, our method is fully automatic and does not make any assumption on the image structure. Local depth ordering arising from T-junctions is incorporated into a region merging process avoiding regions in occlusion to merge. The regions of the final partition and their relative depth relations are then encoded through a DG. Using this formalization, possible conflicting interpretations are easily detected as cycles on the DG and solved leading to a DAG. The depthmap is then obtained as the Hasse Diagram corresponding to the transitive reduction of the DAG. Moreover, beside T-junctions, the proposed framework can incorporate several monocular depth cues such as convexity, size, texture gradient, etc. Future work will be devoted to the inclusion of some of these cues to improve the robustness of the proposed method and to obtain a more detailed depthmap.

6. REFERENCES

[1] W. Metzger, *Gesetze des Sehens*, Waldemar Kramer, 1975.

[2] A. Baumberg, "Reliable Feature Matching Across Widely Separated Views," in *CVPR*, 2000, pp. 774–781.

[3] P. Favaro et al., "On exploiting Occlusions in Multiple-view Geometry," in *ICCV*, 2003, pp. 479–486.

[4] N. Apostoloff and A. Fitzgibbon, "Automatic video segmentation using spatiotemporal T-junctions," in *British Machine Video Conference*, 2006, pp. 1–10.

[5] P.J. Kellman and T.F. Shipley, "Visual interpolation in object perception," *Current Directions in Psychological Science*, vol. 1, no. 6, pp. 193–199, 1991.

[6] J. Malik, "Interpreting line drawing of curved objects," *IJCV*, vol. 1(1), pp. 73–103, 1987.

[7] D.L. Waltz, *The Psychology of Computer Vision*, chapter Understanding line drawing of scenes with shadows, pp. 19–91, McGraw-Hill,NY, 1975.

[8] M. Clowes, "On seeing things," *Artificial Intelligence*, vol. 2(1), pp. 79–116, 1971.

[9] K. Sugihara, "An algebraic approach to the shape-from image problem," *Artificial Intelligence*, vol. 23, pp. 59–95, 1984.

[10] X.Y. Stella et al., "A hierarchical markov random field model for figure-ground segregation," in *CVPR*, 2001, pp. 110–133.

[11] R.-X. Gao et al., "Bayesian inference for layer representation with mixed markov random field," *LNCS*, vol. 4679, pp. 213–224, 2007.

[12] X. Ren et al., "Figure/ground assignment in natural images," in *ECCV*, 2006, pp. 614–627.

[13] A. Saxena et al., "Learning 3-d scene structure from a single still image," in *ICCV*, 2007, pp. 1–8.

[14] D. Hoiem et al., "Recovering Occlusion Boundaries from a Single Image," in *ICCV*, 2007, pp. 1–8.

[15] S. Wuerger et al., "On the visual perceived direction of motion, Hans Wallach: 60 years later," *Vision Res.*, vol. 25, pp. 317–367, 1996.

[16] P. Maragos and G. Evangelopoulos, "Leveling cartoons, texture energy markers, and image decomposition," in *8th Int. Symp. on Mathematical Morphology*, 2007, vol. I, pp. 125–138.

[17] S.M. Smith and M. Brady, "Susan - a new approach to low level image processing," *IJCV*, vol. 23, no. 1, pp. 45–78, 1997.

[18] F. Calderero and F. Marques, "General region merging approaches based on information theory statistical measure," in *ICIP*, San Diego (USA), October 2008.

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