SEMANTIC IMAGE SEGMENTATION USING MORPHOLOGICAL TOOLS

*Alvaro Pardo**

IMERL & IIE
Facultad de Ingeniería
Universidad de la República
C.C. 30, Montevideo, Uruguay

ABSTRACT

In this work, we study the extraction of semantic objects using morphological tools. We decompose the image into its level sets and level lines (the borders of the level sets). Specifically, from all the level lines we extract the ones that contain T-junctions, have compact form, and are well-contrasted to obtain the semantic objects in the scene. For all these factors, we establish an estimation procedure.

1. INTRODUCTION

The extraction of semantic objects from images is one of the most important and challenging problems in image analysis. The applications of such systems range from object based compression to image databases.

A large number of algorithms have been proposed in the literature. Although it is difficult to classify all of them, we will describe their general structure [7]. A segmentation algorithm may be composed of the following three basic steps: **Simplification:** Removes from the original image all the irrelevant information for the specific application. **Feature extraction:** The specific features (grey level, texture, etc.) of the data drive the segmentation. The kind of features depends on the application and on the desired segmentation. **Decision:** The final partition is determined using the features obtained in the previous step.

For the decision step we have: the transition-based techniques and the homogeneity-based techniques [7]. The formers intend to estimate the position of the discontinuities in the feature space defining the boundary of the regions. On the other hand, the later ones look for homogenous regions in the feature space.

Here we address the problem of extraction of semantic structures in the image. We will not consider the post-processing of the obtained segmentation. In our case, the basic structures are the level sets and the basic features are the T-junctions. In some way our algorithm falls into the transition-based category, after the detection of the basic elements the ones with "more" features will be set as part of the segmentation.

*(apardo@iie.edu.uy) Supported by CSIC - Universidad de la República and Fondo Clemente Estable 200 No. 6034 - Conicyt.

2. SEGMENTATION ALGORITHM

To obtain a partition with regions matching as well as possible our perception we use elements of mathematical morphology recently developed in [1, 6, 3, 2].

Before going on to discuss the algorithm used in this work, we review some basic concepts and notation. (For details see [1, 6, 3, 2].) Given an image $I : \Omega \to \mathbb{R}$ we define the lower and upper level sets as:

$$L_\lambda = \{x \in \Omega : I(x) \leq \lambda\} \quad (1)$$
$$U_\mu = \{x \in \Omega : I(x) \geq \mu\} \quad (2)$$

Level sets define a decomposition with two remarkable properties. Firstly, it is a complete decomposition; all the image information is contained on its level sets, and we can reconstruct the original image from (1) and (2) by:

$$I(x) = \sup\{\mu : x \in U_\mu\} = \inf\{\lambda : x \in L_\lambda\}$$

Second, the decomposition is contrast invariant. Gestaltists argue that our perception is contrast invariant so such decomposition seems to be a correct one if we want to be compatible with this theory.

Level sets satisfy in addition the following very important property: the family of upper and lower level sets are decreasing and increasing respectively: if $\alpha < \beta$ then $L_\alpha \subset L_\beta$ and $U_\beta \subset U_\alpha$. That means that level sets are included in each other. This relation can be represented as a tree of lower and upper level sets [6].

If we consider connected components of the level sets and the level lines defined by their borders, we can define the *topographic map as the family of all its level lines [1]*. Following [3] we call *morphological edges the level lines that have a perceptive significance*.

The topographic map of an image contains all its information, however, this information is usually somehow redundant; not all the level lines are equally relevant. The basic idea is then to filter the topographic map, the set of all level lines, to obtain a smaller set of morphological edges. To characterize the filtering process we need to define a set of filtering criterions that permit the extraction of the morphological edges. We propose three filtering criterions. A level line is a morphological edge if: contains T-junctions, has a compact form, and it is "well-contrasted."

2.1. T-junctions

According to the Gestalt school of visual perception, T-junction singularities play a major role in our perception. They are of crucial importance regarding the reconstruction of occlusions; T-junctions appear at the borders of two objects that are occluding each other. From these T-junctions our visual perception reconstructs the occluded object while extending its border to join the T-junctions. In [4] Kanizsa presents a lot of concluding examples and beautiful drawings that support the use of T-junctions as effective perceptive features. Based on these observations, Caselles, Morel and Coll developed a framework compatible with Kanisza's ideas for the case of digital images [1]. Their main conjecture is that the topographic map and the junctions are the "atoms" of visual perception. From these ideas, Froment [3, 2] developed a segmentation algorithm that uses level lines joining T-junctions as basic elements.

Intuitively, in the case of a digital image, a T-junction occurs when two level lines meet. Specifically, we have a T-junction when we can define three significant sets on the neighbourhood of an intersection of two level lines: two sets belonging to the occluding objects, and one to the background.

T-junction detection algorithm [1]
Let $\mathcal{N}(x)$ be the neighbourhood of x where two level lines join. We define:

$$\begin{aligned} x_{\lambda_0} &= \text{argmin}\{I(y) : y \in \mathcal{N}(x)\}, \ \ \lambda_0 = I(x_{\lambda_0}) \\ x_{\mu_0} &= \text{argmax}\{I(y) : y \in \mathcal{N}(x)\}, \ \ \mu_0 = I(x_{\mu_0}). \end{aligned}$$

The set $L_{\lambda_0}^{x_{\lambda_0}}$ is a connected component of L_{λ_0}, which contains x_{λ_0}, and similarly we define $U_{\mu_0}^{x_{\mu_0}}$. In order to say when these connected components are relevant, i.e. they define a T-junction, we ask them to have a minimum area. Also we define the minimum and maximum grey levels in those sets:

$$\begin{aligned} \lambda_1 &= \text{argmin}\{\lambda \geq \lambda_0 : \text{Area}(L_{\lambda_0}^{x_{\lambda_0}}) \geq A\} \\ \mu_1 &= \text{argmin}\{\lambda_0 < \mu \leq \mu_0 : \text{Area}(U_{\mu_0}^{x_{\mu_0}}) \geq A\}. \end{aligned}$$

Finally, x is a T-junction if the two sets are well-contrasted, $\mu_1 - \lambda_1 \geq 2G$, and the background set defined by the connected component of $\{y \in \Omega : \lambda_1 + G \leq I(y) \leq \mu_1 - G\}$ containing at least a pixel of $\mathcal{N}(x)$ has area greater than A.

The border of a level set, its level line, is a union of closed curves; a shape is the interior of such curves [6].

The first criterion of the filtering process keeps only the shapes that contain T-junctions. Furthermore, the more T-junctions a shape has in its border the more important it is. The number of T-junctions can be used as a metric to sort shapes. In Figure 1 we show the detected T-junctions for Claire image.

2.2. Compactness

According to the Gestalt theory, in our field of view we distinguish figure from ground. Figures are perceived as a coherent whole in front of the background, which is perceived as less important. In addition, our perception favours objects with simple and compact form to be perceived as foreground [5]. Then, compactness is an important property when segregating the image into foreground and background. To measure the compactness of a shape $\mathcal{S}$ we use, as is classical in the computer vision literature, the isoperimetric ratio:

$$\text{Isoperimetric Ratio}(\mathcal{S}) = \text{Perimeter}(\mathcal{S})^2/\text{Area}(\mathcal{S})$$

This measure roughly says that between two shapes of equal area, the one of least perimeter will be the more compact one; it penalizes shapes with complex oscillating borders. This criterion is both, a filtering tool and a sorting metric. Meaning that, shapes with an isoperimetric ratio above a given threshold are deleted and, the smallest isoperimetric ratio it has the more important it is.

Remark: A stimulus could be important but not compact, otherwise, we would only see simple forms. The point is that compact forms tend to: 1- attract our perception, 2- be part of semantic objects. This point was discussed by Kanizsa in [4], where he pointed out the misunderstanding regarding the so called principle of "good form".

2.3. Contrast

Typically, well-contrasted regions call our attention. Moreover, a well-contrasted shape is likely to be part of the boundary of a real object in the image. Thus, contrast is another important feature to define perceptive objects. To compute the contrast along the shape border we use the magnitude of the gradient. The level line defining a shape may have part of it inside the object (Along this part, the magnitude of the gradient will be smaller than the magnitude along the object border.) If we use, for example, the mean of the gradient along the shape border we could end up with an unreliable estimation of the contrast. To avoid these problems, we use the median of the gradient along the shape border as a robust measure of contrast.

$$\text{Contrast}(\mathcal{S}) = \text{median}\,\{|\nabla I(x)| : x \in \partial\mathcal{S}\}$$

It is worth to note that the contrast is the least important factor among the three proposed. The reason for that is twofold. First, the contrast is considered for compact shapes with T-junctions. Second, the removed shapes are typically the low-contrasted shapes that do not play an important role in the perception of the image, *soft* morphological edges, shapes in smooth areas, or just noisy shapes.

Summary of the algorithm (fully automatic)

1- Compute the lower and upper trees. The computation of these trees is performed with an algorithm similar to the one proposed in [6].

2- Find all the T-junctions in the image using the algorithm discussed in (2.1).

3- Remove all the shapes, $\mathcal{S}$, in both trees that:

- **3.a-** Have less than T T-junctions.
- **3.b-** (Small shapes) Perimeter$(\mathcal{S}) < P$.
- **3.c-** (Complex shapes) Isoperimetric Ratio$(\mathcal{S}) > IR$.
- **3.d-** (Bad-Cotrasted) Contrast$(\mathcal{S}) < C$.

4- Sort the shapes, firstly according to the number of T-junctions and then increasingly with the isoperimetric ratio. That is, if two shapes have equal number of T-junctions the more compact one is selected as the more important.

5- Add the most important lower and upper shapes to the segmentation. If Tj$(\mathcal{S})$ is the set of T-junctions in the shape, remove them from the remaining shapes. (This step step avoids the inclusion of several shapes that contain nearly the same T-junctions and accumulate close to objects borders.)

Sometimes it is unnecessary to include in the segmentation all the shapes with T-junctions. At the end of the day, no matter how complex and accurate the algorithm could be, the user judgement is crucial to define the end of the process. Therefore, another possibility is to add new shapes until the user stops the process. In this case, the algorithm is semi-automatic (the user interaction is minimal).

The last possibility is to include all shapes without removing the already included T-junctions. The drawback is that several level lines accumulate close to the objects boundary. In [2] Froment proposed this as a multiscale image model. The weakness of this algorithm is that too many level lines tend to accumulate close to the objects boundary. This makes the algorithm not very suitable for further region-based processing; it is harder to obtain a simple segmentation from it.

2.4. Parameter tuning

For the parameters A and G, which control the detection of the T-junctions, we found empirically the values $A = 40$ and $G = 4$ for images of dimensions 256x256, and $A = 30$ and $G = 4$ for images in QCIF format. We encountered little changes when moving these parameters close to the previous ones. In some cases, the T-junctions were nearly the same.

The parameter T is like a scale, the more T-junctions a shape has the more important it is. For this reason, it can be used to obtain segmentations at different resolutions. Yet, we set T greater than two as shapes with only one T-junction are possible be due to noise.

Like the number of T-junctions, the perimeter defines also a scale; shapes with small perimeter constitute the fine scale. In our case we set empirically $P = 20$. Sometimes it can be useful to set a maximum allowed perimeter too, in such a case it can be determined in the same way as the isoperimetric ratio (see below.)

As for the parameter IR that controls the isoperimetric ratio of the shapes it is clearly image dependent. Different images have different complexity and therefore different values of IR. Usually for non-complex images, in terms of the shapes it contains, its value is in the range $100 - 200$. This parameter is the most critical one: if IR is set to high then we could end up adding noisy shapes to the segmentation and, if it is too small, we could loose some important shapes. Likewise, the contrast threshold C, is also image dependent. For both we estimate their values using the statistics of all the morphological edges. In what follows, we discuss the procedure to estimate them.

Since edges in images are not perfect, they do not form step functions but smooth transitions, several level lines accumulate close to the object border. In this way, their contrast, perimeter and isoperimetric ratio are similar. They will also have nearly the same T-junctions. Because of this simple property, we have that the perimeter, isoperimetric ration, and contrast features form clusters.

The isoperimetric ratio threshold, IR, is derived from the statistics of the isoperimetric ration. We use upper and lower shapes to obtain the distributions of the isoperimetric ratio $F_u(\mathcal{S})$ and $F_l(\mathcal{S})$ respectively [1]. The conservative heuristic seeks a small value for IR which does not leave important shapes out. We compute it as the maximum of the points where the distributions of lower and upper shapes equal 0.8 (Probability(Isoperimetric Ratio$(\mathcal{S}) > IR) < 0.2$.) That is, we set IR so to leave out the shapes which isoperimetric ratio has a low probability to occur within the image.

$$IR = \max \left\{ F_l^{-1}(0.8), F_u^{-1}(0.8) \right\}$$

Take the distribution of the contrast, F_c(Contrast$(\mathcal{S})$), for all shapes with more than one T-junction and isoperimetric ration below IR (We apply the same methodology to upper and lower shapes.) The first cluster in the distribution corresponds to the shapes with the smallest contrast. If we consider just the contrast, these shapes are the ones of least relevance. Let C_1 and C_2 be the points where the two

[1] to obtain the distribution we apply a standard kernel method.

first maximum of the contrast distribution occur, and C_1^m the first minimum after C_1. We set C to:

$$C = \max\left\{\min\left\{0.5 * (C_1^m + C_2), F_c^{-1}(0.2)\right\}, 10\right\}$$

This is a very conservative strategy; other values larger than this one produce also good, coarser segmentations.

3. RESULTS

In Figure 1 we present the results obtained with the fully automatic algorithm. As we can see, the segmentation correctly captures the perceptive objects in the image. In some cases, the detected objects have a noisy boundary. This is due to the nature of the level lines close to object and to the amount of noise present. In a forthcoming work, we will propose a way to simplify the detected boundaries.

The second experiment intends to test the importance of the first shapes added to the segmentation. We show the first shapes extracted for Claire and Foreman image. In the case of Clare image, the most important shapes match perfectly our perception. On the other hand, for the Foreman image, the most important region does not fully match our perception. According to our perception, the most important object is mans head, and although the most important region dos not segment it completely it contains it. It is important to note that we detect shapes based on rather simple features and, when detecting mans head, the strongest element for our perception is the fact that it is a head.

4. CONCLUSIONS

In this work, we described algorithms for the extraction of semantic objects in images using morphological tools. We used T-junction singularities together with contrast and compactness measures to select the perceptive shapes. We base all this on perceptual considerations linked with the Gestalt school of visual perception.

We presented three algorithms ranging from fully automatic to semi-automatic. In any case, the user interaction is minimal as we gave explicit methods to automatically estimate the algorithm parameters.

The experimental results showed: the importance of T-junctions as perceptive features, the effectiveness of both contrast and isoperimetric threshold, and the good performance of the described algorithms.

5. REFERENCES

[1] V. Caselles, B. Coll, and J.-M. Morel. Topographic Maps and Local Contrast Changes in Natural Images. *Int. Journal of Comp. Vision*, 33(1):5–27, Sep. 1999.

[2] J. Froment. A Compact and Multiscale Image Model Based on Level Sets. In *Proc. of Sec. Int. Conf. Scale-Space'99*, Num. 1682 in Lecture Notes in Computer Science, pages 152–163. Springer, 1999.

[3] J. Froment. A Functional Analysis Model for Natural Images Permitting Structured Compression. *ESAIM : COCV*, 4:473–495, Aug. 1999.

[4] G. Kanisza. *Organization in Vision*. Praeger, 1979.

[5] W. Kohler. *Gestalt Psychology*. Liveright Publishing Corporation, 1947.

[6] P. Monasse and F. Guichard. Fast Computation of Contrast-Invariant Image Representation. *IEEE Trans. Image Processing*, 9(5):860–872, May 2000.

[7] P. Salembier and F. Marqués. Region-Based Representations of Image and Video: Segmentation Tools for Multimedia Services. *IEEE Trans. on Circuits Syst.Video Technology*, 9(8):1147–1169, Dec. 1999.

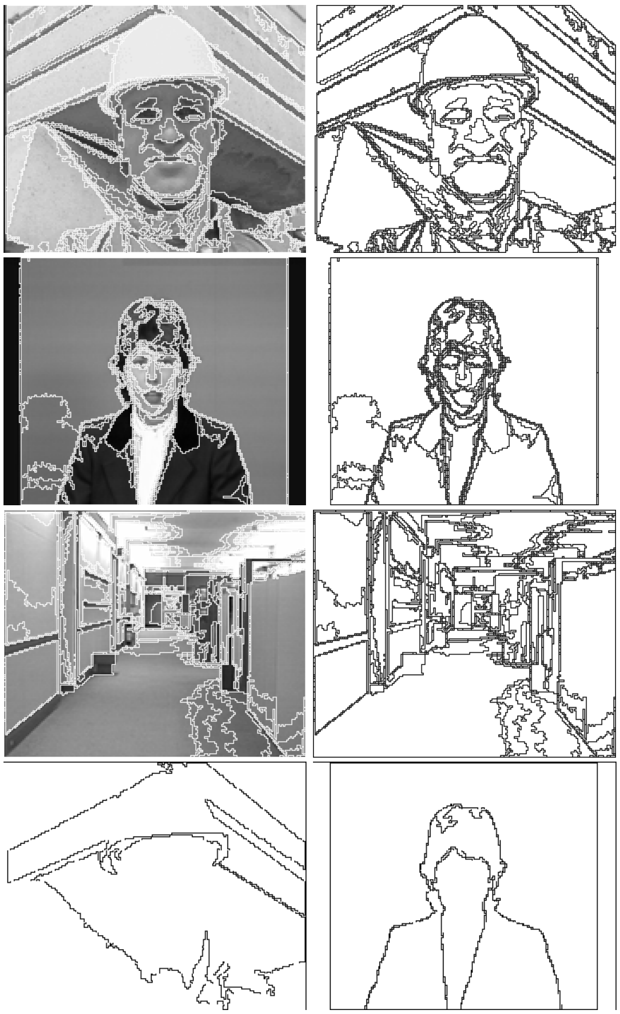

Fig. 1. Detected shapes for Foreman, Claire and Hall images, the first shapes included for Claire and Foreman image. We also show for the first three images the shapes over imposed to the image.