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 from images together with a metric that ranks them according to their perceptual significance. To obtain an initial segmentation we use elements of mathematical morphology (level sets and level lines) and some properties such as T-junctions, contrast and compactness. Then, to refine the initial partition we apply regularization techniques and a standard merging algorithm. Finally, we compute a perceptual metric using factors that influence our perception.
Keywords: Image segmentation, morphological tools, perceptual metric.

1 Introduction

The extraction of semantic objects is one of the most important and challenging problems in image analysis. Nowadays, we have several applications where this kind of processing is utterly needed. A first application of these ideas is compression where different objects are coded with different quality (this is the case of the MPEG4 standard). In a second group we have applications that allow the manipulation and interaction with the objects in the image. For instance, in the case of image databases systems, the objects can be used to do queries in the database. We can establish the need of objects in image analysis from another point of view. Human beings see objects in the real world, neither pixels nor homogenous regions; therefore, systems that pretend to be compatible with human perception must be based on objects.

The organization of our method, and the present paper, is as follows. First, we extract the semantic structures in the image using elements of mathematical morphology. The basic structures in this step are the level sets and the basic features used to obtain the ones with semantic meaning are: T-junctions, compactness and contrast. This initial segmentation is then refined using a first step of regularization followed by a standard merging algorithm. Finally, using the final segmentation we compute the perceptual metric using elements that influence our perception.

2 Extraction of semantic structures

The first step of the algorithm deals with the computation of a good perceptual partition of the image domain. For that end, we use elements of mathematical morphology recently developed

*(apardo@iie.edu.uy) Supported by CSIC - Universidad de la República and Fondo Clemente Estable 2000 - Conicyt - Proyect No. 6034.

in [1, 2, 8, 4, 3]. Before going on to discuss the algorithm used in this work, we shall first review some basic concepts and notation. (For a more detailed discussion of the following topics, we refer the reader to [2, 8, 4, 3].) Given an image $I : \Omega(\subset \mathbb{R}^2) \rightarrow \mathbb{R}$ we define lower and upper level sets as:

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

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) 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 do 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 [8].

If we consider connected components of the level sets and the level lines defined by their borders, we can define the *topographic map* of an image as *the family of all its level lines.* Also, we define each connected component of the level sets as a *shape.* Note that the level line of a level set is the union of closed curves and the interior of each curve is a shape. Finally, following [4] we call *morphological edges* the *level lines that have a perceptive significance.* In other words, morphological edges are level lines enclosing regions of perceptive importance.

The topographic map of an image contains all its information, however, this information is usually somehow redundant as 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 only the morphological edges. We propose three criterions to define the filtering process, a level line is a morphological edge if: contains T-junctions, has a compact form, and it is "well-contrasted." In the remainder of this section, we describe these three elements in detail.

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. As a matter of daily experience, which was extensively analysed by Gestaltist's, from these T-junctions our visual perception reconstructs the occluded object while extending its border to join the T-junctions. In [6] 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 in a series of papers [1, 2] developed a framework compatible with Kanisza's ideas for the case of digital images. Their main conjecture is that the topographic map and the junctions are the "atoms" of visual perception. From these ideas, Froment [4, 3] developed a segmentation algorithm that uses level lines joining T-junctions as basic elements.

To obtain the level lines that contain T-junctions we must first detect them. 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 [2]: Let $\mathcal{N}(x)$ be the neighbourhood of x where two level lines join. We define:

$$x_{\lambda_0} = \text{argmin}\{I(y) : y \in \mathcal{N}(x)\}, \lambda_0 = I(x_{\lambda_0})$$

$$x_{\mu_0} = \text{argmax}\{I(y) : y \in \mathcal{N}(x)\}, \mu_0 = I(x_{\mu_0}).$$

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:

$$\lambda_1 = \text{argmin}\{\lambda \geq \lambda_0 : Area(L_{\lambda_0}^{x_{\lambda_0}}) \geq A\}$$

$$\mu_1 = \text{argmin}\{\lambda_0 < \mu \leq \mu_0 : Area(U_{\mu_0}^{x_{\mu_0}}) \geq A\}.$$

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 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 2 we show the detected T-junctions for Claire image.

2.2 Compactness

Once again, we base the use of this criterion on the Gestalt theory. According to this 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 [7]. 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 [15], 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 removed and, the smallest isoperimetric ratio it has the more important it is. By no means we are saying that our perception always tends to reach the “good form”. In other words, a stimulus could be important but not compact, otherwise, we would only see simple forms. It is clear that this is not the case since we do perceive complex objects. 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 [6], 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, the contrast seems to be another important feature to define perceptive objects. To compute the contrast along the shape border we use the magnitude of the gradient. As we said, 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, i.e. already morphological edges. 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:

1. Compute the lower and upper trees to find all the shapes in the image. The computation of these trees is performed with an algorithm similar to the one proposed in [8].

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:
   a) Have less than T T-junctions over its border.
   b) (Small shapes) Perimeter($\mathcal{S}$) $< P$.
   c) (Complex shapes) Isoperimetric Ratio($\mathcal{S}$) $> IR$.
   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, however, the user interaction it is still minimal. In this case in the last step we add the most important lower and upper shapes to the segmentation while the number of T-junctions is greater than T. If Tj($\mathcal{S}$) is the set of T-junctions in the shape, remove them from the remaining shapes.

The last possibility is to include all shapes without removing the already included T-junctions (We add all lower and upper shapes with more than T T-junctions to the segmentation). In [3] 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.

Let us now discuss a property of the level lines that is useful to understand the proposed estimation of the algorithm parameters. 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. Each one of these clusters containing shapes with similar features, which roughly characterizes objects in the image.

The isoperimetric ratio threshold, IR, is derived from the statistics of the isoperimetric ration. We use upper and lower shapes (connected components of level lines of U^μ and L_λ) 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, $F_c(\text{Contrast}(\mathcal{S}))$, 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 first maximum of the contrast distribution occurs, and C_1^m the first minimum after C_1. We set C to:

$$C = \max\left\{\min\left\{(C_1^m + C_2)/2, F_c^{-1}(0.2)\right\}, 10\right\}$$

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

---

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

3 Segmentation Regularization

We add coherence to the initial segmentation given by the partition induced by morphological edges using the vector probability diffusion scheme (VPD) [13] by adding spatial coherence to the posteriors probabilities of classes present in the image.

We say that a given region from the initial segmentation ,$\{R_1, ..., R_n\}$, is a valid class if its area is bigger than a given threshold. Each class $c_i \in \mathcal{C} = \{c_i : i = 1, ..., m\}$ is represented with the mean of its members: μ_i. For every pixel x we have a probability vector $p(x) \in \mathcal{P} = \{p \in \mathbb{R}^m : \|p\|_1 = 1, p_i \geq 0\}$ where $p_i(x)$ equals the probability of pixel x to belong to the class c_i:

$$p_i(x) = \frac{1}{|I(x) - \mu_i|} \left(\sum_{j=1}^{m} \frac{1}{|I(x) - \mu_j|} \right)^{-1} \tag{2}$$

To add spatial coherence into the classification process VPD diffuses the distance between points in $\mathcal{P}$ with the following diffusion equations:

$$\frac{\partial p_i}{\partial t} = \nabla . \left(\frac{\nabla p_i}{\sqrt{\sum_{i=1}^{m} \|\nabla p_i\|^2}} \right) \quad i = 1, ..., m.$$

The important property of these diffusion equations is that they guarantee that $p(t) \in \mathcal{P}$ for all t and is therefore not necessary to project them into the probability hyper-plane. This is true for the numerical implementation as well. For further details and implementation see [13].

4 Region Merging Algorithm

The region merging algorithm works on the Region Adjacency Graph (RAG) [15]. The RAG is a set of nodes, each one representing a region of the partition, and a set of links connecting neighbouring nodes (regions). The RAG is a good data structure for region merging processes because it easily codes the connectivity, and the distance between nodes.

To apply the merging algorithm we need to define the region model and the merging criterion, which depends on a distance between regions.

The region model μ_i is defined as the mean grey level of the pixels in the region R_i. When two regions are merged, the new model must be computed. To make the model estimation robust, the new model equals the one of the bigger region [5].

$$\mu_{R_1 \cup R_2} = \begin{cases} \text{if Area}(R_1) < \text{Area}(R_2) & \mu_{R_1 \cup R_2} = \mu_{R_2} \\ \text{if Area}(R_1) > \text{Area}(R_2) & \mu_{R_1 \cup R_2} = \mu_{R_1} \\ \text{if Area}(R_1) = \text{Area}(R_2) & \mu_{R_1 \cup R_2} = (\mu_{R_2} + \mu_{R_1})/2 \end{cases}$$

The merging criterion minimizes the cost of each merging. That is, en each step we minimize the cost function (3) merging the pair of nodes with minimum cost.

$$C(R_1, R_2) = P(R_1)D(R_1, R_1 \cup R_2) + P(R_2)D(R_2, R_1 \cup R_2) \tag{3}$$

D is the distance between regions: $D(R_i, R_j) = |\mu_i - \mu_j|^2$, and $P(R_i)$ is the probability of region R_i: $P(R_i) = \text{Area}(R_i)/\text{Area}(\Omega)$. This cost function measures the error between the given partition and the new model.

5 Perceptual Metric

In this section, we present a perceptual metric to automatically determine the perceptual importance of different regions in the image. This metric is based on Osberger's works [10, 9] and uses several features that influence human visual attention. For each region in the image a set of features is computed and then combined to obtain the importance map. This map classifies each region of the image with respect to its perceptual importance.

Firstly, to apply this idea we need a segmentation of the image. This point is crucial; regions in the segmentation should represent semantic regions or part of them, otherwise, the perceptual metric will not correlate with our perception. For this reason, we do not use the initial segmentation to compute the importance map as it contains too many small regions. Instead, we compute the perceptual metric using a coarser segmented image, the one obtained after some steps of the merging algorithm. We will come back to this point later.

5.1 Factors which influence our attention

These factors can be classified into: low level and high level. Among low-level factors we have: contrast, size, shape, and colour. High-level factors are of course more difficult to model. For instance, the presence of people in the image is a strong factor; our attention is drawn to their eyes, mouth, and hands. In our case, we use location and the distinction of foreground and background as high-level factors.

Contrast: Region contrast is a very strong factor; regions with high contrast with their neighbour regions attract our attention, and therefore they might belong to regions of perceptual importance. The contrast of a region R_i which has a set of neighbours $\{R_{i_1}, ..., R_{i_N}\}$ is computed as:

$$\text{Contrast}(R_i) = \frac{1}{N}\sum_{j=1}^{N} \alpha_j |\mu_i - \mu_{i_j}| \quad \alpha_j = \frac{\text{Length}(\partial R_i \cap \partial R_{i_j})}{\text{Perimeter}(R_i)}$$

where μ_{i_j} are the means of the regions R_{i_j} and α_j is a factor that weights the contribution of each neighbouring region to the contrast measure. The idea is that, the more contact between regions the more it should contribute to the contrast measure.

Osberger measures the contrast as the difference of the mean of a region and the mean of the neighbour regions. This is not a robust measure since the mean of the neighbouring regions is strongly affected by an "outlier" region. For example, take a region with mean 128 and two neighbours with means 255 and 0. In this case the mean of the neighbouring regions equals the mean of the region and therefore according to Osberger metric the contrast will be 0. Obviously, this does not reflect what we perceive. With our definition of contrast (neglecting the factors α_j), the contrast is 128.

Size: It has been found that region size is an important factor. Large regions are more likely to attract our attention than the small ones. The size measure is computed as:

$$\text{Size}(R_i) = \max\{\text{Area}(R_i)/A_{max}, 1\}$$

where A_{max} is set to the 1% of the total area and is used to prevent excessive weighting to very large regions.

Shape: It has been argued that long and thin regions are visual attractors [14], but also that our perception tends to favour compact regions [7]. Osberger applies the first idea an computes the shape factor as: $\text{Perimeter}(R_i)^{1.75}/\text{Area}(R_i)$ trying to capture long and thin regions.

Conversely, we apply the second idea using the isoperimetric ratio of the region which scores compact regions as more important. According to our experiments, this selection performs better. Note that the isoperimetric ratio is nearly the inverse of Osberger's measure.

$$\text{Shape}(R_i) = \text{Area}(R_i)/\text{Perimeter}(R_i)^2$$

Foreground/Background: Typically, objects in the foreground attract our attention. To determine if a region is part of the background we measure the number of pixels of the region border that belong to the image border. In this way the foreground/background measure is computed as:

$$\text{FB}(R_i) = 1 - \min\{\text{Length}(\partial R_i \cap \partial\Omega)/(0.5 * \text{Perimeter}(\Omega)), 1\}$$

Location: Different experiments have shown that typically viewers focus at the centre of the image. To compute this factor we measure the number of pixels of the region which are within the 25% centre of the image: $\text{Centre}(R_i)$. Regions that are entirely in the centre of the image will have the maximum weight.

$$\text{Location}(R_i) = \text{Centre}(R_i)/\text{Area}(R_i)$$

5.2 Importance Map

After normalizing each of the factors presented above to the range $[0, 1]$ the importance map is computed as the sum of their squared values. This assigns higher scores to regions with high scores in some factors. In Figures 3 and 1 we present examples of the importance map.

6 Algorithm

Here we describe the segmentation algorithm using the ideas presented in previous sections.

1. Given the initial partition we apply the VPD to add coherence to the initial segmentation. This step will not only regularize the given segmentation but it could also produce new regions. We find the valid classes, using an area threshold A_{vpd}, and their means μ_i, and the probabilities using equation (2). After the VPD we recompute the partition.

2. Apply the merging algorithm until we obtain the desired number of regions, R_{th_1}. In this step, the user interaction is important to stop the merging when the segmentation captures the semantic objects in the image.

3. Compute the perceptual metric.

7 Results

In the first example, Figures 2 and 3, we present the result for Claire image. In Figure 2 we present the detected Tjunctions and the semantic structures obtained with the algorithm of section 2 (The set of parameters used was ($G = 4$,$A = 30$,$IR = 105$,$P(min) = 20$,$P(max) = 1500$,$C_{lower} = 10$,$C_{upper} = 25$, $T = 2$). As we can see the extracted level lines match the semantic objects in the image. For an exhaustive analysis of this algorithm we refer the reader to [12]. Then, we show the results after regularization. As we can see, we greatly reduce

Figure 1: Results for Hall and Carphone images. From left to right: original images, segmented images obtained with our algorithm, and importance map.

the number of regions without loosing the principal structures. Mainly, we removed the small regions. Finally, in the last row of Figure 2 we show the semantic objects after the merging process. In Figure 3 we display the individual factors and the resulting importance map. As we can see the results the metric correctly captures the most important objects. Although, this may be a simple image, this kind of images is usually encountered in videoconference applications. Therefore this metric can be used to tune the video encoding process and assign more resources to the perceptually relevant parts of the image. For more details related with the simplification and merging process we refer the reader to [11].

Finally, in Figure 1 we show the final segmentation results together with the importance metric for Hall and Carphone image. Although these are more complex images, the algorithm performs fairly well. We believe that the inclusion of colour information in the merging process will improve the results. We are currently working on the inclusion of colour in our algorithms.

8 Conclusions

In this work we described algorithms for the extraction of semantic objects in images. This was done using morphological tools: level sets and level lines as basic structures and T-junction singularities together with contrast and compactness measures as features to detect the semantic structures. We base all this on perceptual consideration linked with the Gestalt school of visual perception. Then, we regularized the initial segmentation and performed a standard region-merging algorithm to obtain the relevant semantic objects in the image.

The algorithms presented succeed on the extraction of semantic objects; however, depending on the complexity of the image the results are of different quality. This can be observed in images Claire and Carphone. The later is more complex and the results are not as good as for former one. As we said, we believe that the inclusion of colour will improve this. Furthermore, as we can see from Figures 1 and 3 the perceptual metric roughly captures our subjective perception of the objects in the image.

References

[1] V. Caselles, B. Coll, and J.-M. Morel. A Kanisza Programme. *Progress in Nonlinear Differential Equations and their applications*, 25:35–55, 1996.

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

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

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

[5] L. Garrido, P. Salembier, and D. Garcia. Extensive operators in partition lattices for image sequence analysis. *Signal Processing:Special Issue on Video Sequence Segmentation*, 66(2):157–180, April 1998.

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

[7] W. Kohler. *Gestalt Psychology.* Liveright Publishin Corporation, 1947.

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

[9] W. Osberger. *Perceptual Vision Models for Picture Quality Assessment and Compression Applications.* PhD thesis, School of Electrical and Electronic Systems Engineering-Queensland University of Technology, Brisbane, Australia, March 1999.

[10] W. Osberger and A. J. Maeder. Automatic Identification of Perceptually Important Regions in an Image. In *International Conference on Pattern Recognition*, volume 1, pages 701–704, 1998.

[11] A. Pardo. Extraction of Semantic Objects from Still Images. In *International Conference on Image Processing ICIP2002*, 2002.

[12] A. Pardo. Semantic Image Segmentation using Morphological Tools. In *International Conference on Image Processing ICIP2002*, 2002.

[13] A. Pardo and G. Sapiro. Vector Probability Diffusion. *IEEE Signal Processing Letters*, 8(4):106–109, April 2001.

[14] J. Senders. Distribution of attention in static and dynamic scenes. In *SPIE*, number 3016, pages 186–194, February 1997.

[15] M. Sonka, V. Hlavac, and R. Boyle. *Image Poressing, Analysis, and Machine Vision.* ITP, second edition, 1999.

Figure 2: Original Claire image. Detected Tjunctions. Initial segmentation with 1994 regions, after two iterations of VPD (area threshold is 100 pixels) with 766 regions, and the image with six regions matching the semantic objects. In the last three rows we show the image where each region is represented by its mean and the boundaries.

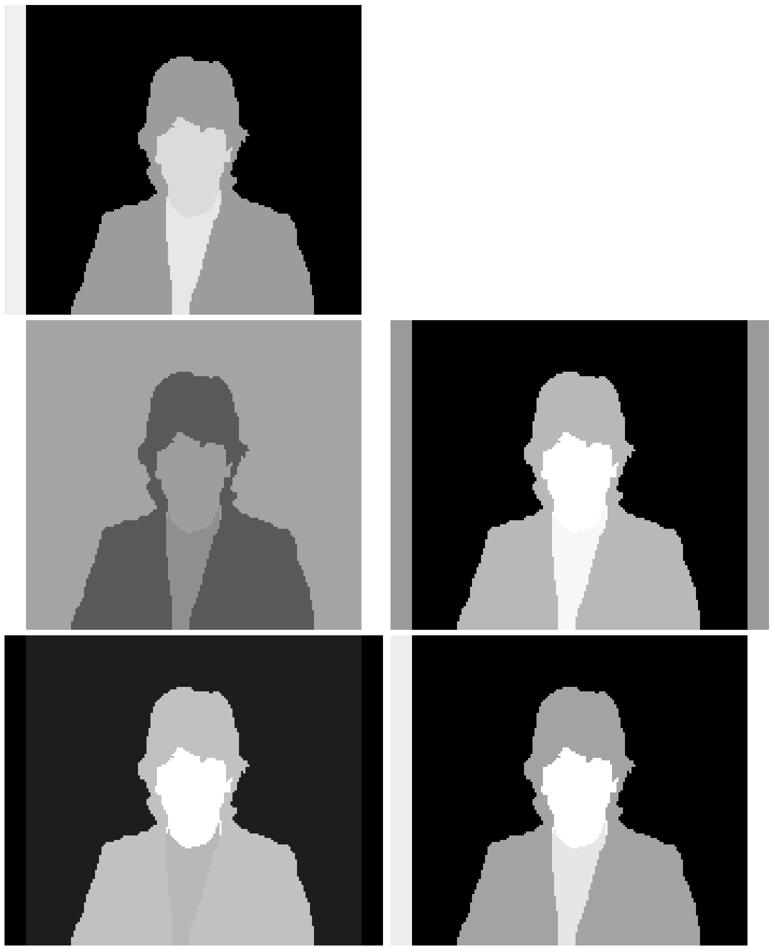

Figure 3: Perceptual metric results for Claire image. From top to bottom: contrast map, size map, shape map, foreground/background map, location map and importance map. Bright indicates an important region. In all cases, the results match our perception. The importance map clearly ranks as the most important region the face followed by the rest of the regions belonging to the woman body and at last the background. In all cases, the images have histogram correction for better visualization.