

# Hybrid Framework to Image Segmentation

Fernando C. Monteiro

Polytechnic Institute of Bragança, Portugal  
monteiro@ipb.pt

**Abstract.** This paper proposes a new hybrid framework to image segmentation which combines edge- and region-based information with spectral techniques through the morphological algorithm of watersheds. The image is represented by a weighted undirected graph, whose nodes correspond to over-segmented regions (atomic regions), instead of pixels, that decreases the complexity of the overall algorithm. In addition, the link weights between the nodes are calculated through the intensity similarities combined with the intervening contours information among atomic regions. We outline a procedure for algorithm evaluation through the comparison with some of the most popular segmentation algorithms: the mean-shift-based algorithm, a multiscale graph based segmentation method, and JSEG method for multiscale segmentation of colour and texture. Experiments on the Berkeley segmentation database indicate that the proposed segmentation framework yields better segmentation results due to its region-based representation.

## 1 Introduction

Image segmentation is an important component in many image understanding algorithms and practical vision systems, and aims at identifying regions that have a specific meaning within images. Another definition of image segmentation is the identification of regions that are uniform with respect to some parameter, such as image intensity or texture. While the latter definition is often used for technical reasons, the former definition should be preferred from an application point of view. Although the effort made in the computer vision community there is no algorithm that is known to be optimum in image segmentation. Much research is being done to discover new methods building up on previous ideas [1, 5, 7].

The main goal of this paper is to develop an algorithm for efficient segmentation of a grey level image that a) identifies perceptually homogeneous regions in the images, b) makes minimal assumptions about the scene, and c) avoids merging of multiple objects into single segments and vice-versa. The definition of a new structure for region-based graph, the presentation of a new similarity function, and the application of multiclass normalized cuts to group atomic regions are the main contributions of this paper.

The combination of watershed and spectral methods solves the weaknesses of each method by using the watershed to provide small prototype regions from which similarity graph can be obtained. Rather than clustering single feature

points we will cluster micro-regions, confident that the underlying primitive regions are reliable. Our approach actually prefers the objects to be over-segmented into a number of smaller regions to ensure that a minimal amount of background is connected to any of the object regions.

The algorithm described in this paper can be well classified into the category of hybrid techniques, since it combines the edge-based, region-based and the morphological techniques together through the spectral-based clustering approach. Rather than considering our method as another segmentation algorithm, we propose that it can be considered as an image segmentation framework within which existing image segmentation algorithms that produce over-segmentation may be used in the preliminary segmentation step.

This paper is organized as follows: in Section 2 the overview of the multiclass spectral segmentation algorithm is given. The details of the proposed framework are explained in Section 3. After presenting the experimental results and evaluation in Section 4, the final section is devoted to the concluding remarks.

## 2 Overview of Multiclass Normalized Cuts

Spectral approach is a global graph-based segmentation method that uses the eigenvectors and eigenvalues of a matrix derived from the pairwise similarities of graph nodes. The problem of image segmentation based on pairwise similarities can be formulated as a graph partitioning problem in the following way: consider the weighted undirected graph  $G = (V, E, W)$  where each node  $v_i \in V$  corresponds to a locally extracted image features, e.g. pixels and the links in  $E$  connect pairs of nodes. A weight  $w_{i,j} \in \mathbb{R}_0^+$  is associated with each link based on some property of the nodes that it connects (e.g., the difference in intensity, colour, motion, location or some other local attribute).

Let  $\Gamma = \{V_i\}_{i=1}^k$  be a multiclass disjoint partition of  $V$  such as  $V = \cup_{i=1}^k V_i$  and  $V_i \cap V_j = \emptyset, i \neq j$ . Image segmentation is reduced to the problem of partitioning the set  $V$  into disjoint non-empty sets of nodes  $(V_1, \dots, V_k)$ , such similarity among nodes in  $V_i$  is high and similarity across  $V_i$  and  $V_j$  is low. The solution in measuring the goodness of the image partitioning is the minimization of the normalized cut as a generalized eigenvalue problem.

Shi and Malik [8] introduced the normalized cut (*NCut*) criterion for bipartitioning segmentation. Let  $V_A, V_B$  be two disjoint sets of the graph  $V_A \cap V_B = \emptyset$ . We define  $links(V_A, V_B)$  to be the total weighted connections from  $V_A$  to  $V_B$ :

$$links(V_A, V_B) = \sum_{i \in V_A, j \in V_B} w_{i,j} . \quad (1)$$

The intuition behind the normalized cut criterion is that not only we want a partition with small link cut but we also want the subgraphs formed between the matched nodes to be as dense as possible. This latter requirement is partially satisfied by introducing the normalizing denominators in the Eq. (1). The

normalized cut criterion for a bipartition of the graph is then defined as follows:

$$Ncut(A, B) = \frac{links(A, B)}{links(A, V)} + \frac{links(B, V)}{links(B, V)} . \quad (2)$$

By minimizing this criterion we simultaneously minimize the similarity across partitions and maximize the similarity within partitions. This formulation allows us to decompose the problem into a sum of individual terms and formulates a dynamic programming solution finding a partition  $\Gamma$  that minimizes the multiclass normalized cut ( $kNCut$ )

$$kNCut(\Gamma) = \frac{links(V_1, \overline{V_1})}{links(V_1, V)} + \frac{links(V_2, \overline{V_2})}{links(V_2, V)} + \dots + \frac{links(V_k, \overline{V_k})}{links(V_k, V)} , \quad (3)$$

where  $\overline{V_i}$  represents the complement of  $V_i$ .

For a fixed  $k$  partitioning of the nodes of  $G$ , reorder the rows and columns of  $W$  accordingly so that the rows of  $W$  correspond to the nodes in  $V_i$ . Let  $D = diag(D_1, \dots, D_k)$  be the  $n \times n$  diagonal matrix so that  $D_i$  is given by the sum of the weights of all links on node  $i$ :  $D_i = \sum_{j=1}^k W_{ij}$ . It is easy to verify that

$$links(V_i, \overline{V_i}) = D_i - W_{ii} \quad \text{and} \quad links(V_i, V) = D_i . \quad (4)$$

A multiclass partition of the nodes of  $G$  is represented by an  $n \times k$  indicator matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_k]$  where  $X(i, l) = 1$  if  $i \in V_l$  and 0 otherwise [9]. Since a node is assigned to one and only one partition there is an exclusion constraint between columns of  $X$ :  $XI_k = I_n$ . It follows that

$$links(V_i, \overline{V_i}) = \mathbf{x}_i^T (D - W) \mathbf{x}_i \quad \text{and} \quad links(V_i, V) = \mathbf{x}_i^T D \mathbf{x}_i . \quad (5)$$

Therefore,

$$\begin{aligned} kNCut(\Gamma) &= \frac{\mathbf{x}_1^T (D - W) \mathbf{x}_1}{\mathbf{x}_1^T D \mathbf{x}_1} + \dots + \frac{\mathbf{x}_k^T (D - W) \mathbf{x}_k}{\mathbf{x}_k^T D \mathbf{x}_k} \\ &= k - \left( \frac{\mathbf{x}_1^T W \mathbf{x}_1}{\mathbf{x}_1^T D \mathbf{x}_1} + \dots + \frac{\mathbf{x}_k^T W \mathbf{x}_k}{\mathbf{x}_k^T D \mathbf{x}_k} \right) , \end{aligned} \quad (6)$$

subject to  $X^T D X = I_k$ .

The solution for the generalized Rayleigh quotients that compose Eq. (6) is the set of eigenvectors  $X$  associated with the set of the smallest eigenvalues  $\Phi = \{0 = \nu_1 \leq \dots \leq \nu_k\}$  of the system

$$(D - W) X = \Phi D X . \quad (7)$$

However, this problem is NP-hard [8] and therefore generally intractable. If we ignore the fact that the elements of  $\mathbf{x}_i$  are either zero or one, and allow them to take continuous values, by using the method of Lagrange multipliers, Eq.

(7) can be expressed by the standard eigenvalue problem. Let  $\mathbf{y}_i = D^{1/2}\mathbf{x}_i$  and  $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k]$ .

$$\widetilde{W}Y = Y\Lambda, \quad (8)$$

subject to  $Y^TY = I_k$ , where  $\widetilde{W} = D^{-1/2}WD^{-1/2}$  is the normalized graph Laplacian matrix<sup>1</sup>, with  $\Lambda = \{1 = \lambda_1 \geq \dots \geq \lambda_k\}$  where  $\lambda_i = 1 - \nu_i$ .

If  $Y$  is formed with any  $k$  eigenvectors of  $\widetilde{W}$  then  $\Lambda$  is the  $k \times k$  diagonal matrix formed with the eigenvalues corresponding to the  $k$  eigenvectors in  $Y$ . These  $k$  eigenvectors must be distinct to satisfy  $Y^TY = I_k$ . This means that

$$Y^T\widetilde{W}Y = Y^TY\Lambda = I_k\Lambda = \Lambda, \quad (9)$$

and the *trace* of  $Y^T\widetilde{W}Y$  is the sum of the eigenvalues corresponding to the  $k$  eigenvectors in  $Y$ . So, Eq. (6) becomes equivalent to

$$kNCut(\Gamma) = k - \text{trace}\left(Y^T\widetilde{W}Y\right) = k - \sum_{i=1}^k \lambda_i. \quad (10)$$

It follows from Fan's Theorem that the maximum on the right hand side of Eq. (10) is achieved when  $Y$  is taken to be any orthonormal basis for the subspace spanned by the eigenvectors corresponding to the  $k$  largest eigenvalues of  $\widetilde{W}$ . From this we establish a lower bound  $l(k)$  on  $kNCut(\Gamma)$  as

$$\min_{\Gamma} NCut_k(\Gamma) \geq k - \sum_{i=1}^k \lambda_i, \quad (11)$$

where  $\lambda_1, \dots, \lambda_k$  are the  $k$  largest eigenvalues of  $\widetilde{W}$ .

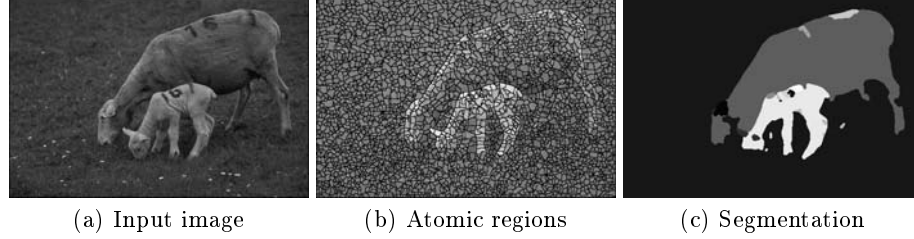
For  $k = 2$  the bound becomes  $l(2) = 2 - (1 + \lambda_2) = 1 - \lambda_2 = \nu_2$  that is the second smallest eigenvalue of the generalized eigensystem of Eq. (7). This is consistent with the bi-partitioning method proposed by Shi and Malik [8].

### 3 Proposed hybrid framework

The proposed methodology has three major stages (see Fig. 1). First, we reduce image noise, as a pre-processing stage, using an anisotropic filter, and create an over-segmented image based on the watershed transform of the gradient image. In the second stage, the over-segmented image will be the input for the region similarity graph (RSG) construction. Finally, we apply a spectral approach on the RSG. This framework integrates edges and region-based segmentation with spectral-based clustering through the morphological watersheds.

---

<sup>1</sup> Although the Laplacian matrix is usually represented by  $I - \widetilde{W}$ , replacing  $\widetilde{W}$  with  $I - \widetilde{W}$  only changes the eigenvalues (from  $\lambda$  to  $1 - \lambda$ ) and not the eigenvectors.



**Fig. 1.** Example of image segmentation. (a) Input image. (b) Atomic regions. Each atomic region is a node in the graph  $G$ . (c) Segmentation (labelling) result.

### 3.1 Watershed over-segmentation

A critical issue in watershed techniques is known to be over-segmentation i.e. the tendency to produce too many basins [1]. Several methods have been proposed in the literature to reduce the spurious boundaries created due to noise and produce a meaningful segmentation.

In this work we provide three methods to overcome this problem. First, bilateral anisotropic filtering can be applied to remove noise from the image. Secondly, some of the weakest edges are removed by a gradient minima suppression process known as *pre-flooding*. This concept uses a measure of depth of a certain basin. Prior to the transform, each catchment basin is flooded up to a certain height above its bottom, i.e. the lowest gradient magnitude and it can be thought as a flooding of the topographic image at a certain level (flooding level). This process will create a number of lakes, grouping all the pixels that lie below the flooding level.

The third one, handles to control over-segmentation eliminating spurious tiny regions associated with uniform regions through a merging step. This eliminates tiny regions which have similar adjacent regions, while maintaining the accuracy of the partition. This stage is required to reduce the computational complexity in the graph partitioning. Another advantage of these steps is to prevent large homogeneous (flat) regions from being split in the graph-based segmentation (a common problem with balanced graph cut methods).

### 3.2 Region similarity graph

Spectral-based methods use the eigenvectors and eigenvalues of a matrix derived from the pairwise similarities of features (pixels or regions). This effect is achieved by constructing a fully connected graph. Considering all pairwise pixel relations in an image may be too computational expensive. Unlike other well known clustering methods [8, 9] which use down-sampling pixel-based to construct the graph, our method is based on selecting links from a weighted undirected graph  $G = (V, E, W)$  based on a region similarity graph where each node corresponds to an atomic region.

Some characteristics of the RSG model that yield to relevant advantages with regard to the region adjacency graph model are: i) the RSG allows the existence of links between pairs of non-adjacent regions. ii) it is defined once and it does not need any dynamic updating when merging regions.

The proposed RSG structure takes advantage on region-based representation. The set of nodes  $V$  are represented by the centroid of each micro-region. The sets of links  $E$  and link weights  $W$  represent, respectively, relationships and similarity measures between pair of regions.

**Pairwise spatial similarity** Each region  $r_i$  represents a small group of pixels where the centroid  $\bar{x}_i$  is utilized as a node of the graph. For each pair of nodes, node similarity is inversely correlated with the maximum contour energy encountered along the line connecting the centroids of the regions. If there are strong contours along a line connecting two centroids, these atomic regions probably belong to different segments and should be labelled as dissimilar. Let  $i$  and  $j$  be two atomic regions and the orientation energy  $OE^*$  between them, then the intervening contours contribution to the link weight is given by:

$$w_{ic}(i, j) = \exp \left[ -\frac{\max_{t \in line(i, j)} \|OE^*(\bar{x}_i, \bar{x}_j)\|^2}{\sigma_{ic}^2} \right] , \quad (12)$$

where  $line(i, j)$  is the line between centroids  $\bar{x}_i$  and  $\bar{x}_j$  formed by  $t$  pixels.

The mean intensity of each node contributes for the link weight according to the following function:

$$w_I(i, j) = \exp \left( -\frac{(I_{\bar{x}_i} - I_{\bar{x}_j})^2}{\sigma_I^2} \right) . \quad (13)$$

These cues are combined in a final link weight similarity function, with the values  $\sigma_{ic}$  and  $\sigma_I$  selected in order to maximize the dynamic range of  $\mathbf{W}$ :

$$\mathbf{W}(i, j) = w_{ic}(i, j) \cdot w_I(i, j) . \quad (14)$$

In almost all the graph-based approaches proposed in the literature the spatial distance cue is also used to compute the similarity between graph nodes. However, during our experiments, we note that such cue is responsible for the partition of homogeneous areas in the image - an issue commonly associated to normalized cut algorithm. Instead, we use intervening contours which are equivalent to spatial distance without suffering from the same problems [7].

**Implementation details of the RSG** For a computational consideration it is important to sort and label all the regions created by the watershed segmentation. In the following some implementation details are given about the construction of the RSG. For each region  $r_i$ , spatial location  $\bar{x}_i$  is computed as centroids of their pixels. Two dynamic data structures are used through which it is very convenient to add or remove regions: 1) A label map in which each pixel

value corresponds to the label of the segment that this pixel belongs to; 2) An array of segments where each segment is represented by a linked-list of pixels which correspond to the pixels that belong to the segment. This list includes the location and the grey-level of each pixel.

This dual representation of a partitioned image allows for a very efficient implementation. The label map grants us immediate access to the label of every pixel in the image. The array of lists gives us immediate access to the set of pixels that belong to each segment. Using this representation two different segments can be merged into one by iterating through the corresponding linked-lists and updating the label map. Even more, we can easily obtain the centroid and the mean value of each segment.

To compute the similarity matrix the current approach uses only image brightness and magnitude gradient. Additional features such as texture, could be added to the similarity criterion. This may slow the construction of the RSG but the rest of the algorithm will proceed with no change.

## 4 Experiments and evaluation

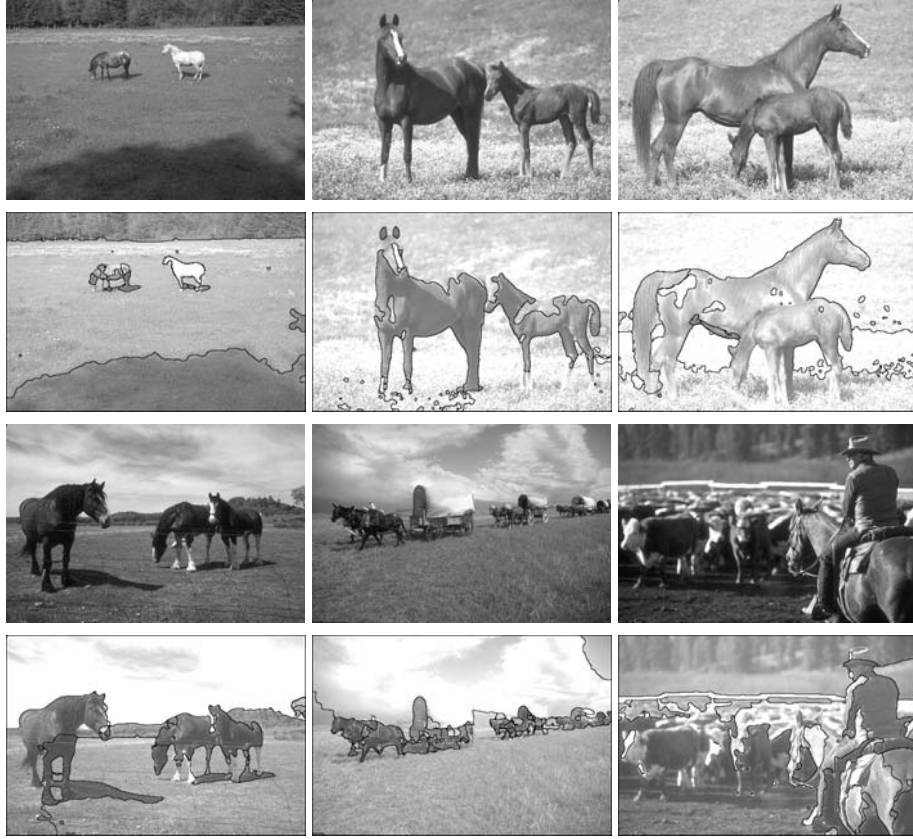
In order to evaluate the performance of the proposed method experiments have been conducted to compare it with some of the most popular algorithms: (i) mean shift (EDISON) [2], (ii) a multiscale graph based segmentation method (MNCUT) [3], and (iii) JSEG [4]. To provide a numerical evaluation measure and thus allow comparisons, the experiments for the evaluation were conducted on the manual segmentations of the publicly available Berkeley Segmentation Dataset<sup>2</sup> [6]. Due to the large number of images, we choose to apply our framework to images with horses on it. The task is cast as a boundary detection problem, with results presented in terms of Precision (P) and Recall (R) measures. The algorithm provides a binary boundary map which is scored against each one of the hand-segmented results of Berkeley Dataset, producing a  $(R, P, F)$  value. The final score is given by the average of those comparisons. The quantitative evaluation of segmentation results (F-measure) is summarized in Table 1.

**Table 1.** Results of quantitative evaluation in terms of F-measure for the comparison between the proposed method (WNCUT), Mean shift (EDISON), JSEG and the multiscale segmentation MNCUT.

Method	28075	113016	113044	197017	216041	220075	291000	361010
WNCUT	0.66	0.64	0.66	0.72	0.65	0.65	0.69	0.74
EDISON	0.61	0.41	0.47	0.69	0.60	0.36	0.35	0.57
JSEG	0.42	0.55	0.53	0.46	0.40	0.32	0.41	0.49
MNCUT	0.25	0.30	0.35	0.48	0.36	0.37	0.24	0.58

<sup>2</sup> <http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>

The proposed approach produces segmentations of high quality and with better results than the other methods for all tested images. This new approach overcomes some limitations usually associated with spectral clustering approaches. Some segmentation results of the proposed method are shown in Fig. 2. Since the F-measure is a boundary-based measure the segmentation results are presented as boundaries over the original images.

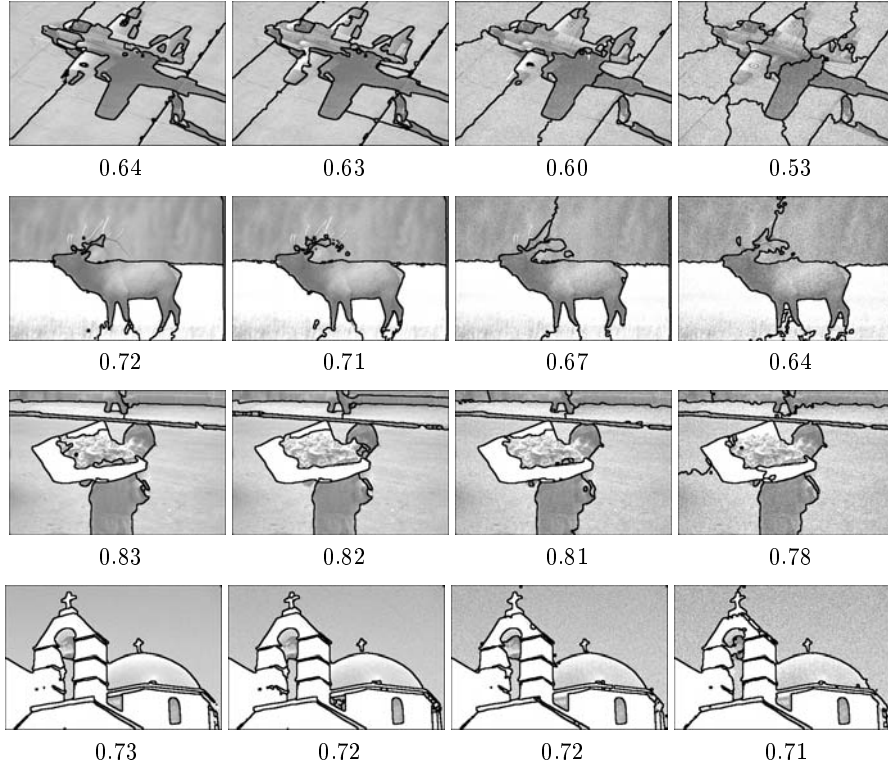


**Fig. 2.** Segmentation results: First and third rows: original images (first six images of Table 1). Second and fourth rows: segmentation results with the proposed method

#### 4.1 Robustness to noise

To analyse the behaviour of the algorithm in presence of noise, the images were corrupted with four levels of Gaussian additive noise with standard deviations  $\sigma = 5, 10, 20, 30$ . The effect of the pre-processing step in reducing the noise, with a reduction on the number of irrelevant regions in the output of the watershed algorithm, can be observed in Fig. 3.





**Fig. 3.** Performance of the proposed approach on noisy images. Results with added Gaussian noise with  $\sigma$ , from left to right, equal to 5, 10, 20, 30. The values below the images are the F-measures.

Our method turned out to be extremely robust to artificially added Gaussian noise. We may notice that segmentation results are not very affected till  $\sigma = 20$ , and it produces a good segmentation even for added Gaussian noise with an amplitude of  $\sigma = 30$ . This amount of noise is greater than would be expected in a normal real image.

## 5 Conclusion

This paper proposes an image segmentation approach which combines edge- and region-based information with spectral techniques through the morphological algorithm of watersheds. Using small atomic regions instead of pixels leads to a more natural image representation - the pixels are merely the result of the digital image discretization process and they do not occur in the real world. Besides producing smoother segmentations than pixel-based partitioning methods, it also reduces the computational cost in several orders of magnitude. The experimental results demonstrate the effectiveness of the proposed approach to compare favourably with some of the most popular image segmentation methods.

## References

- [1] R.J. O' Callaghan and D.R. Bull. Combined morphological-spectral unsupervised image segmentation. *IEEE Transactions on Image Processing*, 14(1):49–62, 2005.
- [2] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Patt. Analysis Machine Intelligence*, 24(5):603–619, 2002.
- [3] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, volume II, pages 1124–1131, June 2005.
- [4] Y. Deng and B. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.
- [5] A. Duarte, A. Sánchez, F. Fernández, and A. Montemayor. Improving image segmentation quality through effective region merging using a hierarchical social meta-heuristic. *Pattern Recognition Letters*, 27(11):1239–1251, 2006.
- [6] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. IEEE Int. Conf. on Computer Vision*, volume II, 2001.
- [7] F.C. Monteiro. *Region-based spatial and temporal image segmentation*. PhD thesis, Faculdade de Engenharia da Universidade do Porto, Portugal, 2007. <http://repositorio.up.pt/aberto/handle/10216/11088?mode=full>.
- [8] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [9] S. Yu and J. Shi. Multiclass spectral clustering. In *Proc. of IEEE International Conference on Computer Vision*, pages 313–319, Nice, France, Oct. 2003.