

Hierarchical Pairwise Segmentation Using Dominant Sets and Anisotropic Diffusion Kernels

Andrea Torsello and Marcello Pelillo

Dipartimento di Informatica
Università Ca' Foscari di Venezia
{torsello,pelillo}@dsi.unive.it

Abstract. Pairwise data clustering techniques are gaining increasing popularity over traditional, feature-based central grouping techniques. These approaches have proved very powerful when applied to image-segmentation problems. However, they are mainly focused on extracting flat partitions of the data, thus missing out on the advantages of the inclusion constraints typical of hierarchical coarse-to-fine segmentations approaches very common when working directly on the image lattice. In this paper we present a pairwise hierarchical segmentation approach based on dominant sets [12] where an anisotropic diffusion kernel allows for a scale variation for the extraction of the segments, thus enforcing separations on strong boundaries at a high level of the hierarchy. Experimental results on the standard Berkeley database [9] show the effectiveness of the approach.

1 Introduction

Proximity-based, or pairwise, data clustering techniques are gaining increasing popularity over traditional central grouping techniques, which are centered around the notion of “feature” (see, e.g., [5,14,15,13]). In many application domains, in fact, the objects to be clustered are not naturally representable in terms of a vector of features. On the other hand, quite often it is possible to obtain a measure of the similarity/dissimilarity between objects. Although such a representation lacks geometric notions such as scatter and centroid, it is attractive as similarity information arising from sources of very different nature can be incorporated very easily, often not requiring more than adding together distances or multiplying similarities calculated from different sources. In contrast, integrating information of different nature within the central clustering framework requires an integrated feature model capable of simultaneously characterizing all information at the feature level.

These approaches have proven very powerful when applied to image segmentation problems [15,8,5,2]. Here, the possibility of easily integrating different

sources of information has been used to incorporate color, texture, and proximity information between pair of pixels. Conversely, feature based segmentation algorithms must explicitly integrate all types of information into a single geometrical model which requires a stronger characterization of the geometry of the image.

Despite the promise of ease of integration of inhomogeneous information, most actual implementations of pairwise segmentation only integrate local appearance based information, with color- and texture-based similarities taking the lion's share over all pairwise measures found in the literature. With few exceptions, little work has been done to integrate locality and boundary information in a pairwise setting. Among these we note Malik and coworkers' proposal to incorporate boundary information in the normalized-cut framework by looking for an *intervening contour* [8]. However, their approach only looks for detected edges in the straight line joining two pixels; hence, it is strongly dependant on the quality of the edge extractor and tends to separate pixels belonging to a single region if this is not convex. The normalized cut framework is relatively forgiving about this problem, but it is particularly severe when using pairwise clustering algorithms that favor "compact" globular clusters such as the dominant sets framework [12]. Furthermore, intervening contour information alone is not able to separate regions with fuzzy or unclear boundaries such as regions delimited by relatively smooth gradients. An alternative is to use the minimal boundary separation along all possible paths [16], however the selection of the optimal path renders this approach not robust with respect to the misdetection of a single boundary point. A more robust path-based segmentation can be achieved using random walks on the image lattice. For example, Grady uses random walks to extract a semi-supervised segmentation [4] where a pixel belongs to the class of whose label is expected to find first on a random walk.

Further, pairwise segmentation algorithms are generally concerned with flat partitions, thus missing out on the advantages of the inclusion constraints typical of hierarchical coarse-to-fine segmentations approaches very common when working directly on the image lattice.

In this paper, we propose a coarse-to-fine segmentation algorithm based on a hierarchical variant of the dominant sets framework [10]. Here, however, the regularizer term is substituted with a heat diffusion kernel [7], which enforces locality and boundary separation based on a limited time random walk on the image lattice. At the beginning of the diffusion process the effects are local and hence the long-range similarity is dominated by the color and texture appearance, while as the time increases the range of the kernel expands thus enforcing a coarser segmentation. To this end we start by segmenting at a high time value, and then we iteratively reduce the time to obtain finer-grain separations. The anisotropic diffusion properties of the heat kernel have been used in the computer graphics and vision communities to perform controlled smoothing [6,1]. Here, however, we are using it to define an explicit scale space on which to base a recursive hierarchical partitioning scheme.

2 Hierarchical Dominant Sets

The dominant set framework [12] is a pairwise clustering approach based on a recursive characterization of the weight $w_S(i)$ of element i with respect to a set S of elements, and characterizes a group as a *dominant set*, i.e., a set that satisfies:

1. $w_S(i) > 0$, for all $i \in S$
2. $w_{S \cup \{i\}}(i) < 0$, for all $i \notin S$.

These conditions correspond to the two main properties of a cluster: the first regards internal homogeneity, whereas the second regards external inhomogeneity.

The main result presented in [12] provides a one-to-one relation between dominant sets and strict local maximizers of the following quadratic program

$$\begin{aligned} & \text{maximize } \mathbf{x}'A\mathbf{x} \\ & \text{subject to } \mathbf{x} \in \Delta \end{aligned} \tag{1}$$

where $A = (a_{ij})$ is the matrix of similarities of the n elements to be grouped,

$$\Delta = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i = 1, \dots, n \text{ and } \mathbf{1}'\mathbf{x} = 1\}$$

is the standard simplex of \mathbb{R}^n , and $\mathbf{1}$ is a vector of appropriate length consisting of unit entries.

Specifically, in [12] it is proven that if \mathbf{x} is a strict local solution of program (1) then its support $S = \sigma(\mathbf{x})$ is a dominant set. Here, the *support* of a vector $\mathbf{x} \in \Delta$ is the set of indices corresponding to its positive components. The local maxima of (1) is found using the replicator equations, a dynamical systems mutated from game-theory. The approach has proven to be a very effective and robust pairwise clustering approach that has in its speed one of its major selling points.

In [10] a hierarchical approach was presented by taking into consideration the regularized quadratic program

$$\begin{aligned} & \text{maximize } \mathbf{x}'(A - \alpha I)\mathbf{x} \\ & \text{subject to } \mathbf{x} \in \Delta \end{aligned} \tag{2}$$

where α is a scale parameter that defines the hierarchy. In [10] was shown that for sufficiently large values of α all elements were clustered into a single group and a recursive divisive algorithm was applied to the data as α was reduced. However, no indication of how to select the relevant values of α was provided.

3 Anisotropic Diffusion Kernel

Let $\mathcal{M} = (V_M, E_M)$ be the regular mesh defined over the image by connecting each pixel to its 4-neighbors. Further, let $\gamma : V_M \times V_M \rightarrow \mathbb{R}_+$ be an edge weight function which reflects how similar two neighboring pixels are. In our boundary-based segmentation approach we set

$$\gamma(i, j) = \begin{cases} e^{-k \frac{\nabla I_i + \nabla I_j}{2}} & \text{if } (i, j) \in E_M \\ 0 & \text{otherwise,} \end{cases}$$

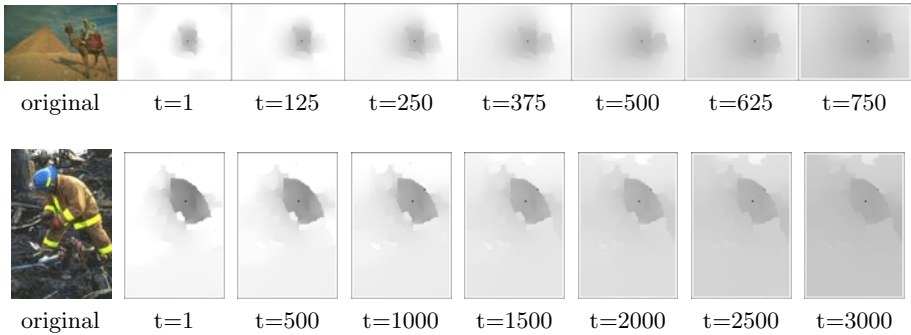


Fig. 1. Location distribution of a walker at time t . the black dot marks the starting position.

where ∇I_i is the image gradient at pixel i . We define $M = (m_{ij})$ as the weighted adjacency matrix of \mathcal{M} , where we set $m_{ij} = \gamma(i, j)$.

An anisotropic diffusion on \mathcal{M} is a lazy random walk from a pixel in the image lattice to other pixels along the mesh connections, where the transition probabilities are proportional to the value of the edge weight function γ . The walk is lazy as at each time step the walker has a non-null probability $1/Z_i$ of remaining at location i and a probability m_{ij}/Z_i of moving to location j , where $Z_i = 1 + \sum_j m_{ij}$ is a normalizing factor. The use of a lazy walk forces the walker to diffuse rapidly on flat image regions, when Z_i is high, while slowing down when there is a complex edge structure around i , i.e., when Z_i is close to 1.

The expected position at time t of a random walker starting from position i is governed by the heat diffusion kernel $e^{-\mathcal{L}t}$, where \mathcal{L} is the Laplacian of \mathcal{M} , with $\mathcal{L} = D - M$ where D is the diagonal matrix with elements $d_{ii} = \sum_j m_{ij}$. At $t = 0$, the diffusion kernel is an identity matrix; as t increases the kernel assigns non-zero values to elements in the vicinity of position i spreading more rapidly along flat locations and stopping on boundaries, while as t becomes very large the support of the kernel is very diffused and far-reaching. Figure 1 shows two images and the location probability of a walker starting on the pixel marked with the black dot. As it can be seen, for small times the support is mostly restricted to a local segment, bound by strong boundaries in the image, while as time increases the support becomes more diffused.

4 Diffusion Regularizers

In the definition of our hierarchical coarse-to-fine segmentation approach we substitute the identity matrix I in the regularizer term of program (2) with the kernel $I - e^{-\mathcal{L}t}$, obtaining the following regularized quadratic program

$$\begin{aligned}
 &\text{maximize} && f_t(\mathbf{x}) = \mathbf{x}' [A - \alpha_t (I - e^{-\mathcal{L}t})] \mathbf{x} \\
 &\text{subject to} && \mathbf{x} \in \Delta
 \end{aligned} \tag{3}$$

where for all $t > 0$ $\alpha_t \geq 0$ is a monotonically increasing function with $\alpha_0 = 0$. Note that in the large time limit this regularizer term becomes equivalent to the one used in [10]. In fact, as program (2) is invariant to constant shifts in the matrix A , subtracting the regularizing term αI is equivalent to subtracting $\alpha(I - \mathbf{1}\mathbf{1}')$. Moreover, we have

$$\lim_{t \rightarrow \infty} (I - e^{-\mathcal{L}t}) = I - \mathbf{1}\mathbf{1}'$$

as \mathcal{L} is positive semidefinite with the only null eigenvalue corresponding to the eigenvector $\mathbf{1}$. Further, for times close to 0, the effect of the kernel vanishes as

$$\lim_{t \rightarrow \infty} (I - e^{-\mathcal{L}t}) = \mathbf{0}.$$

Intuitively we are substituting a regularizing term that increases the support equivalently to all elements, with one that increases the support to neighboring pixels first.

We can now prove that the time parameter indeed produces a scale-space as cluster hierarchy collapses to the full image for sufficiently large times.

Proposition 1. *Let $\lambda_1(A), \lambda_2(A), \dots, \lambda_n(A)$ represent the largest, second largest, ..., smallest eigenvalue of matrix A . If $\alpha_t > \frac{\lambda_1(A)}{1 - e^{-\lambda_{n-1}(\mathcal{L})t}}$, then f_t is a strictly concave function in Δ . Further, if $\alpha_t > \frac{n\lambda_1(A)}{1 - e^{-\lambda_{n-1}(\mathcal{L})t}}$ the only solution of (2) belongs to the interior of Δ .*

Proof. Note that $\mathcal{L}\mathbf{1} = 0$, which implies for all t $e^{-\mathcal{L}t}\mathbf{1} = \mathbf{1}$. Further, all other eigenvalues are in the open interval $(0, 1)$. The function f_t is strictly concave in Δ if for all $\mathbf{y} \in \mathbb{R}^n$ with $\mathbf{y}'\mathbf{1} = 0$, we have $\mathbf{y}' [A - \alpha_t (I - e^{-\mathcal{L}t})] \mathbf{y} < 0$. However,

$$\begin{aligned} \mathbf{y}' (A - \alpha_t (I - e^{-\mathcal{L}t})) \mathbf{y} &\leq \lambda_1(A)\mathbf{y}'\mathbf{y} - \alpha_t (\mathbf{y}'\mathbf{y} - \mathbf{y}'e^{-\mathcal{L}t}\mathbf{y}) \leq \\ \lambda_1(A)\mathbf{y}'\mathbf{y} - \alpha_t (\mathbf{y}'\mathbf{y} - \lambda_2 (e^{-\mathcal{L}t}) \mathbf{y}'\mathbf{y}) &= \mathbf{y}'\mathbf{y} (\lambda_1(A) - \alpha_t (1 - e^{-\lambda_{n-1}(\mathcal{L})t})) < 0. \end{aligned} \tag{4}$$

Since f_t is strictly concave in Δ , program (2), which is a concave with convex constraints, has a unique solution. To prove the second result, suppose by contradiction that this solution \mathbf{x} lies on the boundary of Δ , then we have $x_i = 0$ for some index i . There is a unique $\mathbf{y} \in \mathbb{R}^n$ with $\mathbf{y}'\mathbf{1} = 0$ such that $\mathbf{x} = \frac{1}{n}\mathbf{1} + \mathbf{y}$. Further, since by hypothesis we have $x_i = 0$, then $\mathbf{y}'\mathbf{y} \geq \frac{1}{n(n-1)}$. With this we have

$$\begin{aligned} \mathbf{x}'A\mathbf{x} - \alpha_t (\mathbf{x}'\mathbf{x} - \mathbf{x}'e^{-\mathcal{L}t}\mathbf{x}) &\leq \lambda_1(A)(\mathbf{y}'\mathbf{y} + \frac{1}{n}) - \mathbf{y}'\mathbf{y}\alpha_t (1 - e^{-\lambda_{n-1}(\mathcal{L})t}) = \\ \mathbf{y}'\mathbf{y} \left(\lambda_1(A) \left(1 + \frac{1}{n\mathbf{y}'\mathbf{y}} \right) - \alpha_t (1 - e^{-\lambda_{n-1}(\mathcal{L})t}) \right) &\leq \\ \mathbf{y}'\mathbf{y} (\lambda_1(A)n - \alpha_t (1 - e^{-\lambda_{n-1}(\mathcal{L})t})) &< 0. \end{aligned}$$

Recall that a point $\mathbf{x} \in \Delta$ satisfies the Karush-Kuhn-Tucker (KKT) conditions for problem (2) if

$$\begin{aligned}
 (\mathbf{Ax})_i - \alpha_t (x_i - e^{-\mathcal{L}t\mathbf{x}})_i &= \mathbf{x}'\mathbf{Ax} - \alpha_t (\mathbf{x}'\mathbf{x} - \mathbf{x}'e^{-\mathcal{L}t\mathbf{x}}) \\
 &\quad \text{if } i \in \sigma(\mathbf{x}) \\
 (\mathbf{Ax})_i + \alpha_t (e^{-\mathcal{L}t\mathbf{x}})_i &\leq \mathbf{x}'\mathbf{Ax} - \alpha_t (\mathbf{x}'\mathbf{x} - \mathbf{x}'e^{-\mathcal{L}t\mathbf{x}}) \\
 &\quad \text{otherwise.}
 \end{aligned}
 \tag{5}$$

Then, we have

$$(\mathbf{Ax})_i + \alpha_t (e^{-\mathcal{L}t\mathbf{x}})_i \leq \mathbf{x}'\mathbf{Ax} - \alpha_t (\mathbf{x}'\mathbf{x} - \mathbf{x}'e^{-\mathcal{L}t\mathbf{x}}) .$$

However, this is impossible since $(\mathbf{Ax})_i + \alpha_t (e^{-\mathcal{L}t\mathbf{x}})_i > 0$ and $\mathbf{x}'\mathbf{Ax} - \alpha_t (\mathbf{x}'\mathbf{x} - \mathbf{x}'e^{-\mathcal{L}t\mathbf{x}}) < 0$, thus proving the proposition.

4.1 Selecting Relevant Levels

One of the questions left open in [10] is how to select the values of the regularizer parameter that induce relevant partitions. Indeed, as the scale parameter α varies continuously from its maximum value down to 0, we expect the size of the extracted segments to vary almost as smoothly. Here we adopt an entropy shedding approach to the selection of the relevant levels of the hierarchy: the value of $\mathbf{x} \in \Delta$ that maximizes (2) can be considered as a probability distribution whose entropy is a measure of the size and cohesiveness of the cluster. As the parameter α is decreased we expect the first extracted cluster to steadily become less cohesive, eventually losing a few peripheral elements, until we reach a point where there is a substantial modification in the cluster structure as the current group gets split into multiple parts, thus producing a jump in the entropy value.

Figure 2 shows an example where a set of points generated from three bivariate Gaussian distribution are clustered using the original hierarchical formulation (2). The left image shows the point distribution, while the plot on the right

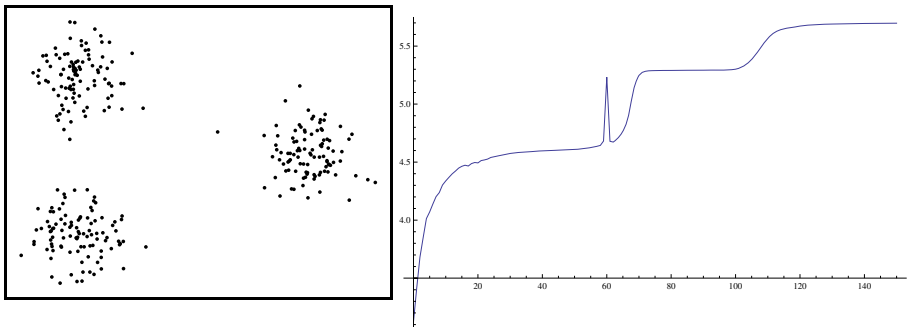


Fig. 2. Entropy value of the distribution associated with first cluster as a function of the regularizing parameter α

show the entropy for different values of α . there are three well defined plateaus corresponding to a single cluster encompassing all the data ($\alpha > 115$), a cluster without the points in the rightmost distribution which is furthest apart from the other ($75 < \alpha < 100$), and cluster encompassing only one of the three point sets ($20 < \alpha < 55$). after that even smaller subsets are extracted.

Accordingly, we start the clustering procedure at a sufficiently high time to obtain a single cluster and we reduce the time until we reach the next entropy plateau. In our implementation we use an exponential schedule for the reduction of the time parameter t , i.e., we multiply t by a constant factor $t_{\text{mult}} = 0.8$. once a new plateau is reached, we partition the data and continue recursively on each cluster, until we reach the final partition at $t = 0$.

4.2 Subsampling

Despite their many advantages, pairwise clustering approaches are computationally very demanding due to their scaling behavior with the quantity of data. On a dataset containing N examples, the number of potential comparisons scales with $O(N^2)$, thereby rendering the approaches unfeasible for problems involving very large data sets as is the case of pixel based segmentation of even moderately large images. It is therefore of primary importance to develop strategies to reduce the number of comparisons required by subsampling the data and extending the grouping to out-of-sample points after the clustering process has taken place.

In [11] was proposed a subsampling approach for the dominant sets framework and a more efficient extension scheme was proposed in [17] in order to adapt the framework to spatio-temporal segmentation. The approach takes an element of a cluster S to act as a cluster centroid, namely it takes the element i which maximizes the weight $w_S(i)$ with respect to the cluster S . This way, the similarity of an out-of-sample point j to a cluster S is simply the similarity to its centroid c_S . The first step of the out-of-sample segmentation is to extract the clusters from the sampled points. With the initial segmentation to hand, each pixel is then assigned to the closest cluster.

While the out of sampling approach certainly helps with the computation of the cluster structure at the various levels, it cannot be used to reduce the complexity of the kernel computation, which requires the computation of the full set of eigenvalues and eigenvectors of the Laplacian matrix \mathcal{L} . Further, note that subsampling techniques require a connection to all the other nodes and while this is not a problem for the matrix A which has connection at all ranges, is unusable with the Laplacian, which has only local connections and subsampling it would break the mesh connectivity structure and severely modify the eigenspaces. However, we can use the locality of the heat kernel to our advantage, as we can down-sample the original boundary information to obtain a smaller mesh from which we can compute the eigenvectors which can then just be up-scaled to the original size with minimal loss of information. Hence, in this work we adopted a mixed strategy for data reduction: we down-sampled the mesh by a factor of 8 in each direction and computed the 10 smallest eigenvectors of \mathcal{L} , which are then up-scaled to reconstruct a least-squares approximation of the heat kernel.

With the full up-scaled kernel to hand, we can use the subsampling procedure described in [17] both on the similarity matrix A and on the kernel $I - e^{-\mathcal{L}t}$.

5 Experimental Results

In order to assess the performance of the proposed kernel-based coarse-to-fine hierarchical segmentation approach, we tested it on the Berkeley database [9] using only color information for the similarity matrix A . Clearly, the final goal is to incorporate more descriptive pairwise similarities, but it would be hard to separate the effect of the diffusion kernel from that of other boundary-based information. In all the experiments the similarity between two pixels i and j is taken to be $a_{ij} = e^{-\frac{1}{2}d(i,j)^2/\sigma^2}$, where $d(i, j)$ is the perceptual distance between the colors of pixels i and j computed as the Euclidean distance on the CIE Luv color space.

Figures 3 and 4 show two images and the computed segment hierarchies. On the top left corner of each group we see the original images, while the other images display the segmentation hierarchy. For each segmented image we show the clusters extracted at the next level, where all pixels belonging to the same cluster are drawn using the average color of the cluster and pixels that have already being eliminated are drawn in black. the segmented image is then linked to the images showing the lower level segmentations. Note how the first separation in the image in Figure 3 is a major figure ground separation with all the details on the camel clustered together. It is only at lower levels of the hierarchy that the details form separate clusters. Further, note how in the image in Figure 4, the

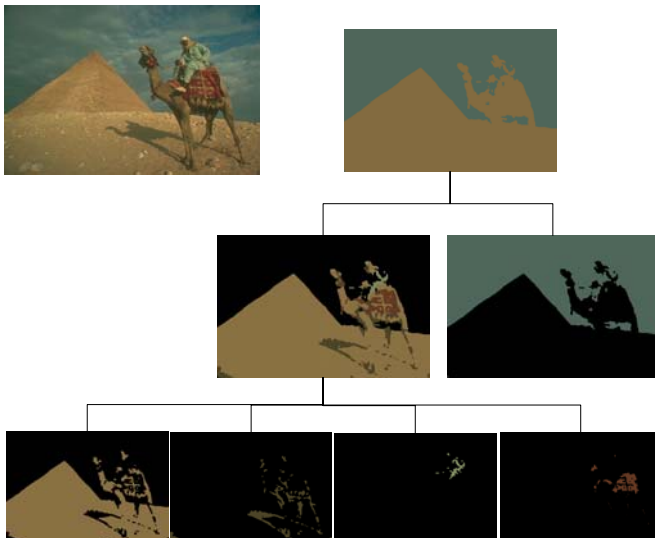


Fig. 3. Segmentation hierarchies extracted from the camel image

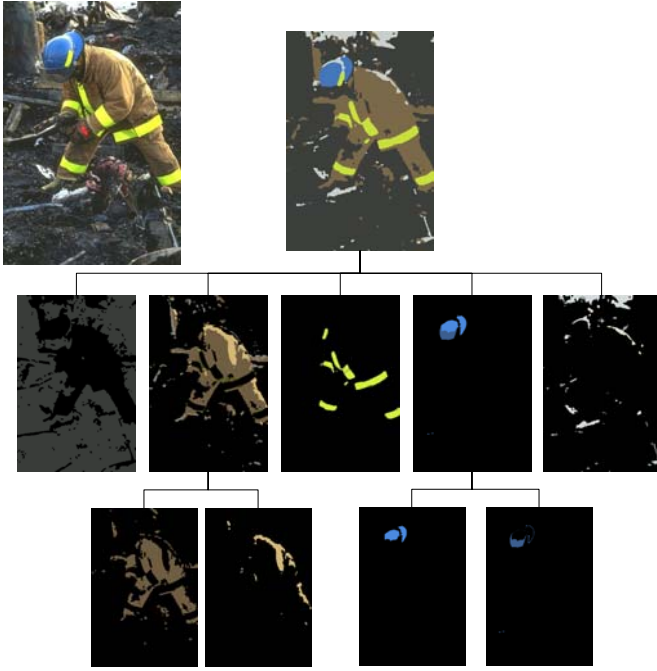


Fig. 4. Segmentation hierarchies extracted from the fireman image

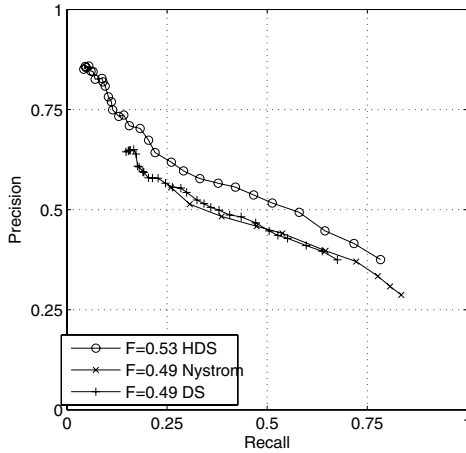


Fig. 5. Precision/Recall on the Berkeley database for our hierarchical method (HDS), the flat dominant set framework (DS) and the Nyström extension (Nystrom)

first partition separates the major image components, while at a lower scale the highlights get separated from the fireman’s suit and helmet.

Next, for a more quantitative analysis, we computed the Precision and Recall for boundary detection on the full Berkeley database. The analysis was performed for our hierarchical approach, the original dominant sets framework with out-of-sample extension, and the Nyström extension [3], an out-of-sample generalization of normalized cut [15]. All pairwise segmentation approaches were based on the same color similarity matrix.

Figure 5 shows the resulting Precision/Recall curves. We can see that the original dominant sets and the Nyström extension performed almost identically, while our proposed approach shows a marginal but very clear advantage. It is worth noting that the standard precision recall curve does not evaluate the quality of the segment hierarchies, but is only concerned with a single flat segmentation. For this reason on our algorithm only the final partition, the one at $t = 0$, was used for the boundary extraction. Hence, the advantage over the standard dominant sets framework is due to the coarse-to-fine nature of the hierarchical divisive algorithm used, which extracts stronger and clearer boundaries first, thus reducing noise and detection errors around clear separations.

6 Conclusions

In this paper we have presented a coarse-to-fine hierarchical segmentation approach which uses an anisotropic diffusion kernel to generate the levels of the hierarchy. The approach is a generalization of the hierarchical dominant sets framework presented in [10] with the addition of a heat kernel-based regularizer term that enforces locality and boundary separation. We have proven that the term does indeed generate a scale-space and proposed an entropy measure to select the relevant scales. Further, we have proposed a mixed strategy for out-of-sample extension in the presence of the diffusion kernels. We compared the performance of the approach to the original flat dominant sets framework and to the Nyström extension applied to normalized cut, showing that our coarse-to-fine approach outperforms both on the standard Berkeley database.

Acknowledgment

We acknowledge the nancial support of the Future and Emerging Technology (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open project SIMBAD grant no. 213250.

References

1. Zhang, F., Hancock, E.R.: Graph Spectral Image Smoothing. In: Escolano, F., Vento, M. (eds.) GbRPR. LNCS, vol. 4538, pp. 191–203. Springer, Heidelberg (2007)
2. Fischer, B., Buhmann, J.M.: Path-based clustering for grouping of smooth curves and texture segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* 25(4), 513–518 (2003)

3. Fowlkes, C., Belongie, S., Chun, F., Malik, J.: Spectral grouping using the Nyström method. *IEEE Trans. Pattern Anal. Machine Intell.* 26, 214–225 (2004)
4. Grady, L.: Random Walks for Image Segmentation. *IEEE Trans. on Pattern Anal. and Machine Intell.* 28(11), 1768–1783 (2006)
5. Hofmann, T., Buhmann, J.M.: Pairwise data clustering by deterministic annealing. *IEEE Trans. Pattern Anal. Machine Intell.* 19, 1–14 (1997)
6. Karni, Z., Gotsman, C.: Spectral compression of mesh geometry. In: *SIGGRAPH 2000: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pp. 279–286. ACM Press, New York (2000)
7. Kondor, R., Lafferty, J.: Diffusion kernels on graphs and other discrete structures. In: *Proceedings of the 19th Intl. Conf. on Machine Learning (ICML)* (2002)
8. Malik, J., Belongie, S., Leung, T., Shi, J.: Contour and texture analysis for image segmentation. *Int. J. of Computer Vision* 43(1), 7–27 (2001)
9. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In: *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, pp. 416–423 (2001)
10. Pavan, M., Pelillo, M.: Dominant sets and hierarchical clustering. In: *9th IEEE International Conference on Computer Vision – ICCV 2003*, vol. I, pp. 362–369. IEEE Computer Society Press, Los Alamitos (2003)
11. Pavan, M., Pelillo, M.: Efficient out-of-sample extension of dominant-set clusters. In: Saul, L.K., Weiss, Y., Bottou, L. (eds.) *Advances in Neural Information Processing Systems 17*, pp. 1057–1064. MIT Press, Cambridge (2005)
12. Pavan, M., Pelillo, M.: Dominant sets and pairwise clustering. *IEEE Trans. Pattern Anal. Machine Intell.* 29(1), 167–172 (2007)
13. Roth, V., Laub, J., Kawanabe, M., Buhmann, J.M.: Optimal cluster preserving embedding of nonmetric proximity data. *IEEE Trans. Pattern Anal. Machine Intell.* 25, 1540–1551 (2003)
14. Sarkar, S., Boyer, K.: Quantitative measures of change based on feature organization: Eigenvalues and eigenvectors. *Computer Vision and Image Understanding* 71, 110–136 (1998)
15. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 888–905 (2000)
16. Torsello, A., Di Ges, M., Pelillo, M.: Integrating Boundary Information in Pairwise Segmentation. In: *International Conference on Image Analysis and Processing - ICIAP 2007*, pp. 23–28. IEEE Computer Society, Los Alamitos (2007)
17. Torsello, A., Pavan, M., Pelillo, M.: Spatio-temporal segmentation using dominant sets. In: Rangarajan, A., Vemuri, B.C., Yuille, A.L. (eds.) *EMMCVPR 2005*. LNCS, vol. 3757, pp. 301–315. Springer, Heidelberg (2005)