A Scale-Space Based Hierarchical Representation of Discrete Data

Moncef Hidane, Olivier Lezoray, Abderrahim Elmoataz

A SCALE-SPACE BASED HIERARCHICAL REPRESENTATION OF DISCRETE DATA

M. Hidane, O. Lezoray, A. Elmoataz

Université de Caen Basse-Normandie, ENSICAEN, CNRS, GREYC Image Team,
6 Boulevard Maréchal Juin, F-14050 Caen Cedex France.

ABSTRACT

A new hierarchical representation of general discrete data sets living on graphs is proposed. The approach takes advantage of recent works on graph regularization. The different levels of the hierarchy are discovered as the regularization process is performed. The role of the merging criterion that is common to hierarchical representations is greatly reduced due to the regularization step. This yields a robust representation of data sets. Moreover, the approach is particularly well adapted to the processing of digital images, where nonlocal processing allows to better handle repetitive patterns usually present in natural images.

Index Terms— Hierarchical representations, Scale-space, Discrete regularization.

1. INTRODUCTION

Multiresolution representations of digital images are a well established technique in image processing. We propose in this work a new multiresolution representation of functions defined on discrete sets represented by a graph structure. Our work uses discrete regularization techniques [1] to produce a hierarchical representation of graphs and functions on graphs. The hierarchical representation we propose respects the inclusion and causality principles.

Multilevel representations of discrete data sets usually fall into two categories: multiscale and multiresolution. Scale space theory [2] provides a systematic and now well established theory to provide a multiscale representation of digital images. Linear scale space consists in successive convolutions of an initial image with a Gaussian kernel of increasing width. A unifying view of scale-space theory was given in [3] where it is shown that any scale-space representation can be obtained through the evolution of a general class of parabolic diffusion equations. The time evolution is then seen as a scale evolution. The generalization of these ideas to multispectral images depends on a discretization choice of the underlying partial differential equation (PDE). Due to stability requirements, semi-implicit schemes are preferred, leading to linear systems at each time step. In [4], the authors introduce a multigrid algorithm [5] to produce a scalable algorithm. For unorganized data sets, the formalism of difference equations over graphs [1] allows to adapt the PDE formulation to graphs of arbitrary topologies. Recently, an inverse scale space representations of functions on graphs has been proposed using the same bases [6]. One of the drawbacks of the scale-space theory is that the constructed stack of images does not verify the inclusion principle [7]. Even in nonlinear scale-space where the discontinuities are preserved, they are generally displaced from one level to the other.

Hierarchical representations were first designed to efficiently handle the multiscale nature of images. They generally combine filtering and subsampling. Regular pyramids keep a constant subsampling rate through the hierarchy. Adaptive pyramids [8], were introduced in order to make a global efficient interpretation of images. The focus is shifted from pixels to regions. The construction of the pyramid consists in decimating some cells while attaching them to surviving cells from one level to the other. Surviving cells are selected through an adjacency graph and a merging criterion. The merging criterion generally greatly affects the resulting pyramid. Another class of hierarchical representations are provided within the connected morphological filtering framework (see [9] for a review).

We propose a new hierarchical representation of general data lying on arbitrary graphs. The proposed algorithm allows to have a multiresolution representation of any data set that exhibits a graph structure. Furthermore, the merging part of the algorithm is based on discrete regularization rather than affection heuristics, which makes it more robust. The main contribution of this paper is to show how the construction of a hierarchical representation of data sets can benefit from the scale-space theory.

2. GRAPH REGULARIZATION

Graph regularization techniques [1] provide a common framework and tools to address the problem of regularization of discrete data sets. We begin by briefly recalling some of the notations and definitions that we will use. A weighted graph $G = (V, E, w)$ consists in a set of vertices V, a set of undirected edges $E \subseteq V \times V$, and a non-negative symmetric weight function $w : V \times V \to \mathbb{R}^+$ verifying the condition $w(\alpha, \beta) = 0$ if and only if $(\alpha, \beta) \notin E$. The function w represents a similarity measure between the vertices of the graph G. In the sequel, we will write $\beta \sim \alpha$ if $(\alpha, \beta) \in E$. We denote $\mathcal{H}(V)$ the set of real functions over V and $\mathcal{H}(E)$ the set of real functions over E. The space $\mathcal{H}(V)$ is endowed with the usual inner product : $(f, g)_{\mathcal{H}(V)} = \sum_{\alpha \in V} f(\alpha) g(\alpha)$ for $f, g \in \mathcal{H}(V)$. Similarly $(F, G)_{\mathcal{H}(E)} = \sum_{(\alpha,\beta) \in E} F(\alpha, \beta) G(\alpha, \beta)$ for $F, G \in \mathcal{H}(E)$. The difference operator $d_w : \mathcal{H}(V) \to \mathcal{H}(E)$ is defined as

$$(d_w f)(\alpha, \beta) := \sqrt{w(\alpha, \beta)}\,(f(\beta) - f(\alpha)),\ \forall (\alpha, \beta) \in E\,. \tag{1}$$

At each vertex, a discrete gradient norm can be computed

$$|(\nabla_w f)(\alpha)| = \sqrt{\sum_{\beta \sim \alpha} w(\alpha, \beta)(f(\alpha) - f(\beta))^2}\,. \tag{2}$$

The divergence operator $\mathrm{div}_w : \mathcal{H}(E) \to \mathcal{H}(V)$ is defined in a similar way to the continuous setting: $(d_w f, G)_{\mathcal{H}(E)} = -(f, \mathrm{div}_w G)_{\mathcal{H}(V)}$. Its expression is given by $(\mathrm{div}_w G)(\alpha) = \sum_{\beta \sim \alpha} \sqrt{w(\alpha, \beta)}(G(\alpha, \beta) - G(\beta, \alpha))$. Let $f^0 \in \mathcal{H}(V)$. The regularization procedure is expressed as the minimization of an energy function

$$E(u; f^0, \lambda) = \sum_{\alpha \in V} |(\nabla_w f)(\alpha)|^p + \frac{\lambda}{2} \sum_{\alpha \in V} (u_\alpha - f^0_\alpha)^2\,. \tag{3}$$

Motivated by the analogy to the continuous total variation (TV) minimization [10], we use the energy (3) with $p = 1$. The minimization of the function (3) with $p = 1$ can be viewed as a nonlocal/graph extension of the classical TV minimization. The latter model is strictly convex but non-smooth. A first approach is to introduce a regularization of the non-smooth term and to apply an iterative scheme to approximate the solution of the resulting system of nonlinear equations [1]. An alternative approach is to use a graph extension of Chambolle's algorithm [11]. The extension of the algorithm in [11] to a non-local setting has first been introduced in [12]. In our setting the characterization of the minimizer is the following

$$u = f - \pi_{\lambda^{-1} K}(f^0)\,, \tag{4}$$

$$\begin{aligned} K = \{v \in \mathcal{H}(V) : \exists p \in \mathcal{H}(E), v = \mathrm{div}_w(p), \\ |p(\alpha, .)| \leq 1\,, \forall \alpha \in V\}\,, \end{aligned} \tag{5}$$

where $\pi_{\lambda^{-1}K}$ is the orthogonal projection onto the closed convex set K and $|p(\alpha, .)| = \sqrt{\sum_{\beta \sim \alpha} p(\alpha, \beta)^2}$. The projection is computed through the following fixed-point algorithm

$$\begin{cases} p^0 = 0\,, \\ p^{n+1}(\alpha, \beta) = \dfrac{p^n(\alpha, \beta) + \tau d_w(\mathrm{div}_w p^n - \lambda f)(\alpha, \beta)}{1 + \tau |d_w(\mathrm{div}_w p^n - \lambda f)(\alpha, .)|}\,. \end{cases} \tag{6}$$

Let $\|\mathrm{div}_w\|$ denote the norm of the divergence operator induced by $\|.\|_{\mathcal{H}(V)}$ and $\|.\|_{\mathcal{H}(E)}$:

$$\|\mathrm{div}_w\| = \sup\{\|\mathrm{div}_w(p)\|_{\mathcal{H}(V)}, p \in \mathcal{H}(E), \|p\|_{\mathcal{H}(E)} \leq 1\}. \tag{7}$$

If $0 < \tau \leq \frac{1}{\|\mathrm{div}_w\|^2}$, then the iterative scheme in (6) converges to p^* and we have $u = f + \lambda^{-1} \mathrm{div}_w(p^*)$. In this work we use the algorithm (6)as it avoids to introduce a regularizing term.

3. HIERARCHICAL CLUSTERING

We present in this section our hierarchical clustering algorithm which is based on the tools presented in Section 2. The algorithm operates on a discrete data set whose structure is modeled by a weighted graph $G_0 = (V_0, E_0, w_0)$. The original data is defined as a function $f_0 \in \mathcal{H}(V_0)$. The bottom level of the hierarchy consists in the original graph G_0 and is denoted as level 0. Alternatively, to reduce the computational complexity, one can start with an initial partition such as super-pixels [13] for images. The partition at level $i + 1$ is generated from the partition at level i by first smoothing the data $f_i \in \mathcal{H}(V_i)$ following (6). This data driven diffusion process yields simplification of the current data and thus allows to identify the possible mergings. Several criteria can then be applied. In this paper, we consider adjacent nodes whose smoothed values are within a given tolerance ϵ to be merged in the same cluster. Vertices at level i belonging to a same cluster are then aggregated into *one* vertex at level $i+1$. Two nodes at level $i+1$ are connected by an edge if they have two representatives at level i that are connected by an edge. The function data f_{i+1} is chosen here as the average value of each cluster. This average value is then used as a distance measure between nodes at level $i + 1$. Finally, the distance is transformed into a similarity, typically by using a Gaussian function: $w(u, v) = e^{\frac{d(u,v)^2}{\sigma^2}}$. The process is stopped when a predetermined number of clusters is reached.
We now move one to some remarks concerning the algorithm. First, the identification of clusters is driven by the diffusion process. This has two important consequences. The first one concerns the sensitivity of the merging criterion which is greatly reduced. The second one concerns the nature of the merging decision. Depending on the graph structure, this decision may reflect nonlocal interactions between spatially distant vertices. The approach is summarized in Algorithm 1.

Algorithm 1 Hierarchical representation

```
1: INITIALIZATION : G_current = G_0, f_current = f_0
2: for i = 0 to n do
3:     f_curr ← regularization(f_i, λ)
4:     G_{i+1} ← merging(G_i, f_curr, merging criterion)
5:     f_{i+1} ← average(f_curr, G_{i+1})
6: end for
```

Step 3 of Algorithm 1 is the regularization step described in Section 2, which updates the data on the graph nodes. Step 4 is the merging step described above, based on any criterion. Finally, Step 5 computes the function on the coarse graph, based on the representatives of each vertex.

The hierarchical clustering algorithm is based on fusion criterion, in our case the parameter ϵ, and a scale parameter t through the evolutions described in (6). Both scale-space approach and irregular pyramids can be re-casted into this framework. Namely, when the tolerance parameter $\epsilon = 0$, and the fidelity parameter $\lambda = 0$, no aggregation is performed, and a scale-space stack is constructed. If there is no evolution of the scale parameter, an irregular pyramid is constructed.

4. EXPERIMENTS

In this section, we show the benefits of our proposal for obtaining a scale-space based hierarchical representation of discrete data. Several different types of discrete data sets are considered: digital images, image databases and point clouds. Each data set is represented by a specific graph: grid graph or nonlocal patch-based graphs for digital images, and neighborhood graphs for image databases and point clouds. First, we present results on digital images and show the benefits of using nonlocal patch-based graphs. In the case of digital images, our algorithm yields a hierarchy of partitions. It can be used for region-based segmentation as well as for image filtering. Figure 1 shows the scale-space based hierarchical representation of a color image at different resolution levels. Images in second and third rows represent the colorization of the regions obtained with our algorithm.

The second row of Figure 1 presents results when considering a 8-connectivity grid-graph weighted by a Gaussian kernel ($\sigma = 20$) with distances between pixels measured as the Euclidean distance between their color vectors. The threshold is fixed to $\sqrt{3}$. Two main effects are obtained as one progresses in the hierarchy: the graph topology is modified along the levels with the number of vertices decreasing, and the function attached to vertices is simplified by regularization. However one can remark that such a local processing cannot efficiently manage textured areas. The hierarchical representation degenerates as one progresses in the hierarchy and incorrect groupings are obtained. The third row of Figure 1 presents results with a 10-nearest neighbors graph with 5x5 patches. With our introduction of regularization on non-local patch-based graph (i.e. vertices non spatially close become neighbors and the similarity between vertices is measured using patches), the textured areas are much better handled. Indeed, vertices that represent similar textures are merged resulting in a much compact and accurate hierarchical representation. This stresses the interest of using regularization with non-local patch-based graphs for the generation of a hierarchical representation of images.

Second, we present results on point clouds. Figure 2. presents a 3D-projection of a points dataset. The data is represented by a 7-nearest neighbor graph weighted by $w = 1/d$ where d is the Euclidean distance between feature vectors. To generate the scale-space based hierarchical representation, a threshold of 0.01 is considered and this threshold evolves by

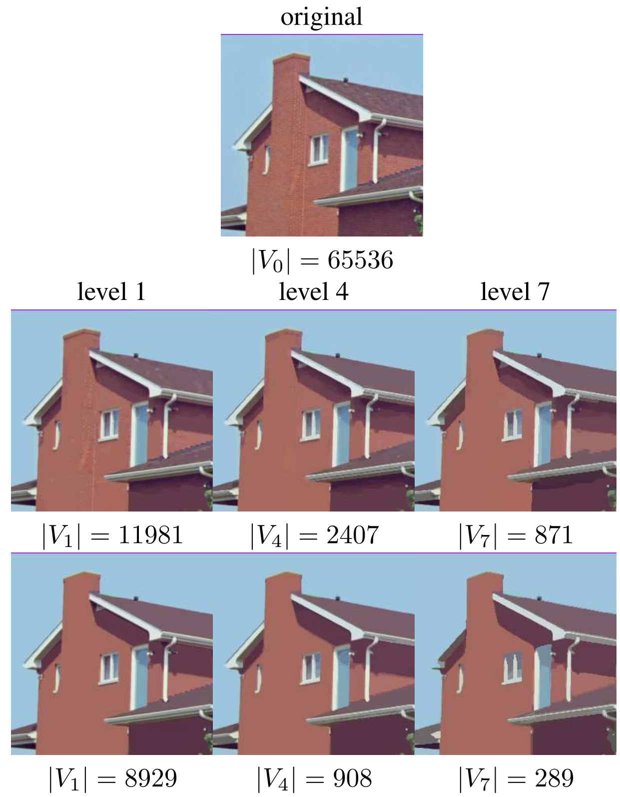

Fig. 1. Hierarchical representation of the top row image with local (second row) and nonlocal (third row) configurations. Numbers under images indicate the number of vertices at each level.

a factor 2 across levels. One can see that the regularization facilitates the grouping of vertices across levels resulting in a small hierarchy that respects the initial repartition of the data.

Finally, we consider image databases. To each image is associated a vertex and one has $F : V \to \mathbb{R}^{28\times 28}$ (which is the size of each image). The same experimental conditions than for point clouds are considered but with an initial threshold of 4. One can again see how our approach succeeds in generating a scale-space based hierarchical representation: images are simplified by regularization and grouped across levels.

5. CONCLUSION

In this paper, we have presented a new algorithm for the hierarchical representation of functions defined over the set of vertices of arbitrary graphs. It builds upon the techniques of discrete regularization to propose a robust approach. We showed the relevance of the approach to the representation of digital images, point clouds and image databases.

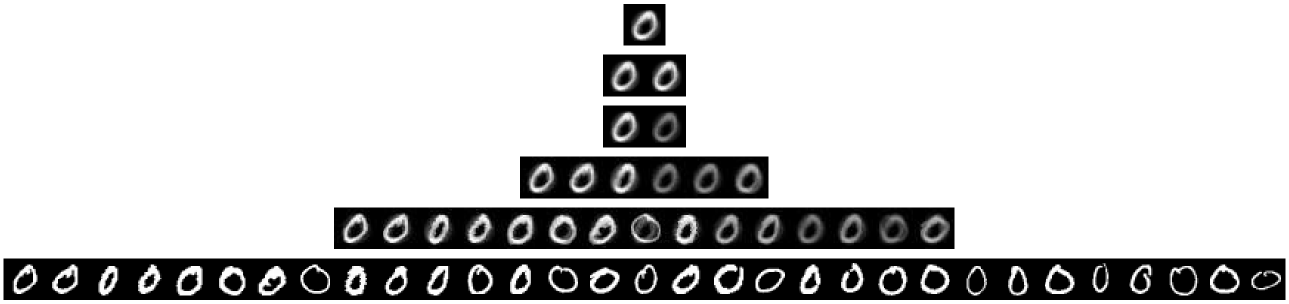

Fig. 2. Hierarchical representation of an image database. The bottom level represents the original data (level 0). The following levels range from 1 to 5.

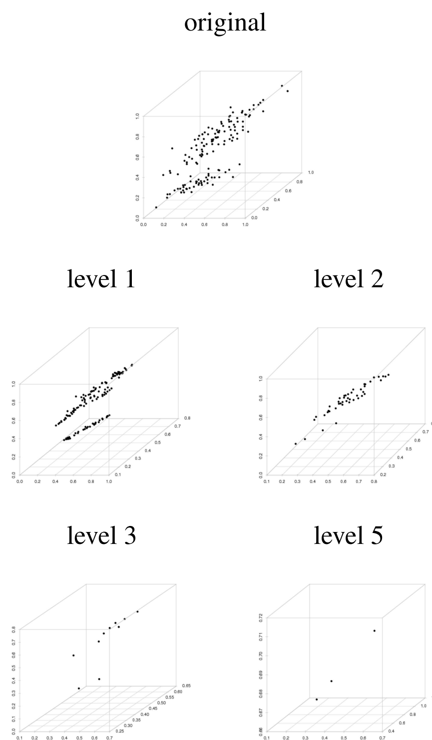

Fig. 3. Hierarchical representation of a point cloud.

6. REFERENCES

[1] A. Elmoataz, O. Lézoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing," *IEEE Transactions on Image Processing*, vol. 17, no. 7, 2008.

[2] A. P. Witkin, "Scale-Space Filtering.," in *8th Int. Joint Conf. Artificial Intelligence*, 1983, vol. 2.

[3] L. Alvarez, P. L. Lions, and J. M. Morel, "Image selective smoothing and edge detection by nonlinear diffusion. ii," *SIAM J. Numer. Anal.*, vol. 29, 1992.

[4] J. M. Duarte-Carvajalino, G. Sapiro, M. Vélez-Reyes, and P. Castillo, "Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 3, 2008.

[5] V.E. Henson W. L. Briggs and S. F. McCormick, *A multigrid tutorial (2nd ed.)*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000.

[6] M. Hidane, O. Lezoray, V.T. Ta, and A. Elmoataz, "Nonlocal multiscale hierarchical decomposition on graphs," in *European Conference on Computer Vision (ECCV)*, 2010, vol. LNCS 6314.

[7] S.L. Horowitz and T. Pavlidis, "Picture segmentation by a tree traversal algorithm," *J. ACM*, vol. 23, April 1976.

[8] J. M. Jolion and A. Montanvert, "The adaptive pyramid: a framework for 2d image analysis," *CVGIP: Image Underst.*, vol. 55, May 1991.

[9] P. Salembier and M. H. F. Wilkinson, "Connected operators: A review of region-based morphological image processing techniques," *IEEE Signal Processing Magazine*, vol. 26, no. 6, 2009.

[10] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, 1992.

[11] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, 2004.

[12] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling and Simulation*, vol. 7, no. 3, 2008.

[13] J. Malik X. Ren, "Learning a Classification Model for Segmentation," in *ICCV*. 2003, IEEE Computer Society.