

Hierarchical Representation of Discrete Data on Graphs

Moncef Hidane, Olivier Lézoray, and Abderrahim Elmoataz

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

{moncef.hidane,olivier.lezoray,abderrahim.elmoataz-billah}@unicaen.fr

Abstract. We propose a new hierarchical representation of discrete data sets living on graphs. The approach takes advantage of recent works on graph regularization. The role of the merging criterion that is common to hierarchical representations is greatly reduced due to the regularization step. The regularization is performed recursively with a decreasing fidelity parameter. This yields a robust representation of data sets. We show experiments on digital images and image databases.

1 Introduction

Multilevel techniques are now well established in image processing. Generally, these techniques fall into two categories : multiresolution and multiscale. The former class yields a stack of successively blurred images and is well understood within the scale-space theory [1], while the latter is well formalized within the Multiresolution Analysis (MRA) framework [2]. MRA generally decomposes a signal into a coarse approximation and a detail or residual part. In graph theory, such an analysis is carried out through a decimation of the vertex set and a reconstruction. We propose in this paper a decimation procedure of functions whose support is the vertex set of a weighted graph. The weighted graph structure is of interest since it encapsulates pairwise interactions between discrete data instances. Furthermore, a graph structure can be obtained after sampling a continuous manifold. Our proposal is based on recent works on graph regularization and difference equations on graphs [3], [4]. The algorithm we propose produces a hierarchy of graphs and functions defined on their vertex sets. Starting with the initial data associated to a graph structure, successive coarsening procedures are applied. The coarsening is based on a preliminary graph partitioning which is mainly driven by a discontinuity-preserving graph regularization. The use of graph regularization yields a more robust representation. We show the applicability of our proposal to digital image hierarchical representation. In this case, the resulting representation can be seen as new adaptive irregular pyramidal representation of images [5]. The paper is organized as follows: in Section 2 we recall the graph regularization framework and present the algorithm we use. Section 3 details the different steps that lead to the representation we seek: regularization, grouping and coarsening. We present experiments and conclude in Section 4.

2 Graph Total Variation

2.1 Definitions

Throughout this section we assume that we are given a weighted graph $G = (V, E, w)$ consisting of a vertex set V , and an edge set $E \subseteq V \times V$. The nonnegative weight function $w : E \rightarrow \mathbb{R}^+$ is supposed symmetric: $w(\alpha, \beta) = w(\beta, \alpha)$ for all $(\alpha, \beta) \in E$. For a given graph edge $(\alpha, \beta) \in E$, the quantity $w(\alpha, \beta)$ represents a similarity or proximity measure between the vertices α and β . This measure is usually computed as a decreasing function of a prior distance measure. For $\alpha, \beta \in V$ we denote $\alpha \sim \beta$ if $(\alpha, \beta) \in E$. The graphs we consider in this paper are undirected with no self loops.

We denote by $\mathcal{H}(V)$ the set of functions that assign a real value to each vertex of the graph G and $\mathcal{H}(E)$ the set of functions that assign a real value to each edge. The sets $\mathcal{H}(V)$ and $\mathcal{H}(E)$ are equipped with the standard inner products denoted $\langle \cdot, \cdot \rangle_{\mathcal{H}(V)}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}(E)}$. The graph difference operator, $d_w : \mathcal{H}(V) \rightarrow \mathcal{H}(E)$, is defined as: $(d_w f)(\alpha, \beta) = \sqrt{w(\alpha, \beta)}(f(\beta) - f(\alpha))$, $f \in \mathcal{H}(V)$, $(\alpha, \beta) \in E$. The graph divergence operator, $\text{div}_w : \mathcal{H}(E) \rightarrow \mathcal{H}(V)$, is related, as in the continuous setting, to the adjoint of d_w :

$$\langle d_w f, G \rangle_{\mathcal{H}(E)} = -\langle f, \text{div}_w G \rangle_{\mathcal{H}(V)}, \quad \text{for all } f \in \mathcal{H}(V), G \in \mathcal{H}(E). \quad (1)$$

Its expression is given by: $(\text{div}_w G)(\alpha) = \sum_{\beta \sim \alpha} \sqrt{w(\alpha, \beta)}(G(\alpha, \beta) - G(\beta, \alpha))$. For $p \in \mathcal{H}(E)$ and $\alpha \in V$, we denote $|p|_\alpha = \sqrt{\sum_{\beta \sim \alpha} p(\alpha, \beta)^2}$. A path joining two vertices $\alpha, \beta \in V$ is a sequence of vertices $(\gamma_1, \dots, \gamma_n)$ such that $\gamma_1 = \alpha$, $\gamma_n = \beta$ and $(\gamma_i, \gamma_{i+1}) \in E$, $i = 1, \dots, n - 1$.

2.2 Minimization

Let $f \in \mathcal{H}(V)$ associated with a graph structure G . The graph total variation (TV) of f is defined as: $TV_w(f) = \sum_{\alpha \in V} \sqrt{\sum_{\beta \sim \alpha} w(\alpha, \beta)(f(\alpha) - f(\beta))^2}$. Let $f_0 \in \mathcal{H}(V)$ be a possibly noisy data. In order to smooth f_0 we seek the minimum of the following functional: $E(f; f_0, \lambda) = TV_w(f) + \frac{\lambda}{2} \sum_{\alpha \in V} (f(\alpha) - f_0(\alpha))^2$. The parameter λ controls the amount of smoothing being applied to f_0 .

Functional E corresponds to the particular case of $p = 1$ in the family of functionals introduced in [3]. It has been studied in [6] and has found numerous applications in image and mesh processing [3], and data filtering [4]. In [7], the authors propose to adapt the penalization to the topology of the underlying function. The same approach has been taken in [8] for motion deblurring. Recently, [9] used functional E as a tool to generate an inverse scale space representation of discrete data on graphs.

Functional E is strictly convex but nonsmooth. Rather than smoothing the penalty term and differentiating, we use an adaption of Chambolle’s projection algorithm [10] to graphs of arbitrary topologies. This adaption first appeared in [6]. In our notations, the solution of $\min \{E(f; f_0, \lambda), f \in \mathcal{H}(V)\}$ is given by :

$f = f_0 - \lambda^{-1} \operatorname{div}_w(p^\infty)$, where p^∞ is the limit of the following fixed point iterative scheme:

$$\begin{cases} p^0 = 0, \\ p^{n+1}(\alpha, \beta) = \frac{p^n(\alpha, \beta) + \tau (d_w(\operatorname{div}_w p^n - \lambda f))(\alpha, \beta)}{1 + \tau |d_w(\operatorname{div}_w p^n - \lambda f)|_\alpha}, \quad (\alpha, \beta) \in E. \end{cases} \quad (2)$$

If $0 < \tau \leq \frac{1}{\|\operatorname{div}_w\|^2}$, where $\|\operatorname{div}_w\|$ is the norm of the graph divergence operator, then (2) converges [6]. We use algorithm (2) in the next section as a tool to detect the possible groupings at different scales. Figure 1 shows the grouping effect yielded by a regularization of a color image and a triangular mesh. One should notice the preservation of discontinuities in the results.

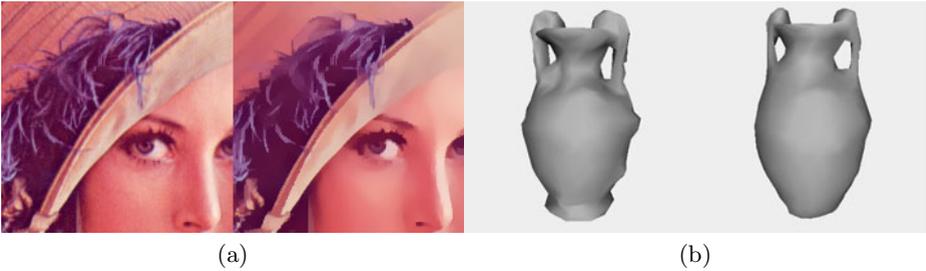


Fig. 1. Grouping effect of TV regularization. (a): left figure: original image; right: result of color components regularization with $\lambda = 0.01$. (b): left : original triangular mesh; right : result of spatial coordinates regularization with $\lambda = 0.01$.

3 Hierarchical Representation

3.1 TV as a Tool for Graph Partitioning

We propose to use TV regularization as a tool to detect the possible partitions in a given graph. The regularization yields a more regular data with respect to the graph TV prior while staying close to the original observations. The degree of closeness is inferred from the parameter λ . Once the TV regularization has been performed, a partitioning can be obtained by considering an equivalence relation on the vertex set. Let $G_i = (V_i, E_i, w_i)$ denote a given weighted graph, $f_i \in \mathcal{H}(V_i)$ and $\lambda > 0$. Let f_i^* be the result of TV regularization of f_i with parameter λ . We associate with each vertex $\alpha \in V_i$ a feature vector $F_i(\alpha)$ whose components are based on f_i^* . For instance, in image processing, the feature vector $F_i(\alpha)$ could consist of the values of $(f_i^*(\beta), \beta \in \mathcal{N}(\alpha))$ where $\mathcal{N}(\alpha)$ is an image patch centered at α . Define the metric d on V_i as the Euclidean distance between feature vectors: $d(\alpha, \beta) = \|F_i(\alpha) - F_i(\beta)\|_2, \alpha, \beta \in V_i$. Consider the following binary relation \mathcal{R}^ϵ on V_i : $\alpha \mathcal{R}^\epsilon \beta$ if $\alpha = \beta$ or if there exists a path $(\gamma_1, \dots, \gamma_n)$ joining α and β such that: $d(\gamma_j, \gamma_{j+1}) < \epsilon$, for all $j = 1, \dots, n - 1, \epsilon > 0$. The relation \mathcal{R}^ϵ is reflexive,

symmetric (G_i is undirected), and transitive. It is an equivalence relation. The quotient set V_i/\mathcal{R}^ϵ of V_i by \mathcal{R}^ϵ is a partition of V_i : $V_i/\mathcal{R}^\epsilon = \{[\alpha], \alpha \in V_i\}$ and $[\alpha] = \{\beta \in V_i : \alpha \mathcal{R}^\epsilon \beta\}$. The partition yielded by \mathcal{R}^ϵ can be seen as a region growing algorithm. The strength of the grouping is controlled by the parameter ϵ . It is important to understand that TV regularization will simplify the initial data and similar vertices will become closer. This will enable us to keep the parameter ϵ fixed in contrast to region merging techniques that do rely on variable thresholds.

3.2 Graph Coarsening

Let $P^\epsilon(V_i) = \{P_{i,1}, \dots, P_{i,n_i}\}$ denote the partition of V_i obtained through the equivalence relation \mathcal{R}^ϵ . In the sequel, we call the elements of $P^\epsilon(V_i)$ parts of V_i . We construct a coarse graph $G_{i+1} = (V_{i+1}, E_{i+1}, w_{i+1})$ by aggregating the nodes belonging to each part. Let $\gamma \in V_{i+1}$ be a vertex in the coarse graph. We denote R_γ^i the set of vertices in V_i which have been aggregated into γ in V_{i+1} . Two nodes $\gamma_1, \gamma_2 \in V_{i+1}$ are connected by a coarse edge if there exists $\alpha_1 \in R_{\gamma_1}^i$, $\alpha_2 \in R_{\gamma_2}^i$ such that α_1 and α_2 are connected by a (fine) edge. In the latter case, we adopt the notation $R_{\gamma_1}^i \sim R_{\gamma_2}^i$.

In order to take account of the volumes of the parts obtained by the partitioning, the edges of the coarse graph should be weighted. We use the ratio-cut measure between two parts in the fine graph as the weight between their aggregates in the coarse graph: $w_{i+1}(\gamma_1, \gamma_2) = \frac{cut(R_{\gamma_1}^i, R_{\gamma_2}^i)}{|R_{\gamma_1}^i|} + \frac{cut(R_{\gamma_1}^i, R_{\gamma_2}^i)}{|R_{\gamma_2}^i|}$, where $cut(A, B) = \sum_{(a,b) \in A \times B} w(a, b)$ is the edge cut between A and B . Once the coarse graph $G_{i+1} = (V_{i+1}, E_{i+1}, w_{i+1})$ has been constructed, we define a new function $f_{i+1} \in \mathcal{H}(V_{i+1})$ by averaging the values of each part: $f_{i+1}(\gamma) = \frac{1}{|R_\gamma^i|} \sum_{\alpha \in R_\gamma^i} f_{i+1}(\alpha)$, $\gamma \in V_{i+1}$.

3.3 Recursive Construction of the Hierarchy

We have showed in the two previous sections how to construct a weighted coarse graph from an input graph and function pair. We now move on to see how this process can be repeated to generate a hierarchy of graphs.

A hierarchical representation can be obtained by varying the ϵ parameter. However, we do not follow this direction in this section and ϵ will remain fixed within all the hierarchy: the partitioning is induced by the TV regularization. We seek to adapt the different levels of the representation to the local properties of the data.

The hierarchical representation is based on recursive partitioning and coarsening as described above. In order to adapt to the local properties of data, the fidelity parameter λ should evolve through the hierarchy. In our case, λ should decrease through the coarsening process, favoring more regularity and less fidelity as the hierarchy evolves. In our experiments, we have chosen a dyadic progression $\lambda_{i+1} = \frac{\lambda_i}{2}$. The initial regularization is responsible for denoising the initial data. It yields a choice for the first fidelity parameter λ_0 which is set

to $\frac{1}{\sigma^2}$ where σ^2 is the variance of the noise, which we suppose Gaussian (see [11]). The standard deviation can be estimated through the standard estimator: $\hat{\sigma} = 1.4826 \text{MAD}(f_0(\alpha), \alpha \in V)$, where MAD is the median absolute deviation estimator [12]. Finally we summarize the algorithm:

Algorithm 1. Hierarchical representation of discrete data on graphs

- 1: INPUT: $G_0 = (V_0, E_0, w_0)$, $f_0 \in \mathcal{H}(V_0)$, $\lambda_0 = \frac{1}{\sigma^2}$, $n \geq 1$, ϵ fixed
 - 2: **for** $i = 0$ **n do**
 - 3: Regularization : $f_i^* \leftarrow \arg \min \{E(f; f_i, \lambda_i), f \in \mathcal{H}(V_i)\}$
 - 4: Partitioning of V_i based on f_i^* through equivalence relation \mathcal{R}^ϵ to yield $P(V_i) = \{P_{i,1}, \dots, P_{i,n_i}\}$
 - 5: Coarse nodes: Aggregate each part to yield $V_{i+1} = \{j_1, \dots, j_{n_i}\}$
 - 6: Coarse edges : $E_{i+1} = \{e_{\alpha,\beta} : \alpha, \beta \in V_{i+1} \text{ and } R_\alpha^i \sim R_\beta^i\}$
 - 7: Coarse weights: $w_{i+1}(\alpha, \beta) = \frac{\text{cut}(R_\alpha^i, R_\beta^i)}{|R_\alpha^i|} + \frac{\text{cut}(R_\alpha^i, R_\beta^i)}{|R_\beta^i|}$, $(\alpha, \beta) \in E_{i+1}$
 - 8: Coarse function $f_{i+1}(\alpha) = \frac{1}{|R_\alpha^i|} \sum_{\beta \in R_\alpha^i} f_i(\beta)$, $\alpha \in V_{i+1}$
 - 9: Update the fidelity parameter : $\lambda_{i+1} = \lambda_i/2$
 - 10: **end for**
-

4 Experiments and Conclusion

We begin by applying our approach to digital images. The algorithm we propose leads to a hierarchy of partitions of an image. Each pixel is represented by a vertex. In the experiments, we have chosen an eight connectivity graph. The edge weights are computed as follows: $w(\alpha, \beta) = \exp(-\frac{d^2(\alpha, \beta)}{\sigma^2})$, where d is the Euclidean distance between two RGB color vectors. The function to regularize is the one that assigns to each pixel its RGB color values. The regularization of multivalued functions is carried on each component but with a common total variation prior. The merging at the first stage is based on the distance between RGB color patches (5x5 in our case). At the following stages, its is based on vertex-wise distance. In all cases, the parameter ϵ was set to one. Figures 2 and 3 show the result of the regions obtained as well as their colorizations based on the original image.

We also show an application of our approach to image databases. Here each vertex represents a given image. The edges are obtained by considering a nearest neighbor graph (NNG) weighted with $w = 1/d$. The number of neighbors was set to 7. Figure 4 shows the hierarchy obtained. One should notice that the graph structure evolves as well as the image data. This yields simplification as well as decimation.

As Figures 2 and 3 show, its is difficult to get rid of outlier pixels in the first levels. Its seems interesting to adopt an approach based on concentration inequalities as used in [13] to replace the equivalence relation \mathcal{R}^ϵ . This will be the subject of a future work.

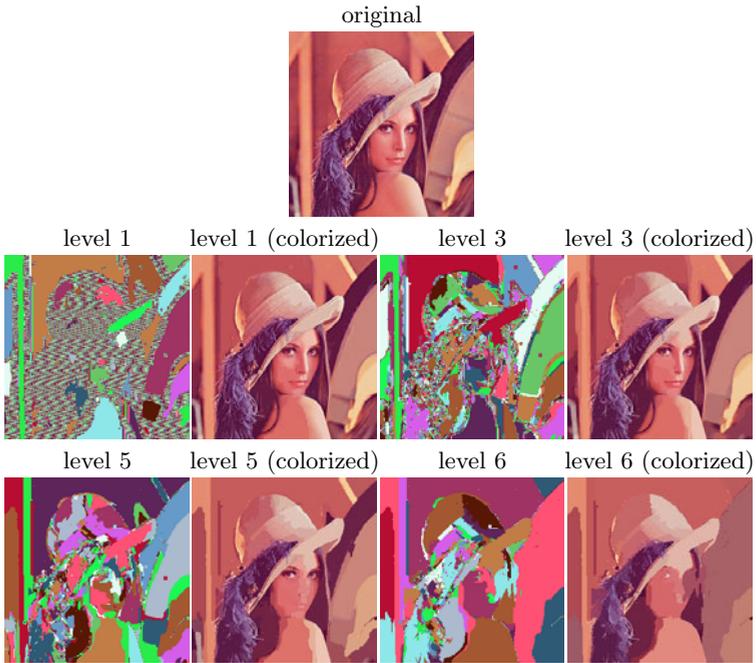


Fig. 2. Hierarchy of partitions and corresponding colorizations. Levels 1, 3, 5 and 6.



Fig. 3. Hierarchy of partitions and corresponding colorizations. Levels 1, 3, 5 and 6.

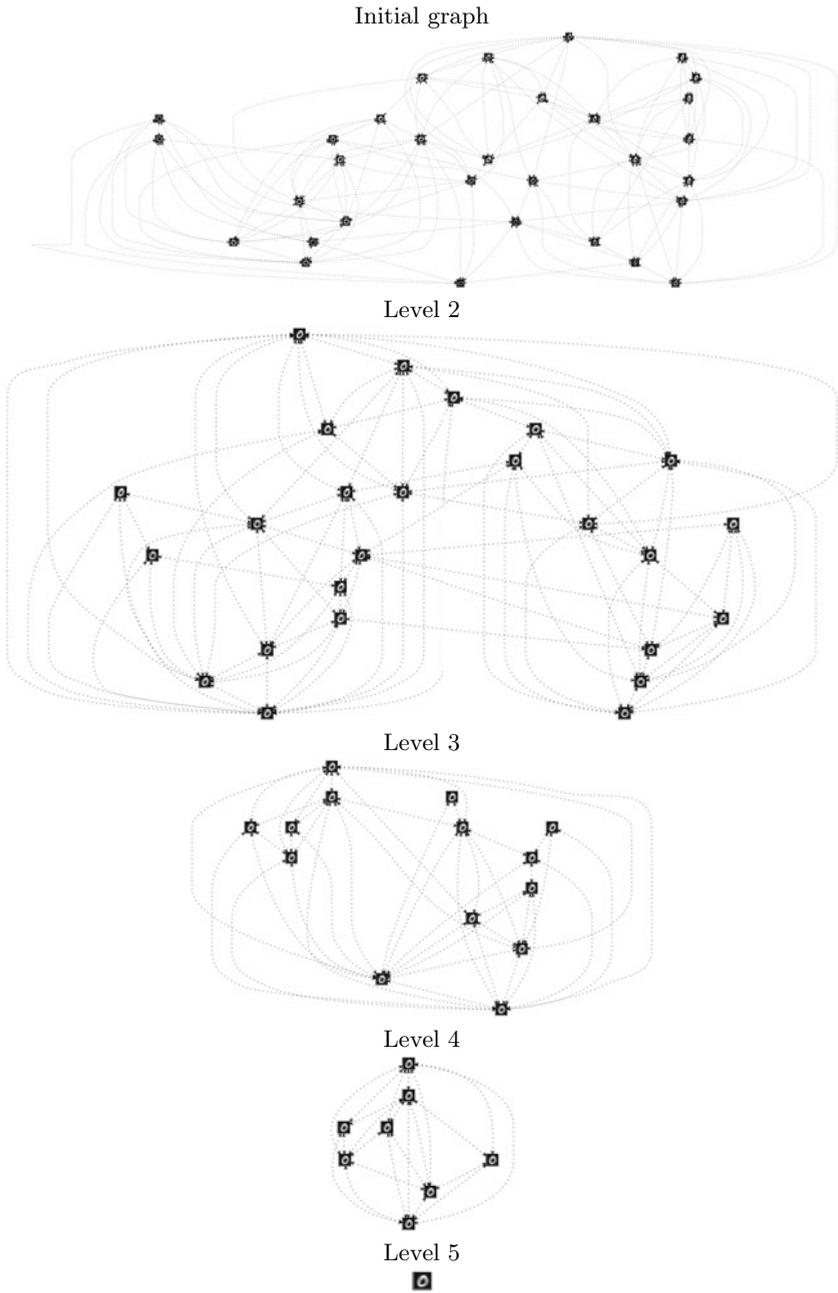


Fig. 4. 0-digits database hierarchical representation

References

1. Weickert, J.: Anisotropic Diffusion in Image Processing. ECMI Series. Teubner (1998)
2. Mallat, S.: Wavelet Tour of Signal Processing, 3rd edn. The Sparse Way. Academic Press, London (2008)
3. Elmoataz, A., Lezoray, O., Bogleux, S.: Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing. *IEEE Transactions on Image Processing* 17, 1047–1060 (2008)
4. Lezoray, O., Ta, V.T., Elmoataz, A.: Partial differences as tools for filtering data on graphs. *Pattern Recognition Letters* 31, 2201–2213 (2010)
5. Brun, L., Kropatsch, W.: Irregular pyramids with combinatorial maps. In: Amin, A., Pudil, P., Dori, D. (eds.) *SPR 1998 and SSPR 1998*. LNCS, vol. 1451, pp. 256–265. Springer, Heidelberg (1998)
6. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. *Multiscale Modeling and Simulation* 7, 1005–1028 (2008)
7. Peyré, G., Bogleux, S., Cohen, L.: Non-local regularization of inverse problems. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part III*. LNCS, vol. 5304, pp. 57–68. Springer, Heidelberg (2008)
8. Yun, S., Woo, H.: Linearized proximal alternating minimization algorithm for motion deblurring by nonlocal regularization. *Pattern Recogn.* 44 (2011)
9. Hidane, M., Lezoray, O., Ta, V., Elmoataz, A.: Nonlocal multiscale hierarchical decomposition on graphs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010*. LNCS, vol. 6314, pp. 638–650. Springer, Heidelberg (2010)
10. Chambolle, A.: An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision* 20 (2004)
11. Chan, T.F., Osher, S., Shen, J.: The digital TV filter and nonlinear denoising. *IEEE Transactions on Image Processing* 10, 231–241 (2001)
12. Black, M., Sapiro, G.: Edges as outliers: Anisotropic smoothing using local image statistics. In: Nielsen, M., Johansen, P., Fogh Olsen, O., Weickert, J. (eds.) *Scale-Space 1999*. LNCS, vol. 1682, pp. 259–270. Springer, Heidelberg (1999)
13. Nielsen, F., Nock, R.: Fast graph segmentation based on statistical aggregation phenomena. In: Ikeuchi, K. (ed.) *Proceedings of the 10th IAPR Conference on Machine Vision Applications* (2007)