

Geometric Image Labeling with Global Convex Labeling Constraints

Artjom Zern^{1,2}, Karl Rohr², and Christoph Schnörr¹

¹ Image and Pattern Analysis Group, Heidelberg University, Germany

² Biomedical Computer Vision Group, BIOQUANT, Heidelberg University, Germany

Abstract. In [2], a smooth geometric labeling approach was introduced by following the Riemannian gradient flow of a given objective function on the so-called assignment manifold. The approach evaluates a user-defined data term and performs spatial regularization by Riemannian averaging of the assignment vectors. In this paper, we extend this approach in order to impose global convex constraints on the labeling results based on linear filter statistics in the label space. The smoothness of the approach is preserved by using logarithmic barrier functions to handle the new constraints. We discuss how suitable filters can be determined from example data of a given image class, and we demonstrate numerically the effectiveness of the constraints in several academic labeling scenarios.

Keywords: Image labeling, assignment manifold, statistical label constraints, Riemannian gradient flow, information geometry.

1 Introduction

The *discriminative* power of *filter statistics* for object detection and classification is well known [7, 8] and has been widely explored in the literature. The *generative* power of filter statistics for representing image structure, on the other hand, has been less explored during the recent years. The present paper focuses on a mathematically sound and numerically tractable approach to impose filter statistics on *labelled* image structure.

Early seminal work on generative aspects of filter statistics includes [14] and [10] and many references in these papers. In the former case, heavy-tailed empirical filter statistics are imposed on the variational problem of learning the parameters of a Gibbs-Boltzmann distribution. In the latter case, several hundred filter constraints form nonlinear submanifolds (level sets) onto which a given image has to be projected. While both works impressively demonstrate the generative power of filter statistics, exploiting these statistics as prior knowledge for inference and reproducibility of results has remained a challenge from the

Acknowledgments: We gratefully acknowledge support by the German Science Foundation, grant GRK 1653.

viewpoint of algorithm design and numerical optimization. This assessment also applies to current mainstream research with a focus on the engineering of deep networks [13], without denying the remarkable quality of corresponding experimental results.

Contribution. The present work conforms to this research direction but deviates in the following aspects:

1. We focus on filter statistics in *label space* rather than in image space. The simplest constraint, for example, imposes lower and upper bounds on the area occupied by some label, without specifying the corresponding locations, of course. More general constraints arise from replacing the ‘identity filter’ by linear filters learned offline through a simple generalized eigenvalue technique, and imposing similar linear statistical constraints. While such statistical moments can be taken into account as constraints using graphical models, in principle, this would again lead like [14] to maximum-entropy distributions in Gibbs-Boltzmann form [5], that are intractable regarding both learning and inference, due to the *global* nature of these constraints. In fact, a recent assessment of approaches to inference with discrete graphical models [6] revealed the limited capability of established state-of-the-art solvers in this respect, i.e. to handle cliques of *large* size of the underlying graph.
2. We focus on a numerically *tractable and reproducible* way to incorporate such constraints into an algorithm for image labeling. To this end, we adopt the recent approach [2] to image labeling based on simple geometric averaging induced by the Fisher-Rao metric on the so-called assignment manifold, i.e. the relative interior of a product of probability simplices, whose vertices represent discrete decisions as is common with graphical models and convex variational relaxations. A key aspect of the approach [2] is that the usual two-step procedure of (i) solving the LP relaxation [12] by some iterative method, and (ii) projecting back the solution to the set of integral solutions, is combined into a *single smooth* process that converges to integral solutions (labelings). The objective of the present paper is to show that *filtered label statistics* can be taken into account in a straightforward and comprehensible way by using standard log-barrier constraints [9, 4] and geometric numerical integration [11].

Organization. Section 2 sketches the works [2, 11] on which the present paper is based. Section 3 details our contribution: learning filters for label statistics and taking corresponding empirical constraints into account during inference for image labeling. We do not focus on any specific application in this paper. Rather, the proof-of-concept experiments discussed in Section 4 are supposed to demonstrate how statistics gathered by linear filters of small support can *enhance image labeling*, represent *primitive shape information* and support *spatial pattern formation*, by extending the geometric non-convex approach [2] through corresponding *convex* constraints.

Basic Notation. Functions and binary operations are applied component-wise to vectors and matrices, i. e. for $u, v \in \mathbb{R}^n$ we have $\sqrt{u} = (\sqrt{u_1}, \dots, \sqrt{u_n})^\top$, $u \cdot v = (u_1 v_1, \dots, u_n v_n)^\top$ and similar for e^u , $\log(u)$ and $\frac{u}{v}$. We set $[n] = \{1, \dots, n\}$

and $\mathbf{1}_n = (1, \dots, 1)^\top \in \mathbb{R}^n$ as well as $\mathbf{1}_{m \times n} = (\mathbf{1}_m, \dots, \mathbf{1}_m) \in \mathbb{R}^{m \times n}$. By $\langle \cdot, \cdot \rangle$ we will denote the Euclidean inner product on \mathbb{R}^n or the Frobenius inner product on $\mathbb{R}^{m \times n}$. For a matrix $W \in \mathbb{R}^{m \times n}$ we will denote the i -th row by $W_i \in \mathbb{R}^n$ and the j -th column by $W^j \in \mathbb{R}^m$. Elements of a sequence are indexed with an upper script index enclosed in brackets, for example, $W^{(k)} \in \mathbb{R}^{m \times n}$.

2 Image Labeling on the Assignment Manifold

We briefly summarize the smooth label assignment approach introduced in [2]. This approach will be extended in the next section in order to handle global constraints imposed on labelings.

Given an image with m pixels and a set $\mathcal{L} = \{l^{(1)}, \dots, l^{(n)}\}$ of n predefined labels, the task is to assign each pixel $i \in [m]$ one label in \mathcal{L} . The labeling problem can be formulated as finding an optimal assignment matrix in

$$\overline{\mathcal{W}}^* = \{W \in \mathbb{R}^{m \times n} : W_i \in \{e^{(1)}, \dots, e^{(n)}\} \subset \mathbb{R}^n, \forall i \in [m]\}, \quad (2.1)$$

where each label $l^{(j)} \in \mathcal{L}$ is represented by a vertex $e^{(j)} \in \mathbb{R}^n$ of the probability simplex. In [2], a smooth geometric approach was presented which is defined on the assignment manifold

$$\mathcal{W} := \{W \in \mathbb{R}_{>0}^{m \times n} : \langle W_i, \mathbf{1}_n \rangle = 1, \forall i \in [m]\} \subset \mathbb{R}_{>0}^{m \times n}, \quad (2.2)$$

that is the set of all row-stochastic matrices with full support. This is a smooth manifold with tangent space at $W \in \mathcal{W}$ given by

$$\mathcal{T} := T_W \mathcal{W} = \{V \in \mathbb{R}^{m \times n} : \langle V_i, \mathbf{1}_n \rangle = 0, \forall i \in [m]\}. \quad (2.3)$$

The assignment manifold \mathcal{W} is turned into a Riemannian manifold by equipping it with the Fisher-Rao metric

$$g_W^{\mathcal{W}}(U, V) = \left\langle \frac{U}{\sqrt{W}}, \frac{V}{\sqrt{W}} \right\rangle \quad \text{for } U, V \in T_W \mathcal{W}, \quad W \in \mathcal{W} \quad (2.4)$$

with *componentwise multiplication* (and subdivision) of vectors and matrices (with strictly positive support).

Input data for the assignment approach are pixel neighborhoods $\mathcal{N}(i) = \{j \in [m] : i \sim j\}$, $i \in [m]$ defined by the adjacency relation (edges) of an underlying graph, and a distance matrix $D \in \mathbb{R}^{m \times n}$ whose components D_{ij} store the application-specific distance between the image data observed at pixel $i \in [m]$ and label $l^{(j)} \in \mathcal{L}$. The goal is to find an assignment $W \in \overline{\mathcal{W}}$ which is spatially consistent with respect to neighborhood assignments, on the one hand, and reflects the data represented by the distance matrix D as closely as possible, on the other hand. This is accomplished by computing a curve $W(t) \in \mathcal{W}$, $t \geq 0$ on the assignment manifold that converges to an *integral* solution and locally minimizes a functional $J(W)$ which accounts for the given data and regularization.

The ingredients for defining a corresponding sequence

$$W^{(k)} = W(t_k), \quad W^{(0)} = W(0) = C := \frac{1}{n} \mathbb{1}_{m \times n} \quad (2.5)$$

are the *barycenter* C of \mathcal{W} , an approximation of the exponential mapping of \mathcal{W} given by

$$\exp_W : \mathcal{T} \rightarrow \mathcal{W}, \quad \exp_W(V)_i = \frac{W_i \cdot e^{V_i}}{\langle W_i, e^{V_i} \rangle}, \quad \forall i \in [m], \quad (2.6)$$

and the orthogonal projection onto the tangent space (2.3)

$$\Pi_{\mathcal{T}} : \mathbb{R}^{m \times n} \rightarrow \mathcal{T}, \quad \Pi_{\mathcal{T}}(D) = D - \frac{1}{n} D \mathbb{1}_{n \times n}. \quad (2.7)$$

The data D is taken into account by the likelihood matrix

$$L(W) = \exp_W(\Pi_{\mathcal{T}}(-D)) \in \mathcal{W}, \quad (2.8)$$

whereas regularization is performed by computing approximate Riemannian means of the assignment vectors $\{L(W)_j\}_{j \in \mathcal{N}(i)}$ over spatial neighborhoods $\mathcal{N}(i)$, for each pixel $i \in [m]$. We refer to [2] for details.

We adopt the general numerical scheme suggested by [11],

$$\dot{V}(t) = -\Pi_{\mathcal{T}}[\nabla J(W(t))], \quad W(t) = \exp_C(V(t)), \quad V(0) = 0, \quad (2.9)$$

which enables to apply standard algorithms for integrating the flow $V(t)$ on the tangent space \mathcal{T} so as to determine a minimizing path $W(t)$ on the manifold \mathcal{W} . For example, combining the simplest integration method, i. e. explicit Euler steps, with smooth rounding to an integral solution leads to a sequence (2.5) given by

$$W_i^{(k+\frac{1}{2})} = \frac{W_i^{(k)} \cdot e^{-h \nabla_{W_i} J(W^{(k)})}}{\langle W_i^{(k)}, e^{-h \nabla_{W_i} J(W^{(k)})} \rangle}, \quad (2.10a)$$

$$W_i^{(k+1)} = \frac{W_i^{(k)} \cdot W_i^{(k+\frac{1}{2})}}{\langle W_i^{(k)}, W_i^{(k+\frac{1}{2})} \rangle}, \quad i \in [m]. \quad (2.10b)$$

We explain in the subsequent section how global labeling constraints can be taken into account within this framework.

3 Label Assignment with Global Constraints

This section details the class of global constraints that we impose on label assignments (Section 3.1), how linear filters defining these constraints are learned offline using basic techniques of numerical linear algebra (Section 3.2), and finally, in Section 3.3, how these constraints are taken into account using the assignment approach of Section 2.

3.1 Global Constraints

In order to incorporate some prior knowledge about the labelings, we consider linear $p \times p$ filters $h \in \mathcal{H} \subset \mathbb{R}^{p^2 \times n}$ operating on assignment matrices $W \in \overline{\mathcal{W}}$: For each label $j \in [n]$, we have a $p \times p$ filter $h^j \in \mathbb{R}^{p^2}$ in the usual sense, and the filter operation is given by

$$h * W := \sum_{j \in [n]} h^j * W^j, \quad (3.1)$$

with the common convolution of the ‘label images’ W^j , $j \in [n]$ with a $p \times p$ filter on the right-hand side. The space of filters \mathcal{H} will be specified in Section 3.2. To avoid complications at and close to the boundary of the image region, we only take into account filter results $(h * W)_i$ at interior pixels i where the $p \times p$ filter support (centered at i) does not overlap with the boundary.

The filter result $(h * W)_i$ at a pixel $i \in [m]$ depends on the assignment within a $p \times p$ neighborhood of i and hence reflects the local spatial relation of the labels. Our objective is to control label assignments by constraining the filter results for a set

$$\{h^{(k)} \in \mathcal{H}: k = 1, \dots, K\} \quad (3.2)$$

of K filters in order to take into account statistical prior information about the local geometry of labelings. Motivated by [3], where the ℓ^1 -norm of the filter results of a grayscale image was considered in connection with non-smooth sparse regularization, we consider here the ℓ^2 -norm of filter results which conforms to our *smooth* geometric label assignment scheme of Section 2.

Specifically, we consider *global convex constraints* of the form

$$c_{\text{low}} \leq \frac{1}{m} W^\top \mathbb{1}_m \leq c_{\text{up}}, \quad (3.3a)$$

$$\|h^{(k)} * W\|_{\ell^2} \leq d^{(k)}, \quad k = 1, \dots, K, \quad (3.3b)$$

where the parameter vectors $c_{\text{low}}, c_{\text{up}} \in \mathbb{R}^n$ impose lower and upper cardinality bounds for the assignment of each label of $\mathcal{L} = \{l^{(1)}, \dots, l^{(n)}\}$ to the range of pixels $[m]$, whereas the parameters $d^{(k)}$ of (3.3b) constrain the output energy of each filter $h^{(k)}$, $k \in [K]$. To ensure that the region of feasible assignments W has a non-empty interior, we require

$$c_{\text{low}} < c_{\text{up}}, \quad \langle c_{\text{low}}, \mathbb{1}_n \rangle < 1 < \langle c_{\text{up}}, \mathbb{1}_n \rangle \quad \text{and} \quad d^{(k)} > 0, \quad k \in [K]. \quad (3.4)$$

As alternative to the ℓ^2 -norm defining (3.3b), we also used a smooth approximation of the ℓ^1 -norm denoted by

$$\|x\|_{\ell_\varepsilon^1} = \sum_{i \in [m]} |x_i|_\varepsilon, \quad |x_i|_\varepsilon = \sqrt{x_i^2 + \varepsilon^2} - \varepsilon. \quad (3.5)$$

3.2 Learning Filters

We discuss how to choose filters for given classes of labelings and corresponding example data. First of all, we restrict the space of all possible $p \times p$ filters in order to eliminate some redundant degrees of freedom. To this end, we consider the decomposition of the space of all filters

$$\mathbb{R}^{p^2 \times n} = \mathcal{H}_0 \oplus \mathcal{H}_1 \oplus \mathcal{H}_2 \quad (3.6)$$

into subspaces given by³

$$\begin{aligned} \mathcal{H}_0 &= \{h \in \mathbb{R}^{p^2 \times n} : h^i = h^j, \text{ mean}(h^j) = 0, \forall i, j \in [n]\}, \\ \mathcal{H}_1 &= \{h \in \mathbb{R}^{p^2 \times n} : \langle h_i, \mathbb{1}_n \rangle = 0, \forall i \in [p^2], \text{ mean}(h^j) = 0, \forall j \in [n]\}, \\ \mathcal{H}_2 &= \{h \in \mathbb{R}^{p^2 \times n} : h_i = h_j, \forall i, j \in [p^2]\}. \end{aligned} \quad (3.7)$$

These spaces are orthogonal to each other with respect to the Euclidean inner product. The space \mathcal{H}_2 consists of all filters h that are constant for each label, i.e. $h^j = c_j \cdot \mathbb{1}_{p^2}$ with $c_j \in \mathbb{R}$ for each $j \in [n]$. The space \mathcal{H}_0 consists of all zero-mean filters, which do not distinguish between the labels. For any filter $h \in \mathcal{H}_0$, we have $h * W = h^1 * \sum_j W^j = h^1 * \mathbb{1}_m = 0$ for all $W \in \overline{\mathcal{W}}$, i.e. the subspace \mathcal{H}_0 does not represent any useful information for our purpose.

Thus, we can choose either $\mathcal{H} = \mathcal{H}_1$ or $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ as the actual space of filters. Our choice is

$$\mathcal{H} = \mathcal{H}_1, \quad \dim \mathcal{H} = \dim \mathcal{H}_1 = (p^2 - 1)(n - 1) \quad (3.8)$$

for two reasons. Firstly, we use this framework for segmentation, where larger homogenous regions occur (e.g., background). Filters which return a small ℓ^2 -norm for such labelings have (approximately) a zero-mean and therefore belong to \mathcal{H}_1 . Secondly, we will use an inner point method for optimization, which requires a feasible initialization. In case of zero-mean filters, we can simply use homogenous assignments as initial assignment. As a result, we do not need an additional initialization process on which the final result might depend. This conforms to the philosophy to start the assignment process without any bias at the barycenter $C \in \mathcal{W}$ – cf. (2.5).

For learning the filters, we assume that sets $\mathcal{I}^+, \mathcal{I}^- \subset \overline{\mathcal{W}}^*$ for favorable and unfavorable label assignments are given. We are looking for filters $h \in \mathcal{H}$ such that $\|h * W\|_{\ell^2}$ is smaller for $W \in \mathcal{I}^+$ than for $W \in \mathcal{I}^-$. For simplicity, we choose

$$\frac{\text{mean}_{W \in \mathcal{I}^+} \|h * W\|_{\ell^2}^2}{\text{mean}_{W \in \mathcal{I}^-} \|h * W\|_{\ell^2}^2} < 1 \quad (3.9)$$

as criterion for filters $h \in \mathcal{H}$, which leads to a generalized eigenvalue problem. Specifically, let $\{e_{\mathcal{H}}^{(i)} : i = 1, \dots, \dim \mathcal{H}\}$ be an orthonormal basis of \mathcal{H} and

³ Notation: Filters h are matrix-valued (image vectors \times labels) with rows h_i and columns h^j . Superscripts in brackets $h^{(k)}$ index members of a collection of filters.

consider the map

$$M : \overline{\mathcal{W}} \rightarrow \mathbb{R}^{\dim \mathcal{H} \times \dim \mathcal{H}}, \quad M(W)_{ij} = \langle e_{\mathcal{H}}^{(i)} * W, e_{\mathcal{H}}^{(j)} * W \rangle_{\ell^2}. \quad (3.10)$$

Then we have

$$\frac{\text{mean}_{W \in \mathcal{I}^+} \|h * W\|_{\ell^2}^2}{\text{mean}_{W \in \mathcal{I}^-} \|h * W\|_{\ell^2}^2} = \frac{x^\top A^+ x}{x^\top A^- x} \quad (3.11)$$

with $A^\pm = \text{mean}_{W \in \mathcal{I}^\pm} M(W)$ and $h = \sum_i x_i e_{\mathcal{H}}^{(i)}$. As a consequence, a set of linearly independent filters satisfying the criterion (3.9) is given by the generalized eigenvectors of the matrix pencil (A^+, A^-) corresponding to eigenvalues less than 1. The filters corresponding to eigenvalues greater than 1 might also focus on useful features as can be seen, for example, in Fig. 4.7 (d), where the last 16 filters correspond to eigenvalues greater than 1. These filters can be used additionally, since they further restrict the assignment and therefore may prevent some assignments, which were not taken into account by \mathcal{I}^- .

Having determined a set of filters as generalized eigenvectors, we normalize them in a post-processing step so as to meet the condition $h * W \in [-1, 1]^m$ for all $W \in \overline{\mathcal{W}}$, i.e.

$$\|h\|_{\mathcal{H}} = 1, \quad \|h\|_{\mathcal{H}} := \max \left\{ - \sum_{i \in [p^2]} \min_{j \in [n]} h_{ij}, \sum_{i \in [p^2]} \max_{j \in [n]} h_{ij} \right\}. \quad (3.12)$$

3.3 Optimization

In order to take into account the constraints (3.3), we use log-barrier functions [9, 4] that have been widely applied (e. g., in [1]). Given the parameters in (3.4) and filters $h^{(k)}$, $k \in [K]$, these functions read

$$\begin{aligned} B_{\text{low}}(W) &= - \langle \mathbb{1}_n, \log(\frac{1}{m} W^\top \mathbb{1}_m - c_{\text{low}}) \rangle, \\ B_{\text{up}}(W) &= - \langle \mathbb{1}_n, \log(c_{\text{up}} - \frac{1}{m} W^\top \mathbb{1}_m) \rangle, \\ B_{\text{filter}}(W) &= - \sum_{k=1}^K \log((d^{(k)})^2 - \|h^{(k)} * W\|_{\ell^2}^2). \end{aligned} \quad (3.13)$$

Summing up these functions yields the overall barrier function

$$B(W) = B_{\text{low}}(W) + B_{\text{up}}(W) + B_{\text{filter}}(W) \quad (3.14)$$

for the constraints (3.3). Complementing the objective function

$$J_\tau(W) = J(W) + \tau B(W) \quad (3.15)$$

with a barrier parameter $\tau > 0$, modifies the flow (2.9) and the sequence (2.5) accordingly. Within the flow, τ is handled as a monotonously decreasing function $\tau : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{> 0}$ with $\lim_{t \rightarrow \infty} \tau(t) = 0$. The flow is initialized at

$$W_i(0) = c_{\text{low}} + \frac{1 - \langle c_{\text{low}}, \mathbb{1}_n \rangle}{\langle c_{\text{up}}, \mathbb{1}_n \rangle - \langle c_{\text{low}}, \mathbb{1}_n \rangle} (c_{\text{up}} - c_{\text{low}}), \quad \forall i \in [m]. \quad (3.16)$$

Since we use zero-mean filters due to (3.8) and (3.7), the initialization (3.16) is strictly feasible for (3.3).

It remains to specify the gradients of the barrier functions that are required to evaluate the vector field, which defines the flow by the first equation of (2.9), with $J(W)$ replaced by $J_r(W)$ due to (3.15). The gradient $\nabla B(W) \in \mathbb{R}^{m \times n}$ of the barrier function (3.14) is given by

$$\nabla B_{\text{low}}(W)_i = -\frac{1}{m} \frac{1}{\frac{1}{m} W^\top \mathbf{1}_m - c_{\text{low}}}, \quad \nabla B_{\text{up}}(W)_i = \frac{1}{m} \frac{1}{c_{\text{up}} - \frac{1}{m} W^\top \mathbf{1}_m} \quad (3.17)$$

for each pixel $i \in [m]$, and by

$$\nabla B_{\text{filter}}(W)^j = 2 \sum_{k=1}^K \frac{h^{(k),j} \star (h^{(k)} * W)}{(d^{(k)})^2 - \|h^{(k)} * W\|_{\ell^2}^2} \quad (3.18)$$

for any label $j \in [n]$ with $h^{(k),j} \in \mathbb{R}^{p^2}$ being the j -th layer (column) of the k -th filter. Here, \star denotes the cross-correlation operation, i.e. convolution with the mirrored filter. This convolution is performed on the whole image with zero-padding.

If the approximated ℓ^1 -norm (3.5) is used instead of the ℓ^2 -norm, the barrier function takes the form

$$B_{\text{filter}}(W) = -\sum_{k=1}^K \log(d^{(k)} - \|h^{(k)} * W\|_{\ell_\varepsilon^1}) \quad (3.19)$$

with the Euclidean gradient given by

$$\nabla B_{\text{filter}}(W)^j = \sum_{k=1}^K \frac{h^{(k),j} \star \nabla \|h^{(k)} * W\|_{\ell_\varepsilon^1}}{d^{(k)} - \|h^{(k)} * W\|_{\ell_\varepsilon^1}}, \quad \nabla \|x\|_{\ell_\varepsilon^1} = \frac{x}{\sqrt{x \cdot x + \varepsilon^2 \mathbf{1}}}, \quad (3.20)$$

where the operations of the latter right-hand side apply componentwise.

4 Experiments

In this section, we investigate the influence of the filter constraints on the labeling result. We test the new approach on several academic labeling scenarios and compare the results with those obtained without using these constraints.

Setup. We represent assignments by choosing for each label $j \in [n]$ some color $c^{(j)} \in [0, 1]^3$ in the RGB color space. Then an assignment $W \in \mathcal{W} \subset \mathbb{R}^{m \times n}$ is represented by the color image $I \in \mathbb{R}^{m \times 3}$ given by $I_i = \sum_{j \in [n]} W_{ij} c^{(j)} \in [0, 1]^3$ for each pixel $i \in [m]$.

We consider *three different data sets* in order to check the effect of constraints on **(a)** primitive shape information, on **(b)** spatial relations (inclusion of regions), and **(c)** on the separation of fore- and background each defined by *several* labels.

The first data set **(a)** contains binary rectangles and ellipses. Filters of size 3×3 were trained for rectangles against ellipses, i. e. all assignments in a set \mathcal{I}^+ represent rectangles, while assignments in the complement set \mathcal{I}^- represent ellipses. The second data set **(b)** comprises three labels (white, orange, black) forming white ellipses overlapped by orange ellipses on black background (see Fig. 4.2 for illustration). All ellipses have varying radii, orientation and position. Filters of size 5×5 were used to separate the positive class \mathcal{I}^+ defined by inclusions of regions, whereas these topological relations are violated in the negative class \mathcal{I}^- . The third data set **(c)** consists of Voronoi diagrams, with each polygon labeled by either one of three foreground labels (red, green, blue) and likewise for the background (black, gray, white). Both foreground and background are connected and the foreground is located in the center of the image domain. The negative class \mathcal{I}^- is defined by randomly labelled Voronoi diagrams.

Implementation details. We solved the gradient flow by the explicit Euler method. A fixed step size Δt was used as long as $W(t_k)$ fulfilled the constraints. Otherwise the step size was reduced by backtracking line search. We did no analysis of the step size but rather used $\Delta t = 100$, which produced satisfying results. As usual for interior point methods, we used $\tau(t_k) = \alpha^{-k} \tau_0$ with $\alpha > 1$ for the barrier parameter. In our experiments, $\alpha = 1.03$ and $\tau_0 = 100$ turned out to be a reliable choice. We terminated the iteration either after 1500 steps or when both the average entropy $-\frac{1}{m} \sum_{i \in [m]} \sum_{j \in [n]} W_{ij}^{(k)} \log W_{ij}^{(k)}$ dropped below a threshold $\epsilon_{\text{entropy}} > 0$ and $\tau(t_k)$ dropped below $\tau_{\text{min}} > 0$. As in [2], we used $\epsilon_{\text{entropy}} = 10^{-3}$. For τ_{min} , we set 10^{-10} .

Results. (a) For the binary data set (rectangles/ellipses), the filter space \mathcal{H} has dimension 8. We used the four filters corresponding to the eigenvalues less than 1. Inspecting these filters reveals discrete versions of the partial derivatives ∂_{xy} , ∂_{xxy} , ∂_{xyy} and ∂_{xxyy} . The upper bounds for the filter constraints were set to $d = 2\|h * W\|$, where W is an assignment representing two rectangles. Fig. 4.1 illustrates that using the filter constraints enables to remove noise, to regularize the rectangle, and to rectify the ellipse by *imposing local shape constraints*.

(b) For the second data set, we used all 48 filters obtained by the generalized eigenvalue problem. The first 24 filters corresponding to eigenvalues less than 1 contribute to separating the orange region from the background. The remaining 24 filters regularize the boundary of the white region (see Fig. 4.3). The upper bounds for the filter constraints were set to $d = \max_{W \in \mathcal{I}^+} \|h * W\|$. The bounds c_{low} and c_{up} were set in a similar way. We used the distance matrix $D_{ij} = \frac{1}{np} \|\tilde{W}_i - e^{(j)}\|_2$, where the matrix $\tilde{W} \in \mathbb{R}^{m \times n}$ was obtained by adding white noise ($\sigma^2 = 4$) to the ground truth assignment. Fig. 4.4 demonstrates that the constraints improve notably the results, and that in addition to the filter constraints, cardinality constraints are essential to reinforce *topological structure*. In order to demonstrate the potential of the constraints for *spatial pattern formation*, we repeated the experiments with *pure noise* as input data. Fig. 4.5 demonstrated the strong regularizing effect of the constraints.

(c) For the Voronoi data set, the filters determined by the eigenvalue problem can be subdivided into three groups: The first 8 filters (eigenvalues ≈ 0.12)

contribute to separating the three foreground labels from three background labels. The next 16 filters (eigenvalues ≈ 0.52) regularize the foreground. The last 16 filters (eigenvalues $\approx 1.46 > 1$) regularize the background (Fig. 4.7). The distance matrix and the parameters for the constraints were set as described above for case (b). The results shown by Fig. 4.8 (b), (f), (g) and (h) demonstrate the effect of the three groups of filters. Repeating the experiments with pure noise as input data illustrates how *spatial patterns* are induced by the constraints (Fig. 4.9).

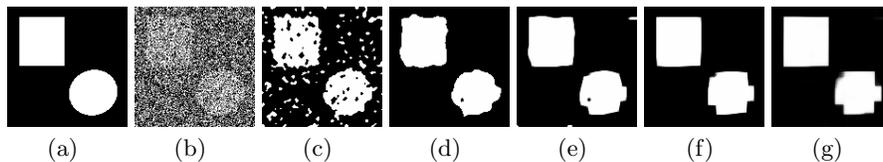


Fig. 4.1. Representing and enforcing rectangular structure. (a) shows the original gray-scale image and (b) shows a noisy version of it, which was used as input data. (c) and (d) show the labeling results *without imposing constraints* obtained through Riemannian averaging over neighborhoods of sizes 3×3 and 7×7 respectively. (e) and (f) show the results of the new approach (without cardinality constraints) with neighborhood size 3×3 and four filters of size 3×3 , which were trained for rectangles against ellipses. These 4 filters prefer horizontal and vertical edges. For (e), the ℓ^2 -norm was used for filter constraints. For (f) and (g), $\|\cdot\|_{\ell^1_\varepsilon}$ with $\varepsilon = 0.1$ and $\varepsilon = 0.01$ was used.

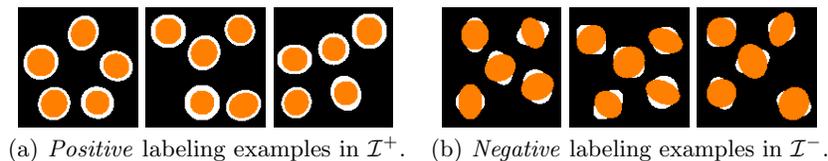


Fig. 4.2. Illustration of the training sets $\mathcal{I}^+, \mathcal{I}^- \subset \overline{\mathcal{W}}^*$. Positive examples (a) are defined by topological relations: orange ellipses are completely contained in the white ones. Negative examples (b) are labelings where this topological relation is violated.

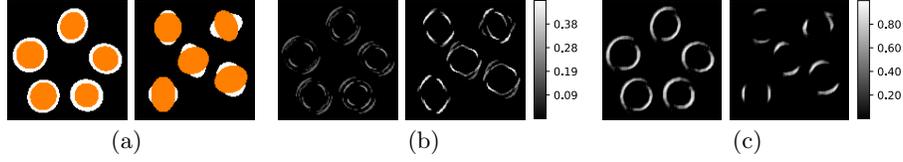


Fig. 4.3. Illustration of the generalized eigenvalue filters. (a) shows one label assignment in \mathcal{I}^+ and one assignment in \mathcal{I}^- , used to illustrate the filter outputs in (b), (c). The computed filters of size 5×5 can be subdivided into two groups. The first 24 filters $h^{(1)}, \dots, h^{(24)}$ respond to the boundary of the orange and black regions, and they have a large response at the border between the orange and black regions. This is illustrated by (b) which shows the absolute value of $h^{(1)} * W$. The last 24 filters $h^{(25)}, \dots, h^{(48)}$ mainly respond to the boundary of the white regions. (c) shows the absolute value of $h^{(48)} * W$.

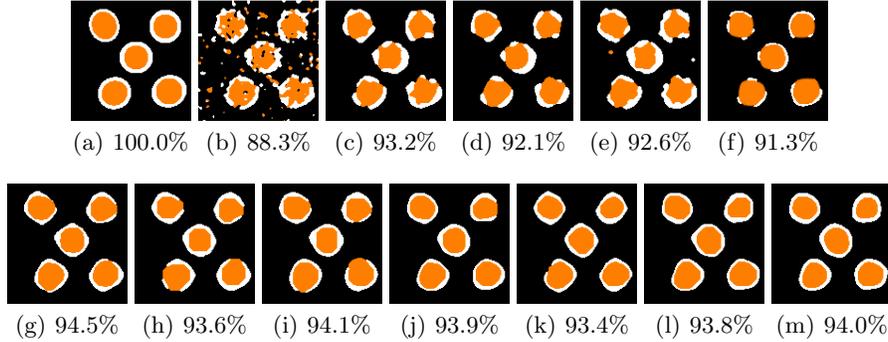


Fig. 4.4. Experimental results obtained with and without constraints. (a) shows the ground truth assignment. A noisy version of this assignment was used as input data. The percentages of correctly labeled pixels are shown below the images. The results obtained *without constraints* are shown in (b)-(d) for neighborhood sizes 3×3 , 5×5 and 7×7 respectively. For the results (e)-(m) of the new approach, 3×3 neighborhoods were used for spatial regularization. For (e), cardinality constraints were only used. For (f), filter constraints were only used (48 filters of size 5×5). For (g)-(m), *both cardinality constraints and filter constraints* were used. (g) and (h) were obtained with 24 filters using the ℓ^2 -norm and the approximated ℓ^1 -norm $\|\cdot\|_{\ell^1_\varepsilon}$ with $\varepsilon = 0.01$ respectively. For (i) and (j), 48 filters were used. For (k)-(m), the distance matrix was rescaled by a factor 0.01, and 48 filters were used with ℓ^2 -norm as well as $\|\cdot\|_{\ell^1_\varepsilon}$ with $\varepsilon = 0.1$ and $\varepsilon = 0.01$ respectively.

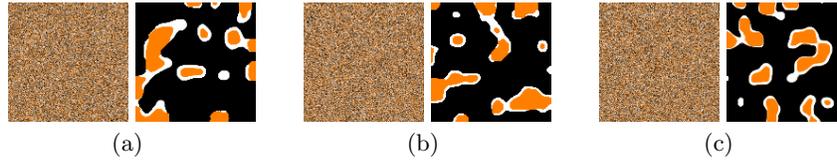


Fig. 4.5. Spatial pattern formation induced by pure noise and convex label constraints. All experiments were done using a 3×3 neighborhood for the spatial regularization, and using both cardinality constraints and filter constraints based on 48 filters and $\|\cdot\|_{\ell_\varepsilon^1}$ with $\varepsilon = 0.01$. Panels (a)-(c) show on the right random spatial labeling patterns induced by the random noise images on the left. These results demonstrate how filter constraints favor local shape and topological spatial structure on image labelings within our geometric approach to label assignments.

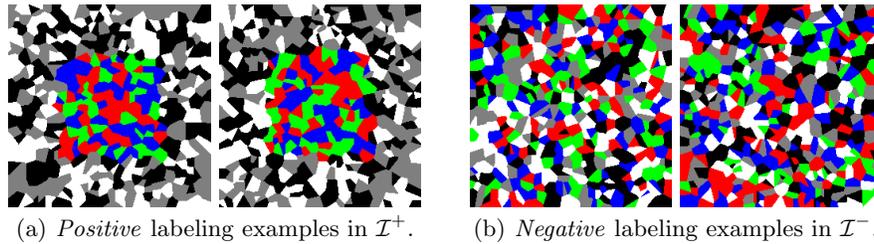
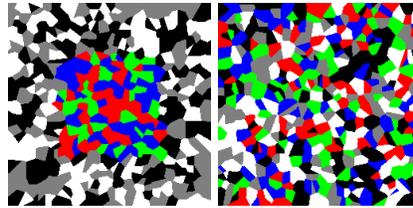
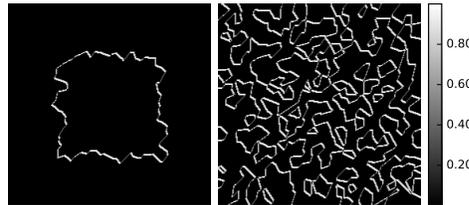


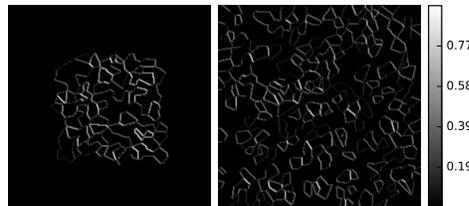
Fig. 4.6. Best viewed in color. Illustration of the training sets $\mathcal{I}^+, \mathcal{I}^- \subset \overline{\mathcal{W}}^*$. Both the foreground region and the background region of these Voronoi tilings are defined by three labels: red, green, blue and black, gray, white, respectively. Positive examples in \mathcal{I}^+ are defined by approximately square-shaped foreground regions that are simply connected and centered in the middle of the image domain. Negative examples \mathcal{I}^- contain polygons that are randomly labeled and distributed over the image domain.



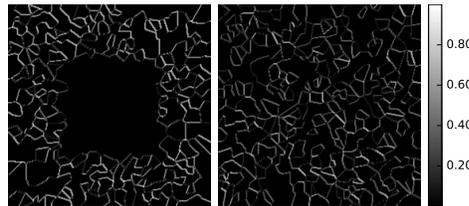
(a)



(b)



(c)



(d)

Fig. 4.7. *Best viewed in color. Illustration of the generalized eigenvalue filters.* (a) shows one assignment in \mathcal{I}^+ and \mathcal{I}^- , respectively, used to illustrate the filter outputs. The computed filters of size 3×3 can be subdivided into three groups. The first eight filters $h^{(1)}, \dots, h^{(8)}$ regularize the boundary *between* foreground (red, green, blue) and background (black, gray, white). (b) shows the absolute value of the filter result $h^{(4)} * W$ as example. Filters $h^{(9)}, \dots, h^{(24)}$ regularize the boundaries *within the foreground* as illustrated by (c), which shows the absolute value of $h^{(10)} * W$. Eventually, filters $h^{(25)}, \dots, h^{(40)}$ regularize the boundaries *within the background* as illustrated by (d).

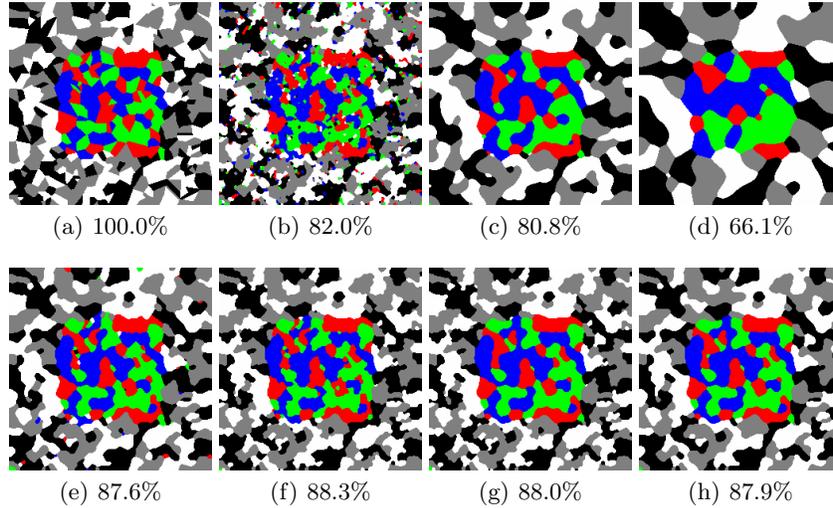


Fig. 4.8. Best viewed in color. Experimental results obtained with and without constraints. (a) shows the ground truth assignment. A noisy version of this assignment was used as input data. The percentages of correctly labeled pixels compared to the ground truth (a) are shown below the images. Panels (b)-(d) show the results obtained *without* constraints using neighborhood sizes 3×3 , 5×5 and 7×7 respectively. (e) is the result for neighborhood size 5×5 , but with a rescaled (factor 100) distance matrix. (f)-(h) show the results *with* constraints using first 8 filters, 24 filters and 40 filters, respectively. These results were computed using 3×3 neighborhoods for spatial regularization and the ℓ^2 -norm for the filter constraints.

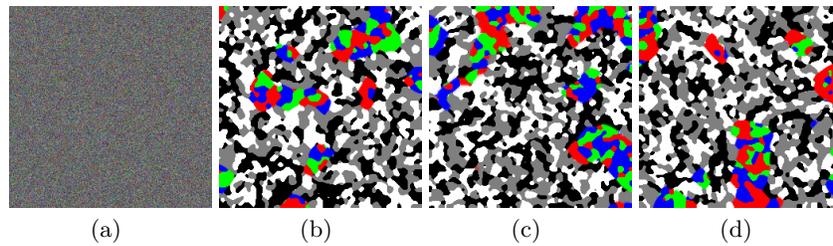


Fig. 4.9. Best viewed in color. Spatial pattern formation induced by pure noise and convex label constraints. (a) shows a random assignment W using the same color coding as for the Voronoi polygons. Each panel (b)-(d) shows the result of labeling a different random input image of type (a). All experiments were done using 3×3 neighborhoods, without cardinality constraints, and with filter constraints based on 40 filters and $\|\cdot\|_{\ell^1_\varepsilon}$ with $\varepsilon = 0.1$. The results demonstrate how the filter constraints enforce both the scale and the spatial structure of fore- and background regions that are randomly located due to the pure noise data.

5 Conclusion

We extended the smooth geometric image labeling approach of [2, 11] in order to incorporate global convex constraints on the labeling result using linear filters in the label space. This extension was mathematically formulated so as to preserve smoothness of the overall approach. We showed how filters can be determined by a generalized eigenvalue problem in order to represent statistical prior knowledge about local shape and spatial relation. Experimental results demonstrate the potential of the approach for imposing these constraints onto labelings of noisy image data.

Our future work will focus on numerical aspects in order to make the approach more efficient for various applications. This includes, in particular, the investigation of how filters of small support can be used to represent and enforce the structure of labelings at multiple spatial scales.

References

1. Alvarez, F., López, J.: Convergence to the optimal value for barrier methods combined with Hessian Riemannian gradient flows and generalized proximal algorithms. *J. Convex Analysis* 17(3&4), 701–720 (2010)
2. Åström, F., Petra, S., Schmitzer, B., Schnörr, C.: Image Labeling by Assignment. *J. Math. Imag. Vision* 58(2), 211–238 (2017)
3. Benning, M., Gilboa, G., Grah, J.S., Schönlieb, C.B.: Learning Filter Functions in Regularisers by Minimising Quotients. In: *Proc. SSVM* (2017)
4. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2004)
5. Cover, T., Thomas, J.: *Elements of Information Theory*. John Wiley & Sons, 2nd edn. (2006)
6. Kappes, J., Andres, B., Hamprecht, F., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., Rother, C.: A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems. *Int. J. Comp. Vision* 115(2), 155–184 (2015)
7. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comp. Vision* 60(2), 91–110 (2004)
8. Morel, J.M., Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM J. Imag. Sci.* 2(2), 438–469 (2009)
9. Nesterov, Y., Nemirovskii, A.: *Interior Point Polynomial Algorithms in Convex Programming*. SIAM (1994)
10. Portilla, J., Simoncelli, E.: A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients. *Int. J. Comp. Vision* 40(1), 49–70 (2000)
11. Savarino, F., Hühnerbein, R., Åström, F., Recknagel, J., Schnörr, C.: Numerical Integration of Riemannian Gradient Flows for Image Labeling. In: *Proc. SSVM. LNCS 10302*, Springer (2017)
12. Werner, T.: A Linear Programming Approach to Max-sum Problem: A Review. *IEEE Trans. Patt. Anal. Mach. Intell.* 29(7), 1165–1179 (2007)
13. Xie, J., Lu, Y., Zhu, S.C., Wu, Y.: A Theory of Generative Convnet. In: *ICML* (2016)
14. Zhu, S., Mumford, D.: Prior Learning and Gibbs Reaction–Diffusion. *IEEE Trans. Patt. Anal. Mach. Intell.* 19(11), 1236–1250 (1997)