SUPERVISED AND TRANSDUCTIVE MULTI-CLASS SEGMENTATION USING p-LAPLACIANS AND RKHS METHODS

S. H. KANG, B. SHAFEI, AND G. STEIDL

Abstract. This paper considers supervised multi-class image segmentation: from a labeled set of pixels in one image, we learn the segmentation and apply it to the rest of the image or to other similar images. We study approaches with p-Laplacians, (vector-valued) Reproducing Kernel Hilbert Spaces (RKHSs) and combinations of both. In all approaches we construct segment membership vectors. In the p-Laplacian model the segment membership vectors have to fulfill a certain probability simplex constraint. Interestingly, we could prove that this is not really a constraint in the case $p = 2$ but is automatically fulfilled. While the 2-Laplacian model gives a good general segmentation, the case of the 1-Laplacian tends to neglect smaller segments. The RKHS approach has the benefit of fast computation. This direction is motivated by image colorization, where a given dab of color is extended to a nearby region of similar features or to another image. The connection between colorization and multi-class segmentation is explored in this paper with an application to medical image segmentation. We further consider an improvement using a combined method. Each model is carefully considered with numerical experiments for validation, followed by medical image segmentation at the end.

1. Introduction

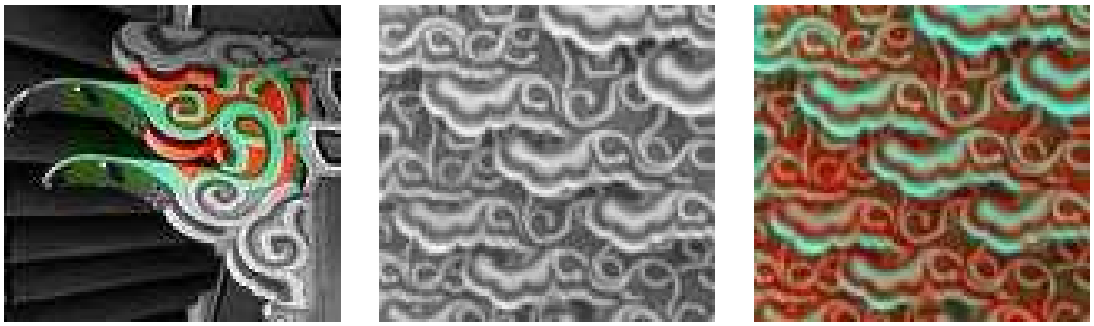
Image segmentation has been extensively studied in recent years. Some of the well-known models include mixture random-field models [23], the Mumford-Shah variational model [37], the Monte-Carlo Markov chain model [52] and the graph-cutting and spectral method [45]. Many of the initial segmentation models focused on two-class segmentation such as the active contour approach [12, 31] and the Chan-Vese model [15]. Later multi-class segmentation models followed such as [4, 8, 14, 17, 30, 41, 51, 33, 44, 53, 58, 61].

In this paper, we focus on supervised multi-class segmentation considering single images as well as collections of images. Given some labeled pixels in one image, we study different models which can find the relevant regions in the remaining parts. Recently, the authors in [32] extended the multi-class segmentation with a fuzzy membership function [35] to the segmentation of a collection of images. We compare our model to this setting.

Our work is motivated by image colorization where the objective is to reconstruct the color of an image from the few color strokes in gray scale images. Starting from an Apollo mission [10], colorization has a wide application to heritage restoration [21] and computer graphics [34, 56, 57]. In particular, in [26] the authors proposed to use a Reproducing Kernel Hilbert Space (RKHS) approach for image colorization and showed successful results, such as Figure 1. This example illustrates the connection between the image colorization (color transfer) and multi-class image segmentation. In image colorization, different colors represent different regions, and in multi-class segmentation, different labels represent different classes. The main idea is to learn the labels from the given information in order to find similar regions for segmentation.

There is a clear connection with the machine learning literature where one of the objectives is to generalize the information based on limited sampled data. RKHS-methods have also made an impact on the machine learning literature both from algorithmic and theoretical perspectives.

Date: June 6, 2012.



(A) (B) (C)

FIGURE 1. Result from [26]. Image (A) has a region with a given color. This color is transferred to the black and white image (B) which gives the colorized image (C).

In learning, there are numerous works using RKHS methods, e.g., techniques to train the labels, classification using Support Vector Machines, probabilistic approaches or hierarchical methods. The objective of this paper is to study the effect of the p-Laplacian model and the RKHS method for multi-class segmentation given a set of labels.

First we explore the p-Laplacian model for multi-class segmentation in Section 3. In particular, the graph 2-Laplacian was applied in image processing and learning for a long time. We refer to [54] and the references therein for an overview of various aspects of spectral clustering up to 2007 and to [45] for the application of 2-Laplacians in image segmentation. Recently, more general p-Laplacians were studied in the context of machine learning, see [9, 28, 50]. The recent work [32] on multi-class segmentation applies also a 2-Laplacian model although Laplacians are not mentioned there. We contribute to the field in Section 3 by (i) showing that the segment membership vector obtained by this approach automatically lies in a probability simplex, and (ii) we extend the approach to p-Laplacians with $p \geq 1$. We demonstrate the properties of the model for $p = 1, \frac{3}{2}, 2$ by a numerical example. Then in Section 4, we apply a (vector-valued) RKHS approach for multi-class segmentation. This approach is based on [26] and we study the similarities and differences to the p-Laplacian model. A combined projection model is proposed in Section 5. In Section 6, we show the performance of our methods applied to collections of medical images. Since medical images typically have a low contrast with a high level of noise, it is challenging to find a method which works well in practice. For example, a series of papers in MICCAI'09 on the left-ventricle segmentation challenge [1] discusses the difficulties of such images. See also [39] for a detailed review of segmentation methods for short axis cardiac MR images.

2. NOTATION

Let Ω be the discrete $n_1 \times n_2$ image domain, where the image $B : \Omega \to \mathbb{R}^d$ is defined ($d = 1$ for gray scale image, and $d = 3$ for color image). For a simple matrix-vector notation, we assume the image to be column-wise reshaped as a vector such that $B \in \mathbb{R}^{N \times d}$, where $N := n_1 n_2$. We retain the notation B for both the original and the reshaped image since its meaning becomes clear from the context.

Let $\mathbb{I}_N := \{1, \ldots, N\}$ be the set of integers from 1 to N (the total number of the pixel), $L \subset \mathbb{I}_N$ be the set of labeled points, and $U \subset \mathbb{I}_N$ be the set of unlabeled points. Of course, the label determines which class the point j belongs to: let $L_k \subset \mathbb{I}_N$ be the set of points belonging to class k so that $L := \bigcup_{k=1}^{c} L_k$. Here, the number of different classes (segments, phases) is denoted by c. In l_L we

collect the characteristic functions of each L_k at every labeled point $j \in L$ given by

$$l_L := \begin{pmatrix} l_L^1 \\ \vdots \\ l_L^c \end{pmatrix}, \quad l_L^k = (l^k(j))_{j\in L}, \quad \text{where} \quad l^k(j) \begin{cases} 1 & \text{if } j \in L_k, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

To indicate which class an arbitrary point j belongs to, we denote the *segment membership vector* by

$$u := \begin{pmatrix} u^1 \\ \vdots \\ u^c \end{pmatrix}, \quad u^k := \left(u^k(j)\right)_{j=1}^N, \quad k = 1, \ldots, c,$$

where each $u^k(j)$ indicates the degree of membership of pixel j to class k. We set $u(j) := (u^k(j))_{k=1}^c$, $j \in \mathbb{I}_N$. When needed, we sort the components of u according to the labeled and unlabeled points such that

$$u = \begin{pmatrix} u_L \\ u_U \end{pmatrix}.$$

Using the notation u_L, the given labels arise as constraint $u_L = l_L$. Further we define the *probability simplex* S_c by

$$S_c := \{x \in \mathbb{R}^c : \sum_{k=1}^c x_k = 1,\ x_k \geq 0\}$$

and its n-fold version by $S_c^n := \underbrace{S_c \times \ldots \times S_c}_{n}$. Finally, the *indicator function* ι_C of a nonempty set C is given by

$$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

For a matrix A with rows and columns indexed by some sets I and J, resp., we use the shorthand notation $A_{IJ} := (a_{i,j})_{i\in I, j\in J}$. Further, I_n is the $n \times n$ identity matrix, $\mathbf{0}_{n,m}$ the $n \times m$ matrix with entries 0, and $\mathbf{1}_n$ the vector with n entries 1. By $A \otimes B$, we denote the *Kronecker product* of the matrices A and B.

3. Supervised Multi-Class Segmentation with p-Laplacians

For the transductive multi-class segmentation, we assume that some labeled points are given ($L \neq \emptyset$). This information needs to be extended to the unlabeled points U for the segmentation. This extension/diffusion process is governed by some similarities and differences among the points, which can be measured in the p-Laplacian model.

First we present the general set-up in Subsection 3.1. In Subsection 3.2, we handle the 2-Laplacian case. Using properties of M-matrices we show in Theorem 3.2 that the resulting segment membership vectors satisfy a probability simplex condition. The p-Laplacian model for $p \geq 1$ with probability simplex constraints is considered in Subsection 3.3. In particular, we focus on $p = 1$. In this case alternating direction algorithms can be used to find a minimizer of the corresponding functional. We present the numerical setting for each case, and compare the effects of p-Laplacian models for $p = 1, \frac{3}{2}, 2$ in Subsection 3.4.

3.1. **General Model.** To define the p-Laplacian we measure the similarity of the features of two pixels $i, j \in \mathbb{I}_N$ by appropriately chosen weights $w_{i,j}$ fulfilling

$$w_{i,j} \geq 0, \quad w_{i,j} = w_{j,i}, \qquad i, j \in \mathbb{I}_N. \tag{2}$$

We restrict our attention to symmetric weights but the approach can be easily generalized to the non-symmetric setting. Since N is large in our applications, only the weights in a 'neighborhood' $\mathcal{N}_i$ of every pixel $f(i)$ will be nonzero, i.e., $\mathcal{N}_i := \{j \in \mathbb{I}_N : w_{i,j} > 0\}$ and $|\mathcal{N}_i| \ll N$. For $p \geq 1$ we consider the objective function

$$Q_p(u) := \frac{2}{p} \sum_{k=1}^{c} \langle u^k, \triangle_p u^k \rangle = \frac{1}{p} \sum_{k=1}^{c} \sum_{i,j=1}^{N} w_{i,j} |u^k(i) - u^k(j)|^p, \tag{3}$$

where $\triangle_p$ denotes the (graph) *p-Laplacian* $\triangle_p : \mathbb{R}^N \to \mathbb{R}^N$ defined by

$$(\triangle_p x)_i := \sum_{j=1}^{N} w_{i,j} \phi(x(i) - x(j)), \quad \phi(t) := |t|^{p-1} \mathrm{sign}(t),$$

see [2, 9]. Then, the segmentation model becomes

$$\operatorname*{argmin}_{u_U} Q_p(l_L, u_U) \quad \text{subject to} \quad u^k(j) \in \{0, 1\}, \ \sum_{k=1}^{c} u^k(j) = 1 \ \forall j \in U.$$

Since the solution of this discrete problem appears to be NP-hard, the binary constraint $u^k(j) \in \{0, 1\}$ is relaxed to $u^k(j) \in [0, 1]$, and we obtain our *general convex model for multi-class segmentation using p-Laplacians*

$$\operatorname*{argmin}_{u_U} Q_p(l_L, u_U) \quad \text{subject to} \quad u_U \in S_c^{|U|}. \tag{4}$$

In the next subsections, we consider the model in more detail.

Remark 3.1. *Graph Laplacians were considered in learning theory and image processing for a long time. Due to* (5)*, the matrix $D - W$ is symmetric, positive semi-definite with smallest eigenvalue 0 corresponding to the eigenspace spanned by its eigenvector 1_N (if we suppose that the matrix is irreducible). It is well-known that the eigenvector corresponding to the second largest eigenvalue of the 2-Laplacian (normalized 2-Laplacian) corresponds to relaxations of the ratio cut (normalized cut) [27, 45] of the corresponding graph. This second largest eigenvector as well as the other ones have found applications in machine learning and image processing [16, 24, 40, 45]. A reformulation of normalized cut segmentation that in a unified way can handle linear equality constraints for an arbitrary number of classes was given in [19]. Motivated by the generalized isoperimetric inequality of Amghibech [2] which relates the second eigenvalues of the graph p-Laplacian to the optimal Cheeger cut, further connections between the Cheeger cut and the second eigenvectors of the graph p-Laplacian were established and applied in machine learning in a couple of recent papers [9, 28].*

3.2. **The 2-Laplacian model.** For $p = 2$, we use the matrix notation (sorted according to the labeled and unlabeled components),

$$W := (w_{i,j})_{i,j=1}^{N} := \begin{pmatrix} W_{LL} & W_{LU} \\ W_{UL} & W_{UU} \end{pmatrix} \quad \text{and} \quad D := \mathrm{diag}(d_i)_{i=1}^{N} \quad \textit{with} \quad d_i := \sum_{j=1}^{N} w_{i,j}. \tag{5}$$

Then the model can be reformulated as

$$Q_2(l_L, u_U) = \frac{1}{2} \sum_{k=1}^{c} \begin{pmatrix} l_L^k \\ u_U^k \end{pmatrix}^{\mathrm{T}} \underbrace{\begin{pmatrix} D_{LL} - W_{LL} & -W_{LU} \\ -W_{UL} & D_{UU} - W_{UU} \end{pmatrix}}_{=:\Delta_2} \begin{pmatrix} l_L^k \\ u_U^k \end{pmatrix}. \tag{6}$$

Note that $W_{LU} = W_{UL}^{\mathrm{T}}$ by the symmetry of our weights. For $c = 1$, the function (6) can be simply written as $Q_2(u) = \frac{1}{2}u^{\mathrm{T}}(D - W)u$.
The following theorem states that under mild conditions on the weights, problem (4) with $p = 2$ has a unique solution which can be obtained by just minimizing the summands in (6) separately. In particular, these minimizers will automatically meet the simplex constraint.
To prove the theorem we need the notation of an M-matrix. A matrix $A \in \mathbb{R}^{n\times n}$ is called an *M-matrix* if $a_{i,j} \leq 0$ for all $i \neq j$ and if it is *inverse isotonic*, i.e., A^{-1} exists and $A^{-1} \geq 0$ componentwise. In general it is hard to see if a matrix is inverse isotonic. However it is for example well-known that any strictly diagonally dominant or irreducible diagonally dominant matrix A which fulfills $a_{i,i} > 0$ for all $i = 1, \ldots, n$ and $a_{i,j} \leq 0$ for all $i \neq j$ is an *M-matrix*, see [29, p.113ff] and [48, p. 303].

Theorem 3.2. *Assume that the weights in Q_2 are chosen such that $\mathcal{N}_i \cap L \neq \emptyset$ for all $i \in U$, i.e., for every $i \in U$ there exists at least one $j \in L$ such that $w_{i,j} > 0$. Then problem (4) with $p = 2$ has a unique solution $\hat{u}_U$ given by the solutions $\hat{u}_U^k$ of the linear systems of equations*

$$(D_{UU} - W_{UU})\hat{u}_U^k = W_{UL}l_L^k, \quad k = 1, \ldots, c.$$

Proof. The function Q_2 can be rewritten as

$$Q_2(l_L, u_U) = \sum_{k=1}^{c}\left(\frac{1}{2}(u_U^k)^{\mathrm{T}}(D_{UU} - W_{UU})u_U^k - (u_U^k)^{\mathrm{T}}W_{UL}l_L^k + \frac{1}{2}(l_L^k)^{\mathrm{T}}(D_{LL} - W_{LL})l_L^k\right).$$

Now $\hat{u}_U$ is a minimizer of this convex function if and only if $\nabla_{u_U}Q_2 = 0$. This condition is fulfilled if and only if

$$(D_{UU} - W_{UU})\hat{u}_U^k = W_{UL}l_L^k \tag{7}$$

for all $k = 1, \ldots, c$. By our assumption on the weights (2) and (5), the matrix $D_{UU} - W_{UU}$ has positive diagonal entries and non-diagonal entries smaller or equal than zero. Moreover, it is strictly diagonally dominant such that it is an M-matrix. Thus $(D_{UU} - W_{UU})^{-1}$ exists and the linear systems in (7) have unique solutions $\hat{u}_U^k$, $k = 1, \ldots, c$. Moreover, since $l_L \geq 0$, $W_{UL} \geq 0$ and $(D_{UU} - W_{UU})^{-1} \geq 0$ we conclude that $\hat{u}_U \geq 0$. To show that $\hat{u}_U \in S_c^{|U|}$ it remains to prove that

$$\sum_{k=1}^{c}\hat{u}_U^k = \mathbf{1}_{|U|}. \tag{8}$$

Summing up the equations in (7) we obtain

$$(D_{UU} - W_{UU})\sum_{k=1}^{c}\hat{u}_U^k - W_{UL}\sum_{k=1}^{c}l_L^k = 0$$

and by the choice of the labeled components in (1) further

$$(D_{UU} - W_{UU})\sum_{k=1}^{c}\hat{u}_U^k - W_{UL}\mathbf{1}_{|L|} = 0.$$

By (5) we know that $(-W_{UL}|D_{UU} - W_{UU})\begin{pmatrix}\mathbf{1}_{|L|}\\ \mathbf{1}_{|U|}\end{pmatrix} = 0$ so that

$$(D_{UU} - W_{UU})\sum_{k=1}^{c}\hat{u}_U^k - (D_{UU} - W_{UU})\mathbf{1}_{|U|} + W_{UL}\mathbf{1}_{|L|} - W_{UL}\mathbf{1}_{|L|} = 0,$$

$$(D_{UU} - W_{UU})\left(\sum_{k=1}^{c}\hat{u}_U^k - \mathbf{1}_{|U|}\right) = 0.$$

Since $D_{UU} - W_{UU}$ is invertible, this implies (8). □

We note that this model is considered in [32]. The authors have chosen sophisticated weights which in particular meet the assumption of this theorem. Their analysis uses duality considerations with Karush-Kuhn-Tucker conditions to prove that the solution fulfills the simplex constraints.

3.3. **The p-Laplacian model, $p \geq 1$.** We consider the model involving general p-Laplacians with $p \geq 1$. We note that a recent preprint [7] considers the 1-Laplacian method, called Mumford-Shah-Potts model for multi-class learning, and that a p-Laplacian which differs from those used in this paper was applied for semi-supervised learning in [59]. Here, we discuss an application to multi-class image segmentation.
For a simple formation, we describe problem (4) in matrix-vector form. For simplicity, we assume that all neighborhoods $\mathcal{N}_i$ have the same cardinality $\nu = |\mathcal{N}_i|$ for all $i \in \mathbb{I}_N$. Let $A \in \mathbb{R}^{\nu N \times N}$ denote the matrix corresponding to the linear mapping

$$x \mapsto \left(\left(w_{i,j}^{1/p} (x(i) - x(j)) \right)_{j \in \mathcal{N}_i} \right)_{i=1}^{N}, \tag{9}$$

and let $A_U \in \mathbb{R}^{\nu N, |U|}$ and $A_L \in \mathbb{R}^{\nu N, |L|}$ denote the matrices containing the columns of A corresponding to the indices in U and L, respectively. Then our minimization problem (4) becomes

$$\operatorname*{argmin}_{u_U} \frac{1}{p} \sum_{k=1}^{c} \left\| A_U u_U^k + A_L l_L^k \right\|_p^p \quad \text{subject to} \quad u_U \in S_c^{|U|}. \tag{10}$$

Using the notation $M_U := I_c \otimes A_U$ and $y := (A_L l_L^k)_{k=1}^c$ and the indicator function $\iota_{S_c^{|U|}}$ of $S_c^{|U|}$ this problem can be further rewritten as

$$\operatorname*{argmin}_{u_U} \frac{1}{p} \left\| M_U u_U + y \right\|_p^p + \iota_{S_c^{|U|}}(u_U)$$

and equivalently as

$$\operatorname*{argmin}_{u_U, v} \frac{1}{p} \left\| v + y \right\|_p^p + \iota_{S_c^{|U|}}(u_U) \quad \text{s.t.} \quad M_U u_U = v. \tag{11}$$

To solve this minimization problem, we apply the primal dual hybrid gradient algorithm with modified (extrapolated) primal variable (PDHGMp) proposed in [13, 41], see also [60, 20]. PDHGMp was proved to converge for our setting if the parameters γ and τ are chosen such that $\gamma\tau \leq 1/\left\|M_U\right\|^2$.

Input: $M_U = I_c \otimes A_U$, $y = (A_L l_L^k)_{k=1}^c$, and two parameters γ and τ.
Output: Segment membership vector u.
Initialization of $u_U^{(1)}$, $v^{(1)}$, $b^{(1)}$ and $b^{(0)}$;
for $r = 1, 2, \ldots$ *until a stopping criterion is reached* **do**

$$u_U^{(r+1)} = \operatorname*{argmin}_{u_U} \iota_{S_c^{|U|}} + \frac{1}{2\tau} \left\| u_U - (u_U^{(r)} - \gamma\tau M_U^{\mathrm{T}} (2b^{(r)} - b^{(r-1)})) \right\|_2^2$$

$$v^{(r+1)} = \operatorname*{argmin}_{v} \frac{1}{p} \left\| v + y \right\|_p^p + \frac{1}{2\gamma} \left\| v - (b^{(r)} + M_U u_U^{(r+1)}) \right\|_2^2$$

$$b^{(r+1)} = b^{(r)} + M_U u_U^{(r+1)} - v^{(r+1)}$$

end

Algorithm 1: PDHGMp for solving (11)

The first minimization step requires a projection of $(u_U^{(r)} - \gamma\tau M_U^{\mathrm{T}}(2b^{(r)} - b^{(r-1)}))$ onto $S_c^{|U|}$ which can be done separately for all $j \in U$. The second minimization step can be rewritten by setting $z = v + y$ as

$$\hat{z} = \operatorname*{argmin}_z \frac{1}{p}\|z\|_p^p + \frac{1}{2\gamma}\left\|z - (y + b^{(r)} + M_U u_U^{(r+1)})\right\|_2^2.$$

In the case $p = 1$ the minimizer $\hat{z}$ can be computed by the component-wise soft shrinkage of $a := y + b^{(r)} + M_U u_U^{(r+1)}$ with threshold γ. For $p > 1$, the minimizer can be computed for every component separately by

$$\hat{z}_i = \operatorname*{argmin}_{x\in\mathbb{R}} \frac{1}{p}|x|^p + \frac{1}{2\gamma}(x - a_i)^2.$$

Setting the derivative to zero, we get

$$\operatorname{sign}(x)|x|^{p-1} + \frac{1}{\gamma}(x - a_i) = 0$$

which can be solved for example by a semi-implicit (Weiszfeld-like) method. For $p = \frac{3}{2}$, the solution is given analytically by

$$\hat{z}_i = \begin{cases} a_i + \frac{\gamma}{2} - \sqrt{(a_i + \frac{\gamma}{2}) - a_i^2} & \text{if } a_i > 0, \\ 0 & \text{if } a_i = 0, \\ a_i - \frac{\gamma}{2} + \sqrt{(a_i - \frac{\gamma}{2}) - a_i^2} & \text{if } a_i < 0. \end{cases}$$

Remark 3.3. *Model* (4) *can be rewritten as*

$$\operatorname*{argmin}_{u\in\mathbb{R}^{cN}} Q_p(u) \quad \textit{subject to} \quad u_L(j) = l_L(j), \ \ \forall j \in L \text{ and } u \in S_c^N.$$

We can approximate this model by

$$\operatorname*{argmin}_{u\in\mathbb{R}^{cN}} \frac{1}{2}\sum_{j\in L}\|l_L(j) - I_c \otimes Ju\|_2^2 + \lambda Q_p(u) \quad \textit{subject to} \quad u \in S_c^N,$$

where $J := \left(I_{|L|}|\mathbf{0}_{|L|,|U|}\right)$. *The minimizer of this convex functional can be found similarly as above. For* $p = 2$ *the minimizer is for example given by the solutions of the linear systems of equations*

$$(J^T J + \lambda L)\hat{u}^k = J^T l_L^k, \quad k = 1, \ldots, c.$$

Assuming that L *is irreducible, we see that* $J^T J + \lambda L$ *is again an* M*-matrix and following the lines of Theorem 3.2 we can conclude that the solution* $\hat{u}$ *automatically fulfills the simplex constraints.*

3.4. **Effects of the different p-Laplacian models via numerical examples.** We are interested in the influence of different values p in multi-class segmentation. Typically, we observe that when a smaller value of p is used, the results are more regular and have smoother boundaries. Figure 2 illustrates this effect, when the same parameters within the weights are used. Image (A) is the given image B, image (E) illustrates the true segmentation superposed with the 9 labeled points. The first row is the segment membership matrix u which is a $n_1 \times n_2 \times 3$-dimensional matrix in this case ($c = 3$). The color red (vector $(1, 0, 0)^{\mathrm{T}}$) represents class 1, the color green class 2, and the color blue class 3. The second row shows the discretization, i.e., the final segmentation which is achieved by taking $\operatorname{argmax}_k u^k$ pixel-wise. Comparing different p-Laplacian models, we see that the 1-Laplacian approach gives smoother boundaries, appears to generate more regular results, and the segment-membership-matrix u is closest to hard clustering.
However, there are many parameters involved in the weights $w_{i,j}$. In the following, we carefully describe the details of how the weights are chosen. We consider geometric weights based on pixel

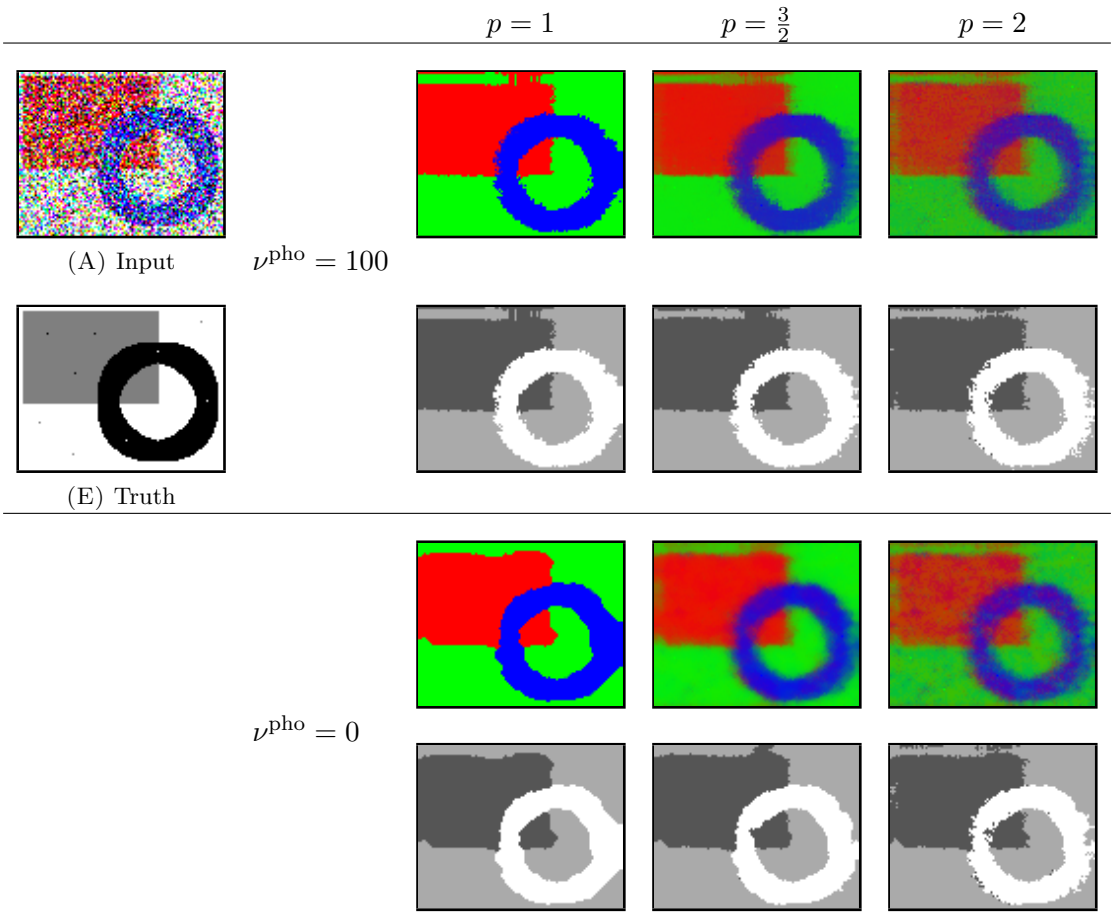


FIGURE 2. Effect of different p and ν^{pho} : (A) Input image. (E) The ground truth superposed with the 9 labeled points. For every ν^{pho}, the first row (color) shows the segment membership matrix u and the second row shows the resulting segmentation which is achieved by taking argmax u^k pixel-wise. The other parameters are $\nu^{\text{lab}} = \frac{1}{2}$, $r = 5$, $\rho^2 = s = 363$. Notice that the 1-Laplacian approach provides smoother boundaries, appears to generate more regular results, and the segment membership matrix u is closest to hard clustering. Decreasing ν^{pho} decreases the influence of the photometric neighborhood and increases the influence of the geometric neighborhood in (13).

locations and photometric weights based on color features similar to [32]. The geometric similarity between two pixels $i, j \in \mathbb{I}_N$ is defined by

$$w_{i,j}^{\text{geo}} := \begin{cases} \dfrac{e^{-\|i-j\|_2^2}}{\sum\limits_{j \in \mathcal{N}_i^{\text{geo}}} e^{-\|i-j\|_2^2}} & \text{if } j \in \mathcal{N}_i^{\text{geo}}, \\ 0 & \text{otherwise,} \end{cases}$$

where the $\mathcal{N}_i^{\text{geo}} := \{j \in \mathbb{I}_N : \|i - j\|_\infty \leq 1\}$ denotes the geometric neighborhood. The photometric neighborhood $\mathcal{N}_i^{\text{pho}}$ of the pixel i is defined to be the 4 most similar pixels j in a 17×17 window around pixel i with respect to the Euclidean norm of the feature vectors $F(i)$ and $F(j)$. The feature vector $F(i) \in \mathbb{R}^s$ is given by a $(2r+1) \times (2r+1)$ window around pixel i where $s = d(2r+1)^2$. The

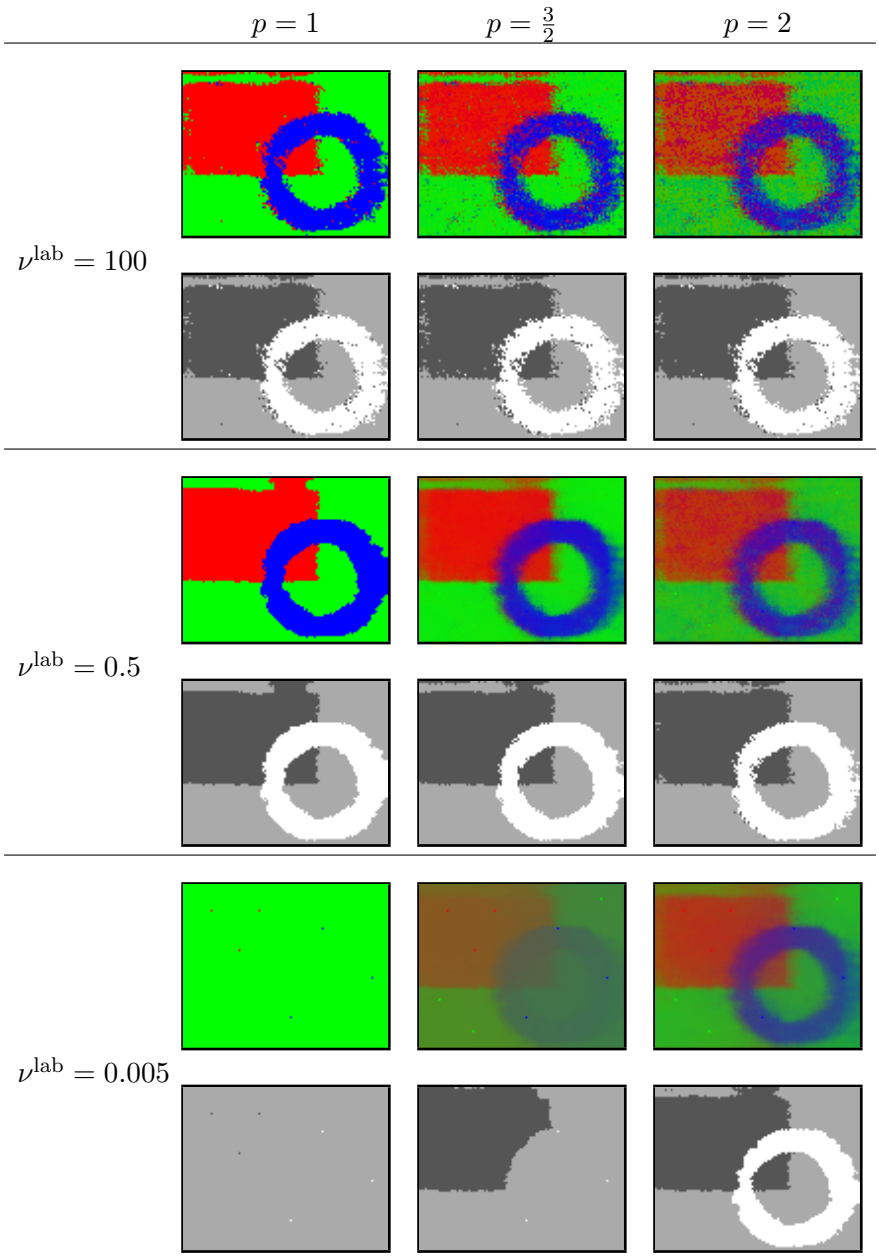


FIGURE 3. Effect of ν^{lab}: the parameters $\nu^{\text{pho}} = 1,\ r = 5,\ \rho^2 = s = 363$ are fixed. For every ν^{lab}, the first row (color) shows the segment membership matrix u and the second row shows the resulting segmentation which is achieved by taking $\operatorname{argmax}_k u^k$ pixel-wise. The resulting segment membership matrices for $p = 1$ are closest to hard clustering. With decreasing ν^{lab}, the influence of the labeled pixels decreases, and we obtain smoother results due to a stronger impact of weights between not necessarily labeled but similar pixels.

weights are given by

$$w_{i,j}^{\text{pho}} := \begin{cases} \kappa_i e^{-\frac{\|F(i)-F(j)\|_2^2}{\rho_i}} & \text{if } j \in \mathcal{N}_i^{\text{pho}}, \\ 0 & \text{otherwise,} \end{cases}$$

where we normalize as in the geometric case with $\kappa_i := 1/\sum_{j\in\mathcal{N}_i^{\text{pho}}} e^{-\frac{\|F(i)-F(j)\|_2^2}{\rho_i}}$. For the parameter ρ_i we use either a constant ρ independent of i or we use the Euclidean norm of the component-wise sample variance of the color features $F(j), j \in \mathcal{N}_i^{\text{pho}}$. The variance is defined by

$$\rho_i := \left\| \left(\frac{1}{|\mathcal{N}_i^{\text{pho}}| - 1} \sum_{j\in\mathcal{N}_i} (F_m(j) - \overline{F_m}) \right)_{m=1}^{s} \right\|_2 \tag{12}$$

where the mean is given by $\overline{F_m} := \sum_{j\in\mathcal{N}_i^{\text{pho}}} F_m(j)/|\mathcal{N}_i^{\text{pho}}|$.

We define the labeled neighborhood $\mathcal{N}_i^{\text{lab}}$ to be the 4 labeled pixels with smallest Euclidean distance to pixel i. The weights are defined analogously to the photometric weights. If there are too many pixels labeled, we choose a random sample of equal size for each segment. The neighborhood of a pixel i is given by $\mathcal{N}_i := \mathcal{N}_i^{\text{geo}} \cup \mathcal{N}_i^{\text{pho}} \cup \mathcal{N}_i^{\text{lab}}$. In summary, the weight matrix is computed as follows: the geometric and photometric weights are added

$$W^* = \frac{1}{1+\nu^{\text{pho}}} W^{\text{geo}} + \frac{\nu^{\text{pho}}}{1+\nu^{\text{pho}}} W^{\text{pho}}. \tag{13}$$

Then, they are compared with the labeled weights via the element-wise maximum

$$\grave{W} = \max\left\{ \frac{\nu^{\text{lab}}}{1+\nu^{\text{lab}}} W^{\text{lab}}, \frac{1}{1+\nu^{\text{lab}}} W^* \right\}. \tag{14}$$

Finally, we use the symmetric weight matrix

$$W = \max\left\{ \grave{W}, \grave{W}^T \right\}.$$

Figure 2 shows the effect of changing p as well as ν^{pho}. Decreasing ν^{pho} decreases the influence of the photometric neighborhood and increases the influence of the geometric neighborhood (13), resulting in smoother results.

A more severe effect can be shown by changing ν^{lab} in (14). We use the input image shown in Figure 2 (A) and (E). In Figure 3, we fix ν^{pho} and vary ν^{lab}. This comparison illustrates that decreasing ν^{lab} decreases the influence of the labeled pixels, which is clear from the definition of the weights (14). With decreasing ν^{lab} we obtain smoother results due to stronger impact of weights between not necessarily labeled but similar pixels.

Comparing among the same parameters, the resulting segment membership matrices u for $p = 1$ are closest to hard clustering and have smoother boundaries. By choosing different weight parameters, p-Laplacians with $p > 1$ can give similar results to $p = 1$: In Figure 3 the segmentation result $\nu^{\text{lab}} = 0.5$ with $p = 1$ is similar to the segmentation result $\nu^{\text{lab}} = 0.005$ with $p = 2$. However, comparing the segment membership matrices u (shown in color) the one for $p = 1$ is sharper.

4. Supervised Multi-class Segmentation with RKHS

In this section, we introduce a multi-class segmentation method based on vector-valued RKHSs. The idea is motivated from image colorization in [26]. Instead of color channels we deal with the c channels of the segment membership functions. This approach is very fast and efficient, in particular when a small region of color is given.

We start this section with a brief introduction to RKHS, the abstract theory for which was developed by Aronszajn in [3]. For the vector-valued extension, we refer to [11, 36, 38, 42]. We consider vector-valued functions $g = (g^k)_{k=1}^c : \mathbb{R}^2 \to \mathbb{R}^c$. A vector-valued RKHS is a Hilbert space $\mathcal{H}$ of vector-valued functions such that the point evaluation operator $\delta_x : \mathcal{H} \to \mathbb{R}^c$ with $\delta_x g := g(x)$ is a linear, bounded operator, i.e., for all $x \in \mathbb{R}^2$ there exists a constant C_x such that

$$\|g(x)\|_2 \leq C_x \|g\|_{\mathcal{H}} \qquad \forall g \in \mathcal{H}.$$

A function $K : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathrm{Sym}_c(\mathbb{R})$ mapping into the set $\mathrm{Sym}_c(\mathbb{R})$ of real-valued, symmetric matrices is called an *operator-valued, positive definite kernel* or *kernel of positive type* if for all $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{R}^2$ and $a_1, \ldots, a_n \in \mathbb{R}$ the relation

$$\sum_{i,j=1}^{n} a_i a_j \langle v, K(x_i, x_j) v \rangle \geq 0 \quad \forall v \in \mathbb{R}^c$$

holds true. As in the scalar case any RKHS canonically defines a kernel K of positive type by $K(x, y) = \delta_x \delta_y^*$. This kernel fulfills $K_x v := K(\cdot, x)v \in \mathcal{H}$ for all $v \in \mathbb{R}^c$ and the *reproducing property*

$$\langle g(x), v \rangle = \langle g, K_x v \rangle_{\mathcal{H}} \quad \forall x \in \mathbb{R}^2, \forall v \in \mathbb{R}^c.$$

Conversely, every kernel of positive type defines a unique vector-valued RKHS. Another way to approach vector-valued RKHSs is to consider for a kernel K of positive type the space

$$\mathcal{H}^{\mathrm{pre}} := \mathrm{span}\{K_x v : x \in \mathbb{R}^2, v \in \mathbb{R}^c\}$$

of all finite linear combinations of $K_x v$ and for $g := \sum_{i=1}^n a_i K_{x_i} v_i$ and $\tilde{g} := \sum_{i=1}^n b_i K_{y_i} w_i$ the inner product

$$\langle g, \tilde{g} \rangle_{\mathcal{H}} := \sum_{i,j=1}^{n} a_i b_j \langle v_i, K(x_i, y_j) w_j \rangle, \quad \|g\|_{\mathcal{H}}^2 = \sum_{i,j=1}^{n} a_i a_j \langle v_i, K(x_i, x_j) v_j \rangle. \tag{15}$$

Then the closure $\mathcal{H} = \overline{\mathcal{H}^{\mathrm{pre}}}$ with respect to this norm is the RKHS associated with K.

4.1. **RKHS segmentation model.** Instead of the segment membership vector u in the previous section, we consider a *segment membership function* $g = (g^k)_{k=1}^c : \mathbb{R}^2 \to \mathbb{R}^c$ belonging to a RKHS $\mathcal{H}$ with kernel K. For all labeled points $i \in L$, we assume that $g(x_i) = l_L(i)$ are given for some $x_i \in \mathbb{R}^2$, where l_L is defined as in (1). We find the segment membership function $\hat{g}$ as the solution of

$$\operatorname*{argmin}_{g \in \mathcal{H}} \sum_{i \in L} \|l_L(i) - g(x_i)\|_2^2 + \lambda \|g\|_{\mathcal{H}}^2, \quad \lambda > 0. \tag{16}$$

Alternatively, we can consider

$$\operatorname*{argmin}_{g \in \mathcal{H}} \|g\|_{\mathcal{H}} \quad \text{subject to} \quad g(x_i) = l_L(i), \ \ \forall i \in L. \tag{17}$$

Then by representer theorem, the minimizers of (16), resp., (17) have the following form (see [26, Proposition 1]):

$$\hat{g} = \sum_{j \in L} K(\cdot, x_j)\alpha(j), \quad \alpha(j) \in \mathbb{R}^c. \tag{18}$$

This is a pleasant generalization property of the minimizer. We need only to determine $\alpha(j)$ for $j \in L$ to get the whole function $\hat{g} \in \mathcal{H}$.

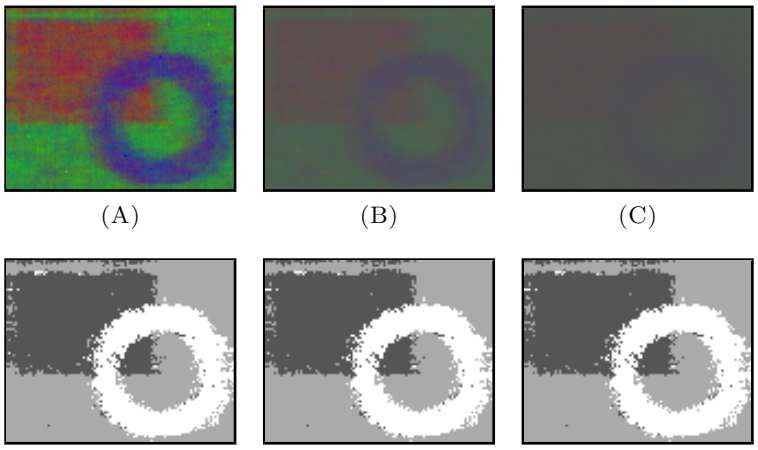


FIGURE 4. Effect of λ in the RKHS model (21): Figure 2 (A) and (E) are the input images. The top row (color) shows the segment membership function $\hat{g}$ and the second row the resulting segmentation achieved by taking $\operatorname{argmax}_k u^k$ componentwise. (A) $\lambda = 0$, (B) $\lambda = 0.1$ and (C) $\lambda = 0.4$. The kernel parameters are $r = 5,\ \sigma_1^2 = 1,\ \sigma_2^2 = \infty$.

To find the vectors $\alpha(j)$, we substitute (18) in (16) and apply the norm definition (15):

$$\operatorname*{argmin}_{\alpha} \sum_{i \in L} \|l_L(i) - K(x_i, x_j)\alpha(j)\|_2^2 + \lambda \sum_{i,j \in L} \alpha(i)^{\mathrm{T}} K(x_i, x_j)\alpha(j). \tag{19}$$

Let $K(x_i, x_j) = \left(K^{\ell,k}(x_i, x_j)\right)_{\ell,k=1}^{c}$. Using the matrix

$$\mathbf{K}_{LL} := \begin{pmatrix} \left(K^{11}(x_i, x_j)\right)_{i,j \in L} & \dots & \left(K^{1c}(x_i, x_j)\right)_{i,j \in L} \\ \vdots & \dots & \vdots \\ \left(K^{c1}(x_i, x_j)\right)_{i,j \in L} & \dots & \left(K^{cc}(x_i, x_j)\right)_{i,j \in L} \end{pmatrix} \in \mathbb{R}^{c|L|, c|L|}, \tag{20}$$

and the notation

$$\alpha := \begin{pmatrix} \alpha^1 \\ \vdots \\ \alpha^c \end{pmatrix}, \quad \alpha^k = (\alpha^k(j))_{j \in L},$$

we can rewrite (19) as

$$\operatorname*{argmin}_{\alpha} \|l_L - \mathbf{K}_{LL}\alpha\|_2^2 + \lambda \alpha^{\mathrm{T}} \mathbf{K}_{LL} \alpha. \tag{21}$$

Setting the gradient with respect to α to zero we verify that the solution $\hat{\alpha}$ of (21) is given by the solution of the linear system of equations

$$(\mathbf{K}_{LL} + \lambda I)\, \hat{\alpha} = l_L. \tag{22}$$

Note that the coefficient matrix is positive definite so that the solution is unique. Similarly, a minimizer of (17) follows by the solution of (22) with $\lambda = 0$, if $\mathbf{K}_{LL}$ is positive definite. Then, the segment membership function $\hat{g}$ can be evaluated at any $x \in \mathbb{R}^2$ by (18).

4.2. **Effects of the RKHS model via numerical examples.** In numerical experiments, we have that $x_i = i \in L$ are labeled image grid points. In general, the reproducing kernel K can generate a fully populated matrix $\mathbf{K}_{LL}$ with entries for all pairwise classes. In our application, we decouple the classes and consider diagonal matrices $K(i,j) := \mathrm{diag}(K^1(i,j), \ldots, K^c(i,j))$, $i,j \in L$. Then $\mathbf{K}_{LL}$ becomes the block-diagonal matrix

$$\mathbf{K}_{LL} = \mathrm{diag}((K^1_{LL}(i,j), \ldots, K^c_{LL}(i,j))$$

which we further simplify by choosing $K_{LL} := K^1_{LL} = \ldots = K^c_{LL}$. Now (21) can be tackled separately for each class $k = 1, \ldots, c$, i.e., we compute

$$\underset{\alpha^k}{\mathrm{argmin}} \left\| l^k_L - K_{LL}\alpha^k \right\|_2^2 + \lambda (\alpha^k)^{\mathrm{T}} K_{LL} \alpha^k$$

by solving

$$(K_{LL} + \lambda I)\, \hat{\alpha}^k = l^k_L. \tag{23}$$

This is just the quadratic support vector machine (SVM) approach [49] applied for each k. By (18), we obtain the values of our segment membership function $\hat{g}$ at the unlabeled points $x_i = i \in U$ by

$$\hat{g}^k = \begin{pmatrix} K_{LL} \\ K_{UL} \end{pmatrix} \hat{\alpha}^k, \tag{24}$$

where $K_{UL} := (K(i,j))_{i \in U, j \in L}$. The entries of the above matrix are values of a positive semi-definite kernel. Thus the matrix is in general fully populated. However, since their number of columns $|L|$ is small both the solution of the linear system (23) and the matrix-vector multiplications in (24) require only few computational time.

This gives the efficiency of this RKHS approach. When the number $|L|$ is small, the computation is very fast and $\hat{\alpha}$ can be reused for different images. To be more precise, let the image with the labeles be the 'first' image and another similar but not labeled image be the 'second' image. One can proceed as follows:

(1) Create $K_{\bar{U}L}$ where $\bar{U}$ contains all pixels of the second image and L are the labeled pixels of the first image.
(2) Compute the segment membership function by $K_{\bar{U}L}\hat{\alpha}$.

In addition, notice from (23) that only the labeled points are used for the segmentation. This is different from the p-Laplacian approach. The result will be more depended on the labeled point such that regularization and smoothing effects will be less apparent compared to the p-Laplacian model. We observe this from the following numerical experiments as well.

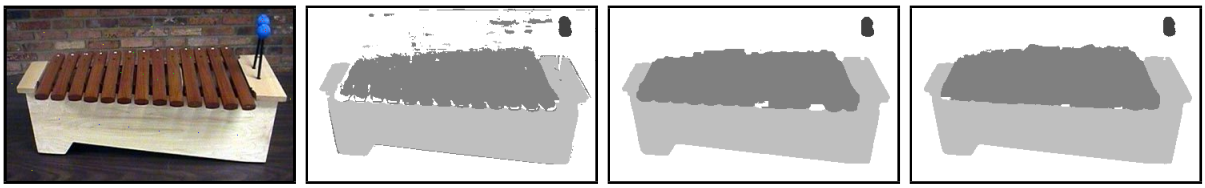
For the elements of K, various weights can be considered. Here we use

$$K(i,j) := \exp\left(-\frac{\|F(i) - F(j)\|_2}{\sigma_1^2 d} \right) \exp\left(-\frac{\left\| \begin{pmatrix} i_x - j_x \\ i_y - j_y \end{pmatrix} \right\|_2}{\sigma_2^2 \sqrt{n_1^2 + n_2^2}} \right)$$

where $1/\infty = 0$ and $(i_x, i_y)^T$ denotes the pixel position of an image in matrix form, i.e., $i = n_1(i_y - 1) + i_x$. Here F is the same feature vector as in Subsection 3.4.

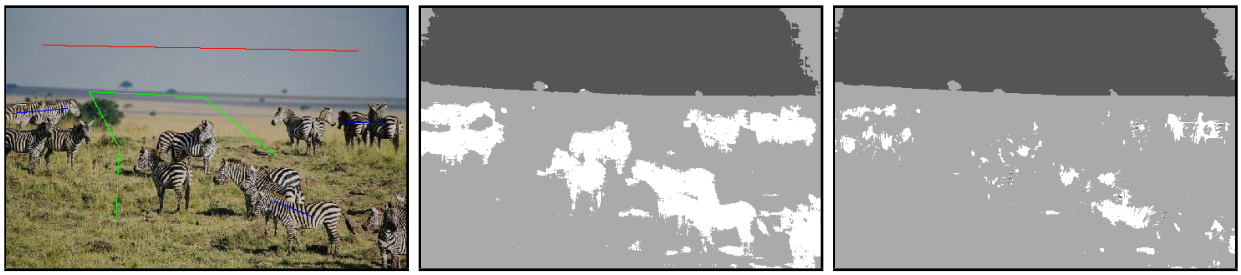
Figure 4 shows the effect of using different regularization parameters λ in the minimization problem (21). For larger λ the contrast of the segment membership function $\hat{g}$ decreases while the segmentation result stays quite similar. Hence, we will use $\lambda > 0$ only if a bad conditioning of the linear system forces us to.

Since the RKHS approach only uses the information from the labeled points L, the regularity of the result can be worse compared to p-Laplacian models, as shown in Figure 5. The p-Laplacian



(A) Input (B) RKHS (C) 1-Lap. (D) 2-Lap.

FIGURE 5. (A) Input image from [25], only 6 pixels per class are labeled. (B) Result of RKHS with $\sigma_1^2 = 1,\ \sigma_2^2 = 250,\ r = 3,\ \lambda = 0$. (C) Result of 1-Laplacian (10) with $\nu^{\text{pho}} = 1,\ \nu^{\text{lab}} = 0.15,\ \rho^2 = 49,\ r = 3$. (D) Result of 2-Laplacian (6) with $\nu^{\text{pho}} = 1,\ \nu^{\text{lab}} = 0.001,\ \rho^2 = 49,\ r = 3$.



(A) Input (B) RKHS (C) 1-Lap.

FIGURE 6. (A) Labeled input image. (B) Segmentation result of RKHS with $\sigma_1^2 = 1,\ \sigma_2^2 = \infty,\ r = 3,\ \lambda = 0$. (C) Result of model (10) with $p = 1$ and $\nu^{\text{pho}} = 1,\ \nu^{\text{lab}} = 1,\ \rho^2 = 49,\ r = 3$.

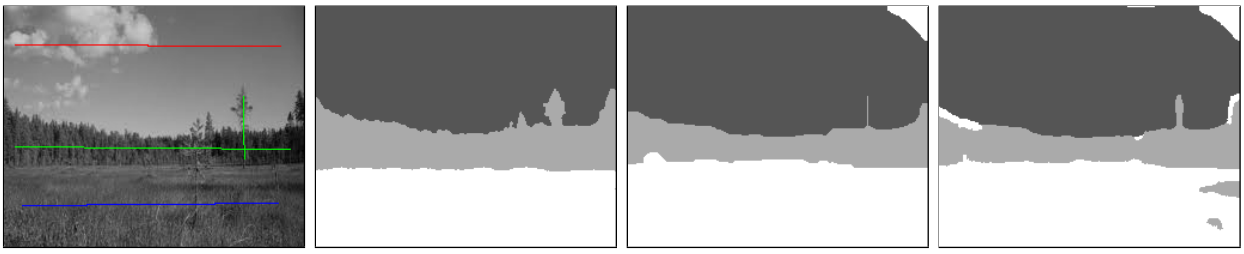
models utilize the information from the unlabeled similar points, which gives more regularization (also compare Figure 4 with Figure 2 and 3).

However, the RKHS approach can provide more accurate results as illustrated in Figure 6 and 7. Figure 6 shows the results for a natural image taken from [25]. The results of the RKHS (21) and the p-Laplacian model (10) for $p = 1$ are depicted in Figure 6B and 6C, respectively. Although Figure 6A has many details and the ground and zebras are hard to distinguish, the RKHS approach recovers the location of many different zebras correctly. Other values for the parameters $\rho, \nu^{\text{lab}}, \nu^{\text{pho}}, r$ did not improve the result of the 1-Laplacian model significantly. Also with the 2-Laplacian model (6) one does not obtain better results. In fact, for the parameters of Figure 6C the segmentations for $p = 1$ and $p = 2$ are quite similar.

Figure 7 shows an example of a landscape. Notice that sky, grass, and the small tree are better captured by the RKHS model. In both Figures 6 and 7 the RKHS model provides a more accurate segmentation. One difference between the RKHS and the Laplacian models is that, for the Laplacian models, computationally we can only cope with few labeled pixels. Each pixel has only 4 labeled pixels in his neighborhood, see Section 3.4 for more details. For a number of labeled pixels a little larger than 4, e.g., 16, we did not observe significant improvements.

5. A combined p-Laplacian and RKHS approach

As noticed from previous sections, the RKHS uses only the information from the labeled points L, while the p-Laplacian model also incorporates the similarity information between unlabeled points. Therefore the p-Laplacian approach can have a stronger regularization effect. On the other hand,



(A) Input (B) RKHS (C) Using 1-Lap. (D) Using 2-Lap.

FIGURE 7. (A) Labeled input image. (B) Segmentation result of RKHS with $\sigma_1^2 = 1,\ \sigma_2^2 = 50,\ r = 5$. (C) Result of the 1-Laplacian model with $\nu^{\text{pho}} = 1,\ \nu^{\text{lab}} = 0.05,\ \rho^2 = 121,\ r = 5$. (D) Result of the 2-Laplacian model with $\nu^{\text{pho}} = 10,\ \nu^{\text{lab}} = 0.001,\ \rho^2 = 121,\ r = 5$.

the RKHS approach leads often to more accurate/detailed segmentation results. We consider a combined model to benefit from both approaches. We will provide a projection approach which makes a combined model practicable for segmentation tasks.

First, we review a straightforward combined approach proposed in [5, 46] for two-class learning which couples the least squares RKHS with 2-Laplacians. In those papers the least squares RKHS was addressed as RLP (regularized least squares). Since only two labels were considered one can restrict the attention to just one segmentation function $f : \mathbb{R}^2 \to \mathbb{R}$ and set a threshold on f to get the two classes. The optimal function $\hat{f}$ was obtained as the minimizer of the following functional

$$\operatorname*{argmin}_{f\in\mathcal{H}} \sum_{j\in L} \frac{1}{2} \|l_L(j) - f(x_j)\|_2^2 + \frac{\lambda}{2} \|f\|_{\mathcal{H}}^2 + \frac{\mu}{2} \sum_{i,j=1}^{N} w_{i,j} |f(x_i) - f(x_j)|^2. \tag{25}$$

Following the idea of the representer theorem [55] it is not hard to check that the minimizer $\hat{f}$ *depends on all values* x_j, $j \in \mathbb{I}_N$, i.e.,

$$\hat{f} = \sum_{j=1}^{N} K(\cdot, x_j)\beta(j). \tag{26}$$

Substituting this expression into (25) and using the definition of the RKHS norm in (15) we conclude that the optimal $\hat{\beta}$ must be a solution of

$$\operatorname*{argmin}_{\beta} \frac{1}{2}\|l_L - JK\beta\|_2^2 + \frac{\lambda}{2}\beta^{\mathrm{T}} K\beta + \frac{\mu}{2}\beta^{\mathrm{T}} K \triangle_2 K\beta,$$

where $K := (K(x_i, x_j))_{i,j=1}^N = \begin{pmatrix} K_{LL} & K_{LU} \\ K_{LU} & K_{UU} \end{pmatrix}$ and J is defined as in Remark 3.3. Setting the gradient with respect to β to zero and using that K is invertible, we obtain $\hat{\beta}$ by solving a linear system of equations

$$(J^{\mathrm{T}} JK + \lambda I_N + \mu \Delta_2 K)\hat{\beta} = J^{\mathrm{T}} l_L.$$

This means that one has to incorporate the whole fully populated matrix K, in particular K_{UU} into the computations. In segmentation tasks the number of unlabeled points is huge (nearly $N = n_1 n_2$ for images of size $n_1 \times n_2$) and the above model is not practicable both with respect to storage and computation time.

5.1. **A projection model.** To avoid the computation with the huge matrix K_{UU}, we propose a combined model which uses a projection idea. In contrast to the generalization ability (26), this model is again transductive with respect to the image grid points $x_i = i$, $i \in \mathbb{I}_N = L \cup U$. Let $\mathbf{K}_{LL}$ be given by (20) and $\mathbf{K}_{UL} := (K(i,j))_{i\in U, j\in L}$. We consider the subspace H of R^{cN} defined by

$$H := \{ \underbrace{\begin{pmatrix} \mathbf{K}_{LL} \\ \mathbf{K}_{UL} \end{pmatrix}}_{\mathbf{K}} \alpha : \alpha \in R^{c|L|} \}$$

with norm

$$\|h\|_H := \alpha^{\mathrm{T}} \mathbf{K}_{LL} \alpha \quad \text{for} \quad h := \mathbf{K}\alpha.$$

We are looking for vectors $\hat{h} \in H$ and $\hat{u} \in R^{cN}$ solving the *combined model*

$$\underset{g\in H, u\in\mathbb{R}^{Nc}}{\operatorname{argmin}} \frac{1}{2} \sum_{i\in L} \|l_L(i) - h(i)\|_2^2 + \frac{\lambda}{2} \|h\|_H^2 + \frac{\mu}{p} \sum_{k=1}^{c} \sum_{i,j\in\mathbb{I}_N} w_{i,j} |u^k(i) - u^k(j)|^p, \quad \text{subject to } h = Pu. \tag{27}$$

Here $P : \mathbb{R}^{cN} \to H$ denotes the *orthogonal projector* from $\mathbb{R}^{cN}$ onto H. More precisely, we expect that $\hat{h}$ has similar properties as $(\hat{g}(x_i))_{i\in\mathbb{I}_N}$ from the RKHS approach and that $\hat{u}$, which is the vector we are really looking for, adopts smoothing effects from the Laplacian regularization.
By definition of H this orthogonal projector is given by $P = \mathbf{K}\mathbf{K}^\dagger$, where $\mathbf{K}^\dagger := (\mathbf{K}^{\mathrm{T}}\mathbf{K})^{-1}\mathbf{K}^{\mathrm{T}}$ is the *Moore-Penrose inverse* of $\mathbf{K}$. Note that $\mathbf{K}^{\mathrm{T}}\mathbf{K}$ is positive definite and thus invertible. Then, for $h := \mathbf{K}\alpha$, the constraint can be written as $\mathbf{K}\alpha = \mathbf{K}\mathbf{K}^\dagger u$. Since $\mathbf{K}$ has full column rank this is equivalent to

$$\alpha = \mathbf{K}^\dagger u.$$

Substituting this into (27) we obtain

$$\underset{u\in\mathbb{R}^{Nc}}{\operatorname{argmin}} \frac{1}{2} \left\| l_L - \mathbf{K}_{LL}\mathbf{K}^\dagger u \right\|_2^2 + \frac{\lambda}{2} u^{\mathrm{T}} (\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL} \mathbf{K}^\dagger u + \frac{\mu}{p} \sum_{k=1}^{c} \sum_{i,j\in\mathbb{I}_N} w_{i,j} |u^k(i) - u^k(j)|^p. \tag{28}$$

For $p = 2$, the solution $\hat{u}$ of (28) can be obtained by setting the gradient of the functional to zero. Then $\hat{u}$ is given by the solution of the linear system of equations

$$\left((\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL} (\mathbf{K}_{LL} + \lambda I_{c|l|}) \mathbf{K}^\dagger + \mu \Delta_2 \right) u = \mathbf{K}^\dagger \mathbf{K}_{LL} l_L. \tag{29}$$

We use the *conjugate gradient method (CG)* to solve (29). Note that the huge matrix K_{UU} does not appear in the above linear system. Therefore, we are able to implement the involved matrix multiplication efficiently with respect to the memory.
For $p = 1$, we can be rewrite (28) with $M := I_c \otimes A$ and A defined in (9) as

$$\min_{u\in\mathbb{R}^{Nc}} \frac{1}{2} \left\| l_L - \mathbf{K}_{LL}\mathbf{K}^\dagger u \right\|_2^2 + \frac{\lambda}{2} u^{\mathrm{T}} (\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL} \mathbf{K}^\dagger u + \mu \|v\|_1 \quad \text{subject to } Mu = v. \tag{30}$$

To solve this problem we apply the alternating direction method of multipliers (ADMM), see, e.g., [6, 18, 20, 43].

Input: $M(= I_c \otimes A$ in (9)) and γ.
Output: segment membership vector u.
Initialization of $u^{(0)} \in \mathbb{R}^{Nc}$ and $b^{(0)}, v^{(0)} \in \mathbb{R}^{N^2 c}$;
for $r = 0, 1, \ldots$ *until a stopping criterion is reached* **do**

$$u^{(r+1)} = \operatorname*{argmin}_u \frac{1}{2} \left\| l_L - \mathbf{K}_{LL}\mathbf{K}^\dagger u \right\|_2^2 + \frac{\lambda}{2} u^{\mathrm{T}} (\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL}\mathbf{K}^\dagger u + \frac{1}{2\gamma} \left\| b^{(r)} + Mu - v^{(r)} \right\|_2^2; \tag{31}$$

$$v^{(r+1)} = \operatorname*{argmin}_v \mu \left\| v \right\|_1 + \frac{1}{2\gamma} \left\| b^{(r)} + Mu^{(r+1)} - v \right\|_2^2; \tag{32}$$

$$b^{(r+1)} = b^{(r)} + Mu^{(r+1)} - v;$$

end

Algorithm 2: ADMM for solving (30).

The minimizer of (32) follows by soft-shrinkage of $b^{(r)} + Mg$ with threshold $\gamma\mu$. The minimizer of (31) can be obtained by setting the derivative of the functional to zero. We have to solve a system similar to (29) namely

$$\left((\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL} \left(\mathbf{K}_{LL} + \lambda I_{c|L|} \right) \mathbf{K}^\dagger + \frac{1}{\gamma} M^{\mathrm{T}} M \right) u = (\mathbf{K}^\dagger)^{\mathrm{T}} \mathbf{K}_{LL} l_L + \frac{1}{\gamma} M^{\mathrm{T}} (v^{(r)} - b^{(r)}).$$

Alternatively, one could use the PDHGMp. The only difference to the ADMM is that the first step reads

$$u^{(r+1)} = \operatorname*{argmin}_u \frac{1}{2} \left\| l_L - \mathbf{K}_{LL}\mathbf{K}^\dagger u \right\|_2^2 + \frac{\lambda}{2} u^{\mathrm{T}} (K^\dagger)^{\mathrm{T}} \mathbf{K}_{LL}\mathbf{K}^\dagger u + \frac{1}{2\tau} \left\| u - \left(\underbrace{u^{(r)} - \tau\gamma M^{\mathrm{T}} (2b^{(r)} - b^{(r-1)})}_{z} \right) \right\|_2^2$$

where $\tau > 0$ and $\tau\gamma \leq 1/\left\|M\right\|^2$.

Remark 5.1. *The various models we considered up to now can be summarized as follows (not considering the simplex condition). The general p-Laplacian model* (4) *can be understood as*

$$\min_{u_U} Q_p(u) \quad \textit{subject to} \quad u_L = l_L$$

where Q_p is defined in (3)*, which can be approximated by a non-constraint problem*

$$\min_u \lambda Q_p(u) + \frac{1}{2} \|l_L - u_L\|^2, \quad \lambda > 0.$$

In the RKHS model in (16) *and* (17)*, the regularization appears from the reproducing kernel Hilbert space norm, i.e.,*

$$\min_{g \in \mathcal{H}} \|g\|_{\mathcal{H}}^2 \quad \textit{subject to} \quad (g(x_i))_{i \in L} = l_L,$$

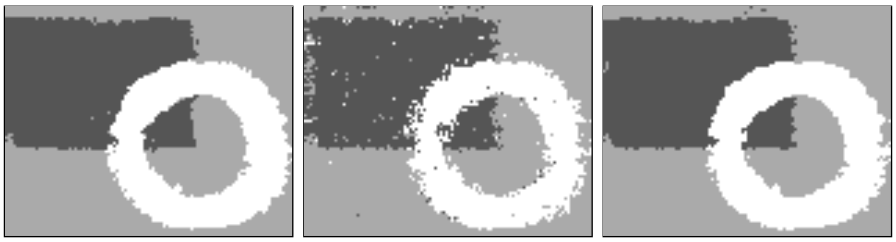
and

$$\min_u \lambda \left\|g\right\|_{\mathcal{H}}^2 + \frac{1}{2} \|l_L - (g(x_i))_{i \in L}\|^2, \quad \lambda > 0,$$

respectively. This can be solved by (21) *simply using* (22)*. The combined model* (27) *has the form*

$$\min_{u,h} \mu Q_p(u) + \lambda \left\|h\right\|_{\mathcal{H}}^2 + \frac{1}{2} \|l_L - h\|^2 \quad \textit{subject to} \quad Pu = h.$$

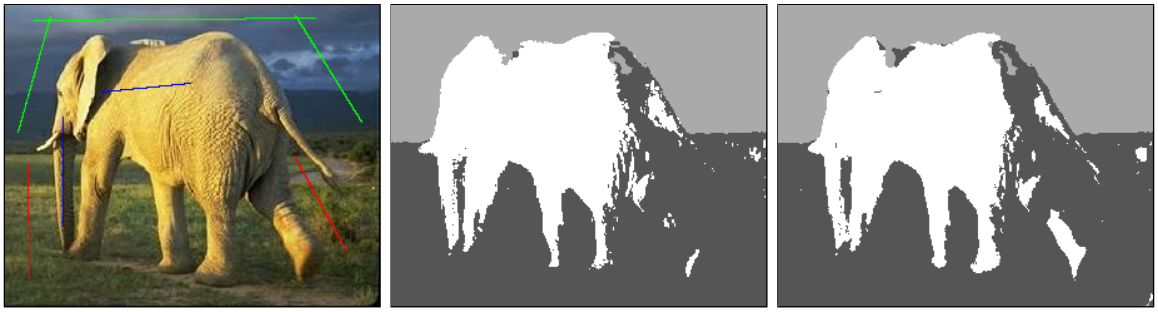
Here u represents the segment membership vector regularized by the p-Laplacian, and h is the segment-membership vector regularized by the RKHS norm with the labeled point constraint. The two vectors u and h are connected by an orthogonal projection $P : \mathbb{R}^{cN} \to H$, which has an



(A) $\hat{u}, (p = 1)$ (B) $P\hat{u}$ $(p = 1)$ (C) $\hat{u}$ $(p = 2)$

(D) $P\hat{u}$ $(p = 2)$ (E) $\hat{g}$ from RKHS

FIGURE 8. The combined model (27) using $\nu^{\text{pho}} = 1,\ \nu^{\text{lab}} = 0.2,\ \sigma_1^2 = 0.5,\ \sigma_2^2 = \infty,\ r = 4,\ \lambda = 0,\ \mu = 0.001$. (A) $\hat{u}$ with $p = 1$. (B) $P\hat{u}$ with $p = 1$. (C) $\hat{u}$ with $p = 2$. (D) $P\hat{u}$ with $p = 2$. (E) RKHS as in Figure 4. By adding the Laplacian regularization, the segmentation result $\hat{u}$ has smoother boundaries compared to $\hat{g}$ from the RKHS model. As expected $P\hat{u}$ is similar to the result $\hat{g}$ from the RKHS method.

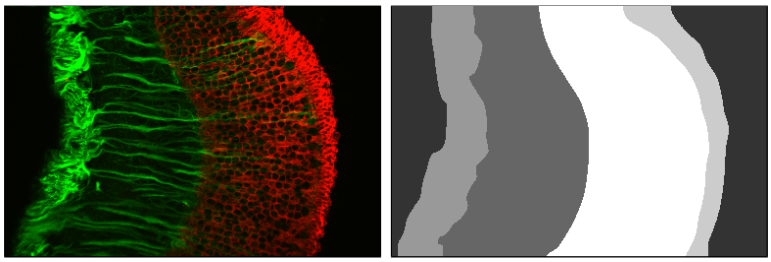


(A) Input (B) RKHS (C) Comb. $p = 2, \mu = 0.001$

FIGURE 9. The combined model compared to the RKHS approach for a natural image with $\nu^{\text{pho}} = 0,\ \nu^{\text{lab}} = 0.5,\ \sigma_1^2 = 1,\ \sigma_2^2 = 250,\ r = 2,\ \lambda = 0$.

explicit form given by the kernel K. Therefore, in the combined model (27)*, we can consider the segmentation result $\hat{u}$ and its projection $\hat{h} = P\hat{u}$.*

5.2. **Numerical experiments for the combined projection model.** Figure 8 shows the results using the combined model (27) compared to the RKHS results. The projection $\hat{h} = P\hat{u}$ computed for both $p = 1$ and $p = 2$ is quite similar to the result of the RKHS approach (not only the depicted segmentation, but also the segment membership function itself). On the other hand, $\hat{u}$ provides for



(A) (B)

FIGURE 10. Input and ground truth.

both $p = 1$ and $p = 2$ significantly smoother segmentations than $\hat{g}$. Figure 9 depicts the result of the combined model (discretization of u) for $p = 2$. The segmentation for $p = 1$ looks quite similar. Note that, besides the smoothing, a larger part of the leg is identified as part of the elephant. The combined model can improve the results, where the RKHS method is not smooth enough but more accurate than the Laplacian model, see Figure 18. The computation time for the combined model is a drawback of the current algorithm. However, there is a lot of potential to speed up computation, e.g., by a parallel implementation on a GPU.

6. Application to Medical Images

Medical images are often very challenging for segmentation since they suffer from low contrast and heavy noise with many fine details. The models discussed in this paper are good candidates for segmenting such images. In this section, we apply our models to show the differences between the results. In contrast to our previous numerical experiments, we apply the segmentation models to a collection of similar images, where only a single input image has got some labeled points in advance. Figures 10 and 11 show a collection of retina images, cf. [32, 22]. The labels for the segmentation were only taken from Figure 10 (A). More precisely, they are randomly sampled from the ground truth in Figure 10 (A). This example illustrates that in cases where the objective of the segmentation (the ground truth of images) has a relative simple structure with smooth boundaries the 1-Laplacian model is most appropriate. As illustrated earlier, the 1-Laplacian model gives the most regularized results and it works well for these applications.

Next images show cardiac MR heart images taken from [62]. The objective is to find the endocardial wall of both right and left ventricles (gray region) which is separated by epicardium walls (darker gray), while discarding the complicated background. This is a 3-phase segmentation: background, surrounding region and inner region. Only the first image, Figure 12 is labeled, and all the images in Figure 13 are segmented using the same labels from the first image.

The images are taken from a stack of the same heart. Hence, the segments can expected to be approximately in the same part of the image. Hence, it useful to work with the spatial parameter $\sigma_2^2 < \infty$. For this example, although the objective is to find relatively smooth objects, the RKHS method gives good results. Figure 14 shows the comparison with the 1-Laplacian model (2-Laplacian behaves similar), which fails in this case: The labeled points are taken from the image in Figure 14A and this image itself is well segmented. However, a similar image in Figure 14C fails to be well segmented with the same labels. This maybe due to the fact that the unlabeled points U bring a lot of similarity details from the background into play which erroneously influences the resulting segmentation, as it was also the case in Figure 6.

For the various models considered in this paper Table 1 shows time comparisons. The algorithms

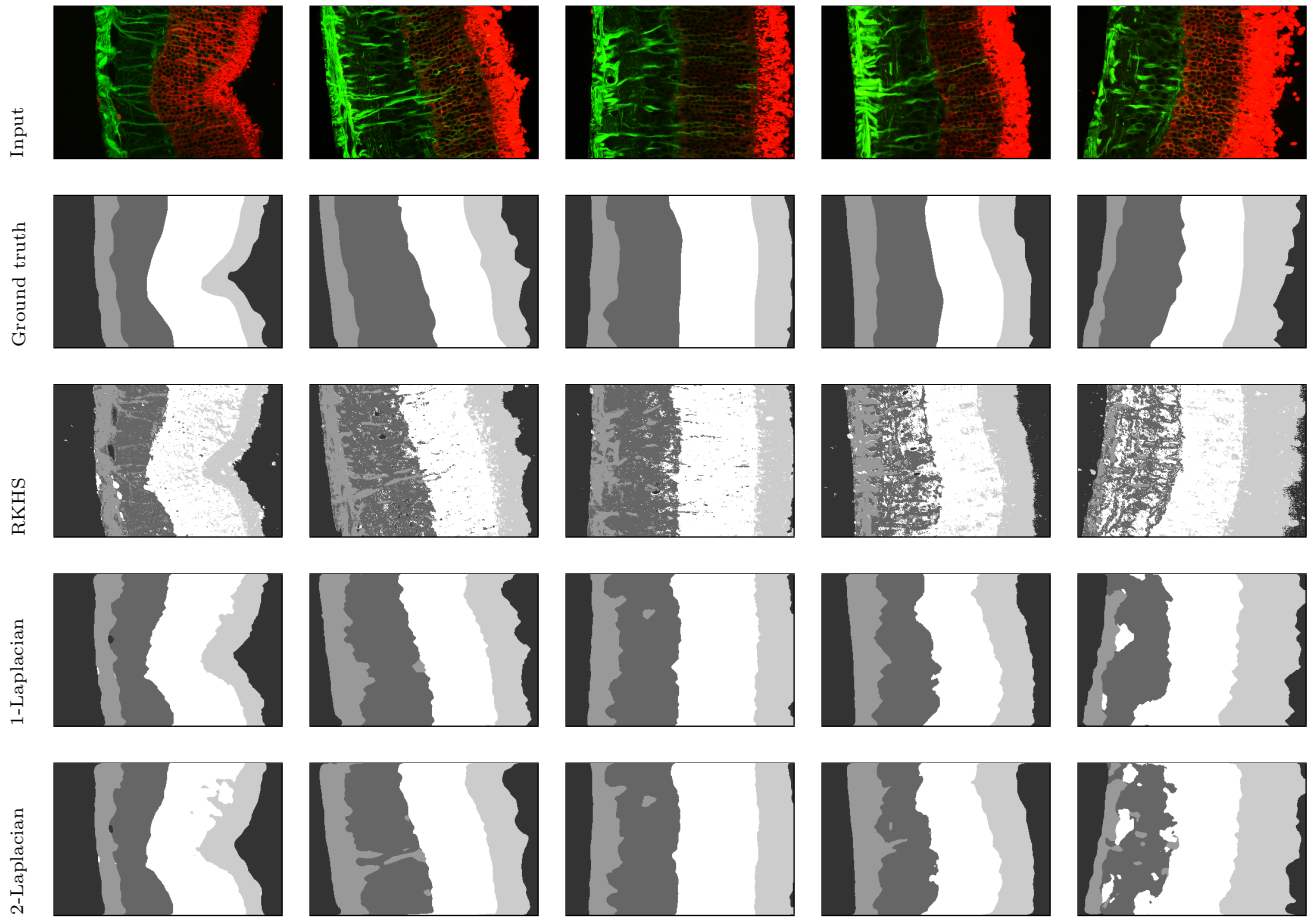


FIGURE 11. Comparison of retina images. The parameter ρ_i was computed by (12). $\sigma_1^2 = \frac{1}{2},\ \sigma_2^2 = \infty,\ \nu^{\text{pho}} = 0.001,\ \nu^{\text{lab}} = 0.01,\ r = 1.$

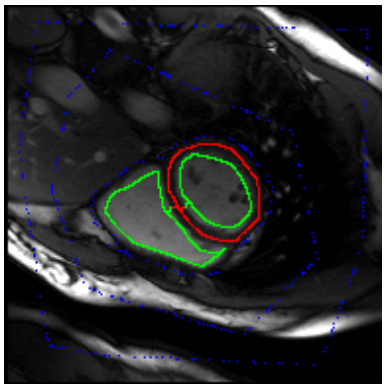


FIGURE 12. Input for the segmentation in Fig. 13 and 14.

were implemented in MATLAB and executed on an Intel Core i7 CPU with 2.93GHz. Our approaches require the computation of the weight (or kernel) before the algorithms can run. We show time comparisons for both the weight computation and the algorithms. The first column states the approximate amount of time needed to create the weight or kernel matrices corresponding to

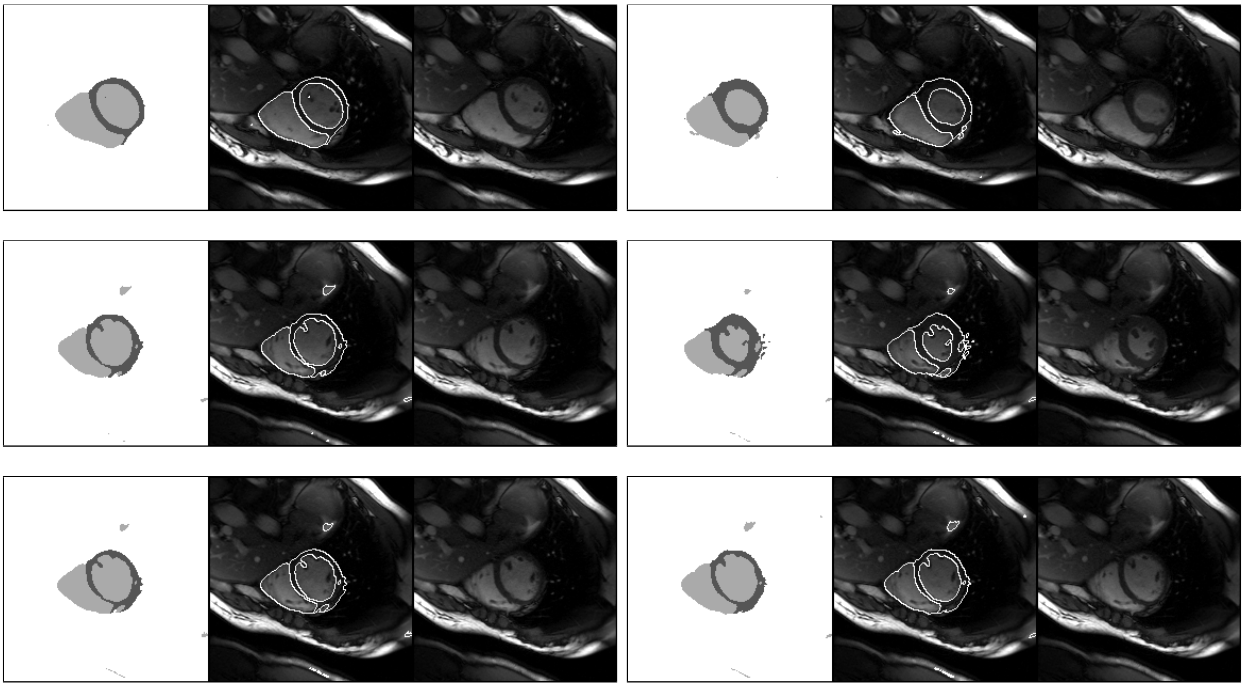
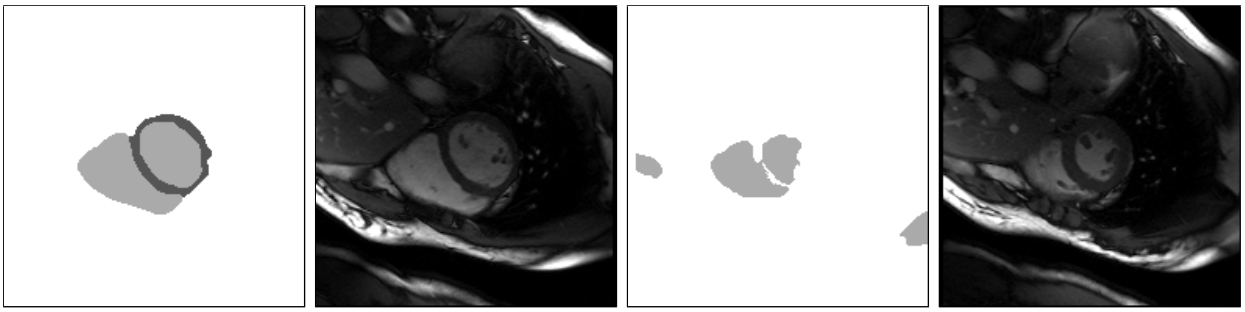


FIGURE 13. Solutions of the RKHS method with parameters $\sigma_1^2 = 1,\ \sigma_2^2 = 5,\ r = 3$ for six images. The left top image is the one from Fig. 12. The segmentation result is shown in the two images next to the given image.

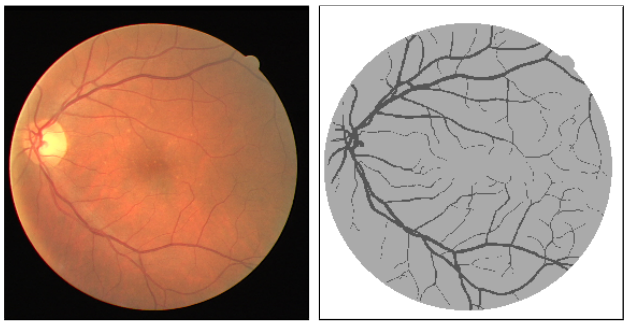


(A) (C)

FIGURE 14. Solutions of 1-Laplacian. $\rho^2 = s = 49,\ \nu^{\text{pho}} = 1\ \nu^{\text{lab}} = 0.1\ r = 3$.

	time for weights comp.	time for actual method
RKHS	24.7s	0.36s
1-Laplacian	32.3s	6.7s
2-Laplacian	28.7s	0.6s
Combined, $p = 1$	49s	196.5s
Combined, $p = 2$	45.5s	33.6s

TABLE 1. Time comparison for single image segmentation using Figure 12 as input.



(A) (B)

FIGURE 15. Input and ground truth.

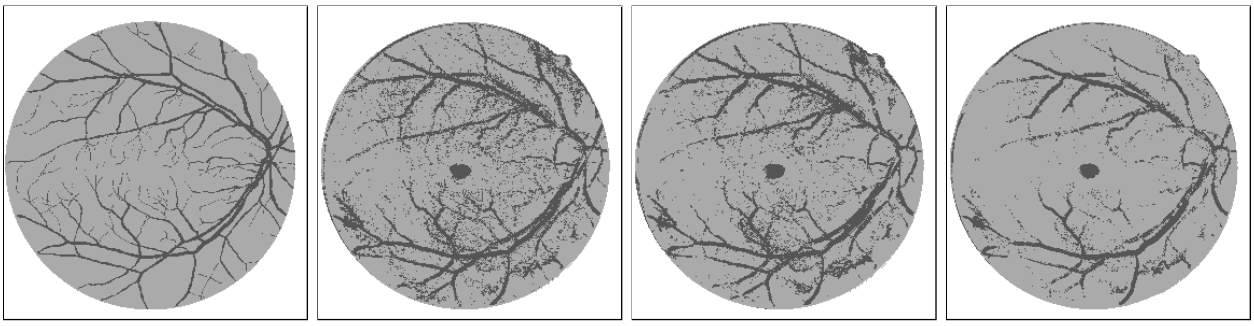
the methods. The second column shows the approximate amount of time the actual minimization process. For the 1-Laplacian and the combined model (27) with $p = 1$, we have used the stopping criterion $\left\|(u^{(r)} - u^{(r-1)})\right\|_2 / \left\|u^{(r-1)}\right\|_2 < 0.001$. This table clearly shows that the RKHS approach is the most efficient one with respect to the computation time.

Figure 15 shows the input and the ground truth images for the test in Figure 18. For the labels L, we have used random elements from the ground truth in Figure 15 (A). In Figure 18, the top rows shows the collection of images similar to Figure 15 (B), but the labels are only taken from Figure 15 (A). Although all the images in the collection stem from iris data they show various differences: notice that in the fourth column the image orientation is opposite from the other images. The second row shows the ground truths. The original images and the ground truth were taken from [32], see also [47]. The third row depicts the result of the RKHS method: although it may look noisy, it keeps a lot of fine details. The fourth and fifth rows show the results of the combined method which are smoother than RKHS results while keeping the details. The last two rows show the results of the p-Laplacian method for $p = 1, 2$. These segmentation results are the smoothest ones. Although very clean, we are missing many small details. Figure 16 shows the zoom-in of some of the images in the second column of Figure 18. Since the RKHS method keeps many of the fine details, one can also post-process these images for further denoising. To show the effect, we simply experimented with a median filter in Figure 17. Of course more sophisticated methods can be applied.

7. Concluding remarks

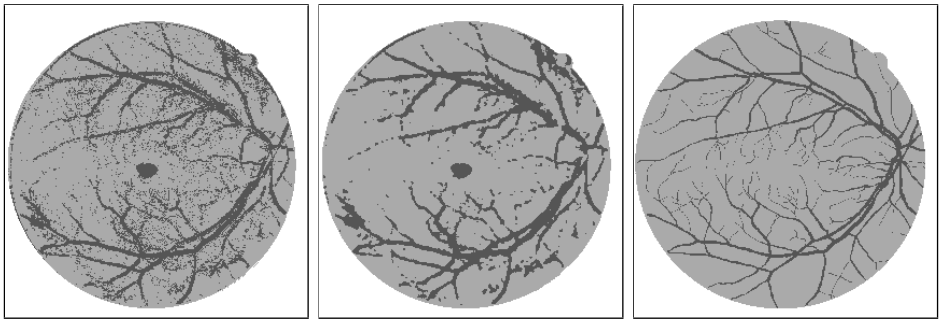
In this paper various methods for multi-class segmentation were developed and studied. The work was inspired by the colorization method based on RKHSs in [26] and has applied the method to multi-class image segmentation. We have explored other approaches as 2-Laplacians which were also considered in [32] without using the Laplacian notation. We have extended this method to p-Laplacian models with $p \geq 1$.

We have observed that p-Laplacians utilize similarity informations between the unlabeled points which can result in a more regular segmentation. Roughly speaking, the 1-Laplacian model gave the smoothest results compared to larger values of $p > 1$. However, there are flexibilities in choosing various weights $w_{i,j}$ and this results in different smoothing effect. The RKHS approach is the most efficient method, only utilizing small amount of labeled pixels. For some complicated images as in Figure 6 and 13, the RKHS method excelled the Laplacian methods. However, since RKHS only utilizes the information from the labeled points, often the result can be less regularized and noisy.



(A) truth (B) RKHS (C) Comb. $p = 1$ (D) 2-Lap

FIGURE 16. These images are zoom-in of the second column results in Figure 18. It is only showing (B) RKHS, where the result keeps the most details. (C) The combined model with $p = 1$, which is more regularized compared to RKHS and the combined model with $p = 2$. (D) 2-Laplacian model which lost fine details. 1-Laplacian even looses more details in this case.



(A) RKHS (B) median filter (C) truth

FIGURE 17. Result of RKHS and its post-processing with median filter of size 1.

We further proposed a combined method which handles this case. Our methods were applied to various collections of medical images.

Acknowledgment. We want to thank the authors of [32] for providing us the input and ground truth images in the Figures 18 and 11.

REFERENCES

[1] Miccai workshop - cardiac MR left ventricle segmentation challenge, 2009. `http://smial.sri.utoronto.ca/LV_Challenge`.

[2] S. Amghibech. Eigenvalues of the discrete p-Laplacian for graphs. *Ars Combinatoria*, 67: 283–302, 2003.

[3] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[4] E. Bae, J. Yuan, and X.-C. Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.

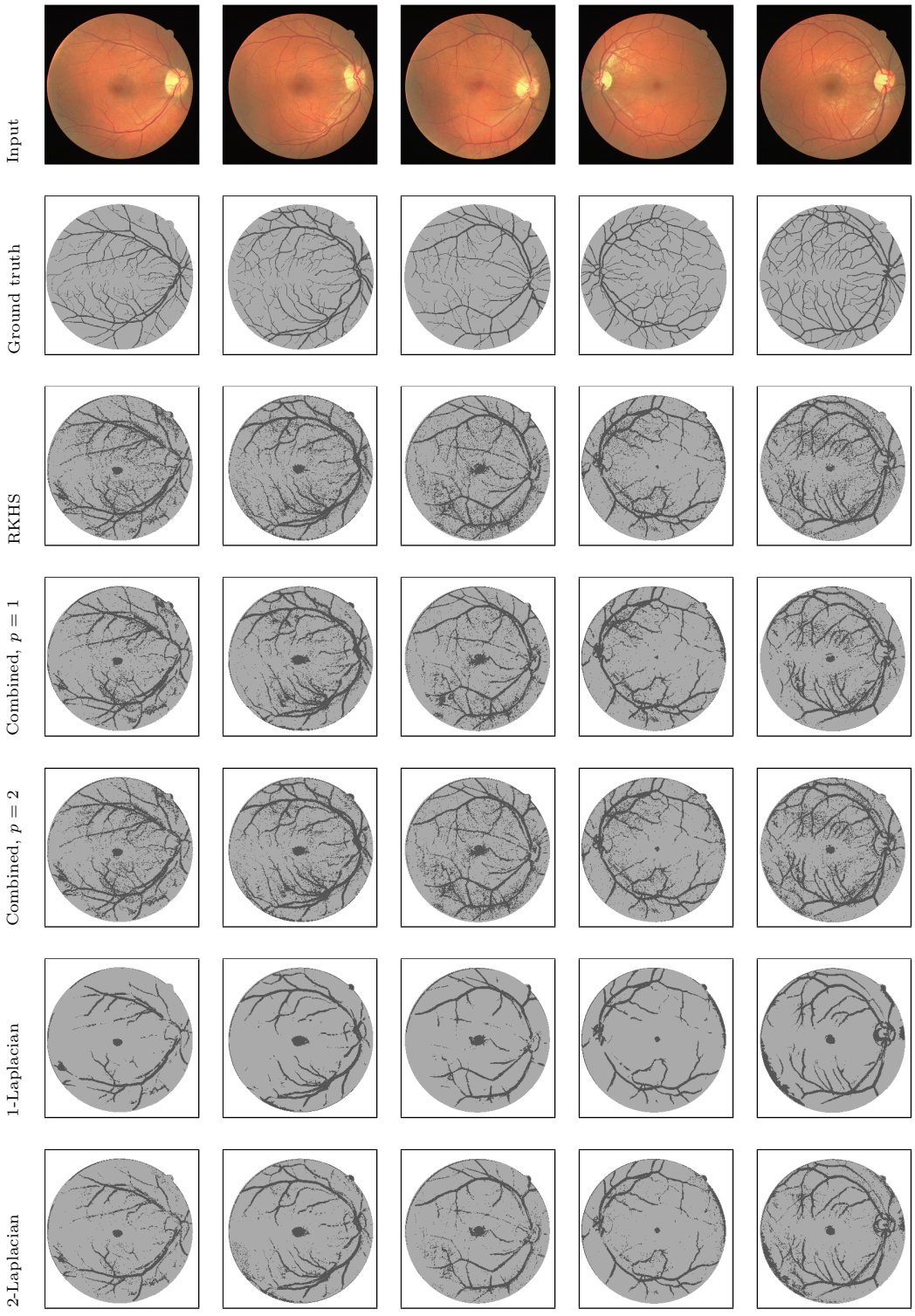


FIGURE 18. Comparison of iris images. In the combined model we used $\mu = 0.0001$ for $p = 1$ and $\mu = 0.01$ for $p = 2$. The parameter ρ_i was computed by (12). $\sigma_1^2 = \frac{1}{2},\ \sigma_2^2 = \infty,\ \nu^{\text{pho}} = 1\ , \nu^{\text{lab}} = 0.8,\ \lambda = 0,\ \ r = 1.$

[5] M. Belkin, P. Nyiogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7: 2399–2434, 2006.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3:1–122, 2011.

[7] X. Bresson, X.-C. Tai, T. Chan, and A. Szlam. Multi-class transductive learning based on 1 relaxations of Cheeger cut and Mumford-Shah-Potts model. CAM-Report 12-03, UCLA, 2012.

[8] T. Brox and J. Weickert. Level set based image segmentation with multiple regions. In *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 415–423. Springer Berlin-Heidelberg, 2004.

[9] T. Bühler and M. Hein. Spectral clustering based on the graph p-Laplacian. In *Proceedings of the 26th International Conference on Machine Learning*, 2009.

[10] G. Burns. Museum of broadcast communications: Encyclopedia of television. World Wide Web electronic publication, 1997. URL `http://www.museum.tv/archives/etv/`.

[11] C. Carmeli and A. T. E. De Vito. Vector valued reproducing kernel Hilbert spaces, integrable functions and Mercer theorem. *Analysis and Applications*, 4(4):377–408, 2006.

[12] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *Journal of Computer Vision*, 22(1):61–79, 1997.

[13] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[14] R. H. Chan. Image segmentation by convex approximation of the Mumford-Shah model. *IEEE Transactions on Image Processing*, to appear.

[15] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 16(2):266–277, 2000.

[16] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21:53 – 94, 2006.

[17] A. Delong and Y. Boykov. Globally optimal segmentation of multi-region objects. In *International Conference on Computer Vision*, pages 285–292, 2009.

[18] J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55:293 – 318, 1992.

[19] A. Erikson, C. Olsson, and F. Kahl. Normalized cuts revisited: a reformulation for segmentation with linear grouping constraints. *Journal of Mathematical Imaging ans Vision*, 39:45–61, 2011.

[20] J. E. Esser. *Primal Dual Algorithms for Convex Models and Applications to Image Restoration, Registration and Nonlocal Inpainting*. PhD thesis, UCLA, 2010.

[21] M. Fornasier. Nonlinear projection digital image inpainting and restoration methods. *Journal of Mathematical Imaging and Vision*, 24(3):359 – 373, 2006.

[22] E. Gelasca, B. Obara, D. Fedorov, K. Kvilekval, and B. Manjunath. A biosegmentation benchmark for evaluation of bioimage analysis methods. *BMC Bioinformatics*, 10:368, 2009.

[23] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[24] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Journal on Multiscale Modeling and Simulation*, 6(2):595 – 630, 2007.

[25] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007. URL `http://authors.library.caltech.edu/7694`.

[26] M. Ha Quang, S. H. Kang, and T. M. Le. Image and video colorization using vector-valued reproducing kernel Hilbert spaces. *Journal of Mathematical Imaging and Vision*, 37:49–65, 2010.

[27] L. Hagen and A. Kahng. Fast spectral methods for ratio cut partitioning and clustering. In *Computer-Aided Design, 1991. ICCAD-91. Digest of Technical Papers., 1991 IEEE International Conference on*, pages 10–13. IEEE, 1991.

[28] M. Hein and T. Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA. In *Proceedings of Advances in Neural Information Processing Systems 23*, 2010.

[29] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1991.

[30] Y. Jung, S. Kang, and J. Shen. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM Applied Mathematics*, 67:1213–1232, 2007.

[31] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[32] Y. N. Law, H. K. Lee, M. K. Ng, and A. M. Yip. A semi-supervised segmentation model for collections of images. *IEEE Transactions on Image Processing*, to appear.

[33] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling with simplex-constrained total variation. In X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods, volume 5567 of LNCS*, volume 5567 of *Lecture Notes in Computer Science*, pages 150–162. Springer, 2009.

[34] A. Levin, D. Lischinski, and Y. Weiss. Colorization using optimization. *Proceedings of the 2004 SIGGRAPH Conference*, 23(3):689–694, 2004.

[35] F. Li, M. Ng, T. Zeng, and C. Shen. A multiphase image segmentation method based on fuzzy region competition. *SIAM Journal on Imaging Sciences*, 3:277–299, 2010.

[36] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17: 177–204, 2005.

[37] D. Mumford and J. Shah. Optimal approximation by piecewise-smooth functions and associated variational problems. *Communications in Pure Applied Mathematics*, 42:577–685, 1989.

[38] G. Pedrick. *Theory of preproducing kernels for Hilbert spaces of vector-valued functions*. University of Kansas Department Of Math., Lawrence, Kansas, 1957.

[39] C. Petitjean and J. Dacher. A review of segmentation methods in short axis cardiac MR images. *Medical Image Analysis*, pages 169–184, 2011.

[40] G. Peyré. Image processing with nonlocal spectral bases. *SIAM Journal on Multiscale Modeling and Simulation*, 7(2):703 – 730, 2008.

[41] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *ICCV Proceedings, LNCS*. Springer, 2009.

[42] L. Schwartz. Sous-espaces Hilbertiens d'espaces vectoriels topoligiques et noyaux associés (noyaux reproduisants). *Journal Analyse Mathematique*, pages 115–256, 1964.

[43] S. Setzer. Operator splittings, Bregman methods and frame shrinkage in image processing. *International Journal of Computer Vision*, 92(3):265–280, 2011.

[44] B. Shafei and G. Steidl. Segmentation of images with separating layers by fuzzy c-means and convex optimization. *Journal of Visual Communication and Image Representation*, 23: 611–621, 2012.

[45] J. Shi and J. Szlam. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[46] V. Sindhwani, P. Niyogi, and M. Belkin. Beyond the point cloud: from transductive to semi-supervised learning. In *Proceedings of the 22nd International Conference on Machine learning*, ICML '05, pages 824–831, New York, NY, USA, 2005. ACM.

[47] J. Staal, M. Abramoff, M. Niemeijer, M. Viergever, and B. Ginneken. Ridge based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4): 501–509, 2004.

[48] J. Stoer and R. Bulirsch. *Numerische Mathematik 2*. Springer, Berlin-Heidelberg-New York, 1990.

[49] J. A. K. Suykens, T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.

[50] A. Szlam and X. Bresson. Total variation and Cheeger cuts. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[51] X.-C. Tai and T. Chan. A survey on multiple level set methods with applications for identifying piecewise constant functions. *International Journal of Numerical Analysis and Modelling*, 1 (1):25–48, 2004.

[52] Z. W. Tu and S. Zhu. Image segmentation by data-driven Markov chain Monte Carlo. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.

[53] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[54] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[55] G. Wahba. *Spline Models for Observational Data*. SIAM, New York, 1990.

[56] T. Welsh, M. Ashikhmin, and K. Mueller. Transferring color to greyscale images. *SIGGRAPH Conference Proceedings*, 21, 2002.

[57] L. Yatziv and G. Sapiro. Fast image and video colorization using chrominance blending. *IEEE Transactions on Image Processing*, 15(5):1120–1129, 2006.

[58] C. Zach, D. Gallup, J.-M.Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. *Vision, Modeling, and Visualization Workshop*, 2008.

[59] D. Zhou and B. Schölkopf. Regularization on discrete spaces. *Pattern Recognition*, pages 361–368, 2005.

[60] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report 08-34*, 2008.

[61] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

[62] W. Zhu, S. H. Kang, and G. Biros. A geodesic active contour based variational model for short axis cardiac-MR image segmentation. *International Journal of Computer Mathematics*, accepted, 2012.

Georgia Institute of Technology, School of Mathematics, Atlanta, USA

Fraunhofer ITWM, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany

University of Kaiserslautern, Department of Mathematics, Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany