Image and Video Colorization Using Vector-Valued Reproducing Kernel Hilbert Spaces

Minh Ha Quang · Sung Ha Kang · Triet M. Le

Abstract Motivated by the setting of reproducing kernel Hilbert space (RKHS) and its extensions considered in machine learning, we propose an RKHS framework for image and video colorization. We review and study RKHS especially in vectorial cases and provide various extensions for colorization problems. Theory as well as a practical algorithm is proposed with a number of numerical experiments.

Keywords Function extension · Vector-valued reproducing kernel Hilbert spaces · Nonlocal kernels · Image colorization · Least square regression · Color inpainting

1 Introduction and Motivation

Let $D \subset \Omega \subset \mathbb{R}^n$ be nonempty sets, and $\mathcal{W}$ a Hilbert space (for now assume $\mathcal{W} = \mathbb{R}^n$). Suppose that we are given an $f : D \to \mathcal{W}$ with f belonging to some function space $X_1(D)$.

M. Ha Quang (✉)
Humboldt University of Berlin, Berlin, Germany
e-mail: minh.ha.quang@staff.hu-berlin.de

S.H. Kang
School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA
e-mail: kang@math.gatech.edu

T.M. Le
Department of Mathematics, Yale University, New Haven, CT, USA
e-mail: triet.le@yale.edu

An important problem in mathematics is to construct an $F : \Omega \to \mathcal{W}$ such that F belongs to some function space $X_2(\Omega)$ with $F \approx f$ on D. The choice of $X_2(\Omega)$ imposes a certain regularity on F. We refer to this problem as an extension problem. Image colorization can be viewed as an instance of this extension problem. The term "colorization" was introduced by Wilson Markle who first processed the gray scale moon image from the Apollo mission. This term was used to describe the process of adding color to grayscale movies or TV broadcasting program [9]. Recently in [22], this colorization problem was motivated by recovering frescoes paintings by A. Mantegna in an Italian church which was destroyed during World War II. There are photos of the full frescoes available in black and white, while only a few real pieces of frescoes with the original colors are remaining. The objective is to reconstruct the original color of the frescoes (image) from the few remaining real pieces of the original (with color) and the full black and white gray scale photos of the frescoes.

In a variational approach, an extension F is computed via minimizing the following functional

$$\inf_{F \in X_2(\Omega)} \{\mathcal{F}(F) = \gamma \mathcal{F}_2(F) + \mathcal{F}_1(F - f)\}, \tag{1}$$

where $\mathcal{F}_1$ and $\mathcal{F}_2$ are functionals defined on $X_1(D)$ and $X_2(\Omega)$ respectively and $\gamma > 0$ is a tuning parameter.

Some variational approaches for image colorization are proposed and mathematically studied in [21–23, 30]. The work in [22] uses a variational functional with a nonlinear function $\mathcal{F}_2$ to fit the grayscale data, and the existence of minimizers is studied in [23]; calibration method is used in [21]; a couple of different variational models using chromaticity and brightness color system are also proposed in [30]. Closely related to these variational methods are partial differential equation based approaches. In [43],

Sapiro recognized the similarity between image colorization and image inpainting [4], and proposed to inpaint the colors by minimizing the difference between the gradient of luminance and the gradient of color. In [56], the authors utilized Dijkstra's shortest path algorithm for fast computation. The idea of adding color to a gray scale image or a movie (even if by hand) is as old as photography itself and many computer-assisted works have been studied in computer vision and graphics literature, such as [25, 27, 41, 55, 56]. More related works using segmentation, matting and probability frameworks can be found in [6, 15, 26, 28, 33, 38, 49, 50] and some other recent works include [19, 34].

A number of other recent approaches for colorization use the similarity information. For example, in [33], the authors recognized that the neighboring pixels in space and time with similar intensities should have similar color, and optimized a quadratic cost function for colorization. In [7], the authors proposed an anisotropic diffusion with an a-priori-defined diffusion direction for conditional color diffusion, where neighborhood filter is proposed for numerical computation. In [35, 40], manifold learning techniques are used. In [35], the authors used locally linear embedding and compared grayscale manifold and color manifold for colorization, and [40] considers geometry of local image patches. In [6], principle component analysis (PCA)-based learning techniques is proposed using probabilistic PCA and regressive PCA.

In machine learning, reproducing kernel Hilbert spaces (RKHS) have recently emerged as a powerful paradigm, both from algorithmic and theoretical perspectives (see for example [52], [45], [46] for comprehensive treatment) . The goal of machine learning is to make inferences and generalizations based on limited sampled data. Thus machine learning algorithms can be very useful for solving the problem of function extension ([17] is one recent RKHS-based approach). In this paper, we will consider RKHS to model $X_2(\Omega)$. We will exploit different choices of the reproducing kernel to obtain different regularity conditions on F. For numerical work, we will employ a version of the well-known regularized least square algorithm in RKHS (see for example [54], [18] for the scalar version).

We start this paper with a brief introduction to RKHS, the abstract theory for which was developed by Aronszajn in [1]. Let D be an arbitrary nonempty set. Let $K : D \times D \to \mathbb{R}$ be a symmetric function, i.e. $K(x, y) = K(y, x)$, satisfying: for any finite set of points $\{x_i\}_{i=1}^N$ in D and real numbers $\{a_i\}_{i=1}^N$,

$$\sum_{i,j=1}^N a_i a_j K(x_i, x_j) \geq 0.$$

K is said to be a **positive definite kernel** on D. Then, there exists a unique Hilbert space $\mathcal{H}_K$ of functions $f : D \to \mathbb{R}$ satisfying:

1. $K_x \in \mathcal{H}_K$ for all $x \in D$, where $K_x(t) = K(x, t)$;
2. span$\{K_x\}_{x \in X}$ is dense in $\mathcal{H}_K$;
3. the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ of $\mathcal{H}_K$ satisfies:

$$f(x) = \langle f, K_x \rangle_{\mathcal{H}_K} \quad \textbf{(reproducing property)},$$

for all $f \in \mathcal{H}_K$ and all $x \in D$. On the dense set span$\{K_x\}$, the inner product is defined by

$$\left\langle \sum_i a_i K_{x_i}, \sum_j b_j K_{y_j} \right\rangle_{\mathcal{H}_K} = \sum_{i,j} a_i b_j K(x_i, y_j).$$

The Hilbert space $\mathcal{H}_K$ is called the **RKHS** with reproducing kernel K and norm $\| \cdot \|_{\mathcal{H}_K}$. The reproducing property means that $\mathcal{H}_K$ is a Hilbert space of functions on D, which are well-defined pointwise. If we apply the Cauchy-Schwarz inequality, we get

$$|f(x)| \leq \|f\|_{\mathcal{H}_K} \|K_x\|_{\mathcal{H}_K} = \sqrt{K(x, x)} \|f\|_{\mathcal{H}_K}.$$

This means that at each point x, the evaluation operator $E_x : f \to f(x)$ is bounded (as an operator from $\mathcal{H}_K$ to $\mathbb{R}$) with norm $\sqrt{K(x, x)}$. *Conversely, if $\mathcal{H}$ is a Hilbert space of functions on D where E_x is bounded for all $x \in D$, then $\mathcal{H}$ is an RKHS.* In fact, by the Riesz Representation Theorem, for each $x \in D$ there is a unique $K_x \in \mathcal{H}$ such that

$$E_x f = f(x) = \langle f, K_x \rangle_{\mathcal{H}}.$$

Then $\mathcal{H}$ is an RKHS with reproducing kernel $K(x, y) = \langle K_x, K_y \rangle_{\mathcal{H}}$. This kernel can also be shown to be unique. There is thus a 1-to-1 correspondence between the category of positive definite kernels on $D \times D$ and that of the RKHS's of functions on D.

The boundedness of the evaluation operators means that in particular, if the kernel is uniformly bounded on D, that is $\kappa = \sup_{x \in D} \sqrt{K(x, x)} < \infty$, then $|f(x)| \leq \kappa \|f\|_{\mathcal{H}_K}$ for all $x \in D$, and thus all functions $f \in \mathcal{H}_K$ are bounded, with $\|f\|_\infty \leq \kappa \|f\|_{\mathcal{H}_K}$. One example is the Gaussian kernel $K(x, y) = \exp(-\frac{\|x-y\|^2}{\sigma^2})$ in $\mathbb{R}^n$ (or similar translation-invariant kernels), where $\kappa = 1$ (or a finite constant, respectively).

Remark 1 The preceding property does not hold in general. If $K(x, y) = \langle x, y \rangle^d$, $d \geq 1$, $d \in \mathbb{N}$, then $K(x, x)$ on $\mathbb{R}^n$ is unbounded, and the functions in $\mathcal{H}_K$ are unbounded, being polynomials of degree d, even if for each fixed $x \in \mathbb{R}^n$ the operator $E_x : \mathcal{H}_K \to \mathbb{R}$ is bounded. In this paper, we will focus on translation-invariant kernels, which induce RKHS of bounded functions.

Remark 2 Note that L^2 spaces are not RKHS in general because they are spaces of equivalence classes of functions which are the same almost everywhere, whereas functions in RKHS are defined everywhere.

Next we would like to present some well-known examples of kernels and RKHS in $\mathbb{R}^n$.

The most popular kernel in practice is the Gaussian kernel $K(x, y) = \exp(-\frac{\|x-y\|^2}{\sigma^2})$ on $\mathbb{R}^n$. One of its generalizations is $K(x, y) = \exp(-\frac{\|x-y\|^p}{\sigma^2})$, which was shown by Schoenberg [44] to be positive definite iff $0 \leq p \leq 2$. We will discuss the cases $p = 1$ and $p = 2$ in more detail below.

The Sobolev space $H^s(\mathbb{R}^n)$, $s > n/2$, is a RKHS. Recall that $f \in H^s(\mathbb{R}^n)$ if

$$
\begin{aligned}
\|f\|^2_{H^s(\mathbb{R}^n)} &= \|(I - \Delta)^{s/2} f\|^2_{L^2(\mathbb{R}^n)} \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} |(1 + |\xi|^2)^{s/2} \hat{f}(\xi)|^2 \, d\xi < \infty,
\end{aligned}
$$

with inner product in $H^s(\mathbb{R}^n)$ defined by

$$
\langle f, g \rangle_{H^s(\mathbb{R}^n)} := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} [(1 + |\xi|^2)^s \hat{f}(\xi) \overline{\hat{g}(\xi)}].
$$

Here $\hat{f}(\xi)$ denotes the Fourier transform of f if $f \in L^1(\mathbb{R}^n)$, which is defined by

$$
\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi, x \rangle} dx.
$$

If $f \in L^2(\mathbb{R}^n)$, then $\hat{f}(\xi)$ denotes the Fourier-Plancherel transform of f (see [29], Chap. 13). Since $s > n/2$, each $f \in H^s(\mathbb{R}^n)$ is continuous and $\hat{f} \in L^1(\mathbb{R}^n)$. By the Fourier Inversion Theorem,

$$
\begin{aligned}
f(x) &= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \hat{f}(\xi) e^{i\langle x, \xi \rangle} d\xi \\
&= \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} (1 + |\xi|^2)^s \hat{f}(\xi) \overline{\left[\frac{e^{-i\langle \xi, x \rangle}}{(1 + |\xi|^2)^s} \right]} d\xi.
\end{aligned}
$$

Under the assumption $s > n/2$, let $k(x) = \frac{1}{(2\pi)^n} \widehat{\frac{1}{(1+|\xi|^2)^s}}(x)$ and $K_x(y) = k(x - y)$, then $\hat{K}_x(\xi) = \frac{e^{-i\langle \xi, x \rangle}}{(1 + |\xi|^2)^s}$ and $K_x(y) \in H^s(\mathbb{R}^n)$. It follows then that for all $f \in \mathcal{H}^s(\mathbb{R}^n)$ and all $x, y \in \mathbb{R}^n$,

$$
f(x) = \langle f, K_x \rangle_{H^s(\mathbb{R}^n)},
$$

$$
\langle K_x, K_y \rangle_{H^s(\mathbb{R}^n)} = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \frac{e^{-i\langle \xi, x-y \rangle}}{(1 + |\xi|^2)^s} d\xi = k(x - y).
$$

Thus $H^s(\mathbb{R}^n)$ is a RKHS with the reproducing kernel $K(x, y) = k(x - y)$.

Remark 3 The reproducing kernel K above is the kernel that corresponds to the Bessel potential $(I - \Delta)^{-s}$. (See [48] Chap. V, Sect. 3 where an explicit formula is given.)

An explicit example for this type of kernels is the Laplacian kernel $K(x, y) = \exp(-a|x - y|) = k(x - y)$, $a > 0$, on $\mathbb{R}^n$. Here $k(x) = e^{-a|x|}$, with

$$
\hat{k}(\xi) = \frac{C(n)a}{(a^2 + |\xi|^2)^{(n+1)/2}}, \qquad C(n) = 2^n \pi^{\frac{n-1}{2}} \Gamma\left(\frac{n+1}{2} \right),
$$

where Γ denotes the gamma function, defined by

$$
\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) dx, \quad 0 < a < \infty.
$$

The RKHS induced by K is

$$
\begin{aligned}
\mathcal{H}_K = \Bigg\{ f \in C_0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) : \\
\|f\|^2_{\mathcal{H}_K} = \frac{1}{(2\pi)^n} \frac{1}{aC(n)} \\
\times \int_{\mathbb{R}^n} (a^2 + |\xi|^2)^{\frac{n+1}{2}} |\hat{f}(\xi)|^2 d\xi < \infty \Bigg\},
\end{aligned}
\tag{2}
$$

which is a Sobolev space of order $s = \frac{n+1}{2}$. Consider the Gaussian kernel $K(x, y) = \exp(-\frac{|x-y|^2}{\sigma^2})$ again. On $\mathbb{R}^n$, the RKHS it induces is

$$
\begin{aligned}
\mathcal{H}_K = \Bigg\{ f \in C_0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) : \\
\|f\|^2_{\mathcal{H}_K} = \frac{1}{(2\pi)^n (\sigma\sqrt{\pi})^n} \int_{\mathbb{R}^n} e^{\frac{\sigma^2|\xi|^2}{4}} |\hat{f}(\xi)|^2 d\xi \\
< \infty \Bigg\}.
\end{aligned}
\tag{3}
$$

Note that for an $f \in \mathcal{H}_K$ with the latter choice of K, $\hat{f}(\xi)$ decays exponentially, which shows that $\frac{\partial^k f}{\partial x^k} \in L^2(\mathbb{R}^n)$ for all $k \geq 0$, hence f is in $C^\infty(\mathbb{R}^n)$. The space $\mathcal{H}_K$ here can be viewed as a Sobolev space of infinite order.

We see that the Laplacian kernel given in (2) provides less smoothing effects than the Gaussian kernel (3). The smoothing properties of functions in RKHS, as seen in these examples, make them particularly suitable for regularization problems. In practice, there are two ways to define a RKHS. The first is to define a kernel K explicitly and then derive the form of the norm and its smoothing properties. The second, as in [54] for smoothing splines problems, is to define the norm first and then compute the kernel. Each approach has its own advantage: the former tends to be more efficient computationally since the kernel has a closed form, the latter tends to be much clearer analytically. More related literature can be found in [3, 37, 42, 53, 54] and the numerous references they contain.

The main contribution of this paper is to apply the theory of RKHS of vector-valued functions and RKHS-based

function extension to image and video colorization. By using the RKHS approach, the kernel (nonlocal) can be chosen appropriately for various applications. We will also give comparisons with non-local diffusion using neighborhood similarities.

This paper is organized as follows: in Sect. 2, we discuss the extension of RKHS to the vector-valued case, in Sect. 3, we present the application to colorization, then various numerical results will be presented in Sect. 4.

2 Vector-Valued Reproducing Kernel Hilbert Spaces

The study of RKHS has been extended to vector-valued functions and further developed and applied in machine learning (see [10, 13, 36] and references therein). In the following, denote by D a nonempty set, $\mathcal{W}$ a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{W}}$, $\mathcal{L}(W)$ the Banach space of bounded linear operators on $\mathcal{W}$.

Let $\mathcal{W}^D$ denote the vector space of all functions $f : D \to \mathcal{W}$. A function $K : D \times D \to \mathcal{L}(\mathcal{W})$ is said to be an **operator-valued positive definite kernel** if for each pair $(x, y) \in D \times D$, $K(x, y) \in \mathcal{L}(\mathcal{W})$ is a self-adjoint operator and

$$\sum_{i,j=1}^{N} \langle w_i, K(x_i, x_j) w_j \rangle_{\mathcal{W}} \geq 0 \qquad (4)$$

for every finite set of points $\{x_i\}_{i=1}^{N}$ in D and $\{w_i\}_{i=1}^{N}$ in $\mathcal{W}$, where $N \in \mathbb{N}$. As in the scalar case, given such a K, there exists a unique $\mathcal{W}$-valued RKHS $\mathcal{H}_K$ with reproducing kernel K (a proof can be found in [13]). The construction of the space $\mathcal{H}_K$ proceeds as follows. For each $x \in D$ and $w \in \mathcal{W}$, we form a function $K_x w = K(., x) w \in \mathcal{W}^D$ defined by

$$(K_x w)(y) = K(y, x) w \quad \text{for all } y \in D.$$

Consider the set

$$\mathcal{H}_0 = \text{span}\{K_x w \mid x \in D, \; w \in \mathcal{W}\} \subset \mathcal{W}^D.$$

For $f = \sum_{i=1}^{N} K_{x_i} w_i$, $g = \sum_{i=1}^{N} K_{y_i} z_i \in \mathcal{H}_0$, we define

$$\langle f, g \rangle_{\mathcal{H}_K} = \sum_{i,j=1}^{N} \langle w_i, K(x_i, y_j) z_j \rangle_{\mathcal{W}}.$$

Taking the closure of $\mathcal{H}_0$ gives us the Hilbert space $\mathcal{H}_K$. The reproducing property is

$$\langle f(x), w \rangle_{\mathcal{W}} = \langle f, K_x w \rangle_{\mathcal{H}_K} \quad \text{for all } f \in \mathcal{H}_K. \qquad (5)$$

As in the scalar case, applying Cauchy-Schwarz inequality gives

$$|\langle f(x), w \rangle_{\mathcal{W}}| \leq \sqrt{\|K(x, x)\|} \, \|f\|_{\mathcal{H}_K} \|w\|_{\mathcal{W}}.$$

Thus for each $x \in D$, each $w \in \mathcal{W}$, the evaluation operator $E_{x|w} : f \to \langle f(x), w \rangle_{\mathcal{W}}$ is bounded as a linear operator from $\mathcal{H}_K$ to $\mathbb{R}$. As in the scalar case, the converse is true by the Riesz Representation Theorem.

Let $K_x : \mathcal{W} \to \mathcal{H}_K$ be the linear operator with $K_x w$ defined as above, then

$$\|K_x w\|_{\mathcal{H}_K}^2 = \langle K(x, x) w, w \rangle_{\mathcal{W}} \leq \|K(x, x)\| \|w\|_{\mathcal{W}}^2,$$

which implies that

$$\|K_x : \mathcal{W} \to \mathcal{H}_K\| \leq \sqrt{\|K(x, x)\|},$$

so that K_x is a bounded operator for each $x \in D$. Let $K_x^* : \mathcal{H}_K \to \mathcal{W}$ be the adjoint operator of K_x, then from (5), we have

$$f(x) = K_x^* f \quad \text{for all } x \in D, f \in \mathcal{H}_K. \qquad (6)$$

From this we deduce that for all $x \in D$ and all $f \in \mathcal{H}_K$,

$$\|f(x)\|_{\mathcal{W}} \leq \|K_x^*\| \|f\|_{\mathcal{H}_K} \leq \sqrt{\|K(x, x)\|} \|f\|_{\mathcal{H}_K},$$

that is for each $x \in D$, the evaluation operator $E_x : \mathcal{H}_K \to \mathcal{W}$ defined by $E_x f = K_x^* f$ is a bounded linear operator. In particular, if $\kappa = \sup_{x \in D} \sqrt{\|K(x, x)\|} < \infty$, then $\|f\|_{\infty} = \sup_{x \in D} \|f(x)\|_{\mathcal{W}} \leq \kappa \|f\|_{\mathcal{H}_K}$ for all $f \in \mathcal{H}_K$. In this paper, we will be concerned with kernels for which $\kappa < \infty$.

2.1 Extension of Vector-Valued Functions

Let $D \subset \Omega$ be closed subsets in a complete separable metric space, and $\mathcal{W}$ be a separable Hilbert space. Our aim is to extend a function $f : D \to \mathcal{W}$ to a function $F : \Omega \to \mathcal{W}$, which is as close to f as possible on D, and at the same time reasonably well-behaved on the larger set Ω. We will describe two algorithms here for function extension using RKHS, the first for a general set D, and the second specifically for the case D is discrete.

2.1.1 Function Extension via Eigenfunctions—the Spectral Algorithm

In [17], Coifman and Lafon discussed scalar-valued function extension using eigenfunctions of the given reproducing kernel. We will extend their approach given in [17] to the vector-valued case here.

Suppose that $K : \Omega \times \Omega :\to \mathcal{L}(\mathcal{W})$ is a positive definite kernel, then K induces a RKHS $\mathcal{H}_K(\Omega)$ of functions $g : \Omega \to \mathcal{W}$. Let further μ be a finite Borel measure on D. Let $L_{\mu}^2(D; \mathcal{W})$ be the space of measurable functions $f : D \to \mathcal{W}$ such that $\|f\|_{\mathcal{W}}^2$ is μ-integrable, with norm

$$\|f\|_{L_{\mu}^2(D; \mathcal{W})} = \left(\int_D \|f(x)\|_{\mathcal{W}}^2 d\mu(x) \right)^{1/2}.$$

Assumption 1 We shall assume throughout the paper that $K(x, x) \in \mathcal{L}(\mathcal{W})$ is a compact operator for each $x \in \Omega$ and that $\kappa = \sup_{x \in \Omega} \sqrt{\|K(x, x)\|} < \infty$.

First consider the integral operator $L_{K,D} : L_\mu^2(D; \mathcal{W}) \to L_\mu^2(D; \mathcal{W})$ defined by

$$L_{K,D} f(x) = \int_D K(x, y) f(y) d\mu(y).$$

Here we have adopted the notation of [18] and [47], where this operator shows its crucial role in learning theory. By Assumption 1, this operator is symmetric, positive, and compact (we refer to [12, 13] for the detailed treatment) so that the eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ with $\lim_{k \to \infty} \lambda_k = 0$. Let $\{\phi_k\}_{k=1}^\infty$ be the corresponding eigenfunctions of $L_{K,D}$, then ϕ_k's can be normalized to form an orthonormal basis for $L_\mu^2(D; \mathcal{W})$.

Assumption 2 We shall assume throughout the remainder of this section (Sect. 2.1.1) that $K_x w \in C(\Omega; \mathcal{W})$ for all $x \in \Omega$, $w \in \mathcal{W}$, where $C(\Omega; \mathcal{W})$ denotes the space of continuous functions mapping Ω into $\mathcal{W}$. If Ω is discrete, then this assumption is not needed.

The following is Mercer's theorem for the vector-valued case (see [12]).

Theorem 1 *Let K be a positive definite kernel satisfying Assumptions 1 and 2. Furthermore, assume also that μ has support $\operatorname{supp}(\mu) = D$, then*

$$K(x, y) = \sum_{k=1}^\infty \lambda_k \phi_k(x) \otimes \phi_k(y), \tag{7}$$

for all $x, y \in D$, where the series converges in the operator norm of $\mathcal{L}(\mathcal{W})$. Here $w_1 \otimes w_2$ denotes the rank-one operator in $\mathcal{L}(\mathcal{W})$, with

$$(w_1 \otimes w_2)v = \langle v, w_2 \rangle_{\mathcal{W}} w_1 \quad \text{for } w_1, w_2, v \in \mathcal{W}.$$

A consequence of Mercer's theorem is that

$$\mathcal{H}_K(D) = \operatorname{Im}(L_{K,D}^{1/2})$$

$$= \left\{ f \in L_\mu^2(D; \mathcal{W}) \,\middle|\, \right.$$

$$\left. \sum_{k=1, \lambda_k > 0}^\infty \frac{\langle f, \phi_k \rangle_{L_\mu^2(D;\mathcal{W})}}{\lambda_k} < \infty \right\}.$$

In particular, this shows that $L_{K,D} f \in \mathcal{H}_K(D)$ for all $f \in L_\mu^2(D; \mathcal{W})$ and that $\{\sqrt{\lambda_k} \phi_k\}_{k=1}^\infty$ is an orthonormal basis for $\mathcal{H}_K(D)$.

To perform an extension from D to Ω, note that by replacing D by Ω, we have $L_{K,\Omega} f \in \mathcal{H}_K(\Omega)$ for all $f \in L_\mu^2(\Omega; \mathcal{W})$. By considering a function $f \in L_\mu^2(D; \mathcal{W})$ as one in $L_\mu^2(\Omega; \mathcal{W})$ with support in D, we have the following well-defined integral operator $L_K : L_\mu^2(D; \mathcal{W}) \to \mathcal{H}_K(\Omega)$, with

$$L_K f(x) = \int_D K(x, y) f(y) d\mu(y), \tag{8}$$

for every $x \in \Omega$. In our context, it defines a pointwise function $L_K f$ on the larger domain Ω from an L_μ^2 function f defined on the smaller domain D, i.e. L_K is an extension operator.

Lemma 1 *The adjoint operator $L_K^* : \mathcal{H}_K(\Omega) \to L_\mu^2(D; \mathcal{W})$ is the restriction operator from $\mathcal{H}_K(\Omega)$ to $L_\mu^2(D; \mathcal{W})$. I.e. in the $L_\mu^2(D; \mathcal{W})$ sense of equality,*

$$L_K^* F = f,$$

for all $F \in \mathcal{H}_K(\Omega)$, where $f = F|_D$.

Proof For every $g \in L_\mu^2(D; \mathcal{W})$, we have

$$\langle L_K^* F, g \rangle_{L_\mu^2(D;\mathcal{W})}$$

$$= \langle F, L_K g \rangle_{\mathcal{H}_K(\Omega)}$$

$$= \langle F, \int_D K(., y) g(y) d\mu(y) \rangle_{\mathcal{H}_K(\Omega)}$$

$$= \int_D \langle F, K(., y) g(y) \rangle_{\mathcal{H}_K(\Omega)} d\mu(y)$$

$$= \int_D \langle F(y), g(y) \rangle_{\mathcal{W}} d\mu(y)$$

(by the reproducing property)

$$= \langle f, g \rangle_{L_\mu^2(D;\mathcal{W})}.$$

This shows that $L_K^* F = f$ as L_μ^2 functions. $\qquad\square$

Given an $f \in L_\mu^2(D; \mathcal{W})$, we are interested in extending f to $F \in \mathcal{H}_K(\Omega)$ by minimizing the following functional (see [47] for a related scalar version):

$$\inf_{F \in \mathcal{H}_K(\Omega)} \|f - L_K^* F\|_{L_\mu^2(D;\mathcal{W})}^2 + \gamma \|F\|_{\mathcal{H}_K(\Omega)}^2, \tag{9}$$

for some $\gamma > 0$. This is a standard least square Tikhonov regularization problem in Hilbert spaces, which has a unique minimizer F_γ satisfying the normal equation (see for example [20])

$$(L_K L_K^* + \gamma I) F_\gamma = L_K f \iff F_\gamma = (L_K L_K^* + \gamma I)^{-1} L_K f. \tag{10}$$

As in the scalar case in [17], for $\lambda_k > 0$, using (8), we can extend the eigenfunction ϕ_k on D to Φ_k on Ω by

$$
\begin{aligned}
\Phi_k(x) &= \frac{L_K \phi_k(x)}{\lambda_k} \\
&= \frac{1}{\lambda_k} \int_D K(x, y) \phi_k(y) d\mu(y), \quad \text{for } x \in \Omega.
\end{aligned}
\tag{11}
$$

For $x \in D$, we have $\Phi_k(x) = \phi_k(x)$. The extension operation gives $L_K \phi_k = \lambda_k \Phi_k$ and $L_K L_K^* \Phi_k = \lambda_k \Phi_k$ as pointwise functions. The restriction operation gives $L_K^* \Phi_k = \phi_k$ and $L_K^* L_K \phi_k = \lambda_k \phi_k$ in the L_μ^2 sense. These relations imply that $\langle \Phi_k, \Phi_j \rangle_{\mathcal{H}_K(\Omega)} = \frac{\delta_{jk}}{\lambda_k}$ so that $\{\sqrt{\lambda_k} \ \Phi_k\}_{k=1}^\infty$ form an orthonormal system in $\mathcal{H}_K(\Omega)$.

Using eigenfunction expansion, the following result is immediate.

Lemma 2 *Let* $f = \sum_{k=1, \lambda_k > 0}^\infty a_k \phi_k \in L_\mu^2(D; \mathcal{W})$ *with* $\sum_{k=1}^\infty a_k^2 < \infty$. *Then the minimizer* F_γ *for* (9) *is given by*

$$
F_\gamma = \sum_{k=1, \lambda_k > 0}^\infty \frac{\lambda_k}{\lambda_k + \gamma} a_k \Phi_k.
\tag{12}
$$

Moreover, we have

$$
\|L_K^* F_\gamma - f\|_{L_\mu^2(D; \mathcal{W})}^2 = \sum_{k=1, \lambda_k > 0}^\infty \frac{\gamma^2}{(\lambda_k + \gamma)^2} a_k^2,
\tag{13}
$$

$$
\|F_\gamma\|_{\mathcal{H}_K(\Omega)}^2 = \sum_{k=1, \lambda_k > 0}^\infty \frac{\lambda_k}{(\lambda_k + \gamma)^2} a_k^2.
\tag{14}
$$

Remark 4 As a multiscale extension, as in [17], we can consider extending f on D to F_δ on Ω, for some $\delta > 0$, where $F_\delta = \sum_{\lambda_k > \delta} a_k \Phi_k$. Thus we have

$$
\|L_K^* F_\delta - f\|_{L_\mu^2(D; W)}^2 = \sum_{\lambda_k \le \delta} a_k^2.
\tag{15}
$$

The multiscale property in (13) is determined by the parameter γ instead of δ as in (15).

Remark 5 In practice, the computation of Φ_k may be numerically difficult when λ_k is small. One should combine equations (11) and (12) and compute directly

$$
F_\gamma(x) = \sum_{k=1}^\infty \frac{a_k}{\lambda_k + \gamma} \int_D K(x, y) \phi_k(y) d\mu(y).
\tag{16}
$$

This formula also takes care of the case $\lambda_k = 0$, when Φ_k is not defined. Note that for ϕ_k with $\lambda_k = 0$, $L_K \phi_k(x) = 0$ for all $x \in D$. In general, eigenfunctions corresponding to very

small eigenvalues tend to be highly oscillatory and their extensions may not be numerically reliable, so one may consider excluding them. For some analytic formulas of kernel spectra, see [37].

We now have the following algorithm for extending $f \in L_\mu^2(D; W)$ to the larger domain Ω using the kernel K and its induced RKHS of functions on Ω.

Function Extension—Spectral Algorithm

(1) Compute the eigenvalues and eigenfunctions $\{(\lambda_k, \phi_k)\}$ of $L_{K,D}$.
(2) Compute the expansion coefficients a_k's of f in the basis $\{\phi_k\}$.
(3) Compute F_γ using equation (16) (or alternatively F_δ as mentioned above).

2.1.2 Regularized Least Square Algorithm

Let us now describe an algorithm with the true flavor of RKHS-based machine learning. Suppose now that D is discrete and is given by $D = \{x_i\}_{i=1}^m$. We are given a set of values $\mathbf{z} = \{(x_i, w_i = f(x_i))\}_{i=1}^m$, $w_i \in \mathcal{W}$. In this case, a big advantage of the kernel method is that the extension of f will be explicitly expressed in terms of basis functions in $\mathcal{H}_K(\Omega)$ at the points x_i's. Here we compute

$$
F_\gamma = \arg \min_{F \in \mathcal{H}_K(\Omega)} \frac{1}{m} \sum_{i=1}^m \|F(x_i) - w_i\|_{\mathcal{W}}^2 + \gamma \|F\|_{\mathcal{H}_K(\Omega)}^2.
\tag{17}
$$

This is the vector-valued version of the well-known regularized least square algorithm in RKHS (see for example [18], [54]). To cast this into the standard Tikhonov form (10), we can consider an approach as in [11]. Consider the sampling operator $S_\mathbf{x} : \mathcal{H}_K(\Omega) \to \mathcal{W}^m$ defined by $S_\mathbf{x}(F) = (F(x_1), \ldots, F(x_m))$. By definition, we have for any $F \in \mathcal{H}_K(\Omega)$ and $\mathbf{w} = (w_1, \ldots, w_m) \in \mathcal{W}^m$,

$$
\begin{aligned}
\langle S_\mathbf{x} F, \mathbf{w} \rangle_{\mathcal{W}^m} &= \sum_{i=1}^m \langle S_{x_i} F, w_i \rangle_{\mathcal{W}} = \sum_{i=1}^m \langle F, S_{x_i}^* w_i \rangle_{\mathcal{H}_K(\Omega)} \\
&= \sum_{i=1}^m \langle F, K_{x_i} w_i \rangle_{\mathcal{H}_K(\Omega)} \\
&= \left\langle F, \sum_{i=1}^m K_{x_i} w_i \right\rangle_{\mathcal{H}_K(\Omega)}.
\end{aligned}
$$

It follows that the adjoint operator $S_\mathbf{x}^* : \mathcal{W}^m \to \mathcal{H}_K(\Omega)$ is given by $S_\mathbf{x}^* \mathbf{w} = S_\mathbf{x}^*(w_1, \ldots, w_m) = \sum_{i=1}^m K_{x_i} w_i$, and the operator $S_\mathbf{x}^* S_\mathbf{x} : \mathcal{H}_K(\Omega) \to \mathcal{H}_K(\Omega)$ is given by $S_\mathbf{x}^* S_\mathbf{x} F =$

$\sum_{i=1}^{m} K_{x_i} F(x_i)$. We can now cast expression (17) into the form

$$F_\gamma = \arg \min_{F \in \mathcal{H}_K(\Omega)} \frac{1}{m} \| S_{\mathbf{x}} F - \mathbf{w} \|_{\mathcal{W}^m}^2 + \gamma \| F \|_{\mathcal{H}_K(\Omega)}^2.$$

This problem has a unique solution, given by

$$F_\gamma = (S_{\mathbf{x}}^* S_{\mathbf{x}} + m\gamma I)^{-1} S_{\mathbf{x}}^* \mathbf{w} = \left(\frac{1}{m} S_{\mathbf{x}}^* S_{\mathbf{x}} + \gamma I \right)^{-1} \frac{1}{m} S_{\mathbf{x}}^* \mathbf{w}. \tag{18}$$

Proposition 1 *The unique solution F_γ of problem (17) has the form*

$$F_\gamma = \sum_{i=1}^{m} K_{x_i} a_i, \quad \text{with } F_\gamma(x) = \sum_{i=1}^{m} K(x, x_i) a_i,$$

where the vectors $a_i \in \mathcal{W}$ satisfy the m linear equations

$$\sum_{j=1}^{m} K(x_i, x_j) a_j + m\gamma a_i = w_i$$

for $1 \leq i \leq m$.

Proof Expression (18) is equivalent to $(S_{\mathbf{x}}^* S_{\mathbf{x}} + m\gamma I) F_\gamma = S_{\mathbf{x}}^* \mathbf{w}$. Using the definition of $S_{\mathbf{x}}$ and $S_{\mathbf{x}}^*$, we obtain

$$\sum_{i=1}^{m} K_{x_i} F_\gamma(x_i) + m\gamma F_\gamma = \sum_{i=1}^{m} K_{x_i} w_i.$$

This implies that

$$F_\gamma = \sum_{i=1}^{m} K_{x_i} \frac{w_i - F_\gamma(x_i)}{m\gamma} = \sum_{i=1}^{m} K_{x_i} a_i,$$

where $a_i = \frac{w_i - F_\gamma(x_i)}{m\gamma}$. We now have

$$F_\gamma(x_i) = \sum_{j=1}^{m} (K_{x_j} a_j)(x_i) = \sum_{j=1}^{m} K(x_i, x_j) a_j.$$

It follows that $a_i = \frac{w_i - \sum_{j=1}^{m} K(x_i, x_j) a_j}{m\gamma}$, or equivalently $\sum_{j=1}^{m} K(x_i, x_j) a_j + m\gamma a_i = w_i$. □

This result was first reported in [36] via a different derivation. Our derivation follows directly from expression (18) and is a natural generalization of the scalar case in [18].

Example 1 Consider the scalar case $\mathcal{W} = \mathbb{R}$. Then, $F_\gamma(x) = \sum_{i=1}^{m} a_i K(x_i, x)$, where $\mathbf{a} = (a_i, \ldots, a_m)$ is the solution of the system of linear equations $(K[\mathbf{x}] + \gamma m I)\mathbf{a} = \mathbf{w}$, where $K[\mathbf{x}]$ is the $m \times m$ matrix defined by $K[\mathbf{x}]_{ij} = K(x_i, x_j)$ (see [18]).

2.1.3 Comparisons between the Two Algorithms

From the theoretical viewpoint, the Spectral Algorithm is more general, since it is for D either continuous or discrete. Let us consider the case D is discrete, of size m, with μ being the uniform probability measure on D. Then the two algorithms are the same analytically, since they both solve the same minimization problem. In fact, we have then $L_K = \frac{1}{m} S_{\mathbf{x}}^*$ and $L_K L_K^* = \frac{1}{m} S_{\mathbf{x}}^* S_{\mathbf{x}}$. From the numerical viewpoint, the Regularized Least Square Algorithm (hereafter referred to as Least Square) is simpler to implement and should be expected to be more stable. For example, for $\mathcal{W} = \mathbb{R}$, it involves solving a well-conditioned system of linear equations, in contrast to the eigenvalues and eigenfunctions that need to be found and extended in the case of the Spectral Algorithm.

2.2 Vector-Valued Diagonal Kernel

Let D be an arbitrary nonempty subset of $\mathbb{R}^m$, and let $\mathcal{W} = \mathbb{R}^n$. In the following, all vectors in $\mathbb{R}^n$ will be treated as column vectors, unless stated otherwise. One example of operator-valued kernels $K : D \times D \to \mathbb{R}^n$ can be defined as

$$K(x, y) = \text{diag}(k_1(x, y), \ldots, k_n(x, y)), \tag{19}$$

where each $k_i(x, y)$ is a positive definite real-valued kernel. In this case, the RKHS induced by K can be described explicitly in terms of those induced by the scalar components of K. As we shall see below, in this case, the solution of the minimization problem (17) has a particularly simple representation. Each basis function in $\mathcal{H}_K$ is defined by

$$K_x w(y) = K(x, y)w = (w_1 k_1(x, y), \ldots, w_n k_n(x, y)),$$

for any $w \in \mathbb{R}^n$ and any $x, y \in D$. Let D be closed and μ be a finite Borel measure on D, with support $\text{supp}(\mu) = D$. Assume that $\kappa = \max_{1 \leq i \leq n} \kappa_i < \infty$ where $\kappa_i = \sup_{x \in D} k_i(x, x)$. We have the Hilbert space

$$L_\mu^2(D; \mathbb{R}^n) = \left\{ f = (f_1, \ldots, f_n) : D \to \mathbb{R}^n \mid \right.$$
$$\| f \|_{L_\mu^2(D; \mathbb{R}^n)}^2$$
$$\left. = \sum_{i=1}^{n} \int_D |f_i(x)|^2 d\mu(x) < \infty \right\},$$

and the integral operator $L_K : L_\mu^2(D; \mathbb{R}^n) \to L_\mu^2(D; \mathbb{R}^n)$, defined by

$$L_K f(x) = \int_D K(x, y) f(y) d\mu(y)$$
$$= \left(\int_D k_i(x, y) f_i(y) d\mu(y) \right)_{i=1}^{n}$$
$$= (L_{k_i} f_i(x))_{i=1}^{n},$$

which is self-adjoint, compact, and positive (since each component L_{k_i} is).

Lemma 3 *Assume that ϕ^i is an eigenfunction of L_{k_i} with corresponding eigenvalue λ^i, then $\psi = (0, \ldots, \phi^i, \ldots, 0)$ is an eigenfunction of L_K corresponding to the same eigenvalue.*

Proof This follows from $L_K \psi = (0, \ldots, L_{k_i} \phi^i, \ldots, 0) = (0, \ldots, \lambda^i \phi^i, \ldots, 0) = \lambda^i \psi$. □

The following theorem is a version of Mercer's theorem adapted to the setting of a diagonal kernel.

Theorem 2 *Assume that each k_i is continuous (the continuity assumption is not needed if D is discrete). Let $\{\lambda_k^i, \phi_k^i\}_{k=1}^{\infty}$ be an $L_\mu^2(D)$ orthonormal spectrum of $L_{k_i}: L_\mu^2(D) \to L_\mu^2(D)$. For each $1 \leq i \leq n$ and $k \in \mathbb{N}$ fixed, let $\psi_k^i = (0, \ldots, \phi_k^i, \ldots, 0)$, with $\psi_k^i(x) \in \mathbb{R}^n$ considered as a column vector for each $x \in D$. Then the system $\{\{\lambda_k^i, \psi_k^i\}_{k=1}^{\infty}\}_{i=1}^{n}$ form an orthonormal spectrum of $L_K: L_\mu^2(D; \mathbb{R}^n) \to L_\mu^2(D; \mathbb{R}^n)$. Furthermore,*

$$K(x, y) = \sum_{i=1}^{n} \sum_{k=1}^{\infty} \lambda_k^i \psi_k^i(x) \psi_k^i(y)^T,$$

where for each pair $(x, y) \in D \times D$, the series converges in the operator norm of $\mathcal{L}(\mathbb{R}^n)$.

Proof Let $e_i = (0, \ldots, 1, \ldots, 0) \in \mathbb{R}^n$ be the column vector with only one nonzero entry at the ith position. Then $e_i e_i^T$ is the matrix whose only nonzero entry is at the (i, i) position, so that $K(x, y) = \sum_{i=1}^{n} k_i(x, y) e_i e_i^T$. Mercer's theorem for the scalar case states that $k_i(x, y) = \sum_{k=1}^{\infty} \lambda_k^i \phi_k^i(x) \phi_k^i(y)$, therefore $k_i(x, y) e_i e_i^T = \sum_{k=1}^{\infty} \lambda_k^i \psi_k^i(x) \psi_k^i(y)^T$, from which the series summation follows. □

Corollary 1 *For*

$$f = \sum_{i=1}^{n} \sum_{k=1}^{\infty} a_k^i \psi_k^i \in \mathcal{H}_K,$$

$$g = \sum_{i=1}^{n} \sum_{k=1}^{\infty} b_k^i \psi_k^i \in \mathcal{H}_K,$$

the inner product in $\mathcal{H}_K$ is given by

$$\langle f, g \rangle_{\mathcal{H}_K} = \sum_{i=1}^{n} \sum_{k=1}^{\infty} \frac{a_k^i b_k^i}{\lambda_k^i}.$$

Proof Let $e_i = (0, \ldots, 1, \ldots, 0) \in \mathbb{R}^n$. By definition

$$K_x e_i(y) = K(x, y) e_i = \left(\sum_{j=1}^{n} \sum_{k=1}^{\infty} \lambda_k^j \psi_k^j(y) \psi_k^j(x)^T \right) e_i$$

$$= \sum_{k=1}^{\infty} \lambda_k^i \psi_k^i(y) \phi_k^i(x)$$

for all $x, y \in D$, from which we have

$$K_x e_i = \sum_{k=1}^{\infty} \lambda_k^i \phi_k^i(x) \psi_k^i.$$

For $w = \sum_{i=1}^{n} w^i e_i \in \mathbb{R}^n$,

$$K_x w = \sum_{i=1}^{n} w^i \sum_{k=1}^{\infty} \lambda_k^i \phi_k^i(x) \psi_k^i,$$

so that the Hilbert space $\mathcal{H}_K$ is

$$\mathcal{H}_K = \overline{\text{span}\{K_x e_i : x \in D, 1 \leq i \leq n\}}.$$

For $x, y \in D$ by definition we have

$$\langle K_x e_i, K_y e_i \rangle_{\mathcal{H}_K} = \langle e_i, K(x, y) e_i \rangle_{\mathbb{R}^n} = k_i(x, y)$$

$$= \sum_{k=1}^{\infty} \lambda_k^i \phi_k^i(x) \phi_k^i(y).$$

From this we infer that if $f = \sum_{k=1}^{\infty} a_k^i \psi_k^i$ and $g = \sum_{k=1}^{\infty} b_k^i \psi_k^i$ are in $\mathcal{H}_K$, then $\langle f, g \rangle_{\mathcal{H}_K} = \sum_{k=1}^{\infty} \frac{a_k^i b_k^i}{\lambda_k^i}$. The general formula follows similarly. □

Corollary 2 *The Hilbert space $\mathcal{H}_K$ is the direct sum of n orthogonal complementary subspaces:*

$$\mathcal{H}_K = \oplus_{i=1}^{n} \mathcal{H}_{K,i},$$

where $\mathcal{H}_{K,i} = \overline{\text{span}\{K_x e_i : x \in D\}}$.

Remark 6 Corollary 2 also follows directly from the definition of the inner product in $\mathcal{H}_K$, as we have

$$\langle K_x e_i, K_y e_j \rangle_{\mathcal{H}_K} = \langle e_i, K(x, y) e_j \rangle_{\mathbb{R}^n} = \delta_{ij} k_i(x, y)$$

for all $x, y \in D$. However, the eigendecomposition arising from Mercer's theorem is of interest in its own right and is useful if we wish to use the Spectral Algorithm.

For the Least Square Algorithm, for $f = (f_1, \ldots, f_n) \in \mathcal{H}_K$ and $w_i = (w_i^1, \ldots, w_i^n) \in \mathcal{W}$, we have

$$\| f(x_i) - w_i \|_{\mathcal{W}}^2 = \sum_{j=1}^{n} |f_j(x_i) - w_i^j|^2,$$

$$\| f \|_{\mathcal{H}_K}^2 = \sum_{j=1}^{n} \| f_j \|_{\mathcal{H}_{K,j}}^2.$$

It follows that the minimization problem (17) becomes

$$F_\gamma = \arg \min_{f \in H_K(\Omega)} \sum_{j=1}^n \left(\frac{1}{m} \sum_{i=1}^m |f_j(x_i) - w_i^j|^2 \right.$$

$$\left. + \gamma \|f_j\|_{\mathcal{H}_{K,j}(\Omega)}^2 \right).$$

It is clear then that $F_\gamma = (F_\gamma^i)_{i=1}^n$, where

$$F_\gamma^j = \arg \min_{f_j \in H_{K,j}(\Omega)} \left(\frac{1}{m} \sum_{i=1}^m |f_j(x_i) - w_i^j|^2 \right.$$

$$\left. + \gamma \|f_j\|_{H_{K,j}(\Omega)}^2 \right).$$

Thus in the diagonal case, the vector-valued minimizer is obtained by solving the minimization problems for all the scalar components separately, using the same regularization parameter γ.

With the theory on vector-valued kernels and RKHS explored in this section, we set up the function extension for the colorization problem in the following section.

3 Colorization Using Vector-Valued RKHS

Let $\Omega \subset \mathbb{R}$ be the image domain, and $D \subset \Omega$ be a nonempty subset of Ω. Colorization typically assumes that the complete black and white (gray scale) image is given in the entire domain Ω. We denote this gray scale image as $g : \Omega \to \mathbb{R}$. Let the small patches where the color is given be the domain D, and f be the given color image, i.e. $f : D \to \mathbb{R}^3$. We consider color images as RGB (red, green, blue channels) which is a 3 dimensional vector. The objective is to colorize the whole domain Ω: to find $F : \Omega \to \mathbb{R}^3$ such that $F|_D \approx f$, i.e. an extension from $f : D \to \mathbb{R}^3$ to $F : \Omega \to \mathbb{R}^3$.

From the variational approach, we consider the following functional for colorization,

$$\inf_F \{\gamma \|F\|_{\mathcal{H}_K(\Omega)}^2 + \|F - f\|_{L^2(D;\mathbb{R}^3)}^2\}, \tag{20}$$

with $\mathcal{H}_K(\Omega)$ being the RKHS with the reproducing kernel K depending on the grayscale image g. In particular, we would like to explore the kernel which utilizes the non-local similarity information in a multiscale fashion. For example for each $x, y \in \Omega$ and some $t > 0$ and $0 < p \leq 2$, the scalar kernel function $k : \Omega \times \Omega \to \mathbb{R}$ is defined as

$$k(x, y) = \exp\left(-\frac{|g(x) - g(y)|^p}{4t} \right), \quad \forall x, y \in \Omega. \tag{21}$$

Here two pixels are similar if they have similar grayscale levels and the parameter $t > 0$ acts as a weight factor influencing the degree of similarity.

We will also consider kernels inspired by those in [8], defined as

$$k(x, y) = \exp\left(\frac{-(G_r * |g(x - \cdot) - g(y - \cdot)|)^p}{4t} \right), \tag{22}$$

where

$$G_r * |g(x - \cdot) - g(y - \cdot)|$$

$$= \left[\frac{1}{|B_r|} \int_{B_r(x)} |g(x - z) - g(y - z)|^2 \, dz \right]^{1/2}. \tag{23}$$

In the multiscale case, let $\{g_1, \ldots, g_m\}$ be a multiscale representation of g. Here, we pick a few meaningful discrete scales g_i, $i = 1, \ldots, m$. For each $x, y \in \Omega$ and some $t_1, \ldots, t_m > 0$, the kernel $k : \Omega \times \Omega \to \mathbb{R}$ is defined as

$$k(x, y) = \exp\left(-\sum_{i=1}^m \left[\frac{|g_i(x) - g_i(y)|^p}{4t_i} \right] \right), \quad \forall x, y, \in \Omega, \tag{24}$$

or a similar variation as in (22).

Since the color image is a vector function, we consider the vector-valued kernel: $K : \Omega \times \Omega \to \mathbb{R}^3$ which depends on the gray-scale image g,

$$K(x, y) := \text{diag}(k(x, y), k(x, y), k(x, y)) = k(x, y)I_{3\times3}, \tag{25}$$

where $I_{3\times3}$ is an identity matrix of size 3 by 3, and $k(x, y)$ is as in (21) or (24). For different applications, one may want to define $K = (k_1, k_2, k_3)$, where k_i is different for each color channel. This is our proposed colorization model, and in the following we present the details of how to compute F numerically.

3.1 Numerical Algorithm

Following the theory developed in Sect. 2, in particular 2.1, we solve the minimizing functional (20) using the Least Square and Spectral Algorithms. In this paper, even within the general framework of operator-valued kernels, we mostly consider the diagonal vector-valued kernels assuming that the three channels, red, green, and blue, are independent to each other and that each channel can be computed separately.

Let $D = \{x_1, \ldots, x_m\}$ be a discrete domain and $D \subset \Omega$. Then similar to (17), we are interested in the solution

$$F_\gamma = \arg \min_{F \in \mathcal{H}_K(\Omega)} \frac{1}{m} \sum_{i=1}^m \|F(x_i) - f(x_i)\|_{\mathbb{R}^3}^2 + \gamma \|F\|_{\mathcal{H}_K(\Omega)}^2.$$

Least-Square Algorithm

- Input: gray-scale image g, domain D, and the given color f.
- Compute the partial kernels K_D and K_{cD} of the full kernel K:

  1. Get $\vec{B}(x)$:
     $\vec{B}(x)$ is the neighborhood vector for each $x \in \Omega$, storing the intensity values of the neighborhood patch of size $(2r + 1) \times (2r + 1)$ centered at x.
  2. Get K_{cD} and K_D:
     Use $\vec{B}(x)$ to compute the kernel using (25), via (21), (24) or others.
     $K_{cD}(x, y)$ is the $NM \times m$ matrix for $x \in \Omega$ and $y \in D$, and $K_D(x, y)$ is $m \times m$ for $x, y \in D$.

- Solve the linear system $(K_D + \gamma m I_{m \times m}) A^j = f^j$.
  Here f^j is the jth channel of the given color $f \in \mathbb{R}^3$ on D, as a column vector.
  A^j is the $m \times 1$ coefficient vector representing the jth-channel.
- Compute the explicit solution F_γ:
  Compute for the j channel $F_\gamma^j = K_{cD} A^j$.

Using Proposition 1, the explicit solution can be computed as

$$F_\gamma = \sum_{i=1}^m K(x, x_i) a_i \qquad (26)$$

where a_i's are the solutions of

$$\left\{ \sum_{j=1}^m K(x_i, x_j) a_j \right\} + m\gamma a_i = f(x_i). \qquad (27)$$

For practical computation, notice that the index i (or j) is from 1 to m which is the size of the domain D with the given color. We need only to compute two kernel matrices here: $K_D(x, y)$, where $(x, y) \in D \times D$, for solving the system of linear equations, and $K_{cD}(x, y)$, where $(x, y) \in \Omega \times D$, for evaluating the result. We introduce these new notations to clearly show the difference in the domains and simplify the notation in the algorithm. The kernel matrix K_D is of size $m \times m$ and K_{cD} is of size $NM \times m$, where the size of the discrete domain Ω is $N \times M$. Notice that this significantly reduces the computational cost, since there is no need to compute the $NM \times NM$ full kernel matrix for colorization. In addition, by using the Least Square Algorithm in RKHS, the solution (26) is computed explicitly without any iteration which also helps to reduce the computational cost.

We have also experimented with using the Spectral Algorithm discussed in Sect. 2.1.1 and found that the numerical results are quite similar to the Least Square Algorithm. We will present the various numerical experiments in Sect. 4.

4 Various Applications in Colorization

We present various numerical results in this section. Let Ω be a discrete image domain with size $N \times M$, and D be

the region where the color is given with cardinality m. The given image is denoted by $F_o : \Omega \to \mathbb{R}^3$: where $F_o |_D = f$, and for $x \in \Omega \backslash D$, $F_o^1(x) = F_o^2(x) = F_o^3(x)$, i.e. all three channels are equal and represent the gray scale. Let $(2r + 1) \times (2r + 1)$ be the size of a square patch used to represent the neighborhood of a point: for each $x \in \Omega$ and a positive integer $l = (2r + 1)^2$, $\vec{x} = (x_1, \ldots, x_l) \in \mathbb{R}^l$ as in (21) and (22). When $r = 0$, this represents using only the intensity value at the point. We experimented with different kernels such as

$$k(x, y) = \exp\left(-\frac{|g(\vec{x}) - g(\vec{y})|^p}{2\sigma_1(2r+1)^p} \right) \exp\left(-\frac{|x - y|^p}{\sigma_2 \rho^p} \right), \quad (28)$$

here ρ is $\sqrt{N^2 + M^2}$. We experimented with $0 < p \le 2$ and various σ_1 and σ_2 values. When $0 < p < 2$, the results can be sharper and less blurry compared to $p = 2$. This is consistent with the smoothing properties of the kernels as described by the mathematical theory: for $p = 2$, the RKHS consist of functions which are smoother than those when $0 < p < 2$, as we saw in Sect. 1, thus the resulting images tend to be more blurry.

4.1 Texture Colorization and Color Transfer

One of the benefits of using the RKHS function extension is in its flexibility of choice of kernel, and as seen in Sect. 6, this approach is related to nonlocal diffusion. These methods can perform very well for texture colorization. Figures 1 and 2 show typical results using the proposed model. Figure 1 shows a complicated textured image with only 2% of real color given (randomly chosen), and it gives a good colorization result. The initial color points are chosen randomly, which shows a possibility to perform color compression. Figure 2 shows another example of real image colorization. Notice that less than 0.5% of color is given from the original image and the colorization result is realistic.

Fig. 1 (**a**) The given image. (**b**) The colorization result with $r = 10$, $p = 1.5$ $\sigma_1 = 0.4$, $\sigma_2 = 10$. Here a small set of color, less than 2% compared to the size of the image, is given **randomly**
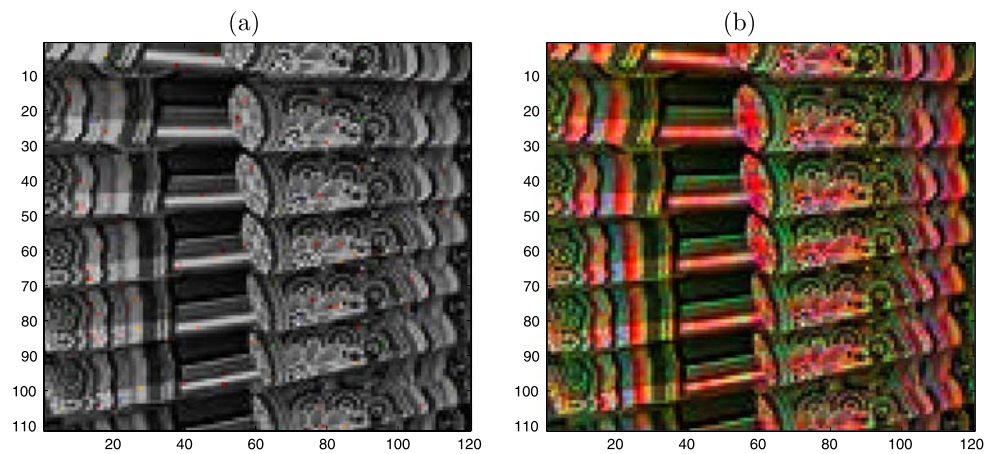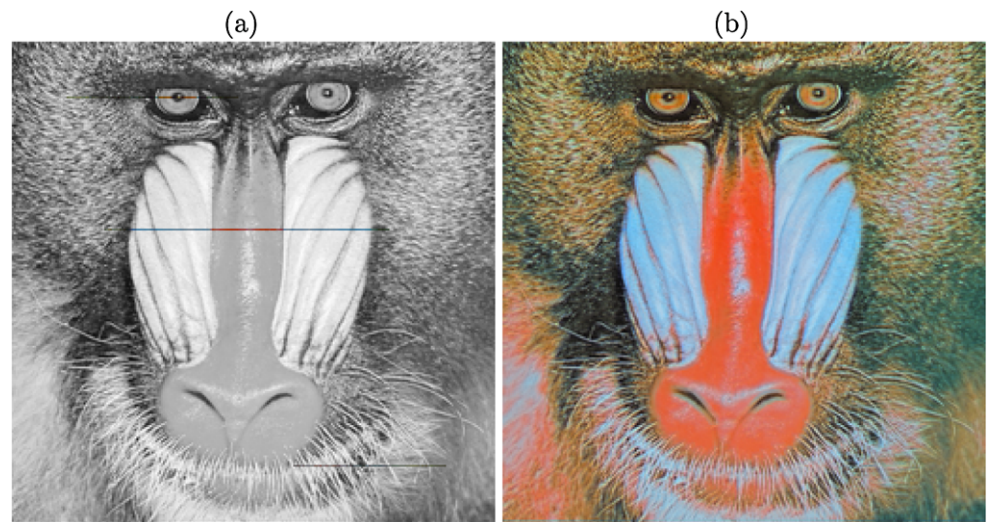

Fig. 2 (**a**) The given image. (**b**) The colorization result with $p = 1$, $r = 2$, $\sigma_1 = 0.5$, $\sigma_2 = 10$. Less than 0.5% of color is given: around the left eye, middle of the nose, and right bottom corner. The small D makes the numerical computation efficient, and the colorization result is realistic

Note also due to the small size of given colored region, the computation is quite efficient numerically (see the discussion after (27)).

Another interesting application of colorization is color transfer [41, 55]. From a given reference image, the color information is transferred to a different gray-scale image. In [55], the authors proposed to match the luminance of two images, using texture information as a guide. Our work along this line is an extension from [41] where the authors matched two colored images. Our model is somewhat different from ordinary colorization methods in that a typical diffusion-based colorization will fail to diffuse the color from one image to another. However, since we propose using RKHS function extension, as long as we define the relation (via Kernel) between the given colored image and any other image, color transfer becomes a natural extension.

Figure 3 shows one such an example and it shows this method can be generalized to video sequence colorization. One of the easiest generalization is to extend the domain. Let F_1 be a given image defined on Ω with a small region of color, and F_2 be another image defined on Ω totally black

and white. One can consider the new image $F_o = [F_1, F_2]$, where F_1 and F_2 are next to each other with the image size of $N \times 2M$ (in fact, the size of F_2 does not have to be the same as that of F_1). Then, apply the vectorial RKHS function extension on the extended image domain. In Fig. 3, from the given image (a) with partial color information, image (a) and image (d) are both colorized at the same time. Notice that images (a) and (d) are quite different and yet the method gives a reasonable colorization.

4.2 Cartoon Colorization and Color Transfer

One typical application of colorization is cartoon colorization as considered in [38, 55]. The proposed method can be also applied to piece-wise constant images, not only images with complicated textures. Figure 4 shows an example of cartoon image colorization given tiny regions of colors. For this example, we used the intensity of each pixel, that is $r = 0$, with $p = 2$, $\sigma_1 = 0.001$, and $\sigma_2 = 10$. Notice that among many regions, only four dab of colors are given (white background, one pink, yellow and green), and this
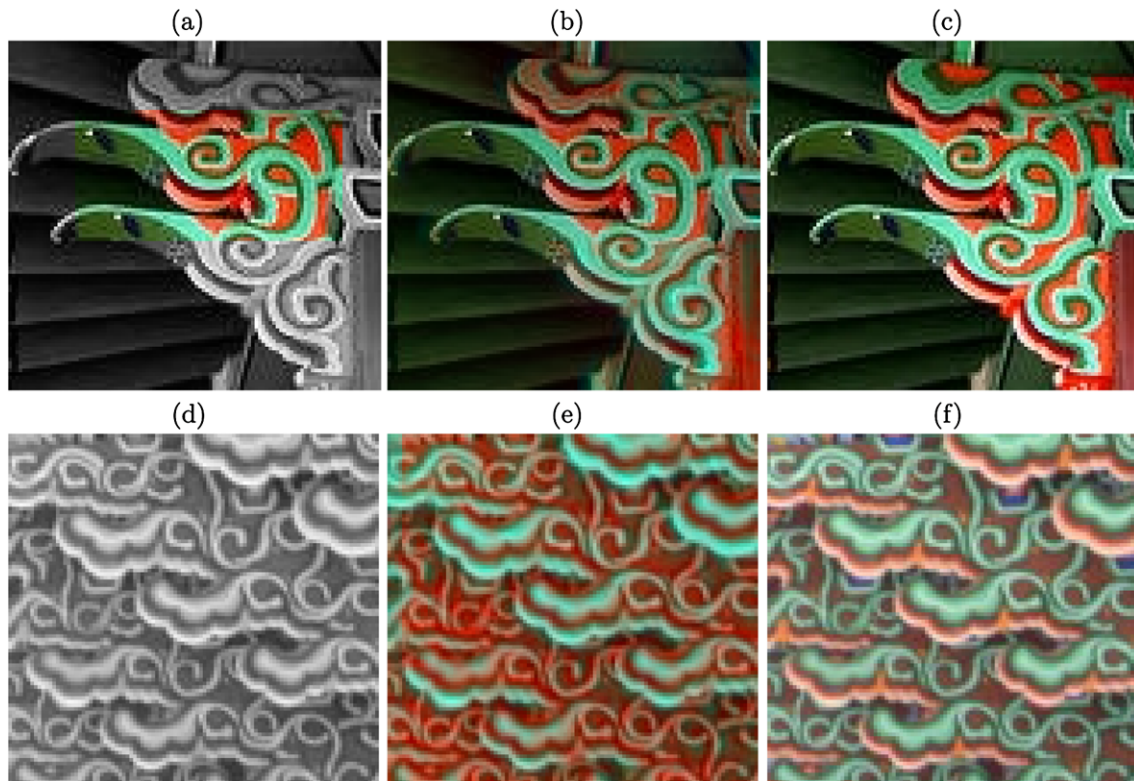
(a)   (b)   (c)

(d)   (e)   (f)

Fig. 3 Image (**a**) and image (**d**) are the two given images: only some part of image (**a**) is given as color and image (**d**) is totally gray scale. (**b**) The colorization result of image (**a**). (**c**) The true image of (**a**). (**e**) The colorization result of image (**d**). (**f**) The true image of (**d**). Both images are colorized at the same time using $r = 4$, $p = 1$ $\sigma_1 = 1$, $\sigma_2 = 10$. Even if the two images are quite different the colorization results are reasonable

(a)   (b)

Fig. 4 (**a**) The given grayscale image with small regions with color. (**b**) The colorization result using the proposed method ($r = 0$, $p = 2$, $\sigma_1 = 0.001$, and $\sigma_2 = 10$). Notice that only one dab of color is given for each four different colors, and all the regions are colorized according to the intensity similarity

proposed method is able to color all the regions which have similar intensity.

This proposed method can be applied to color transfer as before. Figure 5 shows such a result of color transfer, colorizing both image (a) and image (b) at the same time. Notice the new small flower without any color information in image (b), which is also colorized automatically by blending the given color information. From the brightness of this new flower, it is reasonable to guess that its color could be close to yellow yet different. The new color is given by the extension function via a weighted mixture of the given colors.

5 Chromaticity-Brightness Model and Stereographic Projection

Since we are dealing with color images, we mention that RGB is not the only color system available (see [25]). Typical linear models such as RGB (Red, Green and Blue channels) and CMY (Cyan, Magenta and Yellow) are widely used, but in standard color TV broadcasting Luminance separated color systems such as YIQ (Luminance, Hue and Saturation) are used. For digital video, YCbCr (Luminance, two color-difference components) is widely used. There are also nonlinear color representations closer to human color perception such as HSV (Hue, Saturation and Value) and in mathematical settings, color images can also be treated as 3-dimensional vectorial functions [5] as well as tensor products of different color components such as Chromaticity and Brightness (CB). Many related literature can be found in [14, 31, 39, 51].

For our proposed model, we also considered the nonlinear color model Chromaticity and Brightness. From a given
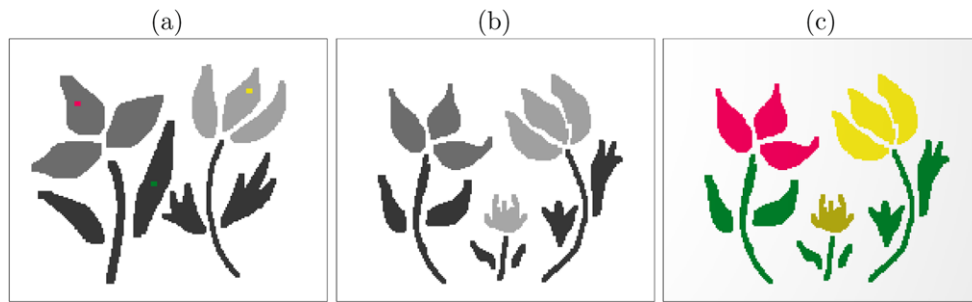
(a)   (b)   (c)

Fig. 5 (**a**) The same as Fig. 4(a), the given image. (**b**) A totally gray scale image. (**c**) The colorization result using extended image domain: $r = 0$, $p = 2$, $\sigma_1 = 0.001$, and $\sigma_2 = 10$. Notice that the new small flower, without any color information, is also colorized automatically by blending the given color information. From the brightness intensity of this new flower, it is reasonable to guess that its color could be close to Yellow but is different, i.e. this method gives a weighted mixture of the given colors

RGB color image $F(x) = (r(x), g(x), b(x))$, the brightness is typically defined by $B(x) = \sqrt{r^2 + g^2 + b^2}$, and the chromaticity by $C(x) = \frac{F(x)}{B(x)}$, i.e. $\|C(x)\|_{\ell^2} = 1$ and $C : \Omega \to S^2$. The motivation is from [14], where the authors found that color denoising is best achieved when it is treated in the Chromaticity and Brightness model (or similarly, when the color is represented by one component and the brightness separately to give added flexibility for keeping details).

In this setting of colorization, we assume the brightness B is given, and we compute the kernel K from B. Since we consider the image F as the multiplication of the brightness B and the Chromaticity C, we only need to find $C : \Omega \to S^2$ from the given color $c : D \to S^2$ in the region D. The difficulty comes from the fact that the Chromaticity component lies on a unit sphere, and the set of S^2-valued functions is not a vector space, so the RKHS model cannot be directly applied.

To resolve this issue, we apply the stereographic projection which maps points from S^2 one-to-one onto the extended complex plane $\mathbb{C} \cup \{\infty\}$. This allows us to get rid of the normalized constraint $\|C\|_{\ell^2} = 1$ of the Chromaticity and directly work on $\mathbb{R}^2$ space. Since the color values are all nonnegative, and to keep the symmetry of the colors, we apply the stereographic projection with the projection point being $(-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}})$ and the projection plane being $x + y + z = 0$. Then, from the sphere $x^2 + y^2 + z^2 = 1$ onto the plane $X + Y + Z = 0$, the projection is given by:

$$X = \frac{3x - (x + y + z)}{\sqrt{3}(x + y + z + \sqrt{3})},$$

$$Y = \frac{3y - (x + y + z)}{\sqrt{3}(x + y + z + \sqrt{3})},$$

$$Z = \frac{3z - (x + y + z)}{\sqrt{3}(x + y + z + \sqrt{3})}.$$

The inverse projection from the plane $X + Y + Z = 0$ onto the sphere $x^2 + y^2 + z^2 = 1$ is:

$$x = \frac{2\sqrt{3}X + 1 - (X^2 + Y^2 + Z^2)}{\sqrt{3}(1 + X^2 + Y^2 + Z^2)},$$

$$y = \frac{2\sqrt{3}Y + 1 - (X^2 + Y^2 + Z^2)}{\sqrt{3}(1 + X^2 + Y^2 + Z^2)},$$

$$z = \frac{2\sqrt{3}Z + 1 - (X^2 + Y^2 + Z^2)}{\sqrt{3}(1 + X^2 + Y^2 + Z^2)}.$$

Therefore, from the given color $c : D \to S^2$, we stereographically project this image onto $x + y + z = 0$ to get $c_p(x) : D \to \mathbb{R}^2$. Then we solve (20)

$$\inf_{C_p \in \mathcal{H}_K(\Omega)} \{\gamma \|C_p\|^2_{\mathcal{H}_K(\Omega)} + \|C_p - c_p\|^2_{L^2(D;\mathbb{R}^3)}\},$$

for two channels, to get the extension $C_p(x)$. Project back this $C_p(x)$ onto S^2 to get $C(x) : \Omega \to S^2$, then the colorization result becomes $F = BC$.

Figure 6 shows this approach. Compared to using RGB vector, especially if three channels are all independently treated, the Chromaticity and Brightness model can give much sharper results. This is due to keeping the sharp brightness information and only recovering the color (see [30], which also uses Chromaticity and Brightness model for colorization).

Another good feature of using Chromaticity and Brightness model is the automatic color blending in the color space. Figure 7 shows such a result blending the color naturally. Notice that by using RKHS function extension, the results are more realistic compared to other methods which assume homogeneous colorization (cf. [30]). Some colorization approaches use a look-up table or a color palette for more natural colorization [25, 27, 56].
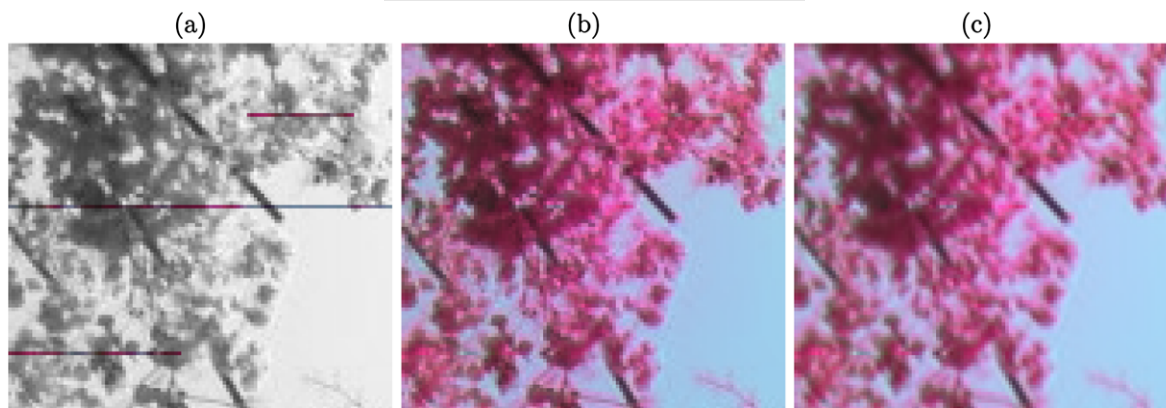
Fig. 6 (**a**) The given image. (**b**) The colorization result using Chromaticity and Brightness model via Stereographic Projection. (**c**) The colorization result using RGB channel. For both experiments $p = 1$, $r = 2$, $\sigma_1 = 0.5$, and $\sigma_2 = 10$ are used. Notice the sharper detail recovery in image (**b**)
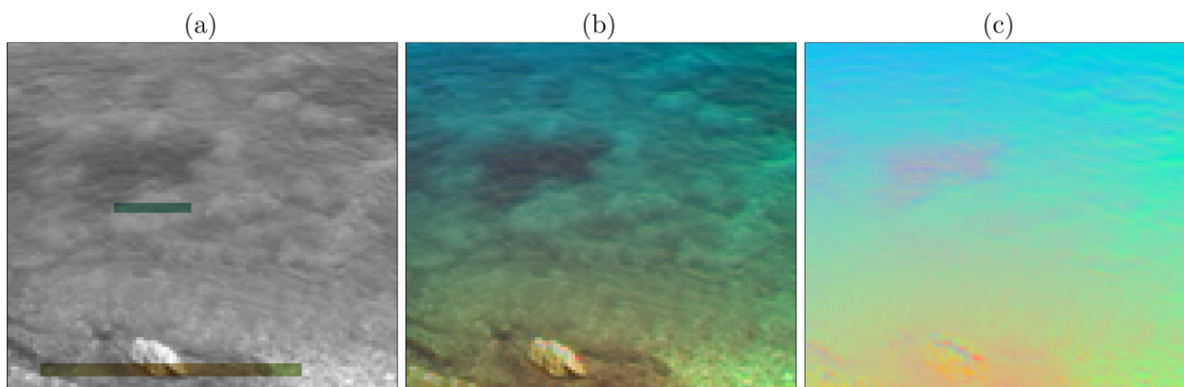

Fig. 7 (**a**) The given image. (**b**) The colorization result using Chromaticity and Brightness model via Stereographic Projection: $p = 2$, $r = 2$, $\sigma_1 = 0.1$, and $\sigma_2 = 10$. (**c**) The Chromaticity of the result. Note that color blending is automatically achieved and from (**c**) the result is more natural compared to typical homogeneous colorization

6 Connection with Nonlocal Methods

One benefit of using the kernel method for colorization is that the kernel information is already fully given by the brightness $g : \Omega \to \mathbb{R}$. So depending on the different possible choices of the kernel function, one can experiment with different effects of the diffusion process.

In this section, we explore the connection with nonlocal methods, motivated from the nonlocal mean filter proposed by Buades-Coll-Morel [8], and nonlocal methods from Kindermann-Osher-Jones [32] and Gilboa-Osher [24], among others. For each component, we consider the following functional $\mathcal{J}$ as a regularization term:

$$\mathcal{J}(F) = \frac{1}{4} \int_\Omega \int_\Omega K(x, y)(F(x) - F(y))^2 \, dx dy,$$

and the minimization for colorization to be

$$\inf_F \left\{ \mathcal{F}(F) = \gamma \mathcal{J}(F) + \frac{1}{2} \| f - F \|^2_{L^2(D)} \right.$$

$$= \gamma \mathcal{J}(F) + \frac{1}{2} \int_\Omega (P_D f(x) - P_D F(x))^2 \, dx \right\}, \quad (29)$$

where

$$P_D F(x) = \begin{cases} F(x) & \text{if } x \in D, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 7 Here we only assume that K is non-negative pointwise. This is different from the RKHS setting, where K is assumed to be positive definite (hence symmetric) but not necessarily non-negative pointwise. Also, we do not want the kernel K to be symmetric. Since the color is already known for $x \in D$ and the unknown colors should not affect the known colors, one should have $K(x, y) = 0$ for all $y \in D^c$. On the other hand, if $x \in D^c$, then to approximate the color at x, we need the color information from $y \in D$. Thus for each $x \in D^c$, $K(x, y) \neq 0$ for some $y \in D$. For example if we do not want to change the colors for $x \in D$,

then

$$K(x, y) = \begin{cases} 0, & \text{if } y \neq x \\ 1, & \text{if } y = x. \end{cases} \tag{30}$$

If one wants some denoising on the colors at $x \in D$, then $K(x, y)$ should be non-zero for some $y \in D$. (Cf. in the RKHS setting, the values $K(x, y)$ for $x, y \in D^c$ were never needed, and hence never computed.)

The differential $\frac{\partial \mathcal{J}(F)}{\partial F}$ in the direction of a test function v is given by

$$\frac{\partial \mathcal{J}(F)}{\partial F}(v) = \frac{1}{2} \int_\Omega \int_\Omega K(x, y)(F(x) - F(y))$$
$$\times (v(x) - v(y)) \, dx \, dy$$
$$:= \int_\Omega [C(x)F(x) - \overline{L}F(x)]v(x) \, dx,$$

where

$$C(x) = \int_\Omega \overline{K}(x, y) \, dy,$$

$$\overline{L}F(x) = \int_\Omega \overline{K}(x, y)F(y) \, dy, \quad \text{and}$$

$$\overline{K}(x, y) = \frac{(K(x, y) + K(y, x))}{2}.$$

Thus a minimizer F_γ of $\mathcal{F}$ defined in (29) satisfies

$$0 = \frac{\partial \mathcal{F}(F_\gamma)}{\partial F} = \gamma(CF_\gamma - \overline{L}F_\gamma) - P_D^*(f - P_D F_\gamma)$$
$$= \gamma(CF_\gamma - \overline{L}F_\gamma) - P_D(f - F_\gamma). \tag{31}$$

Here we assume $f = 0$ in D^c, and used $P_D^* = P_D$, and $P_D^2 = P_D$.

As mentioned in Remark 7, $K(x, y) = 0$ for all $x, y \in D^c$, so that $\overline{K}(x, y) = 0$ for all $x, y \in D^c$. Then, for $x \in D^c$, if F_γ is a minimizer from (31) and using the definition of P_D, $\gamma(CF_\gamma(x) - \overline{L}F_\gamma(x)) = 0$. In other words,

$$F_\gamma(x) = \frac{1}{C(x)} \int_\Omega \overline{K}(x, y)F_\gamma(y) \, dy$$
$$= \frac{1}{C(x)} \int_D \overline{K}(x, y)F_\gamma(y) \, dy. \tag{32}$$

If $x \in D$, then from (31),

$$\gamma(C(x)F_\gamma(x) - \overline{L}F_\gamma(x)) - f(x) + F_\gamma(x) = 0 \quad \Rightarrow$$

$$F_\gamma(x) = f(x) - \gamma C(x)F_\gamma(x) + \gamma \int_\Omega \overline{K}(x, y)F_\gamma(y) \, dy.$$

We have

$$\int_\Omega \overline{K}(x, y)F_\gamma(y) \, dy$$

$$= \int_D \overline{K}(x, y)F_\gamma(y) \, dy + \int_{D^c} \overline{K}(x, y)F_\gamma(y) \, dy$$

$$= \int_D \overline{K}(x, y)F_\gamma(y) \, dy$$

$$+ \int_{D^c} \overline{K}(x, y) \left[\frac{1}{C(y)} \int_D \overline{K}(y, z)F_\gamma(z) \, dz \right] dy,$$

where in the last term we use (32) for $F_\gamma(y)$, $y \in D^c$. Further defining

$$\widetilde{K}(x, z) = \int_{D^c} \frac{\overline{K}(x, y)\overline{K}(y, z)}{C(y)} \, dy$$

for the last term, the minimizer $F_\gamma(x)$ for $x \in D$ can be expressed as

$$F_\gamma(x) = f(x) - \gamma C(x)F_\gamma(x) + \gamma \int_D \overline{K}(x, y)F_\gamma(y) \, dy$$

$$+ \gamma \int_D \widetilde{K}(x, y)F_\gamma(y) \, dy. \tag{33}$$

Combining equations (32) and (33), we have

$$F_\gamma(x) = \begin{cases} \frac{1}{1 + \gamma C(x)}[f(x) + \gamma \int_D (\overline{K}(x, y) \\ \quad + \widetilde{K}(x, y))F_\gamma(y) \, dy] & \text{if } x \in D, \\ \frac{1}{C(x)} \int_D \overline{K}(x, y)F_\gamma(y) \, dy & \text{if } x \in D^c. \end{cases} \tag{34}$$

Discretely, we can solve it in the following way. Suppose $D = \{x_1, \ldots, x_m\}$ and denote $f_i = f(x_i)$, and $C_i = C(x_i)$. Letting $a_i = F_\gamma(x_i)$, we have

$$a_i = \frac{1}{1 + \gamma C_i} \left[f_i + \gamma \sum_{i=1}^m [\overline{K}(x_i, x_j) + \widetilde{K}(x_i, x_j)]a_j \right],$$

which implies that a_i satisfies the linear system

$$(1 + \gamma C_i)a_i - \gamma \sum_{i=1}^m [\overline{K}(x_i, x_j) + \widetilde{K}(x_i, x_j)]a_j = f_i. \tag{35}$$

For all $x \in D^c$,

$$F_\gamma(x) = \frac{1}{C(x)} \sum_{k=1}^m \overline{K}(x, x_i)a_i. \tag{36}$$

Let us compare this nonlocal regularization framework of (35) and (36) with the RKHS setting in Sect. 3, (26) and (27). Here the a_i's represent the recovered values at the known colored pixels in D, which are then used to compute the values at the unknown colored pixels in D^c. In contrast, the a_i's in the RKHS setting represent the coefficient of an explicit function expansion that is valid throughout Ω, which gives the color values at any pixel.

It is important to note here that it is not obvious when the coefficient matrix of system (35) is invertible for any $\gamma > 0$.

It is likely that further assumptions on K are needed, which we will leave for a future work. This is in contrast with the RKHS case, where it is immediate from the assumption of positive definiteness that the coefficient matrix of system (27) is always invertible for any $\gamma > 0$. More importantly, in order to solve system (35), we need to evaluate the matrix $\widetilde{K}$. This essentially involves multiplying two generally very large matrices, each of size $m \times (NM - m)$, and is therefore likely to be highly time consuming. We thus expect that this method is not as numerically efficient as the RKHS framework above.

Remark 8 In the discrete setting, the nonlocal framework just proposed is also related to the literature on the graph Laplacian (see for example [2, 16]). Let G be an undirected graph with N vertices and W be its non-negative symmetric weight matrix. Let D be the diagonal matrix with $D_{ii} = \sum_{j=1}^{N} W_{ij}$. Then the unnormalized graph Laplacian is defined to be $\Delta = D - W$. It is precisely the operator $C - \bar{L}$ above if the set Ω is discrete. For any $\mathbf{y} \in \mathbb{R}^N$ we have

$$\mathbf{y}^T \Delta \mathbf{y} = \frac{1}{2} \sum_{i,j=1}^{N} (y_i - y_j)^2 W_{ij}.$$

The Laplacian Δ always has as eigenvector the constant vector $\mathbf{e}_1 = (1, \dots, 1)$, with corresponding eigenvalue 0. The multiplicity of this eigenvalue is precisely the number of connected components in G.

The matrix Δ is symmetric and positive definite, therefore possesses a non-negative spectrum. Let $\{\mathbf{e}_i\}_{i=1}^{N}$ be an orthonormal basis of $\mathbb{R}^N$ consisting of eigenvectors of Δ, with corresponding eigenvalues λ_i. Then for $\mathbf{y} = (y_1, \dots, y_N)$ in this basis, we have

$$\mathbf{y}^T \Delta \mathbf{y} = \sum_{i=1}^{N} \lambda_i y_i^2.$$

This shows that on the row space $\text{row}(\Delta) = \text{nul}(\Delta)^\perp$, $\mathbf{y}^T \Delta \mathbf{y}$ is a Hilbert space square norm, which is strictly convex. If G is connected—guaranteed if $W_{ij} > 0$ for all i, j—then $\text{row}(\Delta) = \{\mathbf{y} : \sum_{i=1}^{N} y_i = 0\}$. This is consistent with the continuous version in Kindermann-Osher-Jones [32], where (the example in Sect. 4), $J(u)^{1/2}$ is a norm on the subspace of functions satisfying $\int_\Omega u(x)dx = 0$.

7 Concluding Remarks

Motivated by RKHS widely used in machine learning applications, we proposed extension methods for vector-valued functions using vector-valued RKHS. We studied the vectorial setting of RKHS, reformulated RKHS function extensions in terms of operator-valued kernels and considered in detail the diagonal case. One of the advantages of the proposed model is the fact that the solution is given via the minimization of a functional, and an explicit solution can be easily and efficiently computed through solving a simple system of linear equations. The RKHS framework guarantees, via a straightforward manner, that there is a unique global solution and no iteration is required. In addition, the flexibility of different choices of kernel allows texture colorization as well as cartoon.

We see this project as the starting point for exploring vectorial kernel applications. As seen in the example of Sect. 5, the color correlations are complicated and the extension to non-diagonal vectorial kernel is non-trivial. We used the diagonal kernel, and considered both channel by channel case of RGB color as well as chromaticity and brightness color setting to deal with different color treatments and fully explored the benefits of RKHS. We believe there are interesting extensions for the setting of non-diagonal vectorial RKHS and are exploring different possibilities.

References

1. Aronszajn, N.: Theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337–404 (1950)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003)
3. Berlinet, A., Thomas-Agnan, C.: Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer, Berlin (2004)
4. Bertalmio, M., Sapiro, G., Caselles, V., Balleste, C.: Image inpainting. In: Proceedings of SIGGRAPH 2000, New Orleans (2000)
5. Blomgren, P., Chan, T.F.: Color TV: Total variation methods for restoration of vector-valued images. IEEE Trans. Image Process. **7**(3), 304–309 (1998)
6. Bochko, V., Parkkinen, J.: A spectral color analysis and colorization technique. IEEE Comput. Graph. Appl. **26**, 74–82 (2006)
7. Buades, A., Coll, B., Lisani, J.-L., Sbert, C.: Conditional image diffusion. Inverse Probl. Imaging **1**(4), 593–608 (2007)
8. Buades, A., Coll, B., Morel, J.-M.: A review of image denoising algorithms, with a new one. SIAM Multiscale Model. Simul. **15**(3), 490–530 (2005)
9. Burns, G.: Museum of Broadcast Communications: Encyclopedia of Television. World Wide Web electronic publication (1997)
10. Caponnetto, A., Pontil, M., Micchelli, C., Ying, Y.: Universal multi-task kernels. J. Mach. Learn. Res. **9**, 1615–1646 (2008)
11. Caponnetto, A., De Vito, E.: Optimal rates for regularized least-squares algorithm. Found. Comput. Math. **7**(3), 331–368 (2007)

12. Carmel, C., De Vito, E., Toigo, A., Umanità, V.: Vector valued reproducing kernel Hilbert spaces and universality. Anal. Appl. **8**(1), 19–61 (2010)

13. Carmeli, C., De Vito, E., Toigo, A.: Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. Anal. Appl. **4**, 377–408 (2006)

14. Chan, T.F., Kang, S.H., Shen, J.: Total variation denoising and enhancement of color images based on the CB and HSV color models. J. Vis. Commun. Image Represent. **12**(4), 422–435 (2001)

15. Chen, T., Wang, Y., Schillings, V., Meinel, C.: Grayscale image matting and colorization. In: Proceedings of Asian Conference on Computer Vision (ACCV) (2004)

16. Chung, F.: Spectral Graph Theory. American Mathematical Society, Providence (1997)

17. Coifman, R.R., Lafon, S.: Geometric harmonics: a novel tool for multiscale out-of-sample extension of empirical functions. Appl. Comput. Harmon. Anal. **21**, 31–52 (2006)

18. Cucker, F., Smale, S.: On the mathematical foundations of learning. Bull. Am. Math. Soc. **39**(1), 1–49 (2002)

19. Drew, M., Finlayson, G.: Realistic colorization via the structure tensor. In: International Conference on Image Processing, ICIP08 (2008)

20. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Mathematics and Its Applications, vol. 375. Springer, Berlin (1996)

21. Fonseca, I., Leoni, G., Maggi, F., Morini, M.: Exact reconstruction of damaged color images using a total variation model. Preprint (2010)

22. Fornasier, M.: Nonlinear projection digital image inpainting and restoration methods. J. Math. Imaging Vis. **24**(3), 359–373 (2006)

23. Fornasier, M., March, R.: Restoration of color images by vector valued BV functions and variational calculus. SIAM J. Appl. Math. **68**(2), 437–460 (2007)

24. Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. SIAM Multiscale Modeling and Simulation (MMS) **6**(2) (2007)

25. Gonzalez, R., Wood, R.: Digital Image Processing. Addison-Wesley, Reading (1992)

26. Horiuchi, T.: Estimation of color for gray-level image by probabilistic relaxation. In: Proc. IEEE Int. Conf. Pattern Recognition, pp. 867–870 (2002)

27. Horiuchi, T., Hirano, S.: Colorization algorithm for grayscale image by propagating seed pixels. In: Proc. IEEE Int. Conf. Pattern Recognition, pp. 457–460 (2003)

28. Irony, R., Cohen-Or, D., Lischinski, D.: Colorization by example. In: Proceedings of Eurographics Symposium on Rendering, pp. 201–210 (2005)

29. Jones, F.: Lebesgue Integration on Euclidean Space. Jones and Bartlett, Boston (2001), revised edition

30. Kang, S.H., March, R.: Variational models for image colorization via chromaticity and brightness decomposition. IEEE Trans. Image Process. **16**(9), 2251–2261 (2007)

31. Kimmel, R., Sochen, N.: Orientation diffusion or how to comb a porcupine? J. Vis. Commun. Image Represent. **13**, 238–248 (2001)

32. Kindermann, S., Osher, S., Jones, P.: Deblurring and denoising of images by nonlocal functionals. SIAM Multiscale Model. Simul. **4**(4), 2005

33. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Trans. Graph. **23**(3), 689–694 (2004)

34. Lezoray, O., Ta, V.T., Elmoataz, A.: Nonlocal graph regularization for image colorization. In: Proceedings of 19th International Conference on Pattern Recognition, pp. 1–4 (2008)

35. Liu, B., Liu, M., Wang, G.: Colorization based on image manifold learning. In: IEEE Region 10 Conference TENCON, pp. 1–3 (2006)

36. Micchelli, C.A., Pontil, M.: On leaning vector-valued functions. Neural Comput. **17**, 177–204 (2005)

37. Minh, H.Q., Niyogi, P., Yao, Y.: Mercer's theorem, feature maps, and smoothing. In: Proceedings of 19th Annual Conference on Learning Theory, Pittsburg, June 2006. Springer, Berlin (2006)

38. Pan, Z., Dong, Z., Zhang, M.: A new algorithm for adding color to video or animation clips. In: WSCG, pp. 515–520 (2004)

39. Perona, P.: Orientation diffusion. IEEE Trans. Image Process. **7**(3), 457–467 (1998)

40. Qiu, G., Guan, J.: Color by linear neighborhood embedding. In: IEEE International Conference on Image Processing, vol. 3, pp. 988–991 (2005)

41. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Trans. Comput. Graph. Appl. **21**, 34–41 (2002)

42. Saitoh, S.: Integral Transforms, Reproducing Kernels and Their Applications. Pitman Research Notes in Mathematics Series, vol. 369. Longman, Harlow (1997)

43. Sapiro, G.: Inpainting the colors. In: ICIP 2005. IEEE International Conference on Image Processing, vol. 2, pp. 698–701 (2005)

44. Schoenberg, I.J.: Metric spaces and positive definite functions. Trans. Am. Math. Soc. **44**, 522–536 (1938)

45. Schökopf, B., Smola, A.: Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2002)

46. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge (2004)

47. Smale, S., Zhou, D.X.: Learning theory estimates via integral operators and their approximations. Constr. Approx. **26**, 153–172 (2007)

48. Stein, E.M.: Singular Integrals and Differentiability Properties of Functions. Princeton University Press, Princeton (1970)

49. Sýkora, D., Buriánek, J., Žára, J.: Unsupervised colorization of black-and-white cartoons. In: Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering, pp. 121–127. ACM, New York (2004)

50. Tai, Y.W., Jia, J., Tang, C.K.: Soft color segmentation and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1520–1537 (2007)

51. Tang, B., Sapiro, G., Caselles, V.: Color image enhancement via chromaticity diffusion. IEEE Trans. Image Process. **10**, 701–707 (2001)

52. Vapnik, V.: Statistical Learning Theory. Wiley, New York (1998)

53. Wahba, G.: Spline interpolation and smoothing on the sphere. SIAM J. Sci. Stat. Comput. **2**(1), 5–16 (1981)

54. Wahba, G.: Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 59. SIAM, Philadelphia (1990)

55. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: SIGGRAPH Conference Proceedings, p. 21 (2002)

56. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Trans. Image Process. **15**(5), 1120–1129 (2006)