

Computationally efficient data and application driven color transforms for the compression and enhancement of images and video

Massimo Minervini, Cristian Rusu, and Sotirios A. Tsaftaris

Abstract An important step in color image or video coding and enhancement is the linear transformation of input (typically RGB) data into a color space more suitable for compression, subsequent analysis, or visualization. The choice of this transform becomes even more critical when operating in distributed and low-computational power environments, such as visual sensor networks or remote sensing. Data-driven transforms are rarely used due to increased complexity. Most schemes adopt fixed transforms to decorrelate the color channels which are then processed independently. Here we propose two frameworks to find appropriate data-driven transforms in different settings. The first, named approximate Karhunen-Loève Transform (aKLT), performs comparable to the KLT at a fraction of the computational complexity, thus favoring adoption on sensors and resource-constrained devices. Furthermore, we consider an application-aware setting in which an expert system (e.g., a classifier) analyzes imaging data at the receiver's end. In a compression context, distortion may jeopardize the accuracy of the analysis. Since the KLT is not optimal in this setting, we investigate formulations that maximize post-compression expert system performance. Relaxing decorrelation and energy compactness constraints, a second transform can be obtained offline with supervised learning methods. Finally, we propose transforms that accommodate both constraints, and are found using regularized optimization.

Key words: Color space transformation, unsupervised and supervised learning, Karhunen-Loève transform, Foley-Sammon transform, image coding, application-aware compression, JPEG 2000, low computational complexity.

Massimo Minervini · Cristian Rusu · Sotirios A. Tsaftaris
IMT Institute for Advanced Studies, Piazza San Ponziano 6, Lucca, Italy
e-mail: {massimo.minervini, cristian.rusu, sotirios.tsafaris}@imtlucca.it

1 Introduction

The red-green-blue (RGB) color model is currently the standard for acquisition and display of color digital images. However, this representation is not efficient for coding, due to high correlation between color bands of natural images [51].

To reduce spectral redundancy, image and video compression algorithms operate on luminance/chrominance representations of the color information, achieved through linear transformations of the RGB color space [47, 50]. Each color band is coded independently, therein deploying a variety of techniques to address spatial and, for video, also temporal correlation. A family of such color models is the YC_bC_r , defined by the International Telecommunication Union [30] and adopted by many coding standards [55]. However, due to high variability in source image characteristics, a fixed transform may easily result in suboptimal performance, thus motivating the adoption in some contexts of a data-dependent one.

Among the linear transformations, the energy compaction and decorrelation properties of the Karhunen-Loëve transform (KLT) [32] make it desirable for color image compression [19, 37, 51, 52, 64]. The KLT has been adopted in many coding schemes (e.g., for dimensionality reduction [1, 9]), and it was shown to be superior to other approaches in a variety of contexts, both for color [8, 23, 35] and hyperspectral [13] imagery, and has formed the basis for new fixed transforms [24, 36, 42]. However, the computational complexity of calculating the color covariance matrix limits its applicability in real-time video enhancement applications (e.g., denoising [67], contrast and color [20, 21] enhancement, color to gray scale conversion [11]), and sensing environments with low computational power (e.g., surveillance cameras or visual sensor networks operating under low lighting conditions).

A variety of approaches have been proposed to circumvent this bottleneck. Kountchev et al. rely on covariance matrix approximations [38], while [11] uses numerical methods to estimate eigenvalues and eigenvectors of the covariance matrix. Subsampling strategies to limit complexity by reducing the amount of input data are discussed in [14, 48]. Du et al. adopt a learning approach based on neural networks to estimate projection directions [14]. Porikli et al. [49] propose an algorithm based on integral images for fast computation of the covariance matrix. Others focus on schemes that favor parallel implementations [4, 66] of the KLT, or implementations optimized for graphics processing units [4, 34, 41]. In this work, we propose a new data-dependent color transform, the aKLT, rooted in the orthogonal Procrustes problem, that preserves energy compaction and performs similar to the KLT, but is less computationally complex.

Although KLT and aKLT are designed to match the statistical properties of the image data, they are agnostic to the semantics of the scene (e.g., distinction between foreground and background regions). In present days, more often than not, image data are analyzed by computer vision algorithms (e.g., surveillance applications [33], industrial inspection [2], or medical image analysis [7]) and their transmission over channels necessitates their compression, to reduce bandwidth costs. It was shown recently that considering the application and designing data codecs appropriately not to maximize fidelity type criteria (e.g., mean squared error [61]) or

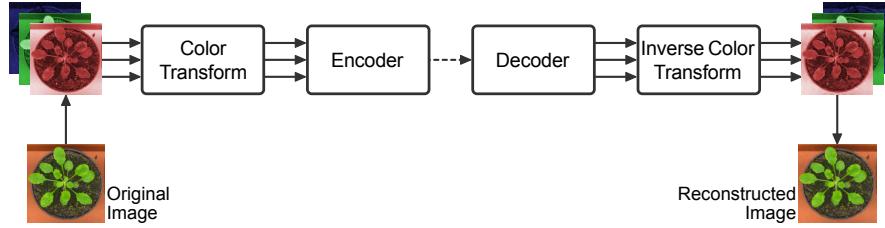


Fig. 1 Schematic of the typical encoding/decoding process of a color image.

psycho-visual criteria (e.g., structural similarity [62]), but considering how would an analysis algorithm (e.g., a classifier) perform on compressed data, is beneficial from a bit rate perspective [57]. This notion was explored in [12] and [57] with respect to quantization, however, as of now the design of color transforms optimized particularly for application accuracy has not been considered yet. In this work we consider such application to be classification, motivated by its broad range of real-world applications. Thus, we propose for the first time to obtain a color transform using supervision (e.g., previously labeled image data), aiming at the preservation of the image features relevant to the application. We formulate our methodology as a supervised learning problem, envisioning two alternative approaches to find a solution, relying either on the Foley-Sammon transform (FST) [17] or on metric learning methods [39]. From an application-aware image compression perspective, it is desirable to achieve: (a) classification accuracy, pursued by enhancing separability of the transformed data, and (b) compression performance, achieved with decorrelation and energy compaction. These two requirements appear conflicting and designing a color transform that optimally accommodates both remains an open challenge. Therefore, we propose to adopt an optimization approach to obtain application-dependent color transforms that while aiming to retain energy compaction properties, also try to maximize separability of the transformed data.

This chapter builds upon our previous work [45], where we first investigate our proposed color transforms and present preliminary results. The rest of the chapter is organized as follows. Section 2 details our methodology to learn color transforms from the data. Section 3 demonstrates the proposed approaches on a variety of different image datasets, using the JPEG 2000 standard to compress test images. Finally, Section 4 offers concluding remarks.

2 Methodology

We represent an RGB image as a $3 \times n$ matrix $\mathbf{X} = (\mathbf{r}, \mathbf{g}, \mathbf{b})^\top$, where \mathbf{r} , \mathbf{g} , and \mathbf{b} are the linearized color components, and n is the number of pixels. Prior to lossy coding, \mathbf{X} is projected into a new color space by $\mathbf{T} \in \mathbb{R}^{3 \times 3}$. Each pixel value $\mathbf{x}_i = (r_i, g_i, b_i)^\top$ in \mathbf{X} is transformed by the linear relation $\mathbf{y}_i = \mathbf{T}\mathbf{x}_i$. Upon reconstruction, the color

transform is inverted, obtaining the approximation $\tilde{\mathbf{x}}_i = \mathbf{T}^{-1}\tilde{\mathbf{y}}_i$ in the RGB domain. To ensure that the output range of \mathbf{y}_i is the same as \mathbf{x}_i (e.g., 0 to 255, for 8-bit unsigned integer representation), we scale the directions (rows) of \mathbf{T} with respect to the ℓ_1 norm [36].

We operate in the general framework depicted in Figure 1, in which color transformation is decoupled from encoding/decoding operations. Therefore, any compression scheme can be adopted (in the experiments we use the JPEG 2000 standard [56]) and the resulting bit stream will be standard-compliant.

In the following, we address the problem of obtaining data-driven color space transformations¹, that change according to the image to be encoded or the application the images will be used for. In Section 2.1, based on a heuristic, we derive a new low-complexity transform (aKLT) that adapts to the content using statistical information from the image being processed. In Section 2.2, we propose a novel approach to obtain transforms that adapt according to the application (here a pixel-level classifier for foreground-background segmentation), relying on supervised learning methods and labeled training data. Finally, in Section 2.3, we combine the unsupervised and the supervised transforms using an optimization approach.

2.1 The aKLT: A low-complexity unsupervised data-dependent transform

The KLT produces an orthogonal transformation, \mathbf{K} , obtained from the eigendecomposition of the color covariance matrix $\Sigma = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^\top$, where $\mu = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ is the mean color vector [51]. The eigenvectors of Σ , sorted in decreasing order of magnitude of the corresponding eigenvalues, define the directions of \mathbf{K} . The KLT achieves complete statistical decorrelation of the color signals and energy compaction in the first channel, thus favoring efficient representation and subsampling of the other two channels [40, 52].

However, estimation of Σ can be computationally demanding in memory and computing power, particularly when images are large, and its application in resource-constrained sensing devices can be problematic. Thus, we seek to find a transform that is close to the KLT but less computationally complex to obtain.

Let $\bar{\mathbf{X}} \in \mathbb{R}^{3 \times n}$ be the matrix obtained by normalizing each column (pixel) of \mathbf{X} with respect to the ℓ_2 norm. We seek an orthogonal transform $\Omega \in \mathbb{R}^{3 \times 3}$ that maps $\bar{\mathbf{X}}$ to a given reference matrix $\mathbf{W} \in \mathbb{R}^{3 \times n}$, and formulate it as:

$$\begin{aligned} & \underset{\Omega}{\text{minimize}} && \|\mathbf{W} - \Omega \bar{\mathbf{X}}\|_F \\ & \text{subject to} && \Omega^\top \Omega = \mathbf{I}, \end{aligned} \tag{1}$$

¹ Approaches to efficiently compute the color space conversion for a known transform are discussed, e.g., in [6, 65].

where $\|\cdot\|_F$ denotes the Frobenius norm, and I is the identity matrix. Let $Z = W\bar{X}^T$, and $Z = USV^T$ be the singular value decomposition (SVD) of Z . The optimization problem of Eq. (1), known as the orthogonal Procrustes problem, admits closed-form solution UV^T [54]. In order to concentrate energy in the first direction, we impose structure to W :

$$W = \begin{pmatrix} 1 & \dots & 1 \\ 0 & \dots & 0 \\ 0 & \dots & 0 \end{pmatrix}.$$

Notably, this leads to a simplified form of Z :

$$Z = \begin{pmatrix} \sum_{i=1}^n r_i & \sum_{i=1}^n g_i & \sum_{i=1}^n b_i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with only a single direction, $\mathbf{a}_1 = \mathbf{z}_1^T / \|\mathbf{z}_1\|_2$, that corresponds to the principal direction, thus making the SVD computation unnecessary. We adopt the vector \mathbf{a}_1 as an approximation of the principal direction of the KLT.

In order to obtain the full transform, we proceed by constructing the 3×3 matrix $A = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, where \mathbf{a}_2 and \mathbf{a}_3 are initialized with random elements, e.g., uniformly distributed on the interval $[0, 1]$ (the effect of randomness on performance is explored in Section 3). Subsequently, we use QR factorization to decompose A into the product $A = QR$, where $Q \in \mathbb{R}^{3 \times 3}$ has orthogonal columns and $R \in \mathbb{R}^{3 \times 3}$ is upper triangular. The aKLT transformation matrix, $\tilde{K} = Q^T$, shares relevant properties with the regular KLT: (a) orthogonality, and (b) energy compaction capabilities. Although there is no guarantee on sorting and relative amount of energy of second and third channel, this is not of concern from a compression standpoint (e.g., chroma subsampling strategies would downsample the lower-energy components using the same scheme).

Complexity comparison between KLT approaches. The computation of the KLT requires $15n$ floating point operations in total, where n is the number of pixels, and is dominated by computation of mean color vector μ ($3n$ operations) and covariance matrix Σ (due to symmetry, $12n$ operations are necessary to compute its 6 distinct entries). Notice that while most authors center the original data on the mean prior to calculating Σ (a step that would require additional $3n$ operations, since it is performed on all image pixels), the covariance matrix can also be defined as $\Sigma = \frac{1}{n} (\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T) - \mu \mu^T$ [3]. Thus, mean subtraction can be performed on the small

Table 1 Comparison of KLT approaches as a function of input size n , where n denotes the number of image pixels.

KLT, Power method [11], ACKLT [38] Penna et al. [48]	IPCA ^a [14]	Porikli et al. [49]	aKLT
$15n$	$\rho 15n$	$15n$	$15n$

^a Complexity reported by the authors of [14] only takes into account multiplications.

3×3 autocorrelation matrix, with fixed computational cost [16]. We also ignore cost of subsequent eigenvalue decomposition of Σ to obtain K , since this step does not depend on n .

Approaches that speed up the eigendecomposition of Σ (e.g., power method [11], or ACKLT [38]) provide negligible benefit in this context (particularly as image resolution increases). As shown in Table 1, the IPCA [14], based on neural networks, achieves an approximation of the principal direction using $15n$ multiplications (additions are not reported by the authors), while [48] necessitates to keep a fraction $\rho = 0.8$ of the data to match the aKLT. Porikli et al. [49] propose a method based on integral images for fast construction of feature covariance matrices of all possible rectangular regions in an image. However, when used in the full image, this approach does not provide any benefit, computing the color covariance matrix using also $15n$ operations.

On the other hand, our proposed aKLT estimates the statistical properties of the source image and computes the color transform with lower computational complexity. It requires $9n$ operations to normalize the input data, i.e. $3n$ to square the pixel values, $2n$ to add the components of each pixel, n to compute the square root for each pixel, and $3n$ to divide each channel by the so-obtained ℓ_2 norm (notice that mean subtraction is not necessary for the aKLT). Furthermore, $3n$ operations are required to calculate \mathbf{z}_1 , and a small fixed cost (27 operations) for the QR decomposition of A , resulting in $12n$ operations in total to obtain \tilde{K} , i.e. 20% reduction in complexity compared to the regular KLT. Computational complexity of the aKLT can be further reduced if combined with the sub-sampling strategies proposed in [48].

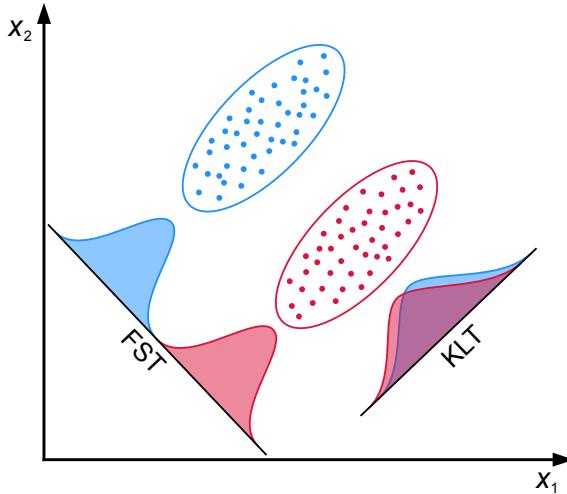
2.2 A supervised approach to an application-dependent color transform using labeled pixels

It is known that projecting to principal components is not always optimal from a pattern recognition perspective: clusters of points belonging to semantically different objects in the scene may overlap now in the projected color space [59] (cf. Figure 2). Introducing distortions due to lossy compression may affect this separability further. With an application-aware compression setting in mind, we seek to identify a transform that maintains (a) class separation as well as (b) decorrelation and energy compaction properties.

We assume that the computation of the supervised color transform will occur in an offline fashion and we will use a training set (pixels partitioned in two classes), thus it is supervised. Compression of newly acquired images at the sensor occurs as before, with the transform now known. The calculation of a new transform is necessary only if the scene conditions change (depending on the process being observed) and if new training data are available.

Let C_1 and C_2 be disjoint sets of pixel values (i.e. $C_1 \cap C_2 = \emptyset$) representative of distinct pattern classes (e.g., foreground and background). We seek an orthogonal transform $D \in \mathbb{R}^{3 \times 3}$ that projects data points belonging to distinct classes, $\mathbf{x}_1 \in C_1$

Fig. 2 The KLT seeks directions of maximum variance in the projected data. On the other hand, the FST as other discriminative methods seeks directions that maximize class separability in a lower-dimensional subspace [59].



and $\mathbf{x}_2 \in C_2$, in a domain where they are maximally separated according to a measure of separability \mathcal{C} :

$$\underset{\mathbf{D}}{\text{maximize}} \quad \mathcal{C}\{\mathbf{D}\mathbf{x}_1, \mathbf{D}\mathbf{x}_2\} \quad (2a)$$

$$\text{subject to} \quad \mathbf{D}^T \mathbf{D} = \mathbf{I}, \quad (2b)$$

$$\|\mathbf{D}\Sigma\mathbf{D}^T - \Lambda\|_F \leq \varepsilon, \quad (2c)$$

where \mathbf{I} is the identity matrix, Σ is the color covariance matrix, Λ is a diagonal matrix whose elements are the eigenvalues of Σ , and $\varepsilon \geq 0$. The objective function of Eq. (2a) accounts for the class separation property of \mathbf{D} , while the constraints of Eq. (2b) and (2c) allow for decorrelation and energy compaction. The parameter ε determines the trade-off between class separation on the one hand and decorrelation and energy compaction on the other hand. Therefore, solving the complete problem of Eq. (2) would lead to an orthogonal transform with the full set of the desired properties. On the other hand, imposing the orthogonality constraint (non-convex in nature) renders the optimization problem of Eq. (2) non-convex, possibly admitting multiple local optima. Non-convex problems are generally difficult to solve (i.e. finding the global optimum), and require to resort to global optimization methods, which are however computationally intensive. In the remainder of this section, we relax the problem by ignoring Eq. (2c), but we revisit the complete problem in the next section.

In the following paragraphs, we discuss two supervised learning approaches to obtaining a color transform with class separation capabilities: (a) the Foley-Sammon transform, based on the linear discriminant analysis, and (b) metric learning approaches.

Foley-Sammon transform. An effective measure of class separability is the Fisher's criterion [15, 59], defined as the difference between the means of the class data points, normalized by a measure of the within-class dispersion. This notion is formally expressed in the form of the generalized Rayleigh quotient:

$$J(\mathbf{d}) = \frac{\mathbf{d}^\top \mathbf{S}_b \mathbf{d}}{\mathbf{d}^\top \mathbf{S}_w \mathbf{d}}, \quad (3)$$

where $\mathbf{d} \in \mathbb{R}^3$, and $\mathbf{S}_b, \mathbf{S}_w \in \mathbb{R}^{3 \times 3}$ are, respectively, between-class scatter matrix and within-class scatter matrix:

$$\begin{aligned} \mathbf{S}_b &= \sum_{i=1}^2 (\mathbf{m}_i - \mu)(\mathbf{m}_i - \mu)^\top, \\ \mathbf{S}_w &= \sum_{i=1}^2 \sum_{\mathbf{x} \in C_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^\top, \end{aligned}$$

where $\mu = \sum_{i=1}^2 P_i \mathbf{m}_i$ is the mean sample vector, and $\mathbf{m}_i \in \mathbb{R}^3$ and $P_i \in \mathbb{R}$ are, respectively, mean and a priori probability of class i .

A closed-form solution to finding orthogonal discriminant vectors that maximize the Fisher's criterion, can be obtained adopting the Foley-Sammon transform (FST) [17]. The first direction, \mathbf{d}_1 , termed Fisher's discriminant vector (or Fisher's linear discriminant [15]), corresponds to the projection direction that yields maximum between-class scatter and minimum within-class scatter, i.e. $\mathbf{d}_1 = \arg \max_{\mathbf{d}} J(\mathbf{d})$, and is obtained as the eigenvector associated with the largest eigenvalue α_1 satisfying $\mathbf{S}_b \mathbf{d}_1 = \alpha_1 \mathbf{S}_w \mathbf{d}_1$ [59]. The following directions to complete the three-dimensional transform are found recursively, by maximizing the Fisher's criterion with an orthogonality constraint. If $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_r)^\top$ is the set of previously obtained directions, \mathbf{d}_{r+1} corresponds to the eigenvector associated with the largest eigenvalue α_{r+1} satisfying $\mathbf{M} \mathbf{S}_b \mathbf{d}_{r+1} = \alpha_{r+1} \mathbf{S}_w \mathbf{d}_{r+1}$, where $\mathbf{M} = \mathbf{I} - \mathbf{D}^\top (\mathbf{D} \mathbf{S}_w^{-1} \mathbf{D}^\top)^{-1} \mathbf{D} \mathbf{S}_w^{-1}$ [31], and \mathbf{I} is the identity matrix. For three-dimensional RGB data, the final color transformation matrix is defined by $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3)^\top$.

In this chapter we consider only a two-class classification problem (i.e. foreground vs. background), however, the FST formulation can be easily generalized to an arbitrary number of pattern classes [18]. According to the class distribution of our test image datasets (cf. Section 3), we use the standard FST formulation to find a linear separation between foreground and background. Kernel formulations of the FST [68] could be adopted to accommodate non-linearly separable classes.

Metric learning approaches. Metric learning methods (see [39] for a comprehensive survey) seek to estimate from supervised information a Mahalanobis distance function over data points, $\mathcal{D}_A(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top A(\mathbf{x}_i - \mathbf{x}_j)$, parametrized by a positive semidefinite matrix $A = L^\top L$. Computing the distance in the input space is equivalent to applying a linear transformation L of the input space, such that data points with small distance according to \mathcal{D}_A are close (in a Euclidean sense) in the projected space, i.e. the matrix L minimizes $\mathcal{D}_A(\mathbf{x}_i, \mathbf{x}_j) = \|L\mathbf{x}_i - L\mathbf{x}_j\|_2^2$, which is

another view of the objective function in Eq. (2a). While the FST matrix D is orthogonal, the L transformation matrix obtained by a metric learning approach is positive semidefinite, and in general not orthogonal. Observe that orthogonality is beneficial for the numerical stability of the color transformation, i.e. errors introduced by compression and decompression operations are not magnified when forward and reverse color space conversions are computed. Therefore, with respect to the framework defined by Eq. (2), metric learning approaches optimize solely for class separability.

Recent metric learning methods include Relevant Component Analysis (RCA) [5], Large Margin Nearest Neighbors (LMNN) [63], and Information Theoretic Metric Learning (ITML) [10], which can all be used to find L .

2.3 Combining unsupervised and supervised approaches

Our approach for finding the supervised transform D in Section 2.2 relaxed the constraint of optimal decorrelation and energy compaction of Eq. (2), finding one that only optimizes for separation. In the previous section we also obtained orthogonality when using the FST, however, this does not guarantee energy compaction, which is achieved by the KLT (or can be approximated by the aKLT). Therefore, we consider now a different approach, removing the orthogonality constraint to obtain a convex relaxation of the problem of Eq. (2).

We seek to find a new transform $D' \in \mathbb{R}^{3 \times 3}$ that is close to D whilst trying to satisfy Eq. (2c), or equivalently, since we know that the KLT (or the aKLT) optimizes Eq. (2c), we can pose the following unconstrained optimization problem:

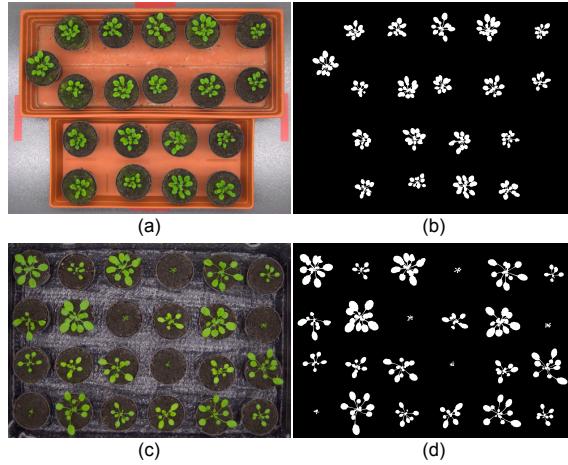
$$\underset{D'}{\text{minimize}} \|D' - D\|_F + \lambda \|D' - \tilde{K}\|_F, \quad (4)$$

thus, finding a transform that is between D (application-aware, obtained offline using labeled data) and the aKLT (obtained at the sensor and computed based on the unseen image), where the trade-off is controlled by the value of the regularization parameter λ (playing here a role similar to ε in Eq. (2)). In the same fashion, the L transform obtained with metric learning methods could be used in Eq. (4) instead of D . Although D and \tilde{K} in Eq. (4) are orthogonal, in general D' will not be orthogonal. Approaches for finding the nearest orthonormal matrix to D' can be adopted, e.g., relying on the polar decomposition [27], or the square root matrix [28] of D' .

While this approach adapts the supervised transform to unseen data on the sensor and is expected to gain decorrelating and compacting capabilities, from a computational perspective may be less attractive. In this setting, with the FST (or the RCA) known, the encoder is required to compute the (a)KLT and then solve Eq. (4) to obtain the final color transform.

The approaches presented in Sections 2.1 and 2.2 admit closed-form solutions, whereas D' is found relying on iterative optimization procedures computing the so-

Fig. 3 (Left) Example images of Arabidopsis plants from different experiments [53], and (right) corresponding ground truth segmentations delineated manually.



lution path along the λ parameter. On the other hand, Eq. (4) involves only color transforms (i.e. small 3×3 matrices), rather than the original image pixels.

3 Results and Discussion

3.1 Experimental settings

The proposed methodology is evaluated on color image data from a variety of classes. We demonstrate the unsupervised approach on standard test images, including natural, aerial, and retinal [58] images (Figure 4). We showcase the supervised transform using images of different size (up to 18 megapixel) downloaded from the Internet², including horses, balloons, and fish (Figure 6). The approaches are also evaluated on a dataset of 20 images (width \times height: 3108×2324 pixels) from a time-lapse sequence of Arabidopsis plant subjects (Figure 3a), arising from plant phenotyping experiments [53]. We use images from this application since they are usually large and due to design requirements they may need to be communicated via the Internet to centralized locations for processing [46]. Thus, any bit rate savings possible are desirable.

We include in the comparison plain RGB (i.e. no color transform) and YC_bC_r (ITU-R BT.601) [30]. KLT and aKLT are computed for each image. We also adopt the Relevant Component Analysis (RCA) [5], a metric learning approach to find a supervised transform L that aims to preserve variability in the data relevant to the classification task at hand. For brevity and clarity of presentation we do not include other popular metric learning approaches, such as LMNN [63] and ITML [10], because they perform similar to the RCA in our image compression context, while

² <http://www.flickr.com/>

being more computationally demanding (they rely on iterative optimization procedures). The supervised transforms (FST, RCA) are estimated on manually labeled training image data (excluded from testing). On the plant dataset, the supervised transforms (D and L) are estimated from the first image of the time-lapse sequence using pixel label information obtained manually. The so-obtained D and L are then applied to all subsequent images of the same sequence and also to a test image of Arabidopsis plants with different scene conditions (Figure 3c). While the other transforms included in the comparison are either fixed (RGB, YC_bC_r) or present closed-form solutions (KLT, aKLT, FST, RCA), to solve the optimization problem of Eq. (4) we use CVX³, a package for specifying and solving convex programs [22].

After color space transformation, the images are compressed at various bit rates (between 0.0625 and 2 bpp) using the JJ2000 software implementation⁴, version 5.1, of the JPEG 2000 coding standard [56]. We implement the proposed methods using Matlab R2011b, and conduct all experiments on a machine with Intel Core 2 Duo CPU E8200 2.66 GHz and 4 GB memory.

The approaches are evaluated according to: (a) reconstruction accuracy, and (b) application error. Reconstruction accuracy is measured using Peak Signal-to-Noise Ratio (PSNR) in RGB image domain, either in the full image or in regions of interest (e.g., foreground regions as in Figure 3). To estimate application error, we adopt the task of plant segmentation for plant phenotyping applications [44, 46], therefore, we first build a rudimentary classifier. Similar to the approach described in [44], we train a Gaussian mixture model, \mathcal{M} , on color features (a^* and b^* components of the CIE $L^*a^*b^*$ color space [29]), using labeled foreground (plant) data from the first uncompressed image of the time-lapse sequence (excluded from testing). At each tested bit rate, we calculate the average application error:

$$E_{\mathcal{M}} = \frac{\sum_{i=1}^n (\mathcal{M}(\tilde{\mathbf{x}}_i) - \mathcal{M}(\mathbf{x}_i))^2}{\sum_{i=1}^n (\mathcal{M}(\mathbf{x}_i))^2}, \quad (5)$$

between the posterior probabilities estimated by \mathcal{M} on the n original, \mathbf{x}_i , and reconstructed, $\tilde{\mathbf{x}}_i$, image pixels. Application error is expressed in percentage, where best possible value of $E_{\mathcal{M}}$ is 0%.

3.2 Results

In this section, we present rate-distortion performance of the proposed approaches. We first compare them in terms of overall reconstruction accuracy. Next, we demonstrate the supervised approach in an application-aware context.

Reconstruction accuracy. On the benchmark images of Figure 4, all of the decorrelating transforms provide considerable PSNR improvement with respect to the

³ <http://cvxr.com/cvx>

⁴ <http://code.google.com/p/jj2000/>

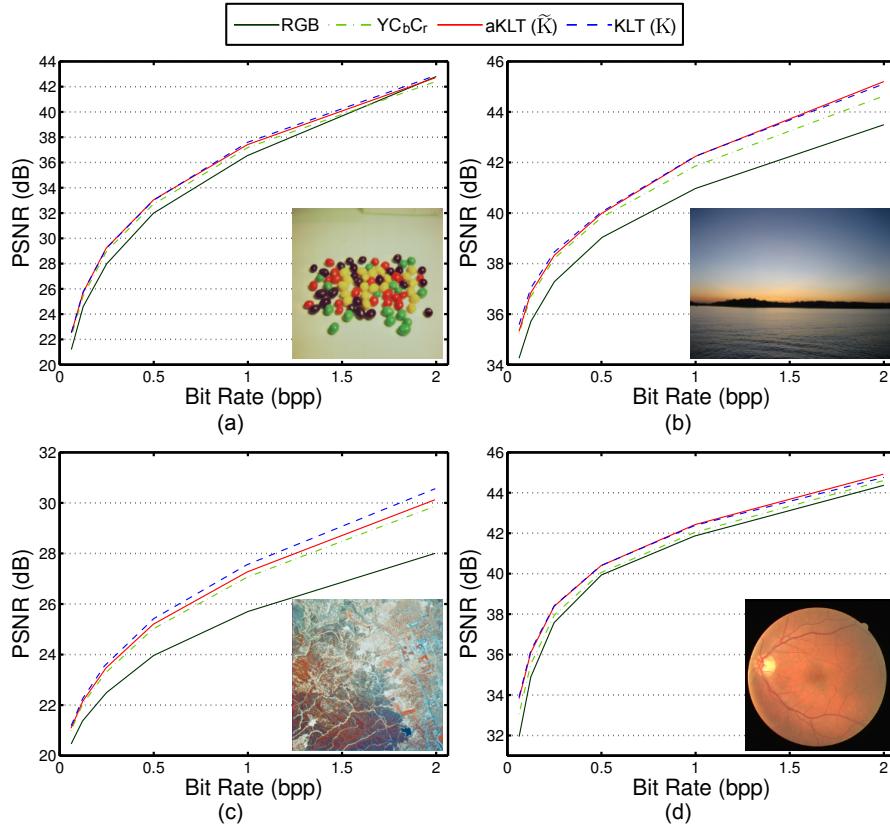


Fig. 4 Reconstruction accuracy of standard test images, using fixed and data-dependent color transforms (proposed transform is shown with solid red curve). (a) Jelly beans (width \times height: 256 \times 256 pixels). (b) Landscape of Bretagne (width \times height: 2592 \times 1944 pixels). (c) Aerial photograph of Woodland Hills, Ca. (width \times height: 512 \times 512 pixels). (d) Human retina [58] (width \times height: 565 \times 584 pixels). For the aKLT, average results are shown, obtained using 100 different initializations (see Section 2.1).

plain RGB color space, with the data-dependent transforms (KLT, aKLT) outperforming the fixed YCbCr. Notably, our proposed low-complexity aKLT, \tilde{K} , exhibits performance very close to the regular KLT, or in some cases slightly superior (cf. red line in Figure 4b and 4d, higher bit rates).

Table 2 reports image fidelity results for the Arabidopsis plant dataset⁵. At low bit rates (< 1 bpp), decorrelating transforms (YCbCr, KLT, aKLT) achieve better performance than RGB (0.25 to 0.6 dB improvement in PSNR). Performance of the aKLT is always superior to the YCbCr, and for bit rates greater than 0.5 bpp it surpasses

⁵ Observe that, in general, major bit rate savings are attained by compression schemes with the combined use of several coding tools. Thus, seemingly small differences in PSNR observed here (i.e. in the order of a fraction of dB) are accounted for by the fact that only the effect of color transformation is tested.

Table 2 Reconstruction accuracy comparison for the plant dataset [53] (cf. Figure 3a).

Bit Rate (bpp)	Average PSNR (dB)					
	RGB	YC _b C _r	KLT	aKLT	FST	RCA
0.0625	26.75	27.07	27.28	27.23	26.81	25.68
0.125	27.86	28.31	28.44	28.39	27.92	26.26
0.25	29.09	29.53	29.58	29.55	29.13	27.21
0.5	30.53	30.78	30.81	30.90	30.49	28.05
1.0	32.39	32.39	32.28	32.46	32.07	29.23
2.0	34.86	34.68	34.48	34.81	34.46	30.42

Table 3 Mean and standard deviation of reconstruction accuracy performance for the images of Figure 4, using the aKLT and 100 different initializations.

Bit Rate (bpp)	Average PSNR (dB)			
	Jelly Beans	Bretagne	Aerial	Retina
0.0625	22.52 ± 0.08	35.32 ± 0.06	21.09 ± 0.03	33.82 ± 0.08
0.125	25.71 ± 0.06	36.85 ± 0.05	22.16 ± 0.07	36.10 ± 0.07
0.25	29.22 ± 0.11	38.33 ± 0.04	23.48 ± 0.09	38.39 ± 0.06
0.5	33.05 ± 0.06	39.98 ± 0.04	25.20 ± 0.10	40.40 ± 0.05
1.0	37.41 ± 0.04	42.25 ± 0.02	27.27 ± 0.12	42.43 ± 0.05
2.0	42.71 ± 0.06	45.20 ± 0.02	30.12 ± 0.17	44.92 ± 0.06

the KLT. As also found by others in some cases [43], at higher bit rates the RGB representation may result in higher performance, due to noise amplification effects of the other transformations and reduced quantization (see solid green line in Figure 4a, in the range of bit rates close to 2 bpp). The supervised FST, D, shows PSNR performance comparable to RGB, with slight improvement only at low bit rates. On the other hand, the supervised RCA, L, performs worse than baseline RGB, probably due to the lack of orthogonality (Gershikov et al. [19] observe a dependence of PSNR performance on the condition number of the color transformation matrix).

Figure 5 offers a visual comparison between the components of the color spaces. The RGB channels appear highly redundant (particularly the first two, i.e. red and green), total signal energy is spread across all channels, and the distributions of intensity values span the entire 0 to 255 range. In the YC_bC_r, the distributions of second and third channel cover a smaller range of values, however signal energy is again dispersed over all three channels. On the other hand, KLT and aKLT present highly similar output, with most of the signal energy (66-70%) compacted in the first channel, and narrow and peaked distributions in second and third channel, containing a relatively low amount of information. On the other hand, the supervised FST concentrates more energy (64%) in the second channel, as the first one (i.e. projection on Fisher's discriminant vector) is purposely designed to exhibit good discrimination capabilities of the plant objects. Such features render the KLT, aKLT, and FST ideal for the coding of color images, because the channels accounting for less energy can be effectively subsampled.

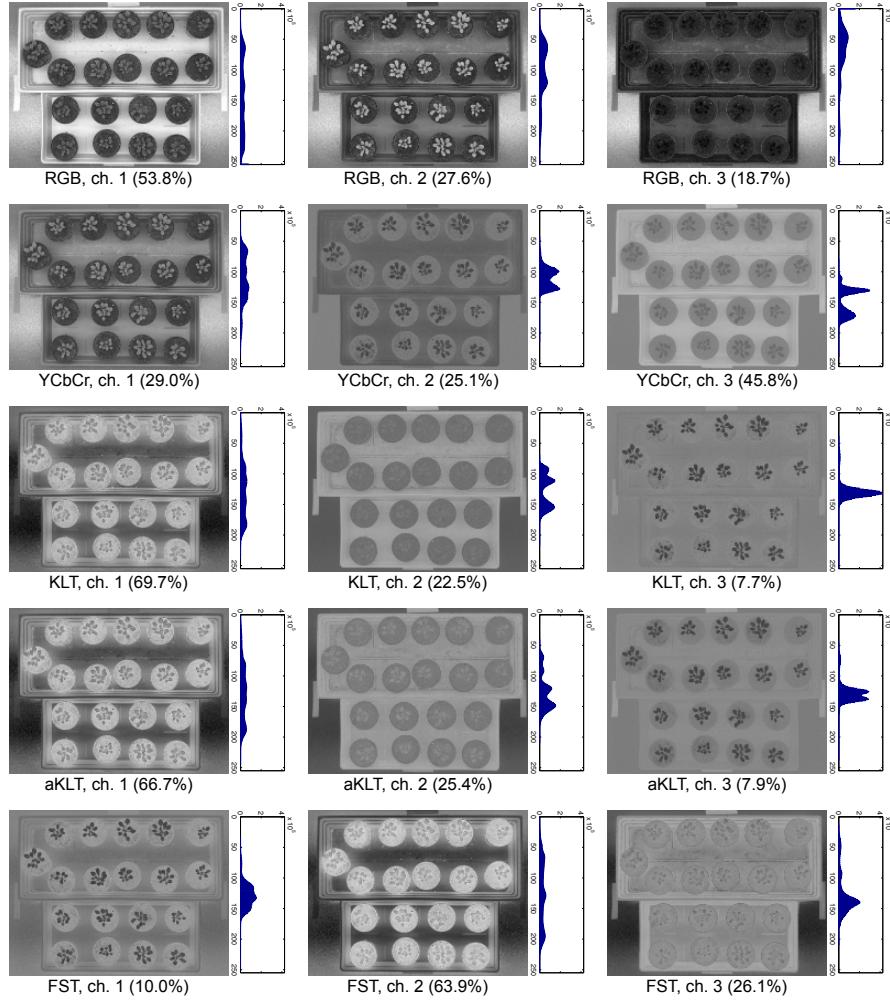


Fig. 5 Projection of the image of Figure 3a in a variety of color spaces. Next to each channel is shown the corresponding histogram of intensity values, and in parentheses the percentage of signal energy contained in that component.

Unsupervised transform. In order to assess the sensitivity of the aKLT to the random initialization of the vectors \mathbf{a}_2 and \mathbf{a}_3 in the matrix \mathbf{A} (cf. Section 2.1), we compute 100 different realizations of $\tilde{\mathbf{K}}$ for each of the test images in Figure 4. As shown in Table 3, the aKLT behaves consistently, and variations in PSNR performance due to different initial values are on average approximately only 0.2%.

Furthermore, the aKLT exhibits good decorrelating capabilities. As shown in Table 4, in the RGB domain, the channels of the test images of Figure 4 present on average strong linear correlation. Inter-channel linear correlation is only moderately

Table 4 Average inter-channel linear correlation of the test images of Figure 4. For the aKLT, average results are shown, obtained using 100 different initializations (see Section 2.1).

Transform	Correlation		
	ch. 1–2	ch. 1–3	ch. 2–3
RGB	0.84	0.71	0.91
YC _b C _r	-0.39	0.11	-0.71
aKLT	0.04	0.09	0.11
KLT	0.00	0.00	0.00

reduced by the YC_bC_r, whereas the aKLT is able to achieve the almost complete decorrelation obtained by the optimal KLT.

Supervised transforms and application-aware compression. Figure 6 provides several visual examples of the supervised transform on a variety of different images, showing its ability to identify the objects of interest in the test images, even when major changes occur in the scene with respect to the training data (e.g., compare background appearance of the images in Figures 6u and 6v). This approach is chiefly based on color information, therefore after learning the transform D on the image of a black horse (Figures 6a and 6g), only the black stripes of the zebra in Figure 6f result in a high response, while the white stripes are regarded as background (cf. Figure 6l). On the other hand, the transform D estimated from training data in Figures 6m and 6q, is able to selectively identify only the red balloons in the image of Figure 6p.

Figure 7 compares the approaches from an application standpoint. Color transformation alone provides up to 1.26 dB improvement in PSNR of the foreground (plant) regions relative to RGB, with the FST now obtaining competitive performance. The supervised transforms do not show remarkable improvements with respect to the other approaches, probably due to lacking decorrelation capabilities for these images, causing losses in bit rate performance.

Supervised transform for ROI detection. The separation property of the supervised color transform can be further exploited in applications in which the objects of interest can be discriminated by color features (e.g., plant objects in our dataset can be separated from the background based on color information). Therefore, we envision the use of the supervised color transform to obtain from the transformed image a region of interest (ROI) estimate, that can be used in an encoder with ROI coding capability (e.g. the JPEG 2000 standard [56]). With respect to other approaches obtaining the ROI information from a detection module external to the encoder [46], we propose for the first time to combine color transformation and ROI estimation in a single framework, identifying potential ROI masks solely on the basis of the class separation capabilities of the supervised transform, thus reducing computational overhead at the encoder.

When using the FST approach, the first channel of the FST domain, $y_i^{(1)} = \mathbf{d}_1^T \mathbf{x}_i$, corresponds to the projection on Fisher's discriminant vector (cf. Figure 5, bottom

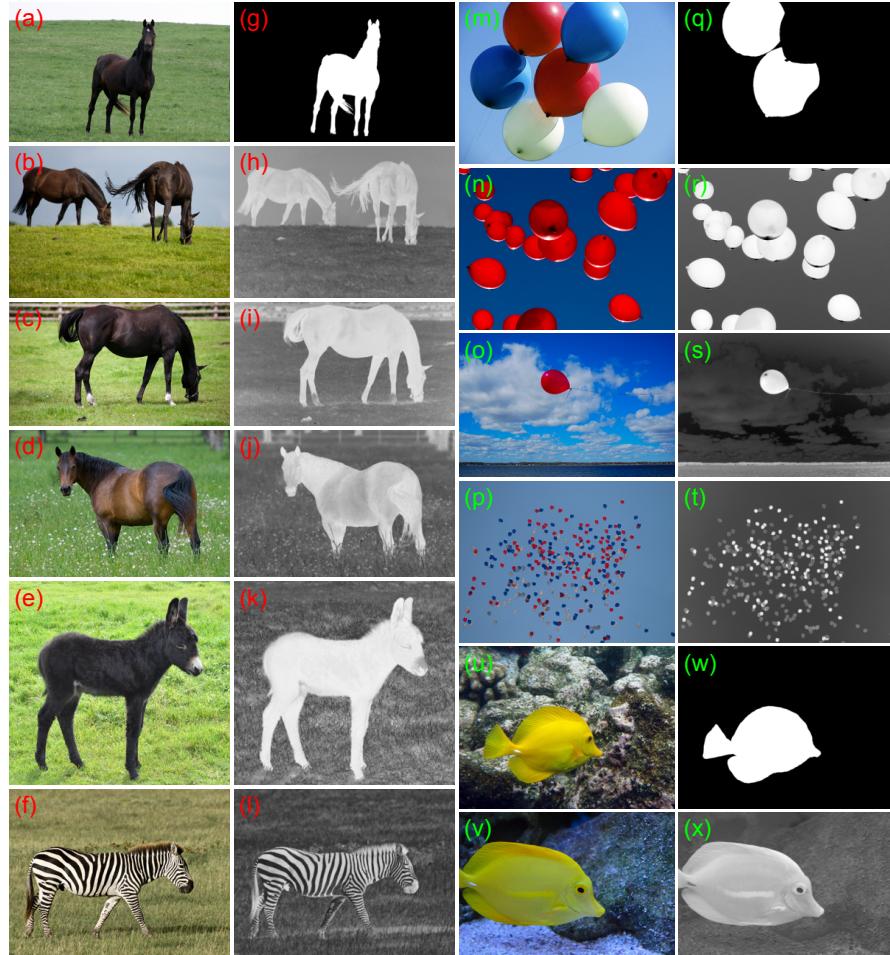


Fig. 6 Demonstration of the supervised transform, using images of: (a)-(f) horses, (m)-(p) balloons, and (u)-(v) fish. For each category, a single FST was obtained, using for training, respectively, images in (a), (m), and (u), and corresponding ground truth segmentations (i.e. (g), (q), and (w), respectively). Images in (h)-(l), (r)-(t), and (x) visualize the projections of the test images on the first component of the FST.

row). In an unseen image, to obtain an ROI estimate, $\Gamma(D, \theta^*) \in \{0, 1\}^n$, we decide the class of a pixel (foreground or background) based on a single threshold θ^* on the values of $y_i^{(1)}$. We estimate θ^* from our training set, maximizing the Dice Similarity Coefficient (DSC):

$$\theta^* = \arg \max_{\theta} \frac{2 \cdot |\Gamma_{\text{GT}} \cap \Gamma(D, \theta)|}{|\Gamma_{\text{GT}}| + |\Gamma(D, \theta)|}, \quad (6)$$

between the ground truth of pixels, Γ_{GT} , and the classification, $\Gamma(D, \theta)$, obtained using D and threshold θ on the training data. Supervised transform D and threshold

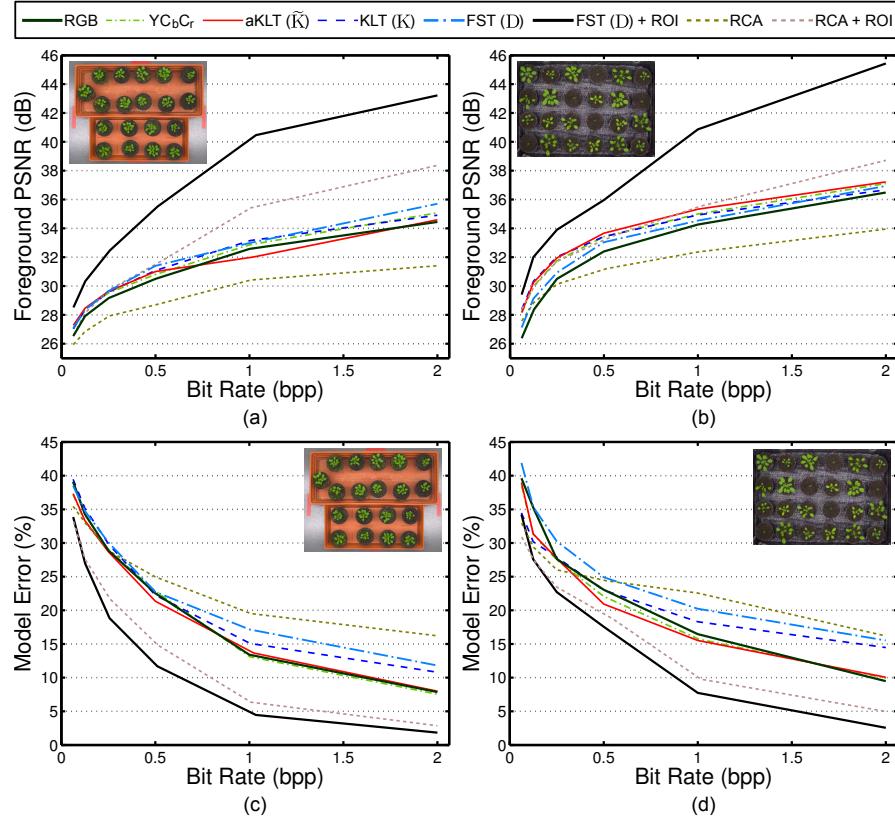


Fig. 7 R-D performance using the proposed transforms (in solid curves) in comparison to others (dashed curves) on the plant image data of Figure 3 using application-aware metrics: (a)-(b) reconstruction accuracy of the objects of interest, and (c)-(d) model error $E_{\mathcal{M}}$ of Eq. (5). Results in (a) and (c) are averaged over 19 test images.

θ^* are generally assumed to be obtained offline, therefore we estimate θ^* using a parameter sweeping strategy. On the other hand, if an application requires that θ^* be obtained at the sensor, statistical assumptions on the distribution of the data (e.g., Gaussian) would lead to closed-form solutions for finding the optimal θ^* efficiently [59]. When using the RCA approach, ROI estimation proceeds analogously.

When used in a spatial decorrelation context to estimate an ROI, combined with the ROI coding feature of JPEG 2000, the FST + ROI approach obtains a major improvement at all bit rates: 2 to 8.8 dB increase in foreground PSNR, and 13 to 77% reduction in application error (cf. black solid line in Figures 7a and 7c). When using the same FST on a test image of Arabidopsis plants acquired under significantly different scene conditions (Figure 3c), the FST + ROI approach proves robust, obtaining again best performance (cf. Figures 7b and 7d). On the contrary, although the RCA approach is capable of detecting the regions of interest in an image in both testing scenarios, when projecting the images in the so obtained color space, the

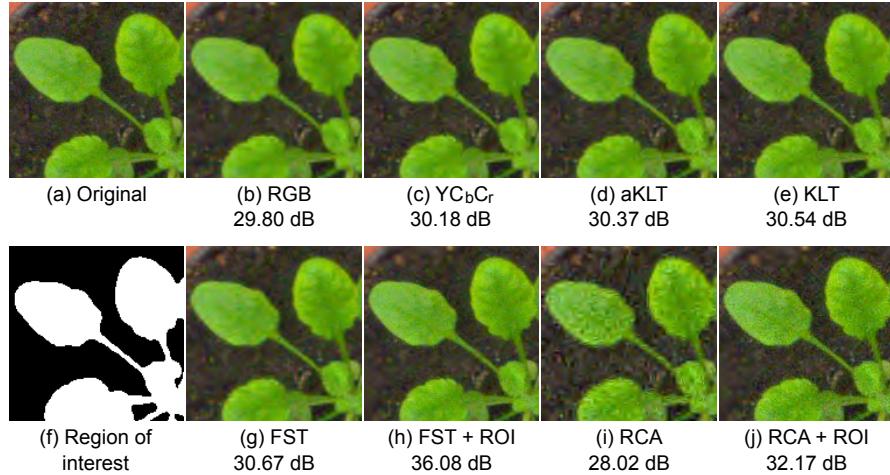


Fig. 8 (a) Detail of the image in Figure 3a. (b)-(e), (g)-(j) Reconstructed images after compression at 0.5 bpp using the JPEG 2000 standard and several color space transformations. Foreground PSNR between (a) and each of the reconstructed versions, calculated for the plant region indicated in (f), is also reported.

new intensity values are altered in a way that the benefits of the application-aware transform are diminished (or surpassed) by numerical errors introduced by the combination of forward and reverse color transformation and compression (cf. yellow dashed line in Figure 7).

A visual comparison of reconstructed images after compression with JPEG 2000 and all color transforms adopted in this work is shown in Figure 8. The RGB image appears oversmoothed, whereas the decorrelating transforms (YC_bC_r , aKLT, and KLT) exhibit higher image fidelity and appear increasingly richer in details (cf. Figures 8b, 8c, 8d, and 8e). The supervised FST alone already provides good reconstruction accuracy, however, the FST + ROI outperforms all other methods (cf. Figures 8g and 8h). The artifacts introduced by the RCA are evident in Figure 8i, and even when coupled with ROI coding the approach produces a noisy image (Figure 8j).

The results envision different use cases for the proposed approaches. The aKLT is general purpose and can be efficiently calculated on a per image basis to target reconstruction accuracy. On the other hand, the supervised approach is best suited for application-aware compression or enhancement scenarios, and since it does require supervision (which can be costly to obtain at the sensor) is assumed to be computed offline. The regularized versions of Eq. (4) are highly dependent on the free parameter λ and their performance is found to lie within the bounds of the other two. When varying the value of λ , the new transform D' exhibits behavior very close to either the supervised or the unsupervised transform, respectively. Therefore, it is best to exploit the classification abilities of the supervised FST to focus bits in appropriate places in the image, which is considerably less computationally demanding.

4 Conclusions

We address the problem of designing image-adaptive color space transformations for coding and enhancement applications. In recognition of the superior data-dependent KLT with respect to fixed transforms such as the YC_bC_r (as also confirmed by our experimental results), we derive a low-complexity approximation, the aKLT, capable of comparable performance. Our proposed aKLT achieves lower computational complexity than other KLT approaches in the literature, which is expected to result in proportionally reduced computation time, when devising optimized implementations. This will ease adoption on resource-constrained devices or in time-critical applications.

We also consider an application-aware compression setting, in which prior knowledge is available on the objects of interest in the scene. We formulate a novel approach to design color transforms with class separation capabilities, using supervised learning methods. Inspired by the linear discriminant analysis, we measure class separability using the Fisher's discrimination criterion, and adopt the Foley-Sammon transform to obtain an orthogonal application-aware color transform. We also adopt metric learning approaches, however they focus only on class separation (renouncing also orthogonality) and are found to result in lower performance in a compression context. The proposed unsupervised and supervised approaches, for which closed-form solutions are presented, address different requirements, therefore we also consider optimization strategies to combine the two approaches. In the experiments, we also showcase the use of the separation property of the supervised transforms to detect regions of interest in an image, and inform the encoder where to focus bit rate spatially.

In an enhancement context, the proposed supervised approach can be used to enhance the contrast of objects of interest in the scene, incorporating also knowledge of the application and the expert system, and facilitate human-computer interaction, or for automatic content-aware cropping or resizing of large images for visualization on small displays [60].

When coupled with quantizer design even greater bit rate savings are possible, but that would in general violate standard compliance. Increased image resolution or video applications are expected to emphasize the benefits of the proposed approaches. While we adopt the JPEG 2000 standard, our methodology is general and can be adapted to other coding schemes. Reversible integer approximations of the proposed transforms can also be obtained [25, 26], for lossless or progressive lossy-to-lossless compression of color images.

Acknowledgements This work was partially supported by a Marie Curie Action: “Reintegration Grant” (grant number 256534) of the EU’s Seventh Framework Programme (FP7).

References

1. Abadpour, A., Kasaei, S.: Color PCA eigenimages and their application to compression and watermarking. *Image and Vision Computing* **26**(7), 878–890 (2008)
2. Agin, G.J.: Computer vision systems for industrial inspection and assembly. *Computer* **13**(5), 11–20 (1980)
3. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis, 3rd edn. Wiley (2003)
4. Andrecut, M.: Parallel GPU implementation of iterative PCA algorithms. *Journal of Computational Biology* **16**(11), 1593–1599 (2009)
5. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D.: Learning a Mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research* **6**, 937–965 (2005)
6. Celebi, M.E., Kingravi, H.A., Celiker, F.: Fast colour space transformations using minimax approximations. *IET Image Processing* **4**(2), 70–80 (2010)
7. Celebi, M.E., Schaefer, G. (eds.): Color Medical Image Analysis, *Lecture Notes in Computational Vision and Biomechanics*, vol. 6. Springer (2013)
8. Chen, Y., Hao, P., Dang, A.: Optimal transform in perceptually uniform color space and its application in image coding. In: International Conference on Image Analysis and Recognition (ICIAR), pp. 269–276 (2004)
9. Clausen, C., Wechsler, H.: Color image compression using PCA and backpropagation learning. *Pattern Recognition* **33**(9), 1555–1560 (2000)
10. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: International Conference on Machine Learning (ICML), pp. 209–216. ACM (2007)
11. Dikbas, S., Arici, T., Altunbasak, Y.: Chrominance edge preserving grayscale transformation with approximate first principal component for color edge detection. In: International Conference on Image Processing (ICIP), vol. 2, pp. 261–264 (2007)
12. Dogane, B.M., Murthi, M.N.: Quantization for classification accuracy in high-rate quantizers. In: Digital Signal Processing Workshop, pp. 277–282 (2011)
13. Du, Q., Fowler, J.E.: Hyperspectral image compression using JPEG2000 and principal component analysis. *IEEE Geoscience and Remote Sensing Letters* **4**(2), 201–205 (2007)
14. Du, Q., Fowler, J.E.: Low-complexity principal component analysis for hyperspectral image compression. *International Journal of High Performance Computing Applications* **22**(4), 438–448 (2008)
15. Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7**(2), 179–188 (1936)
16. Fleury, M., Downton, A.C., Clark, A.F.: Karhunen-Loëve transform: An exercise in simple image-processing parallel pipelines. In: Euro-Par '97 Parallel Processing, *Lecture Notes in Computer Science*, vol. 1300, pp. 815–819. Springer (1997)
17. Foley, D.H., Sammon, J.W.: An optimal set of discriminant vectors. *IEEE Transactions on Computers* **C-24**(3), 281–289 (1975)
18. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press Professional, Inc., San Diego, CA, USA (1990)
19. Gershikov, E., Porat, M.: Does decorrelation really improve color image compression? In: International Conference on Systems Theory and Scientific Computation (ISTASC), pp. 306–309 (2005)
20. Gillespie, A.R., Kahle, A.B., Walker, R.E.: Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches. *Remote Sensing of Environment* **20**(3), 209–235 (1986)
21. Gillespie, A.R., Kahle, A.B., Walker, R.E.: Color enhancement of highly correlated images. II. Channel ratio and “chromaticity” transformation techniques. *Remote Sensing of Environment* **22**(3), 343–365 (1987)
22. Grant, M.C., Boyd, S.P.: Graph implementations for nonsmooth convex programs. In: Recent Advances in Learning and Control, *Lecture Notes in Control and Information Sciences*, vol. 371, pp. 95–110. Springer London (2008)

23. Han, S.E., Tao, B., Cooper, T., Tastl, I.: Comparison between different color transformations for the JPEG 2000. In: PICS 2000, pp. 259–263 (2000)
24. Hao, P., Shi, Q.: Comparative study of color transforms for image coding and derivation of integer reversible color transform. In: International Conference on Pattern Recognition (ICPR), vol. 3, pp. 224–227 (2000)
25. Hao, P., Shi, Q.: Matrix factorizations for reversible integer mapping. IEEE Transactions on Signal Processing **49**(10), 2314–2324 (2001)
26. Hao, P., Shi, Q.: Reversible integer KLT for progressive-to-lossless compression of multiple component images. In: International Conference on Image Processing (ICIP), vol. 1, pp. 633–636 (2003)
27. Higham, N.J.: Functions of Matrices: Theory and Computation. SIAM (2008)
28. Horn, B.K.P., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. Journal of the Optical Society of America A **5**(7), 1127–1135 (1988)
29. International Commission on Illumination (CIE): CIE Colorimetry - Part 4: 1976 L*a*b* Colour Space. Joint ISO/CIE Standard ISO 11664-4:2008(E)/CIE S 014-4/E:2007, CIE Publication (2007)
30. International Telecommunication Union: Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. ITU-R Recommendation BT.601-7 (2011)
31. Jin, Z., Yang, J.Y., Hu, Z.S., Lou, Z.: Face recognition based on the uncorrelated discriminant transformation. Pattern Recognition **34**(7), 1405–1416 (2001)
32. Jolliffe, I.: Principal Component Analysis, 2nd edn. Springer (2002)
33. Jones, G.A., Paragios, N., Regazzoni, C.S. (eds.): Video-Based Surveillance Systems: Computer Vision and Distributed Processing. Kluwer Academic Publishers (2002)
34. Jošt, R., Antikainen, J., Havel, J., Herout, A., Zemčík, P., Hauta-Kasari, M.: Real-time PCA calculation for spectral imaging (using SIMD and GP-GPU). Journal of Real-Time Image Processing **7**(2), 95–103 (2012)
35. Kassim, A.A., Lee, W.S.: Embedded color image coding using SPIHT with partially linked spatial orientation trees. IEEE Transactions on Circuits and Systems for Video Technology **13**(2), 203–206 (2003)
36. Kim, H.M., Kim, W.S., Cho, D.S.: A new color transform for RGB coding. In: International Conference on Image Processing (ICIP), pp. 107–110 (2004)
37. Kouassi, R.K., Devaux, J.C., Gouton, P., Paindavoine, M.: Application of the Karhunen-Loeve transform for natural color images analysis. In: Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp. 1740–1744 (1997)
38. Kountchev, R., Kountcheva, R.: New method for adaptive Karhunen-Loeve color transform. In: International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services (TELSIKS), pp. 209–216 (2009)
39. Kulis, B.: Metric learning: A survey. Foundations and Trends in Machine Learning **5**(4), 287–364 (2013)
40. Limb, J.O., Rubinstein, C.B., Thompson, J.: Digital coding of color video signals – a review. IEEE Transactions on Communications **25**(11), 1349–1385 (1977)
41. Majchrowicz, M., Kapusta, P., Was, L., Wiak, S.: Application of general-purpose computing on graphics processing units for acceleration of basic linear algebra operations and principal components analysis method. In: Man-Machine Interactions 3, *Advances in Intelligent Systems and Computing*, vol. 242, pp. 519–527 (2014)
42. Malvar, H.S., Sullivan, G.J., Srinivasan, S.: Lifting-based reversible color transformations for image compression. In: Proceedings of SPIE, Applications of Digital Image Processing XXXI, vol. 7073 (2008)
43. Marpe, D., Kirchhoffer, H., George, V., Kauff, P., Wiegand, T.: Macroblock-adaptive residual color space transforms for 4:4:4 video coding. In: International Conference on Image Processing (ICIP), pp. 3157–3160 (2006)
44. Minervini, M., Abdelsamea, M.M., Tsafaris, S.A.: Image-based plant phenotyping with incremental learning and active contours. Ecological Informatics **23**, 35–48 (2014). Special Issue on Multimedia in Ecology and Environment

45. Minervini, M., Rusu, C., Tsaftaris, S.A.: Unsupervised and supervised approaches to color space transformation for image coding. In: International Conference on Image Processing (ICIP) (2014)
46. Minervini, M., Tsaftaris, S.A.: Application-aware image compression for low cost and distributed plant phenotyping. In: International Conference on Digital Signal Processing (DSP), pp. 1–6 (2013)
47. Palus, H.: Representations of colour images in different colour spaces. In: The Colour Image Processing Handbook, pp. 67–90. Springer (1998)
48. Penna, B., Tillo, T., Magli, E., Olmo, G.: A new low complexity KLT for lossy hyperspectral data compression. In: International Geoscience and Remote Sensing Symposium (IGARSS), vol. 7, pp. 3525–3528 (2006)
49. Porikli, F., Tuzel, O.: Fast construction of covariance matrices for arbitrary size image windows. In: International Conference on Image Processing (ICIP), pp. 1581–1584 (2006)
50. Poynton, C.: A guided tour of color space. In: New Foundations for Video Technology, pp. 167–180 (1995)
51. Pratt, W.K.: Spatial transform coding of color images. IEEE Transactions on Communication Technology **19**(6), 980–992 (1971)
52. Rubinstein, C.B., Limb, J.O.: Statistical dependence between components of a differentially quantized color signal. IEEE Transactions on Communications **20**(5), 890–899 (1972)
53. Scharr, H., Minervini, M., Fischbach, A., Tsaftaris, S.A.: Annotated image datasets of rosette plants. Tech. Rep. FZJ-2014-03837, Forschungszentrum Jülich GmbH, Jülich, Germany (2014)
54. Schönemann, P.H.: A generalized solution of the orthogonal Procrustes problem. Psychometrika **31**(1), 1–10 (1966)
55. Shi, Y.Q., Sun, H.: Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards, 2nd edn. CRC Press (2008)
56. Skodras, A., Christopoulos, C., Ebrahimi, T.: The JPEG 2000 still image compression standard. IEEE Signal Processing Magazine **18**(5), 36–58 (2001)
57. Soyak, E., Tsaftaris, S.A., Katsaggelos, A.K.: Low-complexity tracking-aware H.264 video compression for transportation surveillance. IEEE Transactions on Circuits and Systems for Video Technology **21**(10), 1378–1389 (2011)
58. Staal, J., Abramoff, M., Niemeijer, M., Viergever, M., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging **23**(4), 501–509 (2004)
59. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Academic Press (2009)
60. Vaquero, D., Turk, M., Pulli, K., Tico, M., Gelfand, N.: A survey of image retargeting techniques. In: SPIE Applications of Digital Image Processing XXXIII, vol. 7798, p. 779814 (2010)
61. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine **26**(1), 98–117 (2009)
62. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
63. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research **10**, 207–244 (2009)
64. Wolf, S.G., Ginosar, R., Zeevi, Y.Y.: Spatio-chromatic model for colour image processing. In: International Conference on Pattern Recognition (ICPR), vol. 1, pp. 599–601 (1994)
65. Yang, Y., Yuhua, P., Zhaoguang, L.: A fast algorithm for YCbCr to RGB conversion. IEEE Transactions on Consumer Electronics **53**(4), 1490–1493 (2007)
66. Zabalza, J., Ren, J., Ren, J., Liu, Z., Marshall, S.: Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging. Applied Optics **53**(20), 4440–4449 (2014)
67. Zhang, L., Dong, W., Zhang, D., Shi, G.: Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition **43**(4), 1531–1549 (2010)
68. Zheng, W., Zhao, L., Zou, C.: Foley-Sammon optimal discriminant vectors using kernel approach. IEEE Transactions on Neural Networks **16**(1), 1–9 (2005)