# Class-driven Color Transformation for Semantic Labeling

Arash Shahriari[1], Jose M. Alvarez[1,2] and Antonio Robles-Kelly[1,2]

[1] School of Engineering, Australian National University, Action ACT 2601, Australia
[2] NICTA*, Locked Bag 8001, Canberra ACT 2601, Australia

**Abstract.** We propose a novel class-driven color transformation aimed at semantic labeling. In contrast with other approaches elsewhere in the literature, our approach is a supervised one employing class information to learn a color transformation. Our method maps image color to a target space with maximum pairwise distances between classes and minimum scattering within each of them. To compute the color transformation, we pose the problem in terms of a composition of two mappings. The first mapping employs a pairwise discriminant cost function minimized through a steepest descent optimization to map the image color data onto a space spanned by the class set. It targets better separability between distinct classes as well as less scattering within each individual class. The second mapping corresponds to subspace projection of this class data to a target space with same dimensionality of image color data. To preserve distances attained by the first of the mappings, this subspace projection is effected making use of metric multi-dimensional scaling. We report our experiments on MSRC-21 and SBD datasets, where our method consistently improves overall and average performances of well-known publicly available TextonBoost and DARWIN multiclass segmentation frameworks at a negligible computational cost. These results confirms our contribution towards reflection of higher distinction in color space by imposing better separability in a novel representation which is learned from class information of the dataset under consideration.

## 1 Introduction

Color has been used as a cue for numerous tasks in computer vision [1]. Due to its importance, a number of color spaces and descriptors have been formulated to address problems spanning from accurate capture and reproduction of images acquired by digital camera sensors [2] to scene and object recognition [1].

Color, as perceived by the observer, is the result of interactions between light sources, material reflectance, surface texture and other photometric effects due to object shape and shadows. Photometric invariance is often achieved by use of

* National ICT Australia (NICTA) is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Center of Excellence program.

surface reflectance as a means toward classification and recognition via a descriptor which is robust to changes in illumination, noise, geometric and photometric effects. For instance, Nayar *et al.* [3] proposed a method of object recognition based on the reflectance ratio between object regions. Dror *et al.* [4] described a vision system that learned the relationship between surface reflectance and certain statistics computed from gray-scale images. Slater *et al.* [5] used a set of Gaussian filters to derive moment invariants for recognition. Jacobs *et al.* [6] employed image ratios for comparing images under variable illumination. Lin *et al.* [7] utilized an eigen-space of chromaticity distributions to obtain illumination direction and color and specularity-invariants for three-dimensional object recognition. Lenz *et al.* [8] deployed perspective projections in canonical space of color signals to separate intensity from chromaticity and recover a three-dimensional color descriptor.

The performance of a number of computer vision methods not only depends on color descriptor used but also space or gamut in which they are defined [9]. This is why the color in an image may be adjusted and adapted using a suitable transformation [10]. The selection of an optimal transformation is not straightforward since, different color spaces may be better adapted to handle textures or complex shapes [11]. For segmentation, the CIE, LUV or Lab spaces [12] are often employed as they map each RGB pixel value in image to a point in color space for which the pairwise deviation in perceived color is equal to Euclidean distance in feature space [13].

Here, we note that color transformations found elsewhere in the literature are often derived by perceptual criteria rather by image information. Thus, we propose a novel approach to find a color transformation for semantic labeling based on class information of image dataset for the application at hand. This transformation is computed via supervised learning from dataset and could be employed as a general preprocessing before feature extraction step in standard pipeline of pixel-wise classifiers.

The core idea is to map primary color space *i.e.* RGB, to a new representation of the same dimensionality *i.e.* target space with maximum class separation. To this end, we decompose the problem into the recovery of two mappings. The former concerns mapping of the primary color space to a class space spanned by the classes in the dataset and the latter corresponds to subspace projection from the class space to above target space. Our proposal in essence follows the practice of many successful learning schemes, where a combination of two mappings could unravel desired structure of data. While our formulation for both mappings is inspired by well-known machine learning techniques, combination in this context is a novel contribution. Moreover, our choice of color space, provides fair comparisons with state-of-the-art which use color features in their experiments. It also achieves a low-cost machinery that could process large amount of images in mobile platforms.

For the mapping of the primary color to the class space, we employ a method akin to the pairwise discriminant analysis developed by Fu and Robles-Kelly [14]. Note that here, the aim is to map the color channels to a higher-dimensional

space spanned by the class set using an approach devoid of matrix inversions, whereas the method in [14] is a subspace projection method. This also contrasts with approaches such as Linear Discriminant Analysis (LDA) [15], where class-specific covariance is used to define within-class scattering while between-class scatter is considered to be uniform for distinct classes. The lack of between-class scatter specificity [16] is compounded by the burden of dimensionality. To deal with high-dimensional data, a number of approaches have been proposed. These include independence rule [17], feature annealed independence rule [18] and nearest shrunken centroid classifier [19]. Nonetheless, these methods may still be potentially unstable due to matrix inversions.

As mentioned earlier, the subspace projection is tackled using metric Multi-dimensional Scaling (MDS) to preserve the distances achieved by the first of our mappings. Given a highly distinct color representation, it is feasible to utilize any sophisticated feature sets in classification process. For instance, in our experiments, TextonBoost [20] uses color, histogram of oriented gradients (HOG), and pixel location features and DARWIN [21] employs RGB color of the pixel, dense HOG, LBP-like features, and averages over image rows and columns. We validate our class-driven color transformation by applying it as a general preprocessing step of semantic labeling process. Experiments conducted on MSRC-21 [22] and Stanford Background Dataset (SBD) [23] for semantic segmentation shows that our color transformation consistently improves the overall and average precisions of both TextonBoost and DARWIN pixel-wise segmentation algorithms at the cost of a simple matrix multiplication.

The rest of paper is organized as follows. In the next section, we expose our color transformation method. In Section 3, we discuss the implementation of our algorithm and its initialization. Finally, in Sections 4 and 5, we report experiments and then conclude on the work presented here respectively.

## 2 Class-Driven Color Transformation

To commence, let us define the generic color transformation problem as treated in this paper. Given the $M$ image color pixels $\mathcal{X} = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_M}\}$ labeled into $N$ distinct classes $\mathcal{L} = \{\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_N}\}$, our goal is to recover a matrix $\mathbf{A} \in \mathcal{R}^{3\times 3}$ that transforms input color pixels $\mathbf{x_i}$ onto a 3-dimensional vector $\mathbf{z_i}$ in the target space such that $\mathbf{z_i} = \mathbf{A}\mathbf{x_i}$ maximizes the separation between color pixels in distinct classes and minimizes the scattering in each individual class.

Posed in this way, we formulate the problem as an optimization one where objective function depends on both, dimensions of the target space and number of classes in the dataset under consideration. As a result, we decompose the matrix $\mathbf{A}$ as follows

$$\mathbf{A} = \mathbf{B}\mathbf{C} \tag{1}$$

where $\mathbf{C} \in \mathcal{R}^{N\times 3}$ is a mapping matrix that transforms the input color space to the $N$-dimensional class space, spanned by classes in the dataset and $\mathbf{B} \in \mathcal{R}^{3\times N}$ is another mapping that projects the class space onto the output target space.

With above decomposition, we can introduce our objective function as a composition of the form

$$\arg \min_{\mathbf{A}=\mathbf{BC}} f(\mathbf{A}) = h(\mathbf{B}) \circ p(\mathbf{C}) \tag{2}$$

where both $h(\cdot)$ and $p(\cdot)$ are cost functions that take the matrices $\mathbf{B}$ and $\mathbf{C}$ as their arguments.

It is worth discussing implications of the formulation above. Note that Equations 1 and 2 are a direct consequence of properties known for decomposition of matrices and composition of functions, where the functions $p(\cdot)$ and $h(\cdot)$ have been composed into objective $f(\mathbf{A})$. Moreover, if $p(\cdot)$ is convex, the optimization in Eq. (1) can be affected by recovering the matrix $\mathbf{C}$ to later optimizing $h(\cdot)$ with respect to $\mathbf{B}$ [24]. The minimization of the cost function $h(\mathbf{B})$ can then be treated in a manner akin to that used by linear feature extraction methods such as LDA or Maximum Margin Criterion (MMC) [25] which are often employed for utilizing label information to learn a linear transformation for classification. Indeed, since the matrix $\mathbf{B}$ is effectively a subspace projection matrix that maps the class space onto the target space, such methods may be used to optimize Eq. (1) in case the matrix $\mathbf{C}$ is at hand.

Further, by minimizing $p(\mathbf{C})$ and $h(\mathbf{B})$ in consecutive steps and choosing cost functions which are convex, $f(\mathbf{A})$ can be shown to be also convex [24]. This follows from the composition of a convex function with a non-increasing one and it is valid since $p(\mathbf{C})$ does not increase once minimized. Our goal of minimizing $h(\mathbf{B})$ is to preserve the class distinctions induced by minimizing $p(\mathbf{C})$ while projecting from the class space to the target space. Our inference on not increasing $p(\mathbf{C})$ comes from the fact that $h(\mathbf{B})$ just tries to fix the class distances resulted by $p(\mathbf{C})$ and hence, we consider minimization of $h(\mathbf{B})$ as an independent optimization while $p(\mathbf{C})$ has been minimized.

Thus, in Section 2.1 we turn our attention to mapping the color space to the class space using a convex cost function and later on, in Section 2.2, we elaborate on subspace projection of the class space to the target space.

### 2.1 Mapping of Color Space to Class Space

In order to map color values to the space spanned by classes in the dataset under consideration, we learn the relevant mapping by using a cost function which accumulates the combination of costs for pairs of binary classes. This is consistent with the notion that any multiclass classification problem can be converted to a number of binary ones by deploying a pairwise fusion framework [26]. This can also be viewed as a process akin to training of a classifier for every two classes and making final prediction based on the combination of decisions yielded by binary classifiers. Moreover, the matrix $\mathbf{C}$ can then be interpreted as a mapping of image color values onto the $N$-dimensional class space.

**Objective Function.** We view recovery of $\mathbf{C}$ as the optimization of function $p(\mathbf{C})$ such that pairwise cluster distances are maximized. As a result, we opt for

objective function presented in [14], given by

$$\arg \min_{\mathbf{C}^T\mathbf{C}=\mathbf{I}} p(\mathbf{C}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \beta_{i,j} g\left(\Phi_{i,j}(\mathbf{C})\right) \tag{3}$$

where $\Phi_{i,j}(C)$ is a class-pair dependent distance function, $\beta_{i,j}$ is a weight that moderates contribution of the class pair $i, j$ to the objective function and

$$g(\Phi_{i,j}(\mathbf{C})) = \frac{1}{1 + \exp(\gamma(\Phi_{i,j}(\mathbf{C}) - \tau))} \tag{4}$$

is a logistic regression function with parameters $\gamma$ and $\tau$ which maps class separability to pairwise costs. Importantly, $g(\cdot)$ takes values in the range $(-\infty, \infty)$ to bounded interval $[0, 1]$. This choice also implies that the function is monotonically decreasing to assign lower costs to increasing class separability values.

It is consistent with the notion that the optimization problem at hand should be solved such that the matrix $\mathbf{C}$ maximizes the costs, *i.e.* separability, for every pair of classes. The target function in Eq. (4) hence aims at maximizing these costs in a cumulative fashion. This equation also reflects the fact that, to derive final target function, we require a criterion to measure separability between two classes in the class space.

Indeed, the choice here is not unique. A straightforward way would be to use the same objective function as in [16] with different definitions for intra and inter-class scatter matrices on every pair of classes. However, it involves a matrix inversion operation for each pairwise within-class scatter matrix. This is undesirable since it can incur in numerical instability when small training sets are available for any of the classes or categories under consideration. Another way would be to assume an underlying Gaussian distribution for each of the above classes and employ information theoretic divergence measures, such as Kullback-Leibler divergence or Bhattacharyya distance with closed form solutions. Unfortunately, this would still require matrix inversion operations and hence may be unstable for applications with small training sets.

In this paper, we follow [14] and employ the distance between class centroids in the color space such that

$$\Phi_{i,j}(\mathbf{C}) = d_{i,j}^{(1)} - d_{i,j}^{(2)} - d_{i,j}^{(3)} \tag{5}$$

where $d_{i,j}^{(1)}$ is the distance between centroids subtracted by $d_{i,j}^{(2)}$ and $d_{i,j}^{(3)}$ which are projections of color scatterings for classes, along direction $\mathbf{C}^T\mu_{i,j}$ given by

$$\begin{aligned} d_{i,j}^{(1)} &= ||\mathbf{C}^T\mu_{i,j}|| \\ d_{i,j}^{(2)} &= \frac{\sqrt{\mu_{i,j}^T\mathbf{C}\mathbf{C}^T\mathbf{S}_i\mathbf{C}\mathbf{C}^T\mu_{i,j}}}{||\mathbf{C}^T\mu_{i,j}||} \\ d_{i,j}^{(3)} &= \frac{\sqrt{\mu_{i,j}^T\mathbf{C}\mathbf{C}^T\mathbf{S}_j\mathbf{C}\mathbf{C}^T\mu_{i,j}}}{||\mathbf{C}^T\mu_{i,j}||} \end{aligned} \tag{6}$$

Here, $\mu_i$ and $\mathbf{S_i}$ are the mean and scatter for the i$^{th}$ class of pixels in the color space and $\mu_{i,j}$ is defined as $\mu_i - \mu_j$.

Note that, we made no assumptions or constraints on standard/spherical cluster distances in our formulation. Both metrics are employed to define class-pair distance in Eq. (5) via aggregation of radial distances between centroids and angular projections of color value scatters in Eq. (6) which latter may create elliptical clusters.

**Steepest Descent Optimization.** At first glance, Eq. (3) appears to be a hard optimization problem. Surprisingly, it can be optimized using steepest descent optimization since it is defined on a Grassmann manifold [27]. In addition to unitary constraint, as a consequence of developments in [27], it can be shown that the objective function is invariant to any rotations in transformed feature space. It also assures that the objective function is a convex one. Thus, by building on recent advances in optimization theory, we can extend unconstrained optimization methods in Euclidean space to Grassmann manifold. Here, we use a projection-based steepest descent with backtracking line search based on [28].

The objective function above can be optimized using a steepest descent method whereby at iteration $t$ the matrix $\mathbf{C}$ can be updated using the rule

$$\mathbf{C}^{(t+1)} = \mathbf{C}^{(t)} + \lambda \Delta \mathbf{C}^{(t)} \tag{7}$$

where $\lambda$ is a step-size variable and the descent direction is given by

$$\Delta \mathbf{C}^{(t)} = -(\mathbf{I} - \mathbf{C}^{(t)T}\mathbf{C}^{(t)})\nabla_{\mathbf{C}} p(\mathbf{C}^{(t)}) \tag{8}$$

One of the main benefits of this steepest descent approach is that, in contrast to traditional subspace projection methods, the optimization of $p(\cdot)$ can be effected without any need for matrix inversions. Moreover, note that the function $\Phi_{i,j}(\mathbf{C})$ is not a metric in the sense that it can be negative. Nonetheless, this is not a problem as we are not using it directly for optimization purpose, rather it is treated as a variable in the objective function.

To appreciate this more clearly, we proceed to compute gradient of the function $p(\cdot)$ *i.e.* $\nabla_{\mathbf{C}} p(\mathbf{C})$ as gradient of the cost function $g(\cdot)$ in Eq. (3) with respect to $\mathbf{C}$. It can be expressed in closed form as follows

$$\nabla_{\mathbf{C}} p(\mathbf{C}) = -\sum_i \sum_j \gamma \beta_{i,j} (\Phi_{i,j}(\mathbf{C}) - \tau) g^2(\Phi_{i,j}(\mathbf{C})) \nabla_{\mathbf{C}} \Phi_{i,j} \tag{9}$$

where

$$\begin{aligned}
\nabla_{\mathbf{C}}\Phi_{i,j}(\mathbf{C}) &= \nabla_{\mathbf{C}} d_{i,j}^{(1)} - \frac{(\Gamma_i + \Gamma_j)\nabla_{\mathbf{C}} d_{i,j}^{(1)}}{{d_{i,j}^{(1)}}^2} - \frac{d_{i,j}^{(1)}(\nabla_{\mathbf{C}}\Gamma_i + \nabla_{\mathbf{C}}\Gamma_j)}{{d_{i,j}^{(1)}}^2} \\
\nabla_{\mathbf{C}} d_{i,j}^{(1)} &= \frac{1}{d_{i,j}^{(1)}}\mu_{i,j}\mu_{i,j}^T\mathbf{C} \\
\nabla_{\mathbf{C}}\Gamma_i &= \frac{2}{\Gamma_i} sym(\mu_{i,j}\mu_{i,j}^T\mathbf{C}\mathbf{C}^T\mathbf{S}_i)\mathbf{C} \\
\nabla_{\mathbf{C}}\Gamma_j &= \frac{2}{\Gamma_j} sym(\mu_{i,j}\mu_{i,j}^T\mathbf{C}\mathbf{C}^T\mathbf{S}_j)\mathbf{C}
\end{aligned} \tag{10}$$

and $sym(\mathbf{\Theta}) = \dfrac{\mathbf{\Theta} + \mathbf{\Theta}^T}{2}$ denotes a symmetry inducing operator for matrix $\mathbf{\Theta}$. In the equations above, we used the shorthand $\Gamma_i = \sqrt{\mu_{i,j}^T\mathbf{C}\mathbf{C}^T\mathbf{S}_i\mathbf{C}\mathbf{C}^T\mu_{i,j}}$.

### 2.2 Subspace Projection of Class Space to Target Space

Once the matrix $\mathbf{C}$ is at hand, we focus our attention on recovery of the matrix $\mathbf{B}$. As mentioned earlier, this can be viewed as a subspace projection matrix which maps the class space onto the target space. This potentially allows for any convex subspace projection method to be employed for computing $\mathbf{B}$. Moreover, literature on multi-dimensional scaling and subspace projection is vast.

Here, we aim at preserving pairwise cluster distances derived from the learned matrix $\mathbf{C}$. This naturally leads to application of linear and non-linear embedding techniques for dimensionality reduction that attempt to preserve global or local properties of original data in low-dimensional representations. Here, we use metric Multi-dimensional Scaling (MDS) due to both, its capacity to preserve pairwise distances in the class space and the fact that, it is a natural generalization of classical approaches elsewhere in the literature. Moreover, metric MDS can employ a wide variety of loss functions. We employ the stress function to measure the error between the pairwise distances in the high-dimensional class space and the low-dimensional target space. Thus, the cost function $h(\mathbf{B})$ becomes

$$h(\mathbf{B}) = -\sum_{\mathbf{x}_i \in \mathcal{X}} (||\mathbf{y}_i - \mathbf{y}_j||^2 - ||\mathbf{B}(\mathbf{y}_i - \mathbf{y}_j)||^2)^2 \tag{11}$$

where $||\cdot||$ is vector norm and we have used shorthand $\mathbf{y}_i = \mathbf{C}\mathbf{x}_i$ to denote instances in the class space corresponding to the pixel color value $\mathbf{x}_i \in \mathcal{X}$. Note that, in the stress function above, the term $||\mathbf{B}(\mathbf{y}_i - \mathbf{y}_j)||$ is effectively the Euclidean distance in the target space whereas $||\mathbf{y}_i - \mathbf{y}_j||$ is the corresponding Euclidean distance in the class space.

As a result of the approach taken in previous section, the separation between pixel pairs belonging to different classes is maximized by $\mathbf{C}$ and hence, the matrix $\mathbf{B}$ is expected to preserve these distances. Moreover, the minimization of $h(\mathbf{B})$ can be performed using various methods such as eigen-decomposition or pseudo-Newton minimization or conjugate gradient which we employed.

## 3 Implementation

Following the previous sections, the training step of our algorithm becomes

1. Compute an initial estimate of matrix $\mathbf{C}$ *i.e.* $\mathbf{C}^{(0)}$.
2. Apply steepest descent of Section 2.1 to optimize the cost function $p(\mathbf{C})$.
3. Once $\mathbf{C}$ is at hand, compute $\mathbf{y}_i = \mathbf{C}\mathbf{x}_i$ for all $\mathbf{x}_i \in \mathcal{X}$.
4. Recover matrix $\mathbf{B}$ using MDS to compute $\mathbf{z}_i = \mathbf{B}\mathbf{y}_i$ for all $\mathbf{y}_i \in \mathcal{Y}$.
5. Train classifier of choice by the transformed training color values $\mathbf{z}_i = \mathbf{A}\mathbf{x}_i$.

whereas for a testing RGB pixel value $\mathbf{x}_i^*$ the step sequence is as follows

1. Compute $\mathbf{z}_i^* = \mathbf{A}\mathbf{x}_i^*$.
2. Feed the transformed testing color value $\mathbf{z}_i^*$ to the classifier.

In the training step sequence above, we commence by computing an initial estimate of $\mathbf{C}$ by employing Fisher's class separability criterion and properties of the matrix span. The reason for our choice hinges in both, the vast amount of work that shows effectiveness of Fisher's criterion for purposes of maximizing class separability and its ease of computation. It is worth noting that Fisher's criterion has been used extensively in LDA.

The literature on LDA is extensive, dwelling into a wide variety of variants of the method itself. For instance, Non-parametric Discriminant Analysis (NDA) incorporates boundary information into between-class scatter. Boudat *et. al.* [29] have proposed a kernel version of LDA that can cope with severe non-linearity of sample set. On the numerical stability and tractability of LDA, there are also a number of methods which aim at overcoming singularity of the inverse intra-class covariance matrix inherent to sub-sampled feature spaces. In a related development, MMC employs an optimization procedure whose constraint is not dependent on the non-singularity of within-class scatter matrix. Here, we use the method of Wang *et. al.* [30] which employs dual subspaces to construct LDA classifiers.

To employ Fisher's criterion, we construct the matrix $\mathbf{D}$. Making use of notation introduced in Section 2.1, we can define entry indexed $i, j$ of the matrix $\mathbf{D}$ as follows

$$D_{i,j} = \frac{||\mu_{i,j}||}{||\mu_{i,j}(\mathbf{S}_i - \mathbf{S}_j)||} \tag{12}$$

which is the distance between class centroids over the projected scatters.

With the matrix $\mathbf{D}$ at hand, we employ its QR decomposition and developments to compute $\mathbf{C}^{(0)}$ by using subspace spanned by the first three columns of $\mathbf{Q}$, *i.e.* $\lfloor \mathbf{C}^{(0)} \rfloor$ as the best second order approximation to $\mathbf{C}^{(0)}$. Being more formal, given $\rho(\mathbf{C}^{(0)}) = \lfloor \mathbf{J}[\mathbf{I} \mid \mathbf{0}]^T \rfloor$, then

$$\rho(\mathbf{C}^{(0)}) = \lfloor \mathbf{C}^{(0)} \rfloor = \lfloor \arg \min_{\mathbf{J}^T\mathbf{J}=\mathbf{I}} ||\mathbf{C}^{(0)} - \mathbf{J}||^2 \rfloor \ . \tag{13}$$

where $\mathbf{J}$ is comprised of the first three singular vectors, computed using SVD of $\mathbf{Q}$ such that $\mathbf{D} = \mathbf{Q}\mathbf{R}$ is QR decomposition of $\mathbf{D}$ and $[\mathbf{I} \mid \mathbf{0}]^T$ is a rectangular matrix conformed by concatenation of the identity matrix $\mathbf{I}$ and the empty matrix $\mathbf{0}$.

Our choice of SVD for our initialization stems from both, the fact that singular value decomposition corresponds to the best second-order approximation to the span and there are efficient methods to compute it.

Once $\mathbf{C}^{(0)}$ has been computed, we minimize $p(\mathbf{C})$ making use of steepest descent optimization in Section 2.1. This is a gradient descent method which consists of two steps interleaved to find a minimum of the objective function. Here, we use interleaved steps of gradient calculation and back-tracking line search along the steepest descent direction until convergence. As a result, we iterate until $\Delta\mathbf{C}^{(t)}$ or $p(\mathbf{C}^{(t)})$ in Eq. (8) are sufficiently small, *i.e.* less than a predefined threshold $\epsilon$.

Also, recall that our steepest descent method employs the step size $\lambda$. Here, we follow Armijo's rule and use the expression

$$\lambda = \begin{cases} 2\lambda & \text{if } p(\mathbf{C}^{(t)}) - p(\mathbf{C}^{(t)} + 2\lambda\Delta\mathbf{C}^{(t)})) \geq \lambda||\Delta\mathbf{C}^{(t)}|| \\ \frac{1}{2}\lambda & \text{if } p(\mathbf{C}^{(t)}) - p(\mathbf{C}^{(t)} + \lambda\Delta\mathbf{C}^{(t)})) < \frac{1}{2}\lambda||\Delta\mathbf{C}^{(t)}|| \end{cases} \tag{14}$$

## 4 Experiments

In this section, we present our experiments on semantic labeling using proposed class-driven color transformation and a number of alternatives. To illustrate the utility of our method, we consider a standard multi-class segmentation pipeline where above transformation is included as a preprocessing step.

Given an input image, we apply our color transformation and use the output as an input to a pixel-level classifier. Here, in order to analyze robustness of the transformation to different number of classes and various classifiers, we consider two publicly available frameworks *i.e.* TextonBoost [20] and DARWIN [21] implementation. The former only gives unary terms for each class as output whereas the latter delivers unary and pairwise terms provided by a post-processing step consisting of a conditional random field (CRF). By using this pipeline, our experiments are conducted on two standard datasets for semantic segmentation *i.e.* MSRC-21 [22] and Stanford Background Dataset (SBD) [23]. To our knowledge, TextonBoost and DARWIN provide competitive results on publicly available frameworks to the state-of-the-art for these datasets, respectively.

For the purposes of quantitative evaluation, we provide pixel-wise comparisons between outputs of the classifiers and ground-truth. To do this, we report global and per-class average accuracies [22]. Here, the global accuracy represents the ratio of correctly classified pixels to total number of pixels in test set. Per-class average accuracy, on the other hand, is computed as the average over all classes for the ratio of correctly classified pixels in a class to total number of pixels in the same class.

In our experiments, both pixel-level classifiers under consideration are learned using training sets of features in [20] for the MSRC-21 and [23] for the SBD. We use all training samples available to learn projection from the input color data to the classes. To apply metric MDS, we use a Markov Chain Monte Carlo (MCMC) random sampler to select a balanced distribution of samples. The class mapping

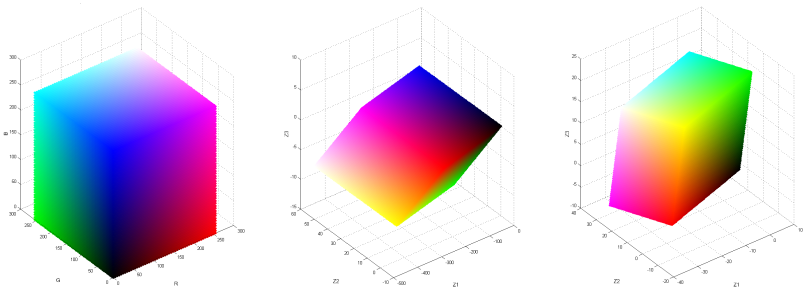

**Fig. 1.** RGB color cube (*l*eft) and transformed color cubes for MSRC-21 (*m*iddle) and SBD (*r*ight) datasets, respectively.

matrix $\mathbf{C}$ and subspace projection matrix $\mathbf{B}$ are then multiplied to compute color transformation matrix $\mathbf{A}$.

In Fig. 1, we show the transformed color cubes for both, MSRC-21 and SBD datasets. As mentioned earlier, note that the mapping induced by $\mathbf{A}$ can potentially yield negative values in the target space. This can be appreciated in the figure, where the transformed cubes have been colored as per the original RGB value at input taken from the cube on the right-hand panel. These transformed cubes do have negative values and moreover, they are consistent with the notion that our approach gives a linear transformation yielded from a convex function.

Recall that both TextonBoost and DARWIN employ positive RGB inputs. To accommodate this requirement, we bound the transformed color cubes using the matrix

$$\mathbf{A}^* = \mathbf{A} \times \mathbf{H} + \mathbf{T} \tag{15}$$

as an alternative to $\mathbf{A}$. In the equation above, $\mathbf{H}$ and $\mathbf{T}$ are diagonal scaling and translation matrices, respectively. This is a straightforward scaling-translation operation in the transformed target space and hence, we do this without any loss of generality. Further, we compute the diagonal matrices $\mathbf{H}$ and $\mathbf{T}$ deploying vertices in the color cubes as an additional training step. This is done by solving a linear equation where six vertices in the RGB color cube are used to obtain six degrees of freedom comprised by three diagonal elements of each of the above two matrices.

For a comprehensive evaluation, we compare accuracy of our method with the results obtained when our color transformation is not included. We have done this to set a baseline that can be employed as an indicator of the contribution of our color transformation to classification performance. As both TextonBoost and DARWIN employ the Lab color space, our baseline results represents this color transformation as well. Hence, for the purposes of comparison, we have employed some other widely used classic color transformations in compute vision including YCrCb, HSL, Luv [12], I1I2I3 [31] and O1O2O3 [32]. TextonBoost and DARWIN compute textons on luminance channel and so, in our experiments, it is given

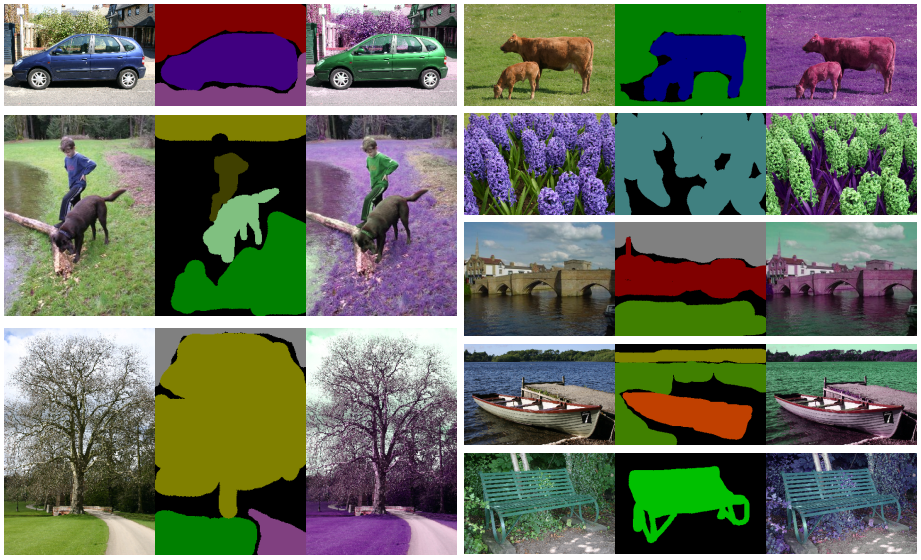

**Fig. 2.** Sample images for MSRC-21 dataset. In each panel, we show the RGB image on the dataset (*l*eft), corresponding labeling (*m*iddle) and transformed image (*r*ight).

by the L-channels of the canonical color transformations, more specifically, I1 and O3 of the illumination invariant color spaces under study *i.e.* I1I2I3 and O1O2O3.

Finally, to further justify our choice of MDS for purposes of subspace projection throughout the section 2.2, we also explore effect of various subspace projection methods as alternatives to metric MDS. The methods used here to that aim are Heteroscedastic Discriminant Analysis (HDA) [16], Maximum Margin Criterion (MMC), Singular Value Decomposition (SVD) and Kernel Principal Component Analysis (KPCA). HDA is based on the heteroscedastic two-class technique using Chernoff criterion and MMC geometrically maximizes the average margin between classes after reducing number of dimensions.

### 4.1 MSRC-21 Dataset

The MSRC-21 dataset [22] consists of 591 color images of size $320 \times 213$ with corresponding ground truth labeling for 21 object classes. As mentioned above, we use the same evaluation procedure as [20] e.g. 276 images for training and 256 images for testing.

To illustrate high-distinct transformed colors in comparison to the original ones with respect to ground-truths on MSRC-21 dataset, we show the transformed color images together with corresponding labeling and RGB color inputs in Fig. 2. To produce these images, we have used the matrix $\mathbf{A}^*$ as an alternative to $\mathbf{A}$. This, in turn has the effect of portraying the images as they would be taken

**Table 1. MSRC-21:** Summary of per-class results on TextonBoost [20]. Bold values indicate the highest accuracies.

| | Bldg. | Grass | Tree | Cow | Sheep | Sky | Aeropl. | Water | Face | Car | Bicycle | Flower | Sign | Bird | Book | Chair | Road | Cat | Dog | Body | Boat | Avg. | Global |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | 68 | **98** | 88 | 85 | 76 | 92 | **86** | 68 | 84 | 77 | 87 | 86 | 57 | 45 | 92 | 60 | 87 | 74 | 36 | 75 | 22 | 73.9 | 82.0 |
| MMC | 72 | **98** | 90 | 86 | 81 | 94 | 83 | 72 | 87 | **84** | **89** | **93** | 66 | 45 | **97** | **68** | 88 | 72 | **42** | 80 | 21 | 76.5 | 84.1 |
| HDA | **73** | **98** | **91** | 85 | 82 | 92 | 81 | 71 | 87 | 77 | 88 | 91 | 64 | 48 | 93 | **68** | **89** | 72 | 41 | 81 | 24 | 76.0 | 83.9 |
| SVD | 71 | **98** | **91** | **87** | 81 | **95** | 82 | 70 | 87 | 81 | **89** | 89 | **70** | 48 | 93 | 65 | 88 | 73 | 37 | 80 | **30** | 76.4 | 84.0 |
| KPCA | 72 | **98** | 89 | 85 | **85** | 94 | **86** | 70 | 87 | 82 | 88 | 90 | 66 | **53** | 96 | 56 | 88 | **78** | 38 | 80 | 24 | 76.3 | 83.9 |
| **Ours** | 70 | **98** | 90 | 86 | 83 | 94 | **86** | **74** | **89** | 81 | 87 | 92 | 63 | 45 | 96 | 65 | 88 | 72 | **42** | **84** | 27 | **76.8** | **84.2** |

**Table 2. MSRC-21:** Summary of results for subspace projections (*l*eft) and color transformations (*r*ight) on DARWIN [21]. Bold values indicate the highest accuracies.

| | Unary | | Pairwise | |
|---|---|---|---|---|
| | Avg. | Gb. | Avg. | Gb. |
| Baseline | 67.1 | 78.9 | 70.2 | 82.8 |
| MMC | 68.1 | 78.5 | 71.3 | 82.7 |
| HDA | 67.4 | 77.9 | 70.8 | 82.2 |
| SVD | 67.3 | 79.8 | 69.1 | 83.5 |
| KPCA | 68.3 | 79.2 | 71.9 | 83.4 |
| **Ours** | **68.9** | **80.0** | **72.1** | **84.0** |

| | Unary | | Pairwise | |
|---|---|---|---|---|
| | Avg. | Gb. | Avg. | Gb. |
| Baseline | 67.1 | 78.9 | 70.2 | 82.8 |
| YCrCb | 65.4 | 77.8 | 69.5 | 81.8 |
| HSL | 65.2 | 77.0 | 70.3 | 82.1 |
| Luv | 65.9 | 78.1 | 70.3 | 82.1 |
| I1I2I3 | 62.6 | 75.2 | 69.6 | 81.3 |
| O1O2O3 | 62.3 | 75.1 | 68.9 | 81.2 |
| **Ours** | **68.9** | **80.0** | **72.1** | **84.0** |

at input by the classifiers. Note that, the transformed colors are somewhat consistent with the labels. This is expected, since the matrix $\mathbf{C}$ is obtained based on the label information and, later on, the metric MDS used to compute $\mathbf{B}$ aims at preserving induced class distances. For instance, the chair and tree are in almost the same green spectrum in the RGB image but in the transformed one, green chair and purple tree provide higher visual separation for different classes.

We summarize our evaluation results in Tab. 1 for the TextonBoost and in Tab. 2 for the DARWIN. As shown in the tables, our color transformation outperforms the alternatives with respect to the baseline when applied as a preprocessing step. This improvement is consistent across a variety of subspace projection methods and color transformations. Note also that in Tab. 2, pattern of improvement holds when a pairwise term is applied after the classifier.

We employed publicly available code of TextonBoost to compare the classification performance with and without our color transformation but due to fine tuning of parameters, preprocessing of images or randomization functions, accuracy on different platforms/compilers were not consistent with those reported in [20]. However, our experiments still confirm the advantages of using the transformed images over original ones for classification.

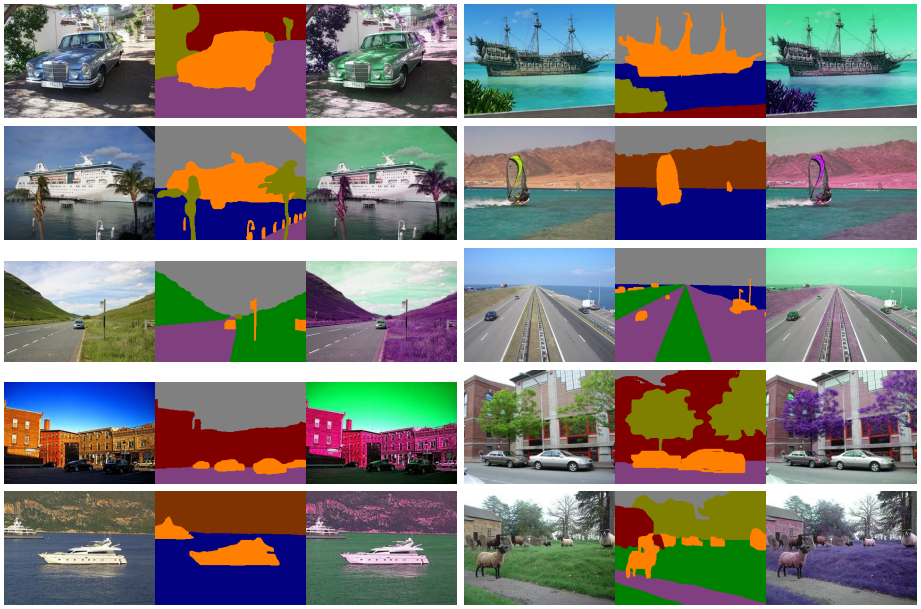

**Fig. 3.** Sample images for SBD dataset. In each panel, we show RGB image on the dataset (*l*eft), corresponding labeling (*m*iddle) and transformed image (*r*ight).

### 4.2 Standford Background Dataset (SBD)

The Stanford Background Dataset [23] consists of 715 color images of size $320 \times 240$ with corresponding ground labeling over 8 classes. In this case, we use the same evaluation procedure as in [23], *i.e.* 5-fold cross validation with 572 images for training and 143 images for testing.

To present the effect of our color transformation on SBD dataset, we show sample images together with their labeling and RGB color inputs in Fig. 3. As previous section, we have used the matrix $\mathbf{A}^*$ as an alternative to $\mathbf{A}$ to produce these images. Note that, colors in the transformed images are also somewhat consistent with the label information. This, as mentioned earlier, is expected due to the manner in which we have computed matrix $\mathbf{A}$.

Table 3 presents outcomes for TextonBoost and Tab. 4 summarize the results for DARWIN for different subspace projections and color transformations under consideration. It is clear that, by using our method as a preprocessing step, we can consistently improve both unary and pairwise classification accuracies. Moreover, using metric MDS as subspace projection method outperforms others, including the baseline.

Therefore, we can conclude that our class-driven color transformation improves classification accuracy when used as a preprocessing step within a multi-class segmentation framework. Importantly, the improvement is independent of dataset and number of classes. Further, this improvement is an additional gain

**Table 3. SBD:** Summary of per-class results on TextonBoost [20]. Bold values indicate the highest accuracies.

| | Sky | Tree | Road | Grass | Water | Bldg. | Mntn. | Forgr. | Avg. | Global |
|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | **86** | 64 | 89 | 65 | 65 | 76 | 03 | 60 | 63.5 | 74.0 |
| MMC | **86** | **68** | 90 | 72 | 62 | 79 | 01 | 63 | 65.2 | 76.3 |
| HDA | **86** | 66 | **91** | 72 | 65 | **80** | 02 | **65** | 65.8 | 76.7 |
| SVD | **86** | 67 | **91** | **73** | 64 | **80** | 06 | 63 | 66.3 | 76.7 |
| KPCA | **86** | **68** | **91** | **73** | 65 | **80** | 03 | 61 | 65.9 | 76.6 |
| **Ours** | **86** | **68** | 90 | 68 | **67** | 79 | **09** | **65** | **66.5** | **76.8** |

**Table 4. SBD:** Summary of results for subspace projection techniques (*l*eft) and color transforms (*r*ight) on DARWIN [21]. Bold values indicate the highest accuracies.

| | Unary | | Pairwise | |
|---|---|---|---|---|
| | Avg. | Gb. | Avg. | Gb. |
| Baseline | 68.3 | 78.5 | 70.2 | 81.5 |
| MMC | 69.6 | 79.9 | 71.7 | 82.9 |
| HDA | 68.9 | 79.3 | 71.2 | 82.4 |
| SVD | 69.2 | 79.8 | 70.9 | 82.8 |
| KPCA | 68.5 | 78.8 | 70.7 | 82.0 |
| **Ours** | **70.4** | **81.4** | **72.5** | **84.2** |

| | Unary | | Pairwise | |
|---|---|---|---|---|
| | Avg. | Gb. | Avg. | Gb. |
| Baseline | 68.3 | 78.5 | 70.2 | 81.5 |
| YCrCb | 68.7 | 77.1 | 70.7 | 80.2 |
| HSL | 68.0 | 76.9 | 70.0 | 80.1 |
| Luv | 68.3 | 76.9 | 70.5 | 80.1 |
| I1I2I3 | 67.1 | 75.9 | 68.6 | 78.7 |
| O1O2O3 | 66.7 | 75.9 | 68.5 | 79.0 |
| **Ours** | **70.4** | **81.4** | **72.5** | **84.2** |

over the classifier output, hence, not being exclusive of other common post-processing steps such as the application of CRFs. Finally, it is worth mentioning that this improvement comes at a negligible computational cost.

## 5 Conclusion

In this paper, we derived an image color transformation based on label information of classes in dataset under study. We did this by posing the problem in terms of a composition of two mappings. The first of these, maps the image color onto a space spanned by the class set. We computed this mapping by optimizing an objective function formulated in terms of the aggregation of pairwise distances within and between the classes using a steepest descent scheme devoid of matrix inversions. The second mapping corresponds to transforming data in the class space to a target space, which we calculated making use of metric multi-dimensional scaling. Our experiments confirm the applicability of our algorithm to enhance global and average precisions of pixel-wise classifiers when it is employed as a general preprocessing step for segmentation and labeling.

**Acknowledgement**. Authors would like to highly appreciate reviewers' efforts and positive feedbacks which improve the quality and readability of this work.

## References


1. Van De Sande, K., Gevers, Th., Snoek, C.: Evaluating color descriptors for object and scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **32:9** 2010 1582–1596
2. Finlayson, G.D., Drew, M.S.: The maximum ignorance assumption with positivity. Society for Imaging Science and Technology Conference on Color and Imaging 1996 202–205
3. Nayar, S.K., Bolle, R.M.: Reflectance based object recognition. International Journal of Computer Vision **17:3** 1996 219–240
4. Dror, R.O., Adelson, E.H., Willsky, A.S.: Recognition of surface reflectance properties from a single image under unknown real-world illumination. 2001
5. Slater, D., Healey, G.: Object recognition using invariant profiles. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1997 827–832
6. Jacobs, D.W., Belhumeur, P.N., Basri, R.: Comparing images under variable illumination. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1998 610–617
7. Lin, S., Lee, S.W.: Using chromaticity distributions and eigenspace analysis for pose-, illumination-, and specularity-invariant recognition of 3D objects, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 1997 426–431
8. Lenz, R., Carmona, P.L., Meer, P.: The hyperbolic geometry of illumination-induced chromaticity changes IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2007 1–6
9. Chong, H.Y., Gortler, S.J., Zickler, T.: A perception-based color space for illumination-invariant image processing. ACM Transactions on Graphics **27:3** 2008 61
10. Strutz, T.: Adaptive selection of colour transformations for reversible image compression. IEEE European Signal Processing Conference (EUSIPCO) 2012 1204–1208
11. Hu, G., Liu, C., Chuang, K., Yu, S., Tsui, T.: General Regression Neural Network utilized for color transformation between images on RGB color space. IEEE International Conference on Machine Learning and Cybernetics (ICMLC) **4** 2011 1793–1799
12. Wyszecki, G., Stiles, W.S.: *C*olor science. Wiley New York (1982)
13. Meyer, G.W., Greenberg, D.P.: Perceptual color spaces for computer graphics. ACM SIGGRAPH Computer Graphics **14:3** 1980 254–261
14. Fu, Z., Robles-Kelly, A.: Learning object material categories via pairwise discriminant analysis. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) 2007
15. McLachlan, G.: *D*iscriminant analysis and statistical pattern recognition. John Wiley & Sons 2004
16. Loog, M., Duin, R., Haeb-Umbach, R.: Multiclass linear dimension reduction by weighted pairwise Fisher criteria. IEEE Transactions on Pattern Analysis and Machine Intelligence **23:7** 2001 762–766
17. Bickel, P.J., Levina, E.: Some theory for Fisher's linear discriminant function,'naive Bayes', and some alternatives when there are many more variables than observations. J. of Bernoulli 2004 989–1010
18. Fan, J., Fan, Y.: High dimensional classification using features annealed independence rules. The Annals of Statistics **36:6** 2008 2605

19. Tibshirani, R., Hastie, T., Narasimhan, B., Chu, G.: Diagnosis of multiple cancer types by shrunken centroids of gene expression. P. of the National Academy of Sciences **99:10** 2002 6567–6572
20. Krähenbühl, Ph., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. arXiv preprint arXiv:1210.5644 (2012)
21. Gould, S.: DARWIN: A framework for machine learning and computer vision research and development. J. of Machine Learning Research **13** (2012) 3533–3537
22. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. International Journal of Computer Vision **81:1** 2009 2–23
23. Gould, S., Fulton, R., Koller, D.: Decomposing a scene into geometric and semantically consistent regions. IEEE International Conference on Computer Vision (ICCV) 2009 1–8
24. Boyd, S., Vandenberghe, L.: *C*onvex optimization. Cambridge university press 2009
25. Li, X., Jiang, T., Zhang, K.: Efficient and robust feature extraction by maximum margin criterion. IEEE Transactions on Neural Networks **17:1** 2006 157–165
26. Hastie, T., Tibshirani, R.: Classification by pairwise coupling. The Annals of Statistics **26:2** 1998 451–471
27. Harris, Ch.: Tracking with rigid models. Active vision 1993 59–73
28. Lin, D., Yan, S., Tang, X.: Pursuing informative projection on Grassmann manifold. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) **2** 2006 1727–1734
29. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Computation **12:10** 2000 2385–2404
30. Wang, X., Tang, X.: Dual-space linear discriminant analysis for face recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) **2** 2004 564
31. Geusebroek, J., Van den Boomgaard, R., Smeulders, A.W.M., Geerts, H.: Color invariance. IEEE Transactions on Pattern Analysis and Machine Intelligence **23:12** 2001 1338–1350
32. Álvarez, J.M., Gevers, T., López, A.M.: Learning photometric invariance for object detection. International Journal of Computer Vision **90:1** 2010 45–61