Color Constancy Via Convex Kernel Optimization

Xiaotong Yuan, Stan Z. Li, and Ran He

Center for Biometrics and Security Research, Institute of Automation,
Chinese Academy of Science, Beijing – China, 100080

Abstract. This paper introduces a novel convex kernel based method for color constancy computation with explicit illuminant parameter estimation. A simple linear render model is adopted and the illuminants in a new scene that contains some of the color surfaces seen in the training image are sequentially estimated in a global optimization framework. The proposed method is fully data-driven and initialization invariant. Nonlinear color constancy can also be approximately solved in this kernel optimization framework with piecewise linear assumption. Extensive experiments on real-scene images validate the practical performance of our method.

1 Introduction

Color is an important feature for many machine vision tasks such as segmentation [8], object recognition [13] and surveillance [4]. However, light sources, shadows, transducer non-linearities, and camera processing (such as auto-gain-control and color balancing) can all affect the apparent color of a surface. Color constancy algorithms attempt to estimate these photic parameters and compensate for their contribution to image appearance. There are a large body of works in color constancy literature. A common approach is to use linear models of reflectance and illuminant spectra [9]. Gray world algorithm [1] assumes the average reflectance of all the surfaces in a scene is gray. White world algorithm [5] assumes the brightest pixel corresponds to a scene point with maximal reflectance. Another widely used technique is to estimate the relative illuminant or mapping of colors under an unknown illuminant to a canonical one. Color gamut mapping [3] uses the convex hull of all achievable RGB values to represent an illuminant. The intersection of the mapping for each pixel in an image is used to choose a "best" mapping. In [14], a back-propagation multi-layer neural network is trained to estimate the parameters of a linear color mapping. In [6], a Bayesian estimation scheme is introduced to integrate prior knowledge, e.g. lighting and object classes, into a bilinear likelihood model motivated from the physics of image formulation and sensor error. Linear subspace learning is used in [12] to develop the *color eigenflows* method to model joint illuminant change. This linear model uses no prior knowledge of lighting condition and surface reflectance and does not need to be re-estimated for new objects or scenes. However, the demanding for large training set and rigorous pixel-wise correspondence between training and test images limits the application of this method.

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 728–737, 2007.

In this work, we build our color constancy study on linear transformation parameter estimation. Recently, [8] presented a diagonal rendering model for outdoor color classification problem. Only one image containing the color samples under a certain "canonical" illuminant is needed for training Gaussian classifiers. The trained colors seen under different illuminations can be robustly recognized via MAP estimation. Due to the advantage of fewer training data requirements, we adopt this diagonal render model as the base model for our study . The main difference between our solution and that of [8] lies in the definition of objective function and the associated optimization method. In [8], image likelihood and model priors are integrated into a MAP formulation and locally optimized with EM algorithm. This algorithm works well when all the render matrices are properly initialized. However, such initializations are not always available and accurate in practice. In this paper, we propose a novel convex kernel based criteria function to measure the color compensation accuracy in a new scene. A sequential global mode-seeking framework is then developed for parameter estimation. The optimization procedure includes following three key steps:

- A two-step iterative algorithm derived by Half-Quadratic optimization is used to find the local maximum.
- A multi-bandwidth method is then used to locate the global maximum by gradually decreasing the bandwidth from an estimated uni-mode promising bandwidth.
- A well designed adaptive re-sampling mechanism is adopted and the above multi-bandwidth method is repeated till the desired number of peak modes are found.

The peak modes obtained in this procedure may be naturally viewed as transformation vectors for apparent illuminants in the scene. Our convex kernel based method is fully data-driven and initialization invariant. Such good numerical properties also leads to our solution for nonlinear color constancy problem based on current linear model. To do this, we make piecewise linear assumption to approximate the general nonlinear cases. Our method can automatically find the transformation vectors for each linear piece. Local optimization methods, such as EM based method in [8], can hardly achieve this goal in practice because of initialization dependency. Some results achieved by our method will be reported.

The reminder of this paper is organized as follows: In Section 2, we model color constancy as linear mapping and estimate the parameters via multi-bandwidth kernel optimization in a fully data-driven way. In Section 3 we show the experimental results that validate the numerical superior of our method to that in [8]. We conclude this paper in Section 4.

2 Problem Formulation

For the benefit of fewer training samples requirement, we adopt the linear render model stated in [8] as the base model for our color constancy study. The key assumptions of are:

- One hand-labeled image is available for training the class-conditional color distributions under the "canonical" illuminant.
- The class-conditional color surface likelihood under the canonical illuminant is a Gaussian density, with mean μ_j and covariance Σ_j
- The illuminant-induced color transformation from test image to training image can be modeled as $F(C_i) = C_i\mathbf{d}$, where $\mathbf{d} = (d_1, d_2, d_3)^T$ is the color render vector to be estimated. $C_i = diag(r_i, g_i, b_i)$ is a diagonal matrix that stores the observed RGB colors for pixel i in the test image.

Suppose we have trained S color surfaces with distributions $y_j \sim \mathcal{N}(\mu_j, \Sigma_j)$,$j = 1, ..., S$. Also, assume given a test image with N pixels C_i, $i = 1, ..., N$, which contains L illuminants linearly parameterized by vectors $\mathbf{d}_l$, $l = 1, ..., L$. Our goal is to estimate the optimal $\mathbf{d}_l$ from image data and then get the assignments of surface class labels $j(i)$ and illuminant type labels $l(i)$ for each pixel i according to:

$$(j(i), l(i)) = \arg\min_{j,l}(dist(C_i\mathbf{d}_l, y_j)) \tag{1}$$

$dist(\cdot)$ is some properly selected distance measurement metric (e.g. Mahalanobis distance in this work).

2.1 Kernel Based Objective Function

To estimate the optimal transformation vectors $\mathbf{d}_l$, we propose do find the L peak modes of following kernel sum function:

$$\hat{f}_k(\mathbf{d}) = \sum_{i=1}^{N}\sum_{j=1}^{S} w_{ij}k(M^2(C_i\mathbf{d}, \mu_j, \eta^2\Sigma_j)) \tag{2}$$

where $k(\cdot)$ is the kernel profile function [2](see sect.2.2 for detailed description), $M^2(C_i\mathbf{d}, \mu_j, \eta^2\Sigma_j) = (C_i\mathbf{d}-\mu_j)^T(\eta^2\Sigma_j)^{-1}(C_i\mathbf{d}-\mu_j)$ is the Mahalanobis distance from compensated color $C_i\mathbf{d}$ to training color surface mean y_j. w_{ij} is the prior weight for pixel i belonging to color surface j. The larger function (2) is, the better test image is compensated by vector $\mathbf{d}$.

In the following subsections 2.2 to 2.4, we will focus on the optimization issues and develop a highly efficient sequential mechanism to find the desired L peak modes of (2) as the optimal $\mathbf{d}_l$.

2.2 Half Quadratic Optimization

In this section, we will use half quadratic technique [10] to optimize objective function (2). The results follow directly from standard material in convex analysis (e.g. [10]) and we will omit the technical proofs for page limit. All the conditions we impose on kernel profile $k(\cdot)$ are summarized as below:

1. $k(x)$ is a continuous monotonously decreasing and strictly convex function
2. $\lim_{x\to 0^+} k(x) = \beta > 0$, $\lim_{x\to+\infty} k(x) = 0$
3. $\lim_{x\to 0^+} k^{'}(x) = -\gamma < 0$, $\lim_{x\to+\infty} k^{'}(x) = 0$, $\lim_{x\to+\infty}(-xk^{'}(x)) = \alpha < \beta$
4. $k'(x)$ is continuous with finite discontinuous points.

The following theorem 1 founds the base for optimizing function (2) in a half quadratic way.

Theorem 1. *Let $k(.)$ be a profile satisfying all above conditions, then there exists a strictly monotonously increasing concave function $\varphi : (0, \gamma) \mapsto (\alpha, \beta)$, such that*

$$k(M^2(C_i\boldsymbol{d}, \mu_j, \eta^2\Sigma_j)) = \sup_p(-pM^2(C_i\boldsymbol{d}, \mu_j, \eta^2\Sigma_j) + \varphi(p))$$

and for a fixed $\boldsymbol{d}$, the supmum is reached at $p = -k^{'}(M^2(C_i\boldsymbol{d}, \mu_j, \eta^2\Sigma_j)$.

To further study criteria (2), we introduce a new objective function $F : \mathbb{R}^3 \times (0, \gamma)^N \mapsto (0, +\infty)$

$$\hat{F}_\eta(\mathbf{d}, \mathbf{p}) = \sum_{i=1}^{N}\sum_{j=1}^{S} w_{ij}(-p_{ij}M^2(C_i\mathbf{d}, \mu_j, \eta^2\Sigma_j) + \varphi(p_{ij})) \tag{3}$$

where $\mathbf{p} = (p_{11}, p_{N1}, ..., p_{NS})$. According to theorem 1, we get

$$\hat{f}_k(\mathbf{d}) = \sup_{\mathbf{p}}(\hat{F}_\eta(\mathbf{d}, \mathbf{p}))$$

It is straight forward to see that

$$\max_{\mathbf{d}} \hat{f}_k(\mathbf{d}) = \max_{\mathbf{d}} \sup_{\mathbf{p}}(\hat{F}_\eta(\mathbf{d}, \mathbf{p})) \tag{4}$$

From (4) we tell that maximizing $\hat{f}_k(\mathbf{d})$ is equivalent to maximizing $\hat{F}_\eta(\mathbf{d}, \mathbf{p}))$, which is quadratic w.r.t. $\mathbf{d}$ when $\mathbf{p}$ is fixed. We propose to use a strategy based on alternate maximization over $\mathbf{d}$ and $\mathbf{p}$ as follows (superscript l denotes the time stamp):

$$p_{ij}^l = -k^{'}(M^2(C_i\mathbf{d}^{l-1}, \mu_j, \eta^2\Sigma_j)) \tag{5}$$

$$\mathbf{d}^l = \left[\sum_{i=1}^{N}\sum_{j=1}^{S} w_{ij}p_{ij}^l C_i^T\Sigma_j^{-1}C_i\right]^{-1}\left[\sum_{i=1}^{N}\sum_{j=1}^{S} w_{ij}p_{ij}^l C_i^T\Sigma_j^{-1}\mu_j\right] \tag{6}$$

2.3 Global Mode-Seeking

Since the above two-step iterations (5) and (6) are essentially a gradient ascending method, it will surely converge to local maximum. In this section, we first derive in the following proposition 1 indicating that if bandwidth parameter η is large enough, then the criterion function (2) is strictly concave, hence is uni-mode. Then we develop a global peak mode seeking method based on this proposition to find the transformation vector $\mathbf{d}$ that best compensates the illuminant in the test image.

Proposition 1. *One sufficient condition for* $\hat{F}_\eta(\boldsymbol{d}, \boldsymbol{p})$ *to be uni-mode is*

$$\eta > Const * \left(2 \sup_v \left(-\frac{k''(v)}{k'(v)} \right) \right)^{\frac{1}{2}} \tag{7}$$

where $Const = \max\{\sqrt{M^2(x, \mu_j, \Sigma_j)} | x \in [0, 255]^3, j = 1, ...S\}$.

The proof is just built on trivial derivative calculation. We give below an example profile to further clarify proposition 1.

Example 1. **(Gaussian profile):** $k(x) = e^{-x/2}$. Then $k^{'}(x) = -\frac{1}{2}e^{-x/2}$, $k^{''}(x) = \frac{1}{4}e^{-x/2}$. $\sup_x \left(-\frac{k''(x)}{k'(x)} \right) = \frac{1}{2}$. By proposition 1, the uni-mode-promising bandwidth can be selected according to $\eta > Const$. In addition, the dual variable function is $\varphi(p) = 2p - 2p \ln 2p$ in theorem 1.

From proposition 1 we can tell that if η is large enough , then from any initial estimation, the two-step iteration algorithm presented in (5) and (6) will converge to a unique maximizer of the over-smoothed density function. When the uni-maximizer is reached, we may decrease the value of η and run the same iterations again, taking the previous maximizers as initializations. This procedure is repeated until a certain termination condition is met (e.g., convergence error is small enough). The final obtained maximizer is very likely to be the global peak mode of the criteria function, since such a numerical procedure is actually deterministic annealing [7]. See algorithm 1 for a formal description of this optimization procedure. We have noticed that this global peak mode-seeking mechanism is similar to what called *annealed mean shift* in [11], which aims to find the global kernel density mode. The key improvement lies in that we give an up-bound of uni-mode promising bandwidth, hence make the algorithm more operable in practice.

Algorithm 1. Global Transformation Vector Seeking

1: $m \leftarrow 0$, Initialize η_m satisfying the condition presented in proposition 1
2: Randomly initialize **d**
3: **while** Terminate condition is not met **do**
4: Run the iteration (5) and(6) till converge.
5: $m \leftarrow m + 1$
6: $\eta_m \leftarrow (\eta_{m-1} * \rho)$.
7: Initialize **d** and **p** with the maximizers obtained in 4.
8: **end while**

In the following subsections, we denote $\mathbf{d}^*$ and $\mathbf{p}^*$ be the convergent points reached in algorithm1, and η^* be the corresponding bandwidth. We also call the global maximizer $\mathbf{d}^*$ reached in algorithm 1 to be the *global transformation vector* (GTV) (associated with current prior weights $\mathbf{w}$).

2.4 Multiple Mode-Seeking

In this section, as an extension of algorithm 1, we develop an adaptive and sequential method, namely *Ada-GTV*, for multiple transformation vector mode-seeking. The core idea of this method is to find the GTVs one after another by adaptively changing the prior weight vector $\mathbf{w}$ and finding the corresponding GTV $\mathbf{d}^*$ via algorithm 1. Suppose that current GTV is estimated , we then search for a local maximizer $\mathbf{d}^{*'}$ around it for the criterion function (2) estimated under equal prior weights (this is because our purpose is to find the peak modes of (2) estimated on original training and test data). Dual variable $\mathbf{p}$ is calculated as $p_{ij} = -k'(M^2(C_i\mathbf{d}^{*'}, \mu_j, \eta^{*2}\Sigma_j)$, $i = 1, ..., N, j = 1, ..., S$. We then reweight all the terms in (2), giving higher weight to the cases that are "worse" compensated (with lower p_{ij}) and repeat the GTV seeking procedure by algorithm 1. This leads to a sequential global mode-seeking algorithm. The formal description of Ada-GTV is given in algorithm 2. The founded GTVs can be naturally viewed as transformation parameters for different illuminations in the scene. Compensation and color classification can be easily done according to (1), as stated in [8]. The running time of Ada-GTV is obviously $\mathcal{O}(L * NS)$ $(L, S \ll N)$, hence it is a linear complexity algorithm w.r.t pixel number N.

Algorithm 2. Ada-GTV

1: **Initialization:** Start with weights $w_{ij}^0 = 1/NS$, $i = 1, ..., N, j = 1, ..., S$
2: **for** $l = 0$ to $L - 1$ **do**
3: **GTV Estimation:** Find GTV $\mathbf{d}^*$ by algorithm 1 with current prior weight $\mathbf{w}^l$.
4: **Mode Refinement :** Starting from $\mathbf{d}^*$, find the local maximum $\mathbf{d}^{*'}$ for $\hat{f}_k(\mathbf{d})$ estimated under η^* and $\mathbf{w}^0$.
5: **Dual Variables:** Get $p_{ij} = -k'(M^2(C_i\mathbf{d}^{*'}, \mu_j, \eta^{*2}\Sigma_j))$.
6: **Sample Re-weight:** Set $w_{ij}^{l+1} \leftarrow w_{ij}^l/(1 + p_{ij})$. Normalize $w_{ij}^{l+1} \leftarrow w_{ij}^{l+1}/\sum_{ij} w_{ij}^{l+1}$
7: **end for**
8: **Color and Illuminant Classification:** Each pixel's illuminant and color label is determined as $(j(i), l(i)) = \arg\min_{j,l}(M^2(C_i\mathbf{d}_l, \mu_j, \Sigma_j))$.

3 Experiments

We present several groups of experiments on color compensation and classification of real scenes to show the performance of the our method.

The first experiment is done to show the global optimization property of our algorithm. For comparison purpose, we adopt one set of image data used in [8]. The training image under "canonical" light (with the manually selected sample colors) and the test image are shown in fig.1(a) and 1(b). Compensation and color classification results by [8] are shown in fig.1(c) $\sim$ 1(f). It is obviously to see that result R_1 from starting point P_1 (fig. 1(c) and 1(d)) is much more satisfying than R_2 from starting point P_2 (fig. 1(e) and 1(f)), hence the algorithm

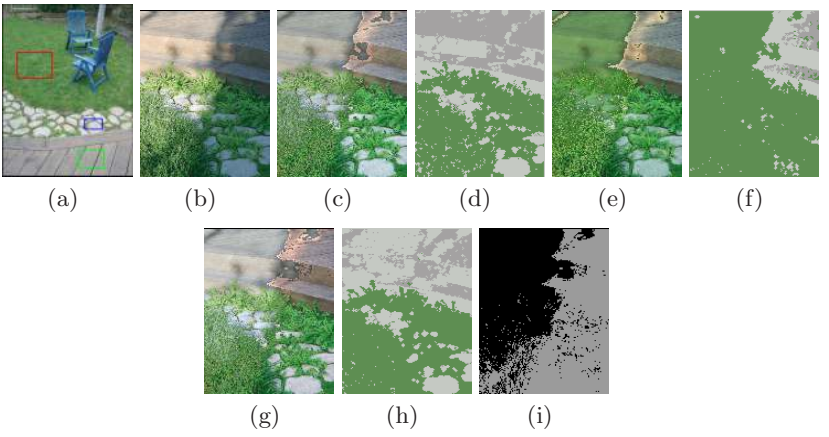

Fig. 1. A comparison example with EM based method [8]. (a): training image (with selected color) under "canonical" light (b) test image. (c) $\sim$ (f): compensation and color classification results by [8]. (c) and (d): R_1 from starting point P_1; (e) and (f): R_2 from starting point P_2. (g) $\sim$ (i): the compensation, color classification and illuminant classification results by Ada-GTV from the starting point either P_1 or P_2.

Table 1. Numerical results, Ada-GTV vs. EM

	$\mathbf{d}_1$	$\mathbf{d}_2$
Starting point $P1$	(1.0,1.0,1.0)	(2.0,2.0,2.0)
Result $R1$ by EM [8]	(0.693,0.773,0.914)	(2.005, 1.636,1.456)
Result by Ada-GTV	**(0.916,0.990,1.053)**	**(2.123,1.614,1.402)**
Starting point $P2$	(0.5,0.5,0.5)	(1.0,2.0,1.0)
Result $R2$ by EM [8]	(0.493,0.748,0.502)	(1.873, 1.557,1.487)
Result by Ada-GTV	**(0.916,0.990,1.053)**	**(2.123,1.614,1.402)**

is highly initialization relevant. The compensation, color classification and illuminant classification results by our Ada-GTV algorithm initialized with either P_1 or P_2 is shown in fig. 1(g) $\sim$ 1(h). Detailed numerical results can be seen in table 1, which clearly indicates the initialization invariant property of our method.

The second experiment will show the ability of our method to handle nonlinear illuminant changes based on current linear render model. To do this, we make piecewise linear assumption to approximate the general nonlinear cases. Our method can automatically find the transformation vectors for each linear piece. We give here one experiment on a pair of "map" images to validate this interesting property. We used Canon A550 DC with automatic exposure, taking care to compensate for the camera's gamma setting. The training image fig.2(a) and test image fig.2(b) are shot under two very different camera settings. The selected 6 sample colors from the training image and their ground truth

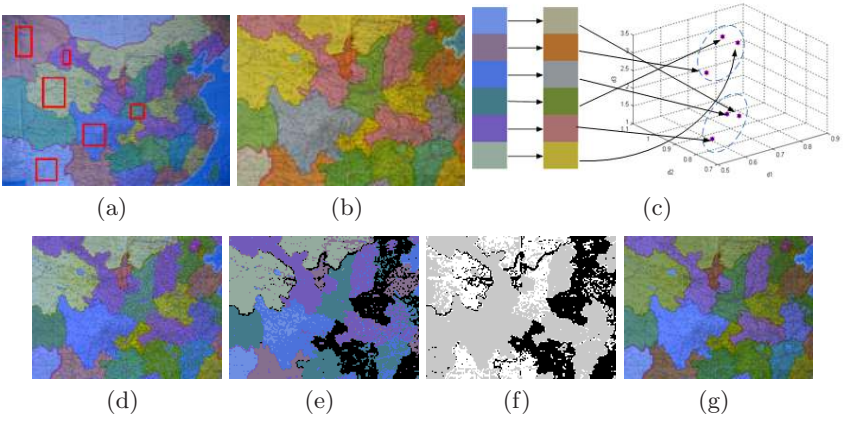

Fig. 2. Piecewise linear color constancy. (a) Training image; (b) Test image; (c) left: 6 selected sample colors and their ground truth counterparts in the test image; right: the ground truth transformation vectors for the 6 sample colors; (e)∼(g) color compensation, color classification and piecewise linear illuminant classification results. The black part in (e) and (f) represents unseen colors in the test image. (h): color compensation result by render vector $\mathbf{d}_1$ only.

Table 2. Initializations and iteration results for render matrices

	$\mathbf{d}_1$	$\mathbf{d}_2$
Initializations	(1,1,1)	(1,1,1)
Iteration results	**(0.649,0.845,1.661)**	**(0.788,1.008,3.014)**
Initializations	(0.5,0.5,0.5)	(0.5,0.5,0.5)
Iteration results	**(0.648,0.843,1.661)**	**(0.788,1.008,3.014)**
Initializations	(5,5,5)	(5,5,5)
Iteration results	**(0.655,0.852,1.661)**	**(0.788,1.008,3.014)**

counterparts in the test image are shown in fig.2(c)(left part). To test whether the illuminant change in the test image is linear or not, we calculate the ground truth transformation vectors for the samples and plot them in fig.2(c)(right part). Obviously two clusters (bounded by dotted ellipses) appear from these vectors, thus the illuminant change is highly nonlinear. One reasonable assumption is that such a change is piecewise linear and we may just feed the image data into Ada-GTV to let it find the transformation vector modes sequentially for each piece, from arbitrary initializations. EM based method [8] can hard to achieve this goal simply because accurate initialization for each linear piece is required, which is not always available beforehand. Here, we properly set the mode number L=2 in Ada-GTV and initialize both render vectors d_1 and d_2 with three different starting points. The convergent points are the same under these initializations, as is shown in table 2 (parameters are set to be $\eta_0 = 1.934$ and $\rho_0 = 0.5$). The image results are shown in fig.2(d)∼ 2(f). fig.2(g) shows the color compensation result by render vector $\mathbf{d}_1$ only, which obviously introduces very

Fig. 3. Some other experimental results. From left to right: training image, test image and color compensated image. (a)"Casia" image pairs, (b)"Comic" image pairs, (c) and (d): "face" image pairs.

large compensation error, visually. Thus, we can see that the adopted piecewise linear assumption greatly improves performance of color constancy.

We have also extensively evaluated our Ada-GTV method on some other real scene image pairs, and selected results are given in fig.3.

4 Conclusion

We introduce in this paper a novel convex kernel based method for color constancy computation with explicit illuminant parameter estimation. A convex kernel sum function is defined to measure the illuminant compensation accuracy in a new scene that contains some of the color surfaces seen in the training image. Render vector parameters are estimated by sequentially locating the peak modes of this objective function. The proposed method is fully data-driven and initialization invariant. Nonlinear color constancy can also be approximately solved in our framework with piecewise linear assumption. The experimental results clearly show the advantage of the our method over local optimization frameowrk, e.g. MAP formulation with EM solution stated in [8].

Acknowledgement

This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Supporting Platform Project #2006BAK08B06, National 863 Program Projects #2006AA01Z192 and #2006AA01Z193, and Chinese Academy of Sciences "100 people project".

References

1. Buchsbaum, G.: A spatial processor model for object color perception. Journal of Franklin Institute 310(1), 1–26 (1980)
2. Comaniciu, D., Meer, P.: Mean shft: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 603–619 (2002)
3. Forsyth, D.A.: A novel algorithm for color constancy. International Journal of Computer Vision 5(1), 5–36 (1990)
4. Gilbert, A., Bowden, R.: Tracking objects across cameras by incremetanlly learning inter-camera colour calibration and patterns of activity. In: European Conference on Computer Vision, vol. 2, pp. 125–136 (2006)
5. Hall, J., McGann, J., Land, E.: Color mondrian experiments: the study of average spectral distributions. J. Opt. Soc. Amer. A(67), 1380 (1977)
6. Finlayson, G., Banard, K., Funt, B.: Color constancy for scenes with varying illumination. Computer Visualization and Image Understanding 65(2), 311–321 (1997)
7. Li., S.Z.: Robustizing robust m-estimation using deterministic annealing. Pattern Recognition 29(1), 159–166 (1996)
8. Manduchi, R.: Learning outdoor color classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1713–1723 (2006)
9. Marimont, D.H., Wandell, B.A.: Linear models of surface and illuminant spectra. J. Opt. Soc. Amer. 9(11), 1905–1913 (1992)
10. Rockfellar, R.: Convex Analysis. Princeton Press (1970)
11. Shen, C., Brooks, M.J., Hengel, V.A.: Fast global kernel density mode seeking with application to localization and tracking. In: IEEE International Conference on Computer Vision, vol. 2, pp. 1516–1523. IEEE, Los Alamitos (2005)
12. Tieu, K., Miller, E.G.: Unsupervised color constancy. In: Thrun, S., Becker, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems 15, MIT Press, Cambridge
13. Tsin, Y., Ramesh, V., Collins, R., Kanade, T.: Bayesian color constancy for outdoor object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 1132–1139 (2001)
14. Funt, B.V., Gardei, V.C., Barnard, K.: Modeling color constancy with neural networks. In: International Conference on Visual Recognition and Action: Neural Models of Mind and Machine (1997)