Exemplar-Based Colour Constancy

Hamid Reza Vaezi Joze
hrv1@sfu.ca

Mark S. Drew
mark@sfu.ca

Simon Fraser University
Vancouver, BC, Canada

Abstract

Exemplar-based learning or, equally, nearest neighbour methods have recently gained interest from researchers in a variety of computer science domains because of the prevalence of large amounts of accessible data and storage capacity. In computer vision, these types of technique have been successful in several problems such as scene recognition, shape matching, image parsing, character recognition and object detection. Applying the concept of exemplar-based learning to the problem of colour constancy seems odd at first glance since, in the first place, similar nearest neighbour images are not usually affected by precisely similar illuminants and, in the second place, gathering a dataset consisting of all possible real-world images, including indoor and outdoor scenes and for all possible illuminant colours and intensities, is indeed impossible. In this paper we instead focus on *surfaces* in the image and address the colour constancy problem by unsupervised learning of an appropriate model for each training surface in training images. We find nearest neighbour models for each surface in a test image and estimate its illumination based on comparing the statistics of pixels belonging to nearest neighbour surfaces and the target surface. The final illumination estimation results from combining these estimated illuminants over surfaces to generate a unique estimate. The proposed method has the advantage of overcoming multi-illuminant situations, which is not possible for most current methods. The concept proposed here is a completely new approach to the colour constancy problem. We show that it performs very well, for standard datasets, compared to current colour constancy algorithms.

1 Introduction

Many computer vision applications as well as image processing problems for both still images and video can make use of colour constancy processing as a prerequisite to ensure that the perceived colour of the surfaces in the scene does not change under varying illumination conditions. The observed colour of the surfaces in the scene is a combination of the actual colour of the surface, i.e., the surface reflection function, as well as the illumination. Estimation of illumination is the main goal of the colour constancy task.

Recently, notwithstanding large amounts of accessible data and of storage capacity, many problems can be simply solved by a search through data instead of applying sophisticated algorithms. Sometimes these methods make use of to nearest neighbour methods. Such use of these techniques occurs in a variety of computer vision problems such as shape matching [3], character recognition [2], human pose estimation [23], image parsing [20], scene recognition [31] and object detection [22]. As an example, Torralba et al. [25] gathered a

large data set of some 80 million tiny 32×32 images, each labelled with a word. They solve different computer vision problems such as scene recognition, object classification, person detection, object categorization, picture orientation determination and even colorization by nearest neighbour methods using this large dataset. Learning based on a previously seen examples is not a new concept. This concept appears in different domains such as exemplar theory in psychology as a model of perception and categorization, case-base reasoning in artificial intelligence and instance-based methods [1] in machine learning.

Many colour constancy algorithms have been proposed (see [16, 17] for an overview). The White-Patch, or Max-RGB, method estimates the light source colour from the maximum response of three different colour channels [18]. Another well-known colour constancy method is based on the Grey-World hypothesis [4], which assumes that the average reflectance in the scene is achromatic. Grey-Edge is a recent version which assumes that the average of the reflectance differences in a scene is achromatic [26]. Shades of Grey [10] is another grey-based method which use the Minkowski p-norm instead of regular averaging. The Gamut Mapping algorithm [12], a more complex and more accurate algorithm, is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours. As [7] mentioned, these methods deal with an image as a bag of pixels and the spatial relation between pixels is not considered.

Applying the concept of exemplar-based or instanced-based learning to the colour constancy problem seems to be an odd idea at first glance since similar or nearest neighbour images are not usually affected by precisely similar illuminants and moreover gathering a dataset consisting of all possible real world images including indoor and outdoor scenes for all possible illuminant colour and intensity is indeed impossible. In contrast, what can we say about surfaces themselves? Every moderate sized dataset of real images includes thousands of surfaces under different viewing and lighting conditions. We can make these surfaces weakly invariant to illumination changes by simple colour constancy algorithms. Therefore, using the exemplar theory concept we can reduce our illumination estimation task into the following steps: (1) finding surfaces in an image; (2) finding a similar surface or surfaces in the training dataset for each of our image surfaces; (3) estimating the illumination for each surface based on comparing the statistics of pixels belonging to similar surfaces and the target surface; (4) combining these estimated illuminants into a unique estimate.

In this paper we present a wholly new line of approach to the colour constancy problem, which we call Exemplar-Based Colour Constancy. We use both texture features and weakly colour-constant three-channel RGB colour values in order to find the nearest neighbour surfaces from training data for each surface. Then we estimate the possible illuminant for each surface based on histogram matching of each surface to its nearest neighbour surfaces from training data. The final step is integrating these estimates into a unique illuminant estimation for the whole image. Since we have no labelled or clustered data for our training process as would be the case for a semantic segmentation task or texture detection task, we lack information for providing confidence for our mapping (such as k-nearest neighbour). Nevertheless, although we find some amount of mismatching for surfaces, the illumination estimation process simply considers these cases as outliers compares to the other estimates. Beside producing excellent results, better than with previous colour constancy algorithms, operating on two standard colour constancy datasets, exemplar-based colour constancy has the advantage of also working in the multi-illuminant situation, which is not possible for most current methods.

The outline of the paper is as follows: we discuss related works in §2 and then in §3 we introduce the proposed method by explaining our surface model and proposed illumination

estimation procedure. We apply our proposed method to two standard colour constancy datasets, comparing to current colour constancy methods in §4 and we conclude the paper and discuss about future works in §5.

2 Related Works

One of the first colour constancy methods which estimate the illuminant by a model that is learned on training images is Gamut Mapping algorithm [12]. It is based on the assumption that in real-world images, for a given illuminant one observes only a limited number of colours; therefore, colours forming a "canonical" gamut which contains possible colours can be observed under a canonical illumination in a training phase, and an estimate of current illuminant can be derived by mapping current pixel colour to that canonical gamut. Several extensions have been proposed for gamut mapping algorithms.

Besides static colour constancy methods such as Max-RGB, Grey-World, Grey-Edge and Shades-of-Grey, which as mentioned before are based on simple assumptions, recently efforts at fusing these algorithms have generated better performance than for the individual algorithms. One of the first attempts in this direction was carried out by Cardei and Funt [6], which applied a weighted committee mechanisms over several of these methods.

More complex methods try to learn to select the best algorithm or combination of algorithms for each image using pixel information as well as spatial information, and hence they do not deal with the image as simply a bag of pixels. As an example, Gijsenij and Gevers [14] clustered the images by a k-means algorithm using natural image statistics to characterize the images on the basis of Weibull distribution parameters. They then correspond each cluster with the best single algorithm for a training image for that cluster. To estimate the illuminant of a test image, they select the algorithm according to its cluster or combination of the individual algorithms according to the distances to neighbouring clusters. In a different approach to selecting best algorithms, Wu et al. [30] introduce a multi-resolution texture descriptor based on an integrated Weibull distribution to extract texture information. They used an image similarity measure derived from the Wiccest feature and spatial pyramid matching to find the K most similar images for a test image from training images, and with these neighbouring images they provide a combination for uniting the data-driven strategy and prior knowledge.

Van de Weijer et al. [27] extend the grey world hypotheses to say: the average reflectance of semantic classes in an image is equal to a constant colour, rather than being just grey. Therefore, for each of the semantic classes present in an image they compute that illuminant that transforms the pixels assigned to that class into the average reflectance colour of that semantic class in the training images. They call this a top-down approach as opposed to bottom-up approaches in many other colour constancy methods. They also make use of high-level visual information to select the best illuminant out of a set of possible illuminants generated by other methods.

Another learning-based approach to illumination estimation problem is the Bayesian approach [5, 13] in which the variability of reflectance of illuminant is modeled as independent random variables. These methods estimate illuminant colour from the posterior distribution condition learned from training images. Here the illuminant prior could be uniform over a subset of illuminants [5] or could be an empirical distribution of illuminants in training images [13].

3   Proposed Method

The main distinctions between this work and other learning bases colour constancy methods that use spatial information by local feature descriptors such as [14, 27, 30] is that they use this information to determined the best or combination of best possible illumination estimation algorithms while we use selected instances for illumination estimation.

Compared to the top-down approach [27] in which they assign a semantic classes to each 20×20 patch of an image based on models learned in the training phase, in our proposed model we assign to segmented regions we call surfaces, from training images to each surface of the test image. As well, [27] used the extended version of grey world assumption to estimate the illuminant whilst we use the ground truth of corresponding surfaces for illumination estimation.

3.1   Surface Model

We find surfaces for both training and test images by mean-shift segmentation, implemented via [9]. Since the pixels in the margin of segmented areas affect texture information, we remove margin pixels of segments by dilating segment edges as well as small segments.

In order to define a model for each surface we use both texture features and colour features. For the purpose of texture features, the MR8 filter bank [28] on three channels is selected for use because of its good performance in texture classification applications [29] and also its fast implementation.

The MR8 filter bank consists of 38 filters (6 orientations at 3 scales for 2 oriented filters, plus 2 isotropic) but only 8 filter responses. The filter bank contains filters at multiple orientations but it records only the maximum filter response across all orientations. We use the normalized histogram of frequency of appearance in that particular surface for each colour channel as our colour features. Since we deal with illumination variation, we divide each channel by its maximum value before computing each histogram. This makes our surface model weakly colour constant. This means that we are not interested in specific colours for our surface matching, but instead on its relative distribution.

In the learning stage, training images are convolved with a filter bank to generate dense filter responses. Exemplar filter responses are chosen as textons via K-Means clustering (with $K = 1000$) and are collected into a dictionary. The histogram of frequency of textons belonging to this dictionary is a common description for texture detection [19] although other local descriptors such as SIFT [21] may used instead of the MR8 filter bank.

Given a surface in a training image, its corresponding model is generated by first convolving it with a filter bank and then labelling each filter response with the Euclidian nearest neighbour texton in the texton dictionary. The histogram of textons, i.e. the frequency with which each texton occurs in that surface, forms the first histogram in the corresponding model for that training surface. We then add a three-channel histogram to that surface's model (we use 10 bins for each channel to be robust again noise). In order to make our model weakly invariant to variation in illuminant colour, we stretch the histogram for each channel to have maximum equal to 1 or, equally, divide the values of each colour channel by its maximum value among pixels within that surface (thus making it colour constant for Max-RGB via using a diagonal transformation). Therefore, each surface model includes four normalized histograms that are then stores in a single vector. We also need to store some meta-data for each model, which consists of ground truth illumination colour for that image as well as the maximum response for each channel used for stretching histograms. Fig. 1

shows a surface and its normalized histogram of textons and three weakly colour constant normalized histograms of colour channels.
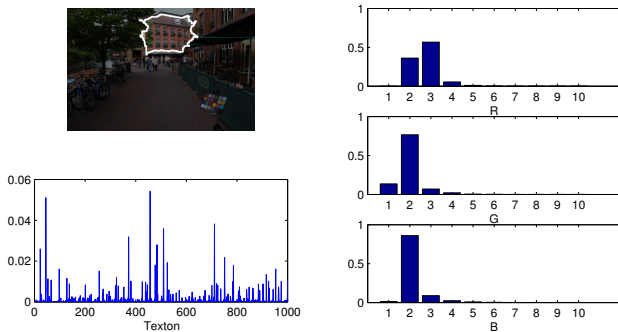

Figure 1:   A surface and its normalized histogram of textons and three weakly colour constant normalized histograms of colour channels.

In summary, the training phase for exemplar-based colour constancy is as follows:

Generate Texton Dictionary:
    convolve with MR8 filter bank to generate dense filter responses for all images
    K-Means clustering with $K = 1000$ for all filter responses
Finding Surface:
    mean-shift segmentation
Surface Model:
    The normalized histogram of textons (vector of 1000)
        the frequency with which each texton (from dictionary) occurs in that surface
    The normalized histogram of colour channel (3-vector of 10)
        divide by the max for each channel
        histogram with 10 bins for each channel

In the test stage, the same procedure is followed to build the model (one histogram of textons and three histograms of colour channels) corresponding to each surface in the test image. This model is then compared with the models corresponding to training surfaces by nearest neighbour classifier with the chi-squared statistic employed to measure distances. We select M nearest neighbours from training surfaces (with $M = 10$). Fig. 2 shows some test image surface examples and their eight nearest surface models from training data. We carry out our experiment on the re-processed version of the ColorChecker dataset but for visualization we use original images (refer to § 4 for details).

Since we have no labelled or clustered data for our training process as would be the case for a semantic segmentation task or texture detection task, we lack information for providing confidence for our mapping. Nevertheless, although we find some amount of mismatching for surfaces as shown in fig. 2, the illumination estimation process simply considers these cases as outliers compares to the other estimates.
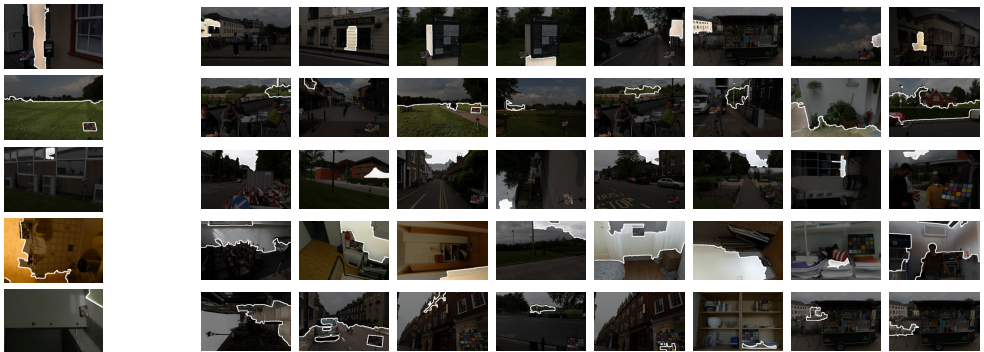
Figure 2: Surfaces from test images (on the left) and their 8 nearest surface models from training images.

3.2 Illumination Estimation

Given a test surface model and its nearest neighbour surface model from training models, we can transfer the test surface's colour to its corresponding training surface's colours linearly by a 3×3 matrix. We can approximate this matrix by a diagonal matrix as discussed in [□] and solve the transformation for each channel separately based on their channel histograms. Therefore, we can write this matrix which transforms test surface to training surface as follows.

$$D = \mathcal{M}_{test}^{-1} D_H \mathcal{M}_{train} \tag{1}$$

where $\mathcal{M}$ is the weakly colour constant diagonal transformation of surface colour from the Max-RGB method and D_H is the transformation of test surface's histograms to training surface's histograms. Since we use this histogram to find similar models these histograms are usually approximately identical or not far from each other; this means that D_H is approximately identity matrix. Finally, since the illumination colour of training surface e_{train} is known, the estimation for test surface illumination colour is:

$$e_{test} = De_{train} = \mathcal{M}_{test}^{-1} D_H \mathcal{M}_{train} e_{train} \tag{2}$$

Given a test image, we will have n large enough surfaces and M nearest neighbour surfaces from training data, or equally M illumination estimates by eq. (2) corresponding to each. The final estimate can be the median or the mean on the three channels separately after removing outliers of all of these estimates in rg chromaticity space ($[rg] = [RG]/(R+G+B)$).

We can also use weighted averaging by defining weights for each estimated illuminant according to the confidence of estimation for each surface, which we compute based on the standard deviation of estimates for that single surface and also similarity which we compute based on chi squared distance between their normalized histograms. Experiments show that none of these techniques outperform others, and therefore for simplicity we estimate the final illuminant by finding the median on the three channels separately over all estimates for three channels separately.

Figure 3 shows the procedure for illuminant estimation for a test image using exemplar-based color constancy. Figure 4 shows a test image and the angular error of estimated illu-
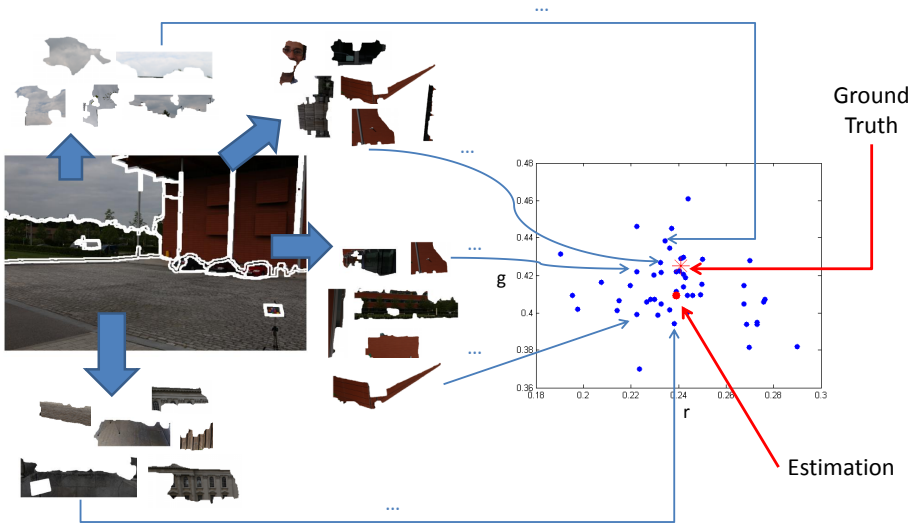
Figure 3: The procedure of estimating illuminant for a test image using exemplar-based color constancy. A test image and its nearest neighbour surface models from training images on left and estimated illuminants according to each model in *rg* chromaticity space on right.

minant for its surfaces; as mentioned, we do not form any estimate for small segments. We see that more textured surfaces such as grass or textured road have more precise estimates compared to smooth road.

In summary, the proposed illumination estimation for a test image is as follows:

Finding Surfaces
For each surface
    generate Surface Models (vector of 1030)
    find $M = 10$ nearest neighbour by chi distance in training surface models
    for each nearest neighbour surface model
        estimate the illuminant by eq. (2)
Estimate unique illuminant
    integrate all estimates in chromaticity space (*rg*) by finding median .

---

4 Experiments

We applied our proposed method to two standard colour constancy datasets of real images of indoor and outdoor scenes. The first dataset is the re-processed version of the Gehler colour constancy dataset [13], denoted the ColorChecker dataset, (provided by Shi and Funt
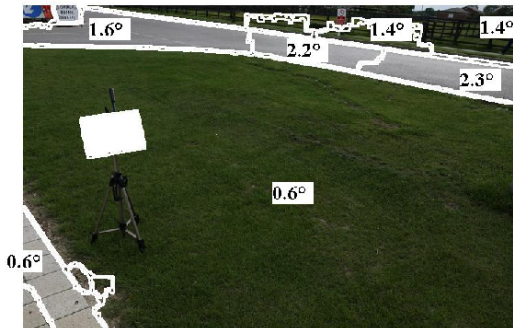
Figure 4: A test image and angular errors of estimated illuminant for its surfaces (computation is not carried out for small segments).

Method	Median Er	Mean Er
White-Patch	5.7°	7.4°
Grey-World	6.3°	6.4°
Grey-Edge ($p = 1, \sigma = 6$)	4.5°	5.3°
Bayesian [13]	3.5°	4.8°
Gamut Mapping pixel ($\sigma = 4$)	2.5°	4.1°
Gamut Mapping 1jet [15] ($\sigma = 9$)	2.5°	4.1°
Bottom-up+Top-down [27]	2.5°	3.5°
Natural Image Statistics	3.1°	4.2°
Exemplar-Based	2.3°	3.1°

Table 1: Angular errors for ColorChecker dataset [24] in terms of mean and median for several colour constancy algorithms.

[24]). We prefer to use the re-processed version because it contains raw image data and attempts to generate linear sensor values which is critical for $3x3$ matrix transformation. This dataset consists of 568 images, both indoor and outdoor. The illuminant ground truth for these images is known because each image has a Macbeth ColorChecker placed in the scene — which must be masked off in tests. For this dataset, we used three-fold cross-validation to learn our models which provide by original dataset, as used by other learning based methods we compared to. Table 1 indicates the accuracy of the proposed methods for the ColorChecker dataset [24], in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. For those methods which need tunable parameters, we utilize optimal parameters for this dataset.

Another dataset, which contains low quality real images, is the GreyBall dataset of Ciurea and Funt [8]; this contains 11346 images extracted from video recorded under a wide variety of imaging conditions. The ground truth was acquired by attaching a grey sphere to the camera, displayed in the bottom-right corner of the image (again masked for experiments). In order to learn our models for this dataset, we use 15 folds each of which represents a recorded video as provided by dataset itself similar to other learning based methods we

Method	Median Er	Mean Er
White-Patch	5.3°	6.8°
Grey-World	7.0°	7.9°
Grey-Edge ($p = 1, \sigma = 1$)	4.7°	5.9°
Gamut Mapping pixel ($\sigma = 4$)	5.8°	7.1°
Gamut Mapping 1jet [15] ($\sigma = 9$)	5.8°	6.9°
Natural Image Statistics [14]	3.9°	5.2°
Exemplar-Based	3.3°	4.4°

Table 2: Angular errors for GreyBall dataset [8] in term of mean and median for several colour constancy algorithms.

compared to. Table 2 shows the performance of our proposed method for the GreyBall dataset [8], in terms of the mean and median of angular errors, for several colour constancy algorithms applied to this dataset. We also utilize optimal parameters for this dataset for those methods which need tunable parameters.

To our knowledge, for these two standard datasets, widely used for testing colour constancy, Exemplar-Based Colour Constancy does best in terms of both mean and median angular error compared to any reported colour constancy methods, even those using a combination of algorithms such as Natural Image Statistics [14].

5 Conclusion

In this paper we present a completely new line of approach to colour constancy problem called Exemplar-Based colour constancy. We use both texture features and weakly colour constant three channel colour values in order to find the nearest neighbour surfaces from training data for each surface, and then we estimate the illuminant for each surface based on histogram matching of each surface to its candidate nearest neighbour surfaces from training data. The final step is integrating these estimates to a unique illuminant estimation for whole image. The proposed method has the advantage of overcoming the difficulty of multi-illuminant situations, which is not possible for most current methods. We show that the proposed method performs very well for two standard datasets compares to current colour constancy algorithms.

In future, we can apply more complex methods of integrating estimated surface illuminants to estimate unique illumination. Also, we should construct a dataset for multi-illuminant colour constancy in order to evaluate Exemplar-Based Colour Constancy for images with more than one light source colour.

References

[1] D. W Aha, D. Kibler, and M. K Albert. Instance-based learning algorithms. *Machine learning*, 6 (1):37–66, 1991.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition, 2000.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

[4] G. Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310: 1–26, 1980.

[5] A. Ladsariya C. Rosenberg, T. Minka. Bayesian color constancy with non-gaussian models, 2003.

[6] V. C. Cardei and B. V. Funt. Committee-based color constancy. In *IS&T/SID Color Imaging Conference*, pages 311–313, 1999.

[7] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy beyond bags of pixels. In *Computer Vision and Pattern Recognition, CVPR IEEE Conference on*, 2008.

[8] F. Ciurea and B. V. Funt. A large image database for color constancy research. In *IS&T/SID Color Imaging Conference*, pages 160–164, 2003.

[9] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[10] G.D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Twelfth Color Imageing Conference: Color, Science, Systems and Applications.*, pages 37–41, 2004.

[11] G.D. Finlayson, M.S. Drew, and B.V. Funt. Color constancy: diagonal transforms suffice. In *Proc. Fourth Int. Conf. on Comp. Vision, Berlin*, pages 164–171. IEEE, 1993.

[12] D.A. Forsyth. A novel approach to color constancy. In *Int. Conf. on Computer Vision '88*, pages 9–18, 1988.

[13] P. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *In Proc. Comp. Vis. and Patt. Rec. (CVPR)*, 2008.

[14] A. Gijsenij and T. Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 33(4):687–698, 2011.

[15] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *Int. J. Comput. Vision*, 86:127–139, 2008.

[16] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20:2475–2489, 2011.

[17] S. D. Hordley. Scene illuminant estimation: past, present, and future. *Color Research and Application*, 31(4):303–314, 2006.

[18] E. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, December 1977.

[19] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[20] C. Liu, J. Yuen, and A. Torralba. Nonparametric scene parsing: Label transfer via dense scene alignment. In *Computer Vision and Pattern Recognition, CVPR IEEE Conference on*, pages 1972–1979, 2009.

[21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. In *Int. J. of Comput. Vision*, volume 60, pages 91–110, 2004.

[22] T. Malisiewicz, A. Gupta, and A. A Efros. Ensemble of exemplar-svms for object detection and beyond. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 89–96, 2011.

[23] G. Shakhnarovich, P. Viola, and T. Darrell. Fast pose estimation with parameter sensitive hashing. In *ICCV*, pages 750–757, 2003.

[24] L. Shi and B. Funt. Re-processed version of the gehler color constancy dataset of 568 images, 2010. http://www.cs.sfu.ca/ colour/data.

[25] A. Torralba, R. Fergus, and W.T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):1958–1970, 2008.

[26] J. van de Weijer and T. Gevers. Color constancy based on the grey-edge hypothesis. In *Int. Conf. on Image Proc.*, pages II:722–725, 2005.

[27] J. van de Weijer, C. Schmid, and J. Verbeek. Using high-level visual information for color constancy. In *Proc. Int. Conf. on Comp. (ICCV)*, 2007.

[28] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the 7th European Conference on Computer Vision-Part III*, ECCV '02, pages 255–271, 2002.

[29] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *Int. J. Comput. Vision*, 62(1-2):61–81, April 2005.

[30] M. Wu, J. Sun, J. Zhou, and G. Xue. Color constancy based on texture pyramid matching and regularized local regression. *J. Opt. Soc. Am. A*, 27(10):2097–2105, 2010.

[31] J. Xiao, J. Hays, K. A Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), IEEE conference on*, pages 3485–3492, 2010.