MIXED POOLING NEURAL NETWORKS FOR COLOR CONSTANCY

D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Trémeau

Laboratoire Hubert-Curien UMR 5516
Saint-Etienne - France

C. Wolf

LIRIS, UMR5205
Lyon - France

ABSTRACT

Color constancy is the ability of the human visual system to perceive constant colors for a surface despite changes in the spectrum of the illumination. In computer vision, the main approach consists in estimating the illuminant color and then to remove its impact on the color of the objects. Many image processing algorithms have been proposed to tackle this problem automatically. However, most of these approaches are handcrafted and mostly rely on strong empirical assumptions, e.g., that the average reflectance in a scene is gray. State-of-the-art approaches can perform very well on some given datasets but poorly adapt on some others. In this paper, we have investigated how neural networks-based approaches can be used to deal with the color constancy problem. We have proposed a new network architecture based on existing successful hand-crafted approaches and a large number of improvements to tackle this problem by learning a suitable deep model. We show our results on most of the standard benchmarks used in the color constancy domain.

***Index Terms*—** Color constancy, Neural networks, Light color estimation, Pooling, Data augmentation

1. INTRODUCTION

RGB outputs of the cameras are the only "color information" we have in many computer vision tasks. However, these values can not be considered as intrinsic features of the observed surfaces since they are the result of the interactions between the current light in the scene, the reflection properties of these surfaces and the camera sensors and post-processing. The color constancy is the ability of the human visual system to perceive constant colors for a surface despite changes in the spectrum of the illumination. For computer vision, many color constancy algorithms have been proposed in the last decades as pre-processing steps [1]. Except few physics-based algorithms [2], most of the approaches are based on empirical assumptions. Starting from the well-known Grayworld approach [3] which assumes that the average reflectance in a scene is constant with respect to the wavelength, a large range of other assumptions have been proposed to found the color constancy algorithms [4]. Since these assumptions are not based on physical rules, one can wonder if the optimal assumption could not be discovered by learning it from real images. Thus, the recent trend consists in learning color constancy algorithms from labeled public datasets [5, 4, 6]. Most of these approaches are learning either a combination of unitary approaches [4] such as gray-world or gray-edge, a correction matrix from the gray-world estimation [6] or they are still using handcrafted features to match patches [5].

Unlike all these approaches, the aim of this paper is to check if an accurate color constancy algorithm can be learned without using any handcrafted features or any unitary algorithm as basis. We start from the assumptions that local filtering seems to improve illumination estimation [7, 8, 9] as well as the combination/pooling of local and global features [10]. Given these requirements, deep networks seem to be perfect tools for this kind of application, since they have shown to provide excellent results in many computer vision and machine learning tasks such as image classification [11]. Thus, we propose two different neural network architectures dedicated to the color constancy task and we show that state-of-the-art results can be obtained thanks to deep networks on the available public datasets. Finally, we propose and test different data augmentation approaches. We will see that this step tends to improve only the results of the architecture with convolutional layers. We claim tha the extensive tests and results provided in this paper can help researchers to design new architectures to tackle the color constancy problem.

2. RELATED WORK

There exist three main categories of algorithms devoted to the color constancy problem. The first one contains the most widely used algorithms in the last decade which are exploiting the statistics of real color images. Among them, we can cite the gray-world [3], the Shades-of-Gray [12], the max-RGB [13] or the gray-edge [7] that have all been unified in a general framework proposed by van de Weijer *et al.* [7]. Recently, some other statistics-based approaches have been successfully applied to this problem [14, 10, 15]. All these methods are based on strong empirical assumptions.

The second category regroups the physics-based algorithms and mainly exploits the dichromatic reflection model from Shafer [2, 16]. Compared to the previous algorithms, these are well founded because they start from accurate re-

flection models but they require to detect the specularities [2] or to segment the images [16].

The third category contains all the learning-based approaches starting from the Gamut Mapping methods [17, 8, 18, 19] or the recent patch-based approach [5] that estimate the light color of a local region from the light colors of a set of ground truth regions that have similar contents. This last approach provides state-of-the-art results but its drawbacks are twofold. First, it requires to store thousands of patches in memory. Second, the estimation procedure for a test image involves many successive steps (segmentation, feature extraction, nearest neighbor search in the training set, local and global estimations). Finlayson proposes a fastest learning-based approach [6]. The idea is to estimate the light color with the classical gray-world approach and then to correct this coarse estimation with a matrix that is learned on a dataset and whose elements depend on the color and edge moments of the image. More recently, Bianco *et al.* [20] proposed to use a convolutional neural network, obtaining state-of-the-art results. The network consists of one convolutional layer with max pooling, one fully connected layer and three output nodes. The approach first learns from local input patches and then uses fine-tuning, minimizing the loss over whole images and not over patches. This approach provides very good results on one specific dataset but no results are reported on other datasets. Finally, very recently Lou *et al.* [21] proposed a deep convolutional neural network that is pre-trained on the big ImageNet dataset with labels evaluated from hand-crafted color constancy algorithms and fine-tuned on each single dataset with groundtruth labels. In this article, we propose an original deep network architecture and assess its quality on all classical datasets. We run extensive tests to measure the impact of parameters and of the proposed data augmentation approaches.

3. PROPOSED NEURAL-NETWORK ARCHITECTURES

We propose two neural network architectures dedicated to color illuminant regression.

3.1. Mixed Max-Minkowski pooling networks

Our neural model, illustrated in Fig. 1, is inspired by the color constancy solution of Gao *et al.* [10] that compared two measures computed at different scales: locally normalized surface reflectance is compared to the global pixel average of an image. Our network emulates different scales with two parallel pooling operations, which we call mixed pooling. One pooling path uses maximum pooling, the second uses average pooling (i.e., downscaling). Using a pooling support of size 20×20, a 180×120 input image produces some 9×6 feature maps. The pooling layer is followed by two fully-connected layers with $Tanh$ activation function and a three-dimensional output layer. The fully connected layers learn to combine information collected from the regions of the image and the different forms of pooling.

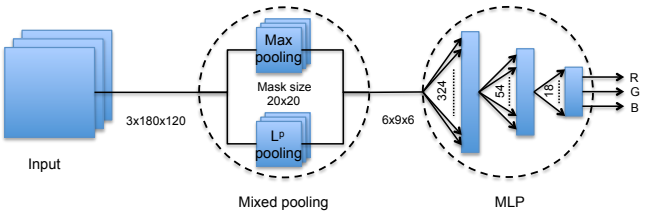

Fig. 1. A neural network model based on Minkowski pooling with large support (20×20) and a fully connected layer.

Combining two pooling operations is inspired by Gao *et al.* [10] and, along the lines of [7], we generalize the second pooling path with a Minkowski pooling using a p-norm (L^p).

3.2. Convolutional Neural Network with mixed pooling

Color illuminant estimation can be improved by adding feature extraction before pooling. In particular, the Gray edge method [7] computes image derivatives that are integrated over image. We decided to add a convolutional layer before the mixed-pooling layer and, since the filters can learn to compute image derivatives, this is a generalization of edge extraction.

We use a convolutional layer with 12 filters of size 7×7 before the mixed pooling (still 20x20). A 180×120 input image produces 24 feature maps (12 filters $\times$ 2 poolings), each of 8×5 pixels (the convolution being undefined at the borders).

4. DATA AUGMENTATION

Convolution networks are not known to be invariant to basic image transformations like translation, rotation, deformations etc. Data augmentation is a frequent means to cope with this issue, basically applying the targeted transformations to the input data instead of integrating invariance into the model. Along the same lines, we resort to a variant of data augmentation adapted to the task at hand. We explored two methods, which we call light transfer augmentation and patch fusion. We will see in the experimental section that these data augmentation approaches only help when convolutional layers are present in the network and we think that it is important to share this experience with the researchers working in this area.

4.1. Light transfer augmentation

Our goal here is to create multiple new training images from a single image in the original training set by artificially chang-

ing the illumination in the image. To this end, we first compute the unbiased image by applying a correction using the (known) color illuminant and the von Kries diagonal transform [22]. This operation is called *light transfer* and it is done through a channel-wise transformation applied to the RGB intensity channels:

$$\begin{bmatrix} R_2 \\ G_2 \\ B_2 \end{bmatrix} = \begin{bmatrix} e2_R/e1_R & 0 & 0 \\ 0 & e2_G/e1_G & 0 \\ 0 & 0 & e2_B/e1_B \end{bmatrix} \times \begin{bmatrix} R_1 \\ G_1 \\ B_1 \end{bmatrix}$$

where $[R_1, G_1, B_1]^T$ ($[R_2, G_2, B_2]^T$ respectively) is the original (transformed, resp.) color vector of a pixel and $[e1_R, e1_G, e1_B]^T$ ($[e2_R, e2_G, e2_B]^T$, resp.) is the original (new, resp.) light color. The objective of data augmentation is to increase the number of labeled training data. However, a domain shift between the original dataset and the new augmented dataset should be avoided (assuming that the distribution of the training samples is close to the distribution of the unknown test situation). We therefore simulate new images by sampling new color illuminants from the original ground truth distribution. In order to avoid changing indoor illuminations to outdoor illuminations (or vice-versa), it is possible to impose a limit on the difference between the original ground truth color illuminant and the simulated one. In practice, experiments showed that this is not necessary.

4.2. Patch fusion

The task of the network is to predict the color illuminant by integrating estimations over the input image. However, this integration does not necessarily need to be done over the full input image, as different parts of the image are supposed to be illuminated with the same color. In the lines of [20], we therefore resort to a patch-wise process, where the network is trained on input patches. During testing, patches are sampled from the test image and the predictions are averaged over the patches.

This patch-wise solution has three advantages over a global method. Firstly, the size and complexity of the network decreases, limiting overfitting. Second, patch-wise training is an explicit form of data augmentation, increasing the number of labeled training samples. And third, a local estimation can tackle the eventual problem of spatially non-uniformly distributed light, by not combining all the local estimates into a global one (not done in this paper).

5. EXPERIMENTAL RESULTS

For the experiments, we use the 5 most used datasets for color constancy which are variations of 2 original ones, namely the Original SFU Gray-ball (GBO) [23] containing 11346 images and the Color-Checker Original (CCO) [19] containing 568 images. The linear version of the GBO dataset is called hereafter the Linear SFU Gray-ball (GBL) and the liear versions of the CCO dataset are called Color-Checker by Shi (CC-Shi) [24] and Color-Checker reprocessed (CCR) [25]. It is worth mentionning that most of the color constancy approaches (statistics-based and physics-based) are designed for linear data, but since many papers present their results on both linear and non-linear data, we also provide results for both.

All models have been implemented using Torch7. We used early stopping and Resilient backpropagation (Rprop [26]) for optimization. Rprop uses Stochastic Gradient Descent (SGSD) but it dynamically adapts the step of each parameter to increase the convergence speed.

5.1. Comparison with the state of the art

Since the intensity of the illuminant cannot be recovered from a single image, color constancy algorithms aim at estimating its chromaticity. Therefore, in order to evaluate the quality of an estimate, the most widely used criterion is the angle in the color space between the estimated illuminant and the ground truth illuminant (angular error). We report in Table 1, the mean and median angular errors, averaged by cross-validating over N-folds.

The table is divided into three parts. The first one contains the methods based on human expertise only, i.e. without any machine learning component; the second part gives the learning based methods and the last one presents our proposed approach with maximum and mixed pooling neural networks without convolutional layers. As recommended by their authors, we have used a -129 offset in the CC-Shi dataset (only for camera 5D) before testing it. Thus, the results of the 6 first unitary approaches (6 first rows) are better than the ones usually presented in color constancy papers, but they are concordant with the ones of the CC-Shi authors [4]. This table shows that our method is better than every static (non learning) method and is comparable to the learning ones. Our method is comparable to the Exemplar-Based [5] method and outperforms it on the Gray-ball linear dataset.

5.2. Network architecture

In Table 2, we notice that the model without convolutional layers turned out to outperform the convolutional one. The convolution layers introduce new parameters and seems to unnecessarily increase the expressive power of the network which leads to overfitting. This is further corroborated by the fact that data augmentation performed on the convolutional network does improve the performance. This confirms the findings of [20], whose deep network does not contain convolutions[1] and validates the intuitions of [14] that local differences do not help for color constancy. We estimate that

[1] To be precise, the paper [20] presents one of the layers as "convolutions of size 1×1", which corresponds to learning a non-linear pixelwise transformation which is shared over all pixels of the image.

Method	CCO	CC-Shi	CCR	GBO	GBL
Gray World [3]	9.8 7.4	4.78 3.63	5.33 3.98	7.9 7.0	13.0 11.0
White Patch [13]	8.1 6.0	5.31 3.15	6.44 4.10	6.8 5.3	12.7 10.5
Shades-of-Gray [12]	7.0 5.3	4.40 2.72	3.98 **2.35**	6.1 5.3	11.6 9.7
general Gray-World [7]	7.0 5.3	4.21 2.70	×	6.1 5.3	11.6 9.7
1st-order Gray-Edge [7]	7.0 5.2	3.72 2.86	5.02 2.88	5.9 4.7	10.6 8.8
2nd-order Gray-Edge [7]	7.0 5.0	3.59 2.64	×	6.1 4.9	10.7 9.0
Local Surface Refl. [10]	×	3.4 2.6	×	6.0 5.1	×
Large Col. diff. [14]	×	3.52 2.14	×	×	×
Grey patches [15]	×	4.6 3.1	×	6.1 4.6	×
Exemplar Based [5]	**5.2** **3.7**	3.1 2.3	×	**4.4** **3.3**	**8.0** **6.5**
Corr.-moments [6]	×	**2.8** 2.0	×	×	×**
SVRC-R [4]	×	× **1.97**	×	×	× 6.81*
CNN [20]	×	**2.63** **1.98**	×	×	×
Single max pooling	6.18 5.03	3.31 2.59	**3.69** **2.70**	5.18 4.51	8.16 7.08
Mixed MaxL^5 pooling	**6.17** **4.92**	3.33 2.63	**3.70** 2.80	**4.94** **4.28**	**7.65** **6.53**

Table 1. Comparison with the state-of-the-art methods. Mean (up) and median (down) angular errors are reported. For each dataset, the 2 best results (for mean and for median independently) are in bold. '*' means tested on a subset of (uncorrelated) data (1135 images among 11346). '**' means that the results provided by [6] can not be fairly compared with the ones of this table since they are evaluated only on 150 images (among 11346) and by using a 3-fold cross validation, which does not respect the 15-fold cross validation used by the other approaches in order to remove the strong correlation between the images within each video. The two last line show our method without using any convolutional layers.

deep color constancy could highly benefit from a very large amount of labeled training data.

The main advantage of the proposed method is its fast inference at test time. As mentioned in the introduction, in order to estimate the light color of a new image, the Exemplar-Based method needs to segment the image, extract color and texture features from all obtained segments and to run a nearest neighbor search in the training set containing thousands of local features. This also means that the training set is required at test time.

5.3. Data augmentation Results

The results of the different data augmentation methods are given in Table 2. As already reported for image classification in [11], data augmentation increases the performance in the presence of convolutional layers. In particular, the light transfer augmentation gives the best results for the convolutional network. For the non-convolutional architecture, increasing the number of samples through patch-wise processing seems to excessively decrease the information contained in each sample and thus gives worse results.

	Without conv. layers		With conv. layers	
Method	Mean	Median	Mean	Median
No augmentation	**3,33**	**2,63**	3,91	3,06
Light transfer augmentation	3,38	2,69	**3,49**	**2,73**
Patch fusion	3,66	3,00	3,78	2,95

Table 2. Effect of different data augmentation methods on the Color-Checker by Shi dataset on neural architectures with and without using convolutional layers.

6. CONCLUSION

In this paper, we have proposed two new deep architectures that are exploiting the available expert knowledge at hand, i.e. our networks are designed so that they can reproduce the processing of the best non-learning methods. The extensive experimental tests actually show that these two networks outperform these non-learning algorithms on all the datasets and are competitive (and sometimes better) than the other learning-based solutions which are much more complex at inference time. In order to be able to exploit more data during the learning step, it could be interesting to exploit several datasets at learning time and take advantage of domain adaptation approaches in order to remove the distribution shift between the different camera sensors.

7. ACKNOWLEDGMENT

Authors acknowledge the support from the ANR project SoL-StiCe (ANR-13-BS02-0002-01).

8. REFERENCES

[1] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *Image Processing, IEEE Transactions on*, vol. 20, no. 9, pp. 2475–2489, Sept 2011.

[2] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Am. A*, vol. 21, no. 3, pp. 321–334, Mar 2004.

[3] G. Buchsbaum, "A spatial processor model for object colour perception," *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1 – 26, 1980.

[4] B. Li, W. Xiong, W. Hu, and B. V. Funt, "Evaluating combinational illumination estimation methods on real-world images," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1194–1209, 2014.

[5] H.R.V. Joze and M.S. Drew, "Exemplar-based color constancy and multiple illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 860–873, May 2014.

[6] G.D. Finlayson, "Corrected-moment illuminant estimation," in *IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 1904–1911.

[7] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2207–2214, Sept 2007.

[8] A. Gijsenij, T. Gevers, and J. van de Weijer, "Generalized gamut mapping using image derivative structures for color constancy," *International Journal of Computer Vision*, vol. 86, no. 2-3, pp. 127–139, 2010.

[9] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 8, pp. 1509–1519, Aug 2012.

[10] S. Gao, W. Han, K. Yang, C. Li, and Y. Li, "Efficient color constancy with local surface reflectance statistics," in *European Conference on Computer Vision (ECCV)*, 2014, vol. 8690, pp. 158–173.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds. 2012, pp. 1097–1105, Curran Associates, Inc.

[12] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy.," in *Color Imaging Conference*. 2004, pp. 37–41, IS&T - The Society for Imaging Science and Technology.

[13] B. Funt and L. Shi, "The rehabilitation of maxrgb," in *18th Color and Imaging Conference*, 2010, pp. 256–259.

[14] D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," *J. Opt. Soc. Am. A*, vol. 31, no. 5, pp. 1049–1058, May 2014.

[15] K.-F. Yang, S.-B. Gao, and Y.-J. Li, "Efficient illuminant estimation for color constancy using grey pixels," 2015.

[16] G. Finlayson and G. Schaefer, "Solving for colour constancy using a constrained dichromatic reflection model," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 127–144, 2001.

[17] D.A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–35, 1990.

[18] G.D. Finlayson, S.D. Hordley, and P.M. HubeL, "Color by correlation: a simple, unifying framework for color constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209–1221, Nov 2001.

[19] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, "Bayesian color constancy revisited," 06 2008, pp. 1–8.

[20] S. Bianco, C. Cusano, and R. Schettini, "Color constancy using cnns," in *DeepVision: Deep Learning in Computer Vision (CVPR workshop)*, 2015.

[21] Z. Lou, T. Gevers, N. Hu, and M. Lucassen, "Color constancy by deep learning," in *British Machine Vision Conference*, 2015.

[22] J. von Kries, "Influence of adaptation on the effects produced by luminous stimuli," in *Sources of color science*, D.L. MacAdam, Ed., Handbook of Stuff I Care About, pp. 120–126. MIT Press, Cambridge, MA, 1970.

[23] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Color Imaging Conference*, 2003, number 1, pp. 160–164.

[24] L. Shi and B. Funt, "Re-processed version of the gehler color constancy dataset of 568 images," *Simon Fraser University*, 2010.

[25] S.E. Lynch, M.S. Drew, and G.D. Finlayson, "Colour constancy from both sides of the shadow edge," in *IEEE ICCV Workshops*. IEEE, 2013, pp. 899–906.

[26] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The rprop algorithm," in *IEEE International Conference on Neural Networks*. IEEE, 1993, pp. 586–591.