Experimental analysis of colour constancy and colour augmentation for painting classification by artistic genre: preliminary results

Francesco Bianconi
Department of Engineering, Università degli Studi di Perugia
Via Goffredo Duranti 93, 06125 Perugia, Italy
E-mail: bianco@ieee.org

Abstract. Automatic painting classification by author, artistic genre and/or other attributes has generated considerable research interest in recent years. Being one of the visual features that mark the difference between artists and artistic genres, colour plays a fundamental role in this process. Colour is the result of the interaction among the intrinsic properties of the material, the illumination conditions and the response of the imaging device. Consequently, the same painting/artwork will look significantly different when imaged under varied conditions, which can be a potential source of bias for automated recognition procedures. One can compensate for such variations either via colour calibration or colour pre-processing. In this work we investigate the latter, and, in particular, evaluate the effectiveness of colour constancy and colour augmentation when coupled with hand-crafted and deep learning features for painting classification by artistic genre. In our experiments neither approach showed a clear advantage compared with no pre-processing at all. Colour constancy brought some improvement in certain cases, whereas colour augmentation virtually provided no benefit despite its adding a significant computational overload to the procedure.

1. Introduction
Automatic painting classification by author, artistic genre and/or other attributes has been attracting increasing research interest in recent years [1, 2, 3]. In this process colour plays a major role, for it is one of the visual features that mark the difference among artists and artistic genres [4]. Colour is the result of the interaction among the intrinsic properties of the material, the illumination conditions and the response of the imaging system. As a consequence, the same painting/artwork will significantly different when imaged under variable conditions, and this represents a source of potential bias for automated recognition procedures.

There are two main strategies to tackle this problem: colour calibration and colour pre-processing. The first consists of converting the colour data into a device-independent space. To this end one needs to know both the illumination conditions and the spectral sensitivity of the imaging device; or, alternatively, it is required to have a colour calibrated rig (for instance a colour checker) in the scene. Both solutions are hard to put into practice, since most painting databases do not provide such kind of information. The second involves the application of some colour pre-processing procedure. These usually come in two varieties: colour constancy and colour augmentation. The first aims at generating a kind of ‘colour neutral’ image from
the original one; the second at creating a number of images that are ‘colour compatible’ with the original one – i.e. that could have been obtained from the same subject if acquired under different illumination conditions and/or with different imaging devices.

The objective of this paper is to investigate whether colour constancy and colour augmentation can improve the accuracy of automatic classification of paintings by art genre.

2. Materials

We considered a collection\(^1\) of \(N = 7724\) images arranged into 12 classes representing the following genres/art movements [5, 6, 7]: Abstract expressionism \((n = 340)\), Baroque \((n = 960)\), Cubism \((n = 920)\), Fauvism \((n = 426)\), High Renaissance \((n = 818)\), Iconoclasm \((n = 665)\), Impressionism \((n = 984)\), old Greek pottery \((n = 350)\), Realism \((n = 307)\), Rococo \((n = 844)\), Romanticism \((n = 874)\) and Surrealism \((n = 242)\). The images are of variable size and come with no additional information regarding the illumination conditions or the imaging device. The dataset includes four predefined folds for assessing the accuracy of classification algorithms.

3. Methods

3.1. Colour constancy

The objective of colour constancy is to obtain a representation of a scene independent of the imaging conditions. Images of the same scene acquired under different lighting conditions should in theory look the same once processed for colour constancy [8]. We considered four colour constancy methods in this study: chromaticity representation \((\text{chroma in the remainder})\), grey world normalisation \((\text{gw})\), max white normalisation \((\text{maxw})\) and histogram equalisation \((\text{stretch})\). We recall the basics of each method here below and refer the reader to [9, 10] for further details. Figure 1 shows the results on a sample image.

**Chromaticity representation** The original \((R, G, B)\) colour coordinates are converted to the normalised ones \((R', G', B')\) in the following way:

\[
R' = \frac{R}{R + G + B}, \quad G' = \frac{G}{R + G + B}, \quad B' = \frac{B}{R + G + B}
\]

**Grey world normalisation** The method is based on the assumption that the average colour of the image is grey. A gain is therefore applied to force the average intensities of the R, G and B channels to be equal:

\[
R' = R \times \left(\frac{\mu_G}{\mu_R}\right) \quad G' = G \quad B' = B \times \left(\frac{\mu_G}{\mu_B}\right)
\]

where \(\mu_R\), \(\mu_G\) and \(\mu_B\) are the average intensities of the respective channels.

**Max white normalisation** In this case the R, G and B values are scaled in such a way that the brightest white point in the image saturates the capacity of the channels:

\[
R' = (L - 1)/R_{\text{max}} \quad G' = (L - 1)/G_{\text{max}} \quad B' = (L - 1)/B_{\text{max}}
\]

where \(L\) is the number of quantization levels; and \(R_{\text{max}}, G_{\text{max}}\) and \(B_{\text{max}}\) the maximum values of each channel.

**Stretch** This approach shifts each colour channel to the left first (Eq. 3), then processes the results through \text{maxw}.

\[
R' = R - R_{\text{min}} \quad G' = G - G_{\text{min}} \quad B' = B - B_{\text{min}}
\]

\(^1\) Pandora7k – Perceptual ANalysis and DescriptiOn of Romanian visual Art
3.2. Colour augmentation

Colour augmentation consists of generating a set of images that are colour-compatible with the original one – i.e. which could represent the same scene as in the original image when acquired under different imaging conditions. The method proposed here is a variation of the one described in [11], the main difference being that the strength of the displacement is deterministic (fixed) here and random in the cited reference.

For a given input image we first compute the principal components of the colour distribution in the RGB space, then apply a translational perturbation along one or more of the principal axes in the following way:

$$
\begin{pmatrix}
R' \\
G' \\
B'
\end{pmatrix} =
\begin{pmatrix}
R \\
G \\
B
\end{pmatrix} +
\begin{bmatrix}
\alpha \rho_1 \\
\beta \rho_2 \\
\gamma \rho_3
\end{bmatrix}
\begin{bmatrix}
v_{1R} & v_{1G} & v_{1B} \\
v_{2R} & v_{2G} & v_{2B} \\
v_{3R} & v_{3G} & v_{3B}
\end{bmatrix}
\begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
$$

(4)

where \(v_i, i \in \{1, \ldots, 3\}\) denotes the unit-length vector of the \(i\)-th principal axis (sorted in descending order of inertia – i.e.: \(i = 1 \rightarrow \) highest inertia, \(i = 3 \rightarrow \) lowest inertia), \(\rho_i\) the radii of gyration along the \(i\)-th principal axis; \(\alpha, \beta\) and \(\gamma\) the user-defined multiplying factors determining the entity of the displacement respectively along \(v_1, v_2\) and \(v_3\). Figure 2 shows the effects of colour augmentation on a sample image for different combinations of \(\alpha, \beta\) and \(\gamma\). In the experiments we evaluated five different configurations as detailed in Tab. 1.

Table 1. Summary table of the configurations used for colour augmentation. \(N\) indicates the number of images in the original dataset. For the meaning of the perturbation values please refer to Fig. 2 and Eq. 4.

| Configuration | Perturbation values | Size of the augmented dataset |
|---------------|---------------------|-------------------------------|
| A             | (5\%, 0, 0), (-5\%, 0, 0) | 3N                            |
| B             | (5\%, 0, 0), (-5\%, 0, 0), (10\%, 10\%, 0) | 4N                            |
| C             | (5\%, 0, 0), (-5\%, 0, 0), (10\%, 10\%, 0), (-10\%, -10\%, 0) | 5N                            |
| D             | (5\%, 0, 0), (-5\%, 0, 0), (10\%, 10\%, 0), (-10\%, -10\%, 0), (20\%, 20\%, 20\%) | 6N                            |
| E             | (5\%, 0, 0), (-5\%, 0, 0), (10\%, 10\%, 0), (-10\%, -10\%, 0), (20\%, 20\%, 20\%), (-20\%, -20\%, -20\%) | 7N                            |

Figure 1. Effects of colour constancy on a sample image (left).
Figure 2. Effects of colour augmentation on a sample image. The values beneath each image indicate the entity of the shift along the first, second and third principal directions of the colour distribution in the RGB colour space. For instance (+15%, +15%, 0) stands for \( \alpha = 0.15, \beta = 0.15 \) and \( \gamma = 0.0 \) (see Eq. 4).

3.3. Image descriptors
We tested the effects of colour constancy and colour augmentation on five pure colour descriptors, five grey-scale texture descriptors and five pre-trained convolutional neural networks (CNN). We summarise the basics of each method in the remainder and again refer the interested reader to the given references for further details.

Colour descriptors
• Mean: Average value of each colour channel (three features [12])
• MeanStd: Average value and standard deviation of each colour channel (six features [13])
• Percentiles: Twentieth, 40th, 60th and 80th percentile of each colour channel (4 × 3 = 12 features [14])
• FullHist: Joint, three-dimensional colour histogram with eight bins per channel (\( 8^3 = 512 \) features [15])
• **MargHists**: Concatenation of the marginal colour histograms of each colour channel with 256 bins per channel ($256 \times 3 = 768$ features [16])

**Grey-scale texture descriptors**

• **DCF**: Average value and standard deviation of the transformed images from a bank of 25 filters obtained by vertical and horizontal combination of one-dimensional discrete cosine filters at five different frequencies ($25 \times 2 = 50$ features [17])

• **Gabor**: Average value and standard deviation of the transformed images from a bank of 25 circular Gabor filters with five frequencies and five orientations ($25 \times 2 = 50$ features [12])

• **Laws**: Average value and standard deviation of the transformed images from a bank of 25 filters obtained by horizontal and vertical combinations of one-dimensional Laws’ masks ($25 \times 2 = 50$ features [18])

• **LBP**: Concatenation of directional local binary patterns histograms computed over eight-pixels digital circles at resolution 1px, 2px ad 3px ($256 \times 3 = 768$ features [19, 20])

• **Zernike**: Average value and standard deviation of the transformed images from a bank of filters from even and odd Zernike polynomials of order $k = \{1, \ldots, 6\}$ ($6 \times 7 \times 2 = 84$ features [21])

Apart from LBP, all the filters were defined over a spatial mask of 5px $\times$ 5px. Conversion to grey-scale was carried out by computing the luminance: $l = 0.3R + 0.59G + 0.11B$ [22].

**Pre-trained convolutional networks** We considered five pre-trained CNN with weights learned on the ImageNet dataset. All the networks were used as ‘off-the-shelf’ feature extractors [23, 24, 25] by taking the $L_1$-normalised output of the next-to-last layer (name of the layer detailed below). No re-training or fine tuning was applied. The input images were preliminary resized via bicubic interpolation to fit the field of view of each net.

• **DenseNet121**: Size = 33MB, layer = ‘avg_pool’, field-of-view = 224px $\times$ 224px (1024 features [26])

• **MobileNet**: Size = 16MB, layer = ‘dropout’, field-of-view = 224px $\times$ 224px (1024 features [27])

• **ResNet50**: Size = 98MB, layer = ‘avg_pool’, field-of-view = 224px $\times$ 224px (2048 features [28])

• **VGG16**: Size = 528MB, layer = ‘fc2’, field-of-view = 224px $\times$ 224px (4096 features [29])

• **Xception**: Size = 88MB, layer = ‘avg_pool’, field-of-view = 299px $\times$ 299px (2048 features [30])

4. Experiments

Each combination colour constancy/image descriptor and colour augmentation/image descriptor was tested for automatic genre classification via nearest-neighbour classification with $L_1$ distance. The accuracy was estimated using the folds that come in bundle with the Pandora dataset. For each experiment four splits were obtained by picking one of the folds for training and the remaining three for testing (train ratio $\approx 25\%$). Preprocessing via either colour constancy and colour augmentation was applied both to the train and test images, this way guaranteeing that the ratio train/test images remained the same as in the original folds. Note that colour augmentation increases the overall number of images by a factor which depends on the number of perturbation used (see Tab. 1).
5. Results and discussion
Table 2 reports the accuracy by image descriptors and colour constancy method. As can be seen, no method clearly emerged as the ‘best option’: chroma outperformed the other approaches in five descriptors out of 15, followed by no colour-preprocessing (four descriptors), gw, mw and stretch (two descriptors each). Notably, the gain that could be achieved via colour constancy over no pre-processing was, in most cases by the order of 2pp or less, therefore rather slim. A significantly higher gain was obtained by coupling chroma with some texture descriptors – specifically Gabor, Laws and Zernike. We conjecture that the benefit, in these cases, comes from the local intensity normalisation that chromaticity representation introduces.

| Descriptor | Colour constancy method |
|------------|-------------------------|
|            | none | chroma | gw | mw | stretch |
| Mean       | 23.31| 18.69  | 14.60 | 24.34 | 24.59 |
| MeanStd    | 25.57| 24.71  | 19.02 | 25.76 | 25.79 |
| Percentiles| 26.28| 23.35  | 21.52 | 26.47 | 26.16 |
| MarginalHists| 28.16| 28.97  | 23.27 | 28.32 | 27.98 |
| FullHist   | 34.24| 26.06  | 33.39 | 32.63 | 32.23 |
| DCF        | 22.51| 22.81  | 22.27 | 22.78 | 22.70 |
| Gabor      | 19.69| 26.65  | 19.10 | 19.91 | 19.68 |
| Laws       | 20.82| 28.49  | 21.02 | 21.38 | 21.26 |
| LBP        | 33.22| 28.06  | 33.49 | 33.38 | 33.33 |
| Zernike    | 20.27| 28.45  | 19.92 | 20.29 | 20.24 |
| DenseNet121| 38.57| 27.73  | 38.98 | 34.93 | 34.96 |
| MobileNet  | 37.80| 28.68  | 37.94 | 34.73 | 34.74 |
| ResNet50   | 54.75| 37.35  | 53.92 | 53.39 | 53.35 |
| VGG16      | 46.94| 32.30  | 46.88 | 46.12 | 46.24 |
| Xception   | 28.15| 30.25  | 26.73 | 27.63 | 27.73 |

As for colour augmentation, the results (Tab. 3) indicate that doing nothing was the best option in eight of the 15 images descriptors considered, followed by configuration E (best option for two descriptors), and configurations A and C (best option for one descriptor each). In those cases when there was a gain, this was very limited, never surpassing 1pp.

On the whole pre-trained networks performed better than the other methods (best accuracy 54.75%, ResNet50), followed by colour descriptors (34.24%, FullHist) and texture descriptors (33.49%, LBP). The overall trend is therefore in agreement with the results reported in [2]. Arguably, higher classification rates could be obtained via more sophisticated classifiers than the simple 1-NN used here, for instance Support Vector Machines and Random Forest. This is an interesting direction for future studies.

6. Conclusions
In this work we have investigated colour constancy and colour augmentation as potential means to improve accuracy in automatic painting classification by artistic genre. On the whole both approaches failed to show a clear advantage compared with no pre-processing at all. Colour constancy brought some improvement in certain cases, whereas colour augmentation virtually provided no benefit at all despite its adding a significant computational overload to the procedure.
Table 3. Average accuracy by image descriptor and colour augmentation configuration. Boldface figures indicate the best results for each colour augmentation configuration; ‘none’ that no colour pre-processing was applied. See Tab. 1 for the meaning of letters A to E.

| Descriptor         | Colour augmentation configuration |
|--------------------|-----------------------------------|
|                    | none    | A       | B       | C       | D       | E       |
| Mean               | 23.31   | 22.22   | 21.61   | 21.21   | 20.75   | 20.33   |
| MeanStd            | 25.57   | 24.72   | 24.27   | 23.86   | 23.58   | 23.32   |
| Percentiles        | 26.28   | 25.68   | 25.35   | 25.05   | 24.80   | 24.61   |
| MarginalHists      | 28.16   | 27.79   | 27.37   | 27.36   | 27.08   | 26.93   |
| FullHist           | 34.24   | 34.21   | 33.85   | 33.78   | 33.33   | 33.20   |
| DCF                | 22.51   | 22.74   | 22.85   | 22.90   | 23.09   | 23.20   |
| Gabor              | 19.69   | 19.60   | 19.65   | 19.64   | 19.67   | 19.79   |
| Laws               | 20.82   | 21.20   | 21.19   | 21.34   | 21.49   | 21.65   |
| LBP                | 33.22   | 32.39   | 32.38   | 32.32   | 32.22   | 32.15   |
| Zernike            | 20.27   | 20.20   | 20.24   | 20.20   | 20.13   | 20.03   |
| DenseNet121        | 38.57   | 38.52   | 38.46   | 38.54   | 38.18   | 38.08   |
| MobileNet          | 37.80   | 38.15   | 38.24   | 38.15   | 38.02   | 37.99   |
| ResNet50           | 54.75   | 54.81   | 54.65   | 54.67   | 54.38   | 54.35   |
| VGG16              | 46.94   | 46.95   | 46.99   | 46.91   | 46.79   | 46.80   |
| Xception           | 28.15   | 28.22   | 28.13   | 28.24   | 28.13   | 28.16   |

This study is not exempt from limitations: among them the limited number of datasets used (one) and that of the methods included in the investigation. The results found here are therefore preliminary and should be validated in future studies.

Acknowledgements

This work was partially supported by the Department of Engineering at the Università degli Studi di Perugia, Italy, under the project ‘Shape, colour and texture features for the analysis of two- and three-dimensional images: methods and applications’ (Fundamental Research Grants Scheme 2019).

References

[1] Agarwal S, Karnick H, Pant N and Patel U 2015 Genre and style based painting classification Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (Waikoloa Beach, United States) pp 588–594 art. no. 7045938
[2] Bianconi F, Bello-Cerezo R and Napoletano P 2018 Journal of Electronic Imaging 27 art. No. 011002
[3] Bianco S, Mazzini D, Napoletano P and Schettini R 2019 Expert Systems with Applications 135 90–101
[4] Ciocca G, Napoletano P and Schettini R 2019 Evaluation of automatic image color theme extraction methods Proceedings of the International Workshop on Computational Color Imaging CCIW 2019 (Lecture Notes in Computer Science no 11418) (Chiba, Japan: Springer) pp 165–179
[5] Florea C, Condorovici R, Vertan C, Butnaru R, Florea L and Vranceanu R 2016 Pandora: Description of a painting database for art movement recognition with baselines and perspectives Proceedings of the 24th European Signal Processing Conference (EUSIPCO) (Budapest, Hungary) pp 918–922 art. no. 7760382
[6] Bianconi F and Bello-Cerezo R 2018 Evaluation of visual descriptors for painting categorisation Proceedings of Florence Heri-Tech – The Future of Heritage Science and Technologies (IOP Conference Series: Materials Science and Engineering vol 364) (IOP Publishing Ltd) article no. 012037
[7] Image Processing and Analysis Laboratory 2016 Perceptual ANalysis and DescriptiOn of Romanian visual Art (PANDORA) available online at http://imag.pub.ro/pandora/pandora_download.html. Last accessed on Nov. 6, 2017
[8] van de Weijer J Color in computer vision available online at: http://lear.inrialpes.fr/people/vandeweijer/research.html. Last accessed on Oct. 1, 2019
[9] Nikitenko D, Wirth M and Trudel K 2018 Journal of Multimedia 3 9–18
[10] Cernadas E, Fernández-Delgado M, E G R and Carrión P 2017 Pattern Recognition 61 120–138
[11] Krizhevsky A, Sutskever I and Hinton G 2012 ImageNet classification with deep convolutional neural networks Proceedings of the 26th Annual Conference on Neural Information Processing Systems (Lake Tahoe, United States)
[12] Kukkonen S, Kälviäinen H and Parkkinen J 2001 Optical Engineering 40 170–177
[13] López F, Miguel Valiente J, Manuel Prats J and Ferrer A 2008 Pattern Recognition 41 1761–1772
[14] Niskanen M, Silvén O and Kauppinen H 2001 Color and texture based wood inspection with non-supervised clustering Proceedings of the 12th Scandinavian Conference on Image Analysis (SCIA 2001) (Bergen, Norway) pp 336–342
[15] Swain M and Ballard D 1991 International Journal of Computer Vision 7 11–32
[16] Pietikäinen M, Nieminen S, Marszalec E and Ojala T 1996 Accurate color discrimination with classification based on feature distributions Proceedings of the International Conference on Pattern Recognition (ICPR) vol 3 (Vienna, Austria) pp 833–838 art. no. 547285
[17] Ahmed N, Natarajan T and Rao K 1974 IEEE Transactions on Computers C-23
[18] Laws K 1980 Rapid texture identification Image Processing for Missile Guidance (SPIE Proceedings vol 0238) ed Wiener T
[19] Ojala T, Pietikäinen M and Mäenpää T 2002 IEEE Transactions on Pattern Analysis and Machine Intelligence 24 971–987
[20] Pardo-Balado J, Fernández A and Bianconi F 2015 Texture classification using rotation invariant LBP based on digital polygons New Trends in Image Analysis and Processing –ICIAP 2015 Workshops (Lecture Notes in Computer Science vol 9281) ed Murino V and Puppo E (Springer) pp 87–94
[21] Lakshminarayanan V and Fleck A 2011 Journal of Modern Optics 58 545–561
[22] Kanan C and Cottrell G 2012 PLoS ONE 7 art. no. e29740
[23] Napoletano P 2017 Hand-crafted vs learned descriptors for color texture classification Proceedings of the 6th Computational Color Imaging Workshop (CCIW’17) (Lecture Notes in Computer Science vol 10213) ed Bianco S, Schettini R, Tominaga S and Tremeau A (Milan, Italy: Springer) pp 259–271
[24] Bello-Cerezo R, Bianconi F, Di Maria F, Napoletano P and Smeraldi F 2019 Applied Sciences 9 article number: 738
[25] Scabini L F S, Condori R H M, Ribas L C and Bruno O M 2019 Evaluating deep convolutional neural networks as texture feature extractors Image Analysis and Processing –ICIAP 2019 (Lecture Notes in Computer Science vol 11752) ed Ricci E, Rota Belo S, Snoek C, Lanz O, Messelodi S and Sebe N (Trento, Italy: Springer) pp 192–202
[26] Huang G, Liu Z, van der Maaten L and Weinberger K Q 2017 Densely connected convolutional networks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Hawaii, USA) pp 2261–2269
[27] Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M and Adam H MobileNets: Efficient convolutional neural networks for mobile vision applications arxiv:1704.04861
[28] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Las Vegas, United States) pp 770–778
[29] Simonyan K and Zisserman A 2015 Very deep convolutional networks for large-scale image recognition Proceedings of the 3rd International Conference on Learning Representations (ICLR) (San Diego, USA)
[30] Chollet F 2017 Xception: Deep learning with depthwise separable convolutions Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (Honolulu; United States) pp 1800–1807