

# Supervised Object Class Colour Normalisation

Ekaterina Riabchenko<sup>1</sup>, Jukka Lankinen<sup>1</sup>, Anders Glent Buch<sup>2</sup>,  
Joni-Kristian Kämäräinen<sup>3</sup>, and Norbert Krüger<sup>2</sup>

<sup>1</sup> Lappeenranta University of Technology (Kouvola Unit),

<sup>2</sup> Mærsk McKinney Møller Institute, University of Southern Denmark

<sup>3</sup> Department of Signal Processing, Tampere University of Technology

**Abstract.** Colour is an important cue in many applications of computer vision and image processing, but robust usage often requires estimation of the unknown illuminant colour. Usually, to obtain images invariant to the illumination conditions under which they were taken, color normalisation is used. In this work, we develop a such colour normalisation technique, where true colours are not important per se but where examples of same classes have photometrically consistent appearance. This is achieved by supervised estimation of a class specific canonical colour space where the examples have minimal variation in their colours. We demonstrate the effectiveness of our method with qualitative and quantitative examples from the Caltech-101 data set and a real application of 3D pose estimation for robot grasping.

## 1 Introduction

Colour is an important cue in many applications of computer vision and image processing. The colours are determined by the intrinsic properties of objects and surfaces as well as the colour of light source. For robustness, the effect of the light source must be filtered out and that is referred to as *colour normalisation* or *colour constancy*. The previously proposed methods for colour constancy aim to estimate the colour of the light source, and then transform image colours to the “world colour space” (white illuminant). In the world colour space, the photometric variations are only due to the natural variation of the scene objects. A survey of the existing methods can be found from Gijsenij et al. [7].

The light source estimation methods in [7] are based on physical formulations and heuristics related to the problem, but for more cognitive vision tasks, such as object recognition and classification, it is unclear how colour normalisation or constancy should be performed. For example, the human visual system (HVS) is remarkably robust to change of colours and even to completely abnormal colours [9], but the mechanism is not fully understood [6]. There is evidence that the recognition and colour normalisation are coupled [8], which supports the finding that supervised learning methods perform best in computational colour constancy [7].

In this paper, we propose a novel computational approach to colour normalisation. We replace the requirement of physically plausible world colours with

estimated transforms making examples of visual object classes look photometrically (chroma and brightness) equivalent. Our approach is supervised utilising common object landmarks. Using the landmarks an optimal “canonical object colour space” is estimated and the all examples transformed to the space (Fig. 1). New examples can be transformed to the space by detecting the same landmarks.

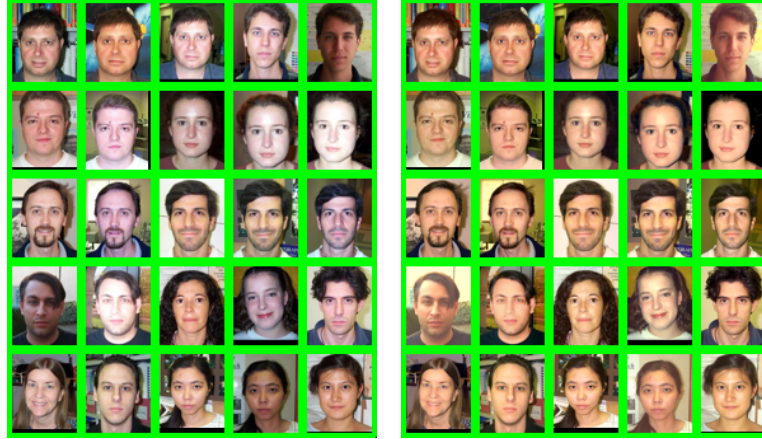


Fig. 1: Original Caltech-101 faces (left) and after transformation to the estimated canonical object colour space.

### 1.1 Related work

The general idea of our colour normalisation is very similar to the traditional colour constancy techniques. The difference is that normalised colors are not the true colours, like in constancy, but the ones from a class specific colour space.

For colour constancy, the learning-based methods perform the best [7], which supports the coupling of the recognition and colour constancy. This can be achieved, by detecting known objects (grass, water, etc.) [18] or on lower levels by computing image statistics which reflect, e.g., common textures [7]. Another similar problem in image processing is colour transfer [16], which does not aim for constancy, but how to utilise another span of colours for an image. The span can also reflect personal preferences [11]. Another related problem is the photometric camera calibration [13]. Object landmarks (local parts) have been used in colour constancy estimation in [15, 14], but these provide constancy for only a single object, not for an object class. Our method utilises linear geometric transformations (non-diagonal similarity transform), which is supported by the recent work of Chong et al. [3]. To the authors’ best knowledge, we are the first

to formulate object class specific colour normalisation which instead of physically correct colours, produces colours which minimise the difference within a provided set of images.

## 2 Estimation of Canonical Object Colour Space

The goal of our method is to estimate a class specific “canonical object colour space”, where colour variance of transferred objects is minimised. We utilise tristimulus colour values and the transformation is rigid 3D geometric transformation, similarity, allowing 3D rotation, translation, and scaling. For 3D similarity, efficient point correspondence based estimation methods using singular value decomposition exist. The colour transfer operator is a linear  $4 \times 4$  transformation  $T_i$  (homogeneous coordinates) which retains photometric information and is invertible. The point correspondence are formed from a set of common landmarks (minimum of 3 landmarks for 3D isometry/similarity) (Fig. 2). Joint colour variation of the landmark regions is minimised during the estimation procedure.

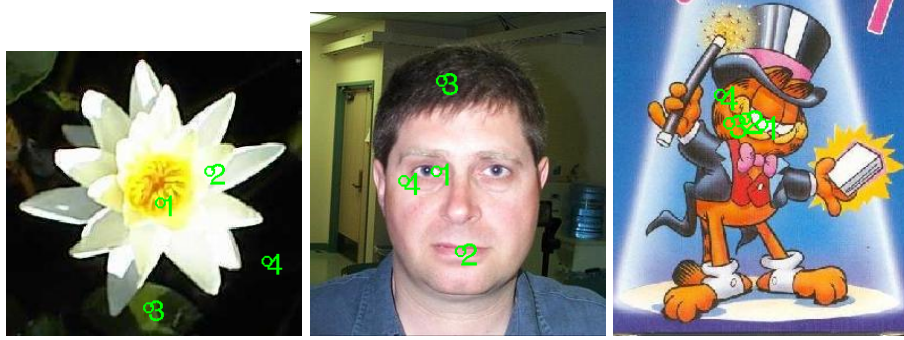


Fig. 2: Caltech-101 examples with annotated landmarks (denoted by the green circles and numbers).

The selection of the tristimulus colour space (RGB, HSV, Lab, etc.) should be made prior the estimation step. Typically, the selection depends on an application, but on the other hand, it has been often pointed out that in computational methods the standard RGB seems to perform the best or there are no differences between the colour spaces [1]. This hold in our preliminary experiments and therefore we report results only for the RGB colour space.

Our canonical colour space estimation method is motivated by the mean shape model by Cootes et al. [4] and the approach was also used by the authors [10] for spatial landmark normalisation. The estimation algorithm is sketched in Alg. 1. Geometric transformation is estimated using the Umeyama method [17]. The estimation procedure is further illustrated in Fig. 3.

**Algorithm 1** Canonical colour space.

- 
- 1: Select a random seed image  $r$  and use  $N$  landmark colours  $\{c_{r,1}, \dots, c_{r,N}\}$  as the initial colour space  $S$ .
  - 2: **for all** images  $i$  **do**
  - 3:   Estimate the geometric transformation  ${}^S T_i$  (3D similarity) from the colours of the  $i$ th image landmarks to the current colour space  $S$ .
  - 4:   Transform the  $i$ th landmarks  $\{c_{i,1}, \dots, c_{i,N}\}$  to the current colour space using  ${}^S T_i$ .
  - 5:   Refine the current canonical colour space by taking the average of the all transformed landmarks  $S \leftarrow \text{avg}(\{c_n\}_i)$ .
  - 6: **end for**
  - 7: Return  $S$  as the canonical colour space.
- 

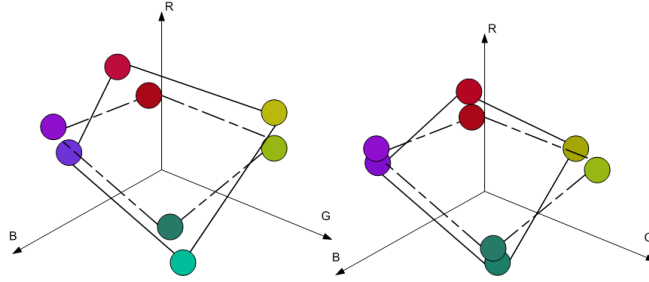


Fig. 3: Example of the Umeyama estimated similarity; original (left) and the transformed (right). Note that the transform is not exact for four points.

### 2.1 Seed selection

One concern that affects the final result of Alg. 1 is the seed selection. It is noteworthy that the seed does not particularly affect colour variance but mean, i.e. average colours of each landmark (Fig. 4). Therefore, for computational methods the result is seed-independent, but for a human viewer it can be undesired that colours change after each run of the algorithm due to a random seed. However, there is a simple procedure to fix seed selection: compute the mean colours of each landmark and then select the image whose landmark colours are closest to the mean values. We adopted this simple procedure in our experiments.

## 3 Experiments

In the following, we report qualitative and quantitative results for classes from the popular Caltech-101 data set for visual object classification [5] and for a real application where our method is used to photometrically normalise images prior to object pose estimation for robot grasping (see [2] for more details).



Fig. 4: Original (left) and colour normalised images using three different random seeds. Note that in all images the skin colours look natural, only biased toward the skin colour of a person in each seed image.

### 3.1 Caltech-101

For testing, we selected the following Caltech-101 categories: *Garfield*, *Water lily*, *Strawberry*, *Sunflower*, *Panda* and *Faces* containing 11-28 images. 3-4 landmarks were manually annotated and the landmarks were intuitively selected with no special attention. The examples of the original and processed images are shown in Fig. 1 and Figs. 5-6.

For quantitative performance, we computed the proportional changes in landmark colour variances by the following formula:

$$\frac{var(\mathbf{c}_{orig}) - var(\mathbf{c}_{canonical})}{var(\mathbf{c}_{orig})} . \quad (1)$$

The computed performance values are given in Table 1. The reduction for all classes was between 0.32-0.72 indicating significant improvement in colour similarity (see also Fig. 7 for illustration).

Table 1: Relative variances of the landmark colours after our colour normalisation.

Cat.	Relative variance				
	LM-1	LM-2	LM-3	LM-4	Avg.
Faces	0.61	0.62	0.15	0.56	0.55
Water lily	0.63	0.57	0.60	0.14	0.55
Garfield	1.08	0.45	0.40	0.50	0.32
Sunflower	0.34	0.47	0.59	0.13	0.44
Panda	0.78	0.42	0.54		0.61
Strawberry	0.72	0.65	0.76		0.72



Fig. 5: *Sunflowers, Panda and Garfield* (originals top).

### 3.2 Colour Feature Based Pose Estimation for Robot Grasping

We demonstrate the use of our approach for a practical vision application in a robotic grasping work cell, which requires accurate pose estimation of objects. We use the Kinect as visual scene input, and we apply the pose estimation system recently proposed in [2]. The task is to find pose of a real object in a captured scene using the KIT 3D model database [12]. The KIT database contains colourful richly textured objects for which colour is an important cue for relating model points to corresponding scene points. However, the very different illumination conditions between our setup and the setup used for capturing the model textures had significant negative impact on the calculation of colour correspondences between the model textures and the observed scene data, making the pose estimation fail. To overcome this, we used small set of landmarks between the textured models and frontal views of the objects in our setup for estimating the colour transformation. After processing the models, we were able to successfully carry out pose estimation with great degree of robustness and accuracy. For more details we refer to the paper where the pose estimation method and full results are published [2]. An example scenario for two objects is given in Fig. 8.





Fig. 6: *Strawberry* and *Water lily* (originals top).

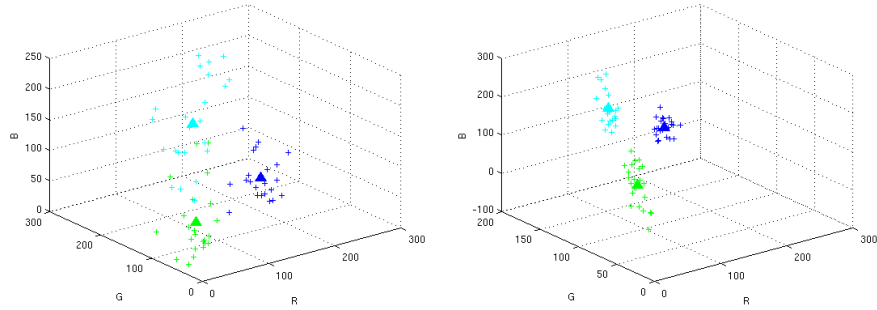


Fig. 7: All sunflower landmark colours as points in the RGB space: original (left) and processed.

## 4 Conclusions

In this work, we proposed a novel computational method for colour normalisation. Instead of the traditional approach of estimating the illuminant colour,



Fig.8: An input Kinect scene (top left), two textured KIT object models (top middle), colour normalised KIT models (top right) and pose estimation results (colour features projected to the scene) (bottom).

we adopted a “subjective” approach where the true physical colours are not of interest but photometric appearance of a same class objects is consistent. Our idea originated from the subjective nature of colour perception in the human visual system, which is evidently able to “normalise” colours if there are recognisable objects in the scene. We proposed a simple yet effective method to find optimal “canonical colour space” where the colour differences are minimised under invertible linear transformation. We validated our method with visual object classes from the Caltech-101 data set for which our method clearly improved the colour similarity visually and quantitatively and our method solved the colour normalisation problem in a real application of pose estimation for robot grasping.



## References

1. Alata, O., Quintard, L.: Is there a best color space for color image characterization or representation based on multivariate gaussian mixture model? *CVIU* 113, 867–877 (2009)
2. Buch, A.G., Kraft, D., Kämäräinen, J.K., Petersen, H.G., Krüger, N.: Pose estimation using local structure-specific shape and appearance context. In: *ICRA* (2013), accepted
3. Chong, H., Gortler, S., Zickler, T.: The von Kries hypothesis and a basis for color constancy. In: *ICCV* (2007)
4. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models – their training and application. *Computer Vision and Image Understanding* 61(1) (1995)
5. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In: *CVPR Workshop on Generative-Model Based Vision* (2004)
6. Foster, D.: Color constancy. *Vision Research* 51, 341–349 (2010)
7. Gijsenij, A., Gevers, T., van de Weijer, J.: Computational color constancy: Survey and experiments. *IEEE Trans. on Image Processing* 20(9) (2011)
8. Hansen, T., Olkkonen, M., Walter, S., Gegenfurtner, K.R.: Memory modulates color appearance. *Nature Neuroscience* 9(11), 1367–1368 (Oct 2006)
9. Ho-Phuoc, T., Guyader, N., Landragin, F., Guerin-Dugue, A.: When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision* 12((2):4) (2012)
10. Kamarainen, J.K., Ilonen, J.: Learning and detection of object landmarks in canonical object space. In: *20th Int. Conf. on Pattern Recognition (ICPR2010)* (2010)
11. Kang, S., Kapoor, A., Lischinski, D.: Personalization of image enhancement. In: *CVPR* (2010)
12. Kasper, A., Xue, Z., Dillmann, R.: The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics. In: *The International Journal of Robotics Research*. vol. 31, pp. 927–934 (2012)
13. Kim, S., Pollefeys, M.: Robust radiometric calibration and vignetting correction. *PAMI* 30(4) (2008)
14. Moerland, T., Jurie, F.: Learned color constancy from local correspondences. In: *IEEE Int. Conf. on Multimedia & EXPO (ICME)* (2005)
15. Obdrzalek, S., Matas, J., Chum, O.: On the interaction between object recognition and colour constancy. In: *ICCV* (2003)
16. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graph. Appl.* 21 (2001)
17. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. *IEEE PAMI* 13(4), 376–380 (1991)
18. van de Weijer, J., Schmid, C., Verbeek, J.: Using high-level visual information for color constancy. In: *ICCV* (2007)