Towards a Psychophysical Evaluation of Colour Constancy Algorithms

Javier Vazquez, Maria Vanrell, Ramon Baldrich and Alejandro Párraga

Centre de Visió per Computador
Computer Science Department
Universitat Autònoma de Barcelona
Edifício O, Campus UAB (Bellaterra), C.P.08193, Barcelona, Spain.

Keywords: Computational colour constancy, psychophysical evaluation

Towards a Psychophysical Evaluation of Colour Constancy Algorithms

1. Introduction

Computational Colour Constancy tries to solve the problem of recovering the illuminant of a scene from its acquired image. It is easy to show that this colour constancy problem is an ill-posed problem, that is, it might not have a unique solution. Several algorithms have been proposed to solve this problem, such as gamut mapping, colour by correlation, grey-world, white-patch, etc. We can classify algorithms into two families, those that consider the information of the acquisition process (calibrated algorithms) and those which do not (un-calibrated algorithms). In all of them, they always select a unique solution.

The performance of the colour constancy algorithms have been shown as not good enough to solve computer vision tasks [Funt98]. Comparison studies ([Barnard02a], [Barnard02b]) have proposed different rankings between large sets of different algorithms, depending on the properties of the dataset of images used to evaluate them. Colour-by correlation [Finlayson01] and C-RULE (gamut mapping) [Forsyth90] usually present the best performance. Evaluation is done by computing the error between the canonical image (that is unknown in real applications of computer vision) and the image selected by the algorithms. Since these evaluations are usually based on the RMS performance of the algorithms, in [Hordley04] authors propose a new evaluation one-to-one of the algorithms on each image. In this way they get a comparison with more independency of the used image dataset and giving a more robust ranking of the performance of the algorithms. However, the problem of defining an algorithm that always selects a solution close to the optimal is not resoluble.

Considering these previous works and the ill-posedness of the colour constancy problem, Tous in [Tous06] proposes to provide multiple solutions having a semantic interpretation, instead of a unique solution based on a specific heuristic on the set of all the feasible solutions. Providing multiple solutions can allow introducing high level considerations in the selection of solutions. In this work, he proposed to weight the solutions of the feasible set by introducing the Nameability assumption; it allows selecting solutions representing images with a high probability of having a semantic interpretation based on the ability of assigning basic colour names. Working on multiple solutions is just a proposal that can overcome the problems of the usual approach on computational colour constancy.

In this paper we propose a new evaluation based on a psychophysical experiment that tries to compare colour constancy algorithms based on perceptual judgements on the adequacy of a solution. It is just a preliminary proposal where we evaluate three different algorithms, C-RULE, Grey-world and MaxName. By MaxName we refer the algorithm presented in [Tous06] taking as a unique solution the one with maximum nameability.

2. Method

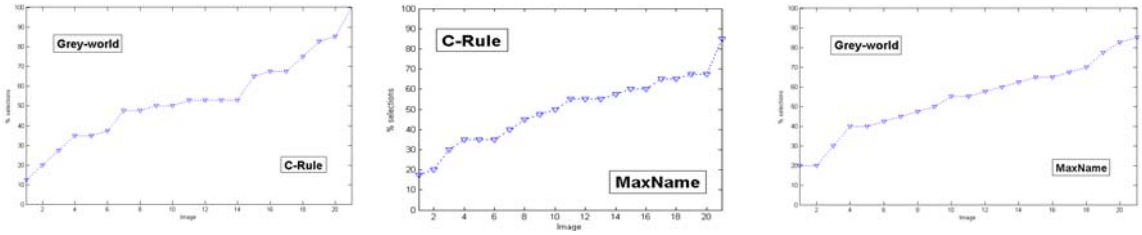
We have used a dataset of 21 scenes, each one acquired under four different illuminants, taken from the Simon Fraser database [Barnard02c]. We have applied the three selected algorithms on them getting one solution in each case.

These solution images were converted from CIE RGB space to CIE XYZ using standard transformations under CIE illuminant E and were presented on a calibrated CRT monitor (Viewsonic P227f) using a digital video processor (Cambridge Research Systems Bits++). Experiments were conducted on a dark room. There were 10 naïve observers recruited among University students and staff (none of the observers had previously seen the picture database). Pairs of pictures (each obtained using one of the two colour constancy algorithms) were presented side by side on a grey background (31 Cd/m^2). Each picture subtended 7.5 x 5.1 deg to the observers. After each presentation, observers were asked to select the picture that seemed most "natural". There was no time limit, although observers were encouraged to spend about 20 second per picture. The presentation order was randomised.

3. Results and Discussion

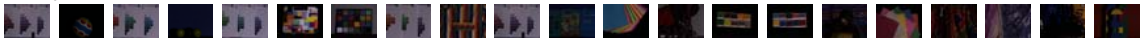
This make us to conclude that none of these algorithms present a better performance over the others, all of them present a similar ratio of selected solutions in average. C-Rule solutions are selected in 53,1% of choices in front of Grey-word solutions, C-Rule solutions are selected in 50,2% of choices in front of Maxname solutions and MaxName solutions present a 54,2%of choice in front of Grey-world solutions.

This first result could be interpreted as a consequence of a bad adequacy on the answered question, which could provide a random response. However, if we order the images according to its rate of selection for each pair of algorithms then we can state that the ability of selecting the illuminant is not random within a given image. In the figures below, we show images ordered by the number of times they are selected under any given illuminant by each pair of algorithm from the 40 selections made by the 10 subjects. In the x-axis we present a rank of the 21 scenes and on the y-axis we give the selection ratio on the solutions provided by each pair of algorithms.

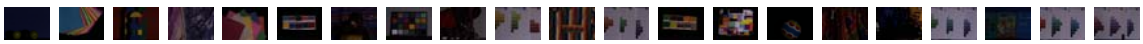


The images ordered on the x-axis are the following for each case:

Grey-world vs. C-Rule:



C-Rule vs. MaxName:



Grey-world vs. MaxName

From this we can see that on each comparison some algorithms are preferred for a subset of images and others subsets are equally selected. It makes us to conclude that different selection heuristics are preferred in front of others depending on the image

content. These results point out the limits of computational colour constancy based on a unique solution. Small improvements on differences in angular error do not give evidence of better performance algorithms in terms of human colour constancy.

The same analysis has been done considering the colour of the illuminant and no correlation can be extrapolated. The image content seems to be related to the preferred heuristic. This open a new way of looking at the evaluation and to the heuristic selection. A deep study on the statistics of the preferred images is needed to correlate image content and best heuristics.

5. Bibliography.

[Tous06] F. Tous. Computacional framework for the white point interpretation base don colour matching. *Phd. Thesis, Computer Vision Center – Universitat Autònoma de Barcelona, 2006.*

[Hordley04] S.D. Hordley and G.D. Finlayson. Re-evaluating Colour Constancy Algorithms. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*

[Funt98] B. Funt, K. Barnard and L. Martin. Is colour constancy good enough?. *European Conference on Computer Vision, 1998, pages 445-459.*

[Forsyth90] D. Forsyth. A Novel Algorithm for Colour Constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

[Finlayson01] G. Finlayson, S. Hordley, and P. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001.

[Barnard02a] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms; part one: Methodology and experiments with synthetic images. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002.

[Barnard02b] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms; part two: Experiments with image data. *IEEE Transactions on Image Processing*, 11(9):985–996, 2002.

[Barnard02c] K. Barnard, L. Martin, B. Funt, and A. Coath, A Data Set for Colour Research. *Color Research and Application*, 27 (3) , pp. 147-151, 2002.