Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

Attention-Based Colour Correction

F. W. M. Stentiford & A. Bamidele

University College London, Adastral Park Campus
Martlesham Heath Ipswich, UK
<f.stentiford@adastral.ucl.ac.uk>
<a.bamidele@adastral.ucl.ac.uk>

Abstract

This paper proposes a new algorithm that extracts colour correction parameters from pairs of images and enables the perceived illumination of one image to be imposed on the other. The algorithm does not rely upon prior assumptions regarding illumination constancy and operates between images that can be significantly different in content. The work derives from related research on visual attention and similarity in which the performance distributions of large numbers of randomly generated features reveal characteristics of images being analysed.

1 Introduction

The colour of illumination of a scene can have a dramatic effect on the performance of image retrieval systems. In addition different imaging devices will produce widely different responses. As there is normally no control over the camera characteristics, the image preprocessing, the brightness or the colour of the illumination, and the surface reflectances, this becomes a serious problem for object recognition and will lead to apparently identical images being assessed as different by the machine. It would not be acceptable for a photo taken in the late afternoon to reside in the same class as an identical photo taken in the morning.

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

Colour constancy seeks a relationship between colours and surface illumination so that the recorded colour can be mapped to the correct one [1]. In the absence of a precise definition of the 'correct' illumination in each case, most approaches to the colour correction problem have to make assumptions about the statistics of the reflectances and the illuminants that will be encountered. The basic Retinex computation achieves a good measure of colour constancy by mapping the maximum value of each channel to white, but sometimes forces scenes dominated by a single colour to become gray [2]. Finlayson et al. [3] used a diagonal model of illumination change and applied histogram equalization with some success but found that performance was lower on images with spatially varying illumination. The Colour by Correlation approach [4] requires a set of training illuminants that encompasses the images in question and furthermore it must also overcome sources of mismatch between the model and the real world. Jackowski [5] corrects colour distortions introduced by a camera by using known colour charts to calibrate transformation functions that relate colours on the chart to those in the image. Reinhard et al.[6] imposes one image's colour characteristics on another by matching the average colours in Lab colour space. The problem of comparing images possessing different compositions is addressed by comparing swatches taken from key parts of each image.

Human vision possesses an ability to interpret scenes without being troubled by the brightness or colour of the illumination and should be an inspiration for new ideas in this field. Visual attention plays a major function in human vision both for our survival and our ability to spot patterns in our environment. Work on modeling the functionality of visual attention is proving fruitful in the areas of compression [7,8], symmetry detection [9], perspective analysis [10], image focusing [11], and visual similarity [12].

Models of attention endeavour to predict which parts of an image or video will attract our attention and promise to find application in sectors where human vision is currently an essential component. Measures of attention within images are closely related to measures of similarity between images because attention normally is drawn towards differences between foreground and background and this is necessarily determined by the large scale similarity of the background with itself. The identification of visual anomalies is therefore in many ways equivalent to the problem of detecting similarities elsewhere in an image and this suggests that attention mechanisms may be used to estimate similarity between different images.

The work outlined in this paper is based on the notion that colour constancy arises in human vision as a result of experiencing visual similarity rather than of some absolute definition of colour correctness. This paper describes a new approach to illuminant

Annals of the BMVA Vol. 2007, No. 5, pp 1−11 (2007)

estimation and correction as part of the need to obtain similarity estimates that are independent of the illumination.

1.1 Similarity Measures

Similarity measures are central to most pattern recognition problems not least in computer vision and the need to access huge volumes of multimedia content now being broadcast and offered on the Internet. These problems have motivated considerable research into content based image retrieval e.g. [12] and many commercial and laboratory systems are described in the literature e.g. [13]. There are several approaches to similarity and pattern matching and much of this is covered in several survey papers e.g. [14]. Many approaches involve the use of pre-determined features such as edges, colour, location, texture and functions dependent on pixel values e.g. [15]. Mikolajczyk et al. [16] use edge models to obtain correspondences with similar objects. The advantages and disadvantages of using 3D histograms in which bins represent location are investigated by Ankerst et al. [17].

The selection of features dependent upon the spatial arrangement of sets of points sampled from shapes is a strategy used by several authors to obtain some of the best results to date e.g. [18]. However, all the approaches use pre-determined point selection rules and metrics that can limit performance on unseen data. Viola et al. [19] restrict themselves to a specific type of rectangle feature which works well in their face recognition application, but may not perform as well on data that is not suited to this feature.

Increasingly research is turning to models of perception in order to reflect the behaviour of the human visual system in measures of similarity. Mojsilovic et al. [20] use perceptually important colours to construct a feature vector for similarity measurement, and overcome the problem of close colours occupying different quantization bins. Law et al. [21] introduce a measure of saliency in their development of a feature selection and clustering algorithm. A feature is deemed irrelevant if its distribution is independent of class labels. Visual attention models by Itti [22] are used by Frintrop et al. [23] to focus computational resource and recognise 3D objects. Shape contours are detected by Grigorescu et al. [24] using a model of human visual surround suppression that identifies perceptually significant edges.

It is almost universal that established approaches to pattern analysis make use of a priori features to distinguish and recognise classes of data [25]. Unless the universe of data is

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

completely understood it will always be possible to produce unseen data on which these systems will fail. The nature of visual data certainly is such that no single fixed set of features appears capable of encompassing the all the complex relationships that exist between each item. In fact the strength of similarity between each pair of patterns in a class could be on a different basis in every case with no two patterns sharing the same features as any other pair. High dimensional feature spaces offer no solution to this problem.

1.2 Cognitive visual attention

Studies in neurobiology [26] are suggesting that human visual attention is enhanced through a process of competing interactions among neurons representing all of the stimuli present in the visual field. The competition results in the selection of a few points of attention and the suppression of irrelevant material. It means that people and animals are able to spot anomalies in a scene no part of which they have seen before and attention is drawn in general to the anomalous object in a scene.

Such a mechanism has been extended [12] to apply to the comparison of two images in which attention is drawn to those parts that are in common rather than their absence as in the case of saliency detection in a single image.

The model of Cognitive Visual Attention (CVA) used in this paper relies upon the matching of large numbers of pairs of pixel groups or n-tuples taken from patterns A and B under comparison.

Let a location x in a pattern correspond to a measurement a where

$$\boldsymbol{x} = (x_1, x_2) \text{ and } \boldsymbol{a} = (a_1, a_2, a_3) \quad (1)$$

Define a function $\boldsymbol{F}$ such that $\boldsymbol{a} = \boldsymbol{F}(\boldsymbol{x})$.

Select an n-tuple of m random pixels S_A in pattern A where

$$S_A = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, ..., \boldsymbol{x}_m\}. \quad (2)$$

Similarly select an n-tuple of m pixels S_B in pattern B

$$S_B = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, ..., \boldsymbol{y}_m\} \quad (3)$$

where $x_i - y_i = \boldsymbol{\delta}_j$

The n-tuple S_A matches the n-tuple S_B if

$$\left|F_j(\boldsymbol{x}_i) - F_j(\boldsymbol{y}_i)\right| \le \varepsilon_j \quad \forall i \quad (4)$$

An example of an m = 4 n-tuple fitting images A and B is shown in Figure 1. In general ε is not a constant and will be dependent upon the colour space of the measurements under comparison i.e.

$$\varepsilon_j = f_j\left(\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{y})\right) \qquad (5)$$

Up to N selections of the displacements $\boldsymbol{\delta}_j$ are used to apply translations to S_A to seek a matching n-tuple S_B. The CVA similarity score C_{AB} is produced after generating and applying T n-tuples S_A.

$$C_{AB} = \sum_{i=1}^{T} w_i \quad where\ w_i = \begin{cases} 1 & if\ S_A\ matches\ S_B\ within\ N\ attempts \\ 0 & otherwise \end{cases} \qquad (6)$$

C_{AB} is large when a high number of n-tuples are found to match both patterns A and B and represents features that both patterns share. It is important to note that if C_{AC} also has a high value it does not necessarily follow that C_{BC} is large because patterns B and C may still have no features in common. The measure is not constrained by the triangle inequality and is able to model a greater degree of complexity than a more restricted metric.

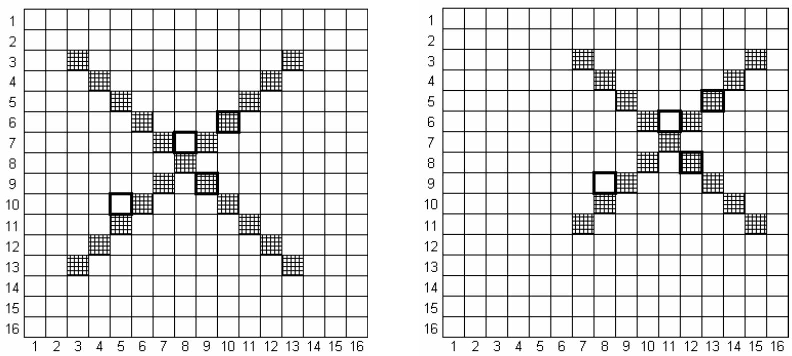

Figure 1: Four pixel n-tuple matching image A and image B with $\boldsymbol{\delta}_j = (3,-1)$.

2 Relative Illuminant Estimation

In this work we have applied random colour shifts during the n-tuple matching process as a means of obtaining illumination independence when calculating the CVA similarity measure. No restriction is placed on the choice of colour shift (which effectively varies brightness) and will change for each new n-tuple generated. In this case the n-tuple S_A matches the n-tuple S_B if

$$\left|F_j(\boldsymbol{x}_i) - F_j(\boldsymbol{y}_i) + \boldsymbol{\alpha}_j\right| \le \varepsilon_j \quad \forall i \quad for\ some\ \boldsymbol{\alpha}_j, \boldsymbol{\delta}_j \quad j = 1,2,\ldots,N \mid \qquad (7)$$

where $\boldsymbol{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \alpha_{j3})$ is a valid random displacement in the pixel colour components in image A. Peaks $\boldsymbol{\mu}_{AB}$ in the distribution of $\boldsymbol{\alpha}_j$ for j corresponding to matches will reflect the overall difference in illumination between images A and B providing sufficient common geometry and n-tuple matches are found. $\boldsymbol{\mu}_{AB}$ being an average does not reflect any variation in illumination across the image although by restricting diameter of n-tuples an estimate of $\boldsymbol{\mu}_{AB}(\boldsymbol{y}_i)$ as it varies over image B may be obtained. This approach has the advantage that no prior assumptions are made about the camera or the nature of the illumination that might impose anomalous results.

The colour shift $\boldsymbol{\mu}_{AB}$ that stimulates the highest frequency of matches represents a measure of the relative illumination of image B with respect to image A. We apply the reverse colour shift $-\boldsymbol{\mu}_{AB}$ to pixels in image B to obtain a transformation that approximates the illumination present in image A. In the results below $\boldsymbol{\mu}_{AB}$ was taken to be the average of the colour shifts $\boldsymbol{\alpha}_{AB}$ that were applied when a match resulted:

$$\boldsymbol{\mu}_{AB} = \sum_{S_A \; matches \; S_B} \boldsymbol{\alpha}_{AB} \Big/ C_{AB} \qquad (8)$$

If the original reference image A is geometrically identical to the colour corrected version B it becomes possible to measure the percentage colour deviation e_{AB} between the two images on a pixel by pixel basis:

$$e_{AB} = \frac{\sum_{ij} \sum_{k=r,g,b} \left| x_{ijk} - y_{ijk} \right|}{\sum_{ijk} 255} \qquad (9)$$

where x_{ijk} and y_{ijk} are corresponding pixel colour component values (0-255) in each image.

3 Results

The number of pixels (m) to be included in each n-tuple is a critical parameter. If m is too great the number of matches decreases leading to poor statistics and higher computation; if m is too small image structure may be overlooked leading to inaccurate colour correction. The image colour correction error e_{AB} was computed for images A and B in Figure 2 [27] for several values of m and averaged over 5 runs with T = N = 1000. The performances are shown in Table 1.

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

m	1	2	3	4	5	6	7	8
e_{AB}	7.86	6.00	4.28	3.99	4.02	4.00	4.00	4.01
Standard deviation	0.74	0.04	0.12	0.05	0.06	0.04	0.04	0.15

Table 1: Colour correction errors for images A and B in Figure 2.

The error rates in Table 1 indicate that little is to be gained by increasing m beyond 4 not least because computation is proportional to m. In the case of m = 1 the correction is heavily dependent upon the relative colour values of the largest areas of co-coloured pixels in each image and gives an approximate result only when the images are geometrically similar. Further work is necessary to confirm this result on other images.
Subsequent images are all processed with T = 500 randomly generated n-tuples each possessing m = 4 pixels with N = 500. The distribution of colour shifts for red green and blue for matching n-tuple s between images A and B in Figure 2 is shown in Figure 3. In this case $\boldsymbol{\mu}_{AB} = (-33,39,-4)$ indicating that a reduction in red values and an increase in green encouraged a greater likelihood of a match. By the same token this same shift was a measure of the change in illumination between image A and Image B and applying a colour shift of $-\boldsymbol{\mu}_{AB}$ produces the corrected version with an error value of 3.89%.

In Figure 4 image A is illuminated with yellow light and image B with blue light [28]. The difference in illumination is obtained with the same parameters as used in Figure 2 and gives $\boldsymbol{\mu}_{AB} = (-27,9,57)$. The reverse shift in pixel colours removes the blue tinge and restores the yellow illumination to image B with an error value of 4.05%.

Figure 5 shows a cropped version of image A in Figure 2 together with the same image B as Figure 2. The resulting colour shift $\boldsymbol{\mu}_{AB} = (-33,38,-9)$ yields an error of 4.16% only marginally greater than that obtained using geometrically identical images.

More significantly the images do not have to be of the same scene. Unlike Reinhard et al. [6] we do not need to select areas within each image for specific comparison because the statistics of matching pairs of n-tuples already take this into account. Matching n-tuples which span important colour adjacencies in different positions in each image contribute towards the estimate of the relative illumination regardless of the detailed composition of the images. Image A in Figure 6 is illuminated with sunlight [29]. Image B [30] has a colour cast which is common in digital photography. This is removed using a colour shift of $\boldsymbol{\mu}_{AB} = (-23,-5,-15)$ and an error of 3.92% and may be compared with the manually treated image on the right.

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

The fact that relative illumination may be calculated from a sample of images exhibiting a preferred set of illuminations means that working images may be processed to appear under a range of lighting conditions without it being necessary to make special arrangements. Furthermore images may be tailored automatically to individual preferences by simply supplying examples of acceptable images and allowing the system to adjust the colour balance accordingly.

Figure 2: Original image A, illuminated version B, and corrected version $\mu_{AB} = (-33,39,-4)$, error = 3.89%.

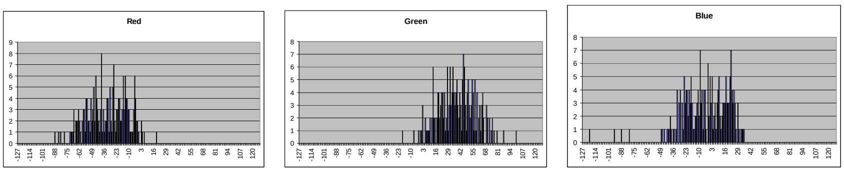

Figure 3: Distributions of R,G,B colour shifts for matching n-tuples for images A and B in Figure 2.

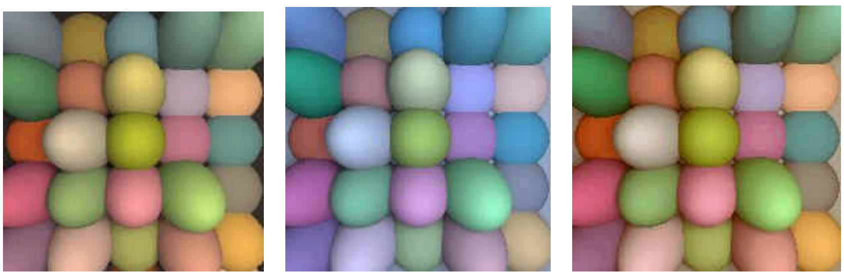

Figure 4: Image A in yellow light, image B in blue light and corrected $\mu_{AB} = (-27,9,57)$, error = 4.05%.

Figure 5: Cropped image A in Figure 2, Image B in Figure 2, and corrected version $\mu_{AB} = (-33,38,-9)$, error = 4.16%.

Figure 6: Image A, Image B with blue colour cast, corrected version $\boldsymbol{\mu}_{AB} = (-23,-5,-15)$, error = 3.92%, and manually corrected version.

4 Discussion

It is important to note that this approach identifies colour shifts that maximise the similarity of two images. In this way it gives a good indication of a possible change in illumination between the two images, but it does not provide an absolute estimate for the illumination in either image. It is possible that with more information gleaned from more images under different illuminations good estimates of illuminants can be obtained.

The colour shifts $\boldsymbol{\mu}_{AB}$ extracted in this work are based on averages obtained from n-tuple matches over the entire area of the image, and when used as a colour adjustment factor, are only strictly correct for those pixels in a matching n-tuple at that precise shift. A correction based on average shifts may be right for most of the image, but will not be appropriate, for example, for darker parts of the image where the eye is more sensitive to changes in brightness. A refined correction process that makes more use of the statistics that are extracted may enable the correction error to be reduced much further.

Shadows are sometimes a problem for object recognition and this approach offers some scope for minimizing their effects. The colour and brightness of objects immersed in shadow have a relationship with those from the same object that are not immersed in shadow. It may be the differences are simply due to brightness but commonly the colours in shadow are shifted by other sources of light. As above it may be reasonable to expect that a more refined colour correction process between similar images with and without shadows will reduce their prominence.

The computational requirements of this technique are independent of the size of the image, but increase as the pair of images under comparison become less similar. Dissimilar images will not yield many matching n-tuples and the colour correction

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

statistics will become less reliable as the number of matches falls. Processing of geometrically similar images takes 12 sec on a 1Ghz machine with T = N = 1000 using Visual Basic software. This rises to 25 sec for dissimilar images. Equivalent code in C++ takes under 100ms and related implementations using some parallelism on the Texas Instruments DM642 DSP platform indicate that processing can take place at video speeds.

Some initial work on establishing the effectiveness of the approach has been carried out and more subjective experiments are underway. 200 cameraphone images were taken under lighting conditions of morning, afternoon, late afternoon, tungsten, and fluorescent. In addition some underwater photographs were included. These images were adjusted using good quality reference images that corresponded to each illumination condition and were presented to subjects for their opinions. Originals and their corrected versions were shown to subjects in pairs but in random positions so that the subjects were unaware which was the original. Opinions were gathered on a 7 step scale ranging from "left much better" (value = 1) to "right much better" (value = 7). 30 subjects viewed the image pairs and the average overall opinion of the quality of the corrected version was 4.6 indicating a small but significant improvement in subjective appearance.

5 Conclusions

This paper has presented a new way of adjusting the perceived illumination between pairs of images. The pairs of images do not have to be geometrically identical, as the reference image can be a portion of the image being corrected or can contain completely different content. More investigations are necessary to explore the performance on a greater range of image content and the extent to which the images can become structurally dissimilar before the correction becomes unreliable. The results, however, are encouraging and should find application in the optimization of camera settings and a colour correction service offered over a mobile network.

This research has been conducted within the framework of the European Commission funded Network of Excellence "Multimedia Understanding through Semantics, Computation and Learning" (MUSCLE) [31].

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

References

[1] K.Barnard, V.Cardei, and B.Funt. A comparison of computational color constancy algorithms – Part 1: methodology and experiments with synthesised data. IEEE Trans. Image Processing, 11(9):972–983, September 2002.

[2] Z.Rahman, D.J.Jobson, and G.A.Woodell. Retinex processing for Automatic Image Enhancement. In Proc. Human Vision and Electronic Imaging VII, SPIE 4662:390–401, 2002.

[3] G.Finlayson, S.Hordley, G.Schaefer, and G.Y.Tian. Illuminant and device invariant colour using histogram equalization," Pattern Recognition, 38(2):179–190, 2005.

[4] G.D.Finlayson, S.D.Hordley, and P.M.Hubel. Color by correlation: a simple, unifying framework for color constancy. IEEE Trans. Pattern Analysis and Machine Intelligence, 23(11):1209–1221, November 2001.

[5] M.Jackowski, A.Goshtasby, S.Bines, D.Roseman and C.Yu. Correcting the geometry and color of digital images. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(10):1152 –1158, October 1997.

[6] E.Reinhard, M.Ashikhmin, B.Gooch, and P.Shirley. Color transfer between images. IEEE Computer Graphics and Applications. 21(5):34 –41, May 2001.

[7] F.W.M.Stentiford. A visual attention estimator applied to image subject enhancement and colour and grey level compression. In Proc. ICPR 2004, Cambridge, 23rd – 26th August, 2004.

[8] F.W.M.Stentiford. An estimator for visual attention through competitive novelty with application to image compression. Proc. Picture Coding Symposium 2001, pages 101–104, April 2001.

[9] F.W.M.Stentiford. Attention based facial symmetry detection. In Proc. ICAPR 2005, August 2005.

[10] F.W.M.Stentiford. Attention-Based Vanishing point detection. In Proc. ICIP 2006, October 2006.

[11] R.Shilston and F.W.M.Stentiford. An attention-based focus control system. In Proc. ICIP 2006, October, 2006.

[12] F.W.M.Stentiford. Attention based similarity. Pattern Recognition, 40(3):771–783, March 2007.

[13] C.Carson, S.Belongie, H.Greenspan, and J.Malik. Blobworld: image segmentation using expectation-maximisation and its application to image querying. IEEE Trans. Pattern Analysis and Machine Intelligence. 24(8):1026 –1038, August 2002.

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

[14] D.Zhang, and G.Lu. Review of shape representation and description techniques. Pattern Recognition, 37(1):1–19, January 2004.

[15] R.Manmatha, S.Ravela, and Y.Chitti. On computing local and global similarity in images. In Proc. of SPIE Human Vision and Electronic Imaging III, SPIE 3299:540–551, 1998.

[16] K.Mikolajczyk, A.Zisserman, and C.Schmid. Shape recognition with edge-based features. In Proc. British Machine Vision Assoc. Conf. Norwich, September, pages 779–788, 2003.

[17] M.Ankerst, G.Kastenmuller, H-P.Kriegel, and T.Seidl. 3D histograms for similarity search and classification in spatial databases. In Proc. Int. Symposium on Advances in Spatial Databases, pages 207–226, 1999.

[18] C.Grigorescu, and N.Petkov. Distance sets for shape filters and shape recognition. IEEE Trans. Image Processing, 12(10):1274 –1286, October 2003.

[19] P.Viola, and M.Jones. Rapid object detection using a boosted cascade of simple features. In Proc. Computer Vision and Pattern Recognition, Hawaii, pages 511–518, 2001.

[20] A.Mojsilovic, J.Hu, and E.Soljanin. Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis. IEEE Trans. Image Processing, 11(11):1238 –1248, November 2002.

[21] M.H.C.Law, M.A.T.Figueiredo, and A.K.Jain. Simultaneous feature selection and clustering using mixture models. IEEE Trans. Pattern Analysis and Machine Intelligence. 26(9):1154 –1166, September 2004.

[22] L.Itti. Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans. Image Processing, 13(10):1304 –1318, October 2004.

[23] S.Frintrop, A.Nuchter, and H.Surmann. Visual attention for object recognition in spatial 3D data. In Proc. IEEE International Conference on Intelligent Robots and Systems, September 2004.

[24] C.Grigorescu, N.Petkov, and M.A.Westenberg. Contour detection based on nonclassical receptive field inhibition. IEEE Trans. Image Processing, 12(7):729 –739, July 2003.

[25] R.O.Duda, P.E.Hart, and D.G.Stork. Pattern Classification. Wiley Interscience, 2000.

[26] R.Desimone. Visual attention mediated by biased competition in extrastriate visual cortex. Phil. Trans. R. Soc. Lond. B, 353(1373):1245 – 1255, 1998.

[27] http://www.cs.sfu.ca/~colour/research/colour-constancy.html

Annals of the BMVA Vol. 2007, No. 5, pp 1–11 (2007)

[28] http://www.staff.ncl.ac.uk/anya.hurlbert/

[29] http://cilab.knu.ac.kr/research/Color/constancy.htm

[30] http://www.panoguide.com/howto/panoramas/colour_correction.jsp

[31] Multimedia Understanding through Semantics, Computation and Learning, Network of Excellence. EC 6th Framework Programme. FP6-507752. http://www.muscle-noe.org/