Colour Segmentation and Recognition post Colour Constancy for Environmental Monitoring

Juliana F. Camapum[1], Mark H. Fisher[1] and Constantin F. Hanes[2]
School of Engineering - Coventry University[1]
Technical University of Cluj Napoca, Romania[2]
jcamapum@coventry.ac.uk

Abstract

This work is motivated by a need to monitor the natural environment by using a remote machine vision system to record animal populations. In this paper we consider the problem of using colour as a classification feature, and describe how Colour Texture angles have been used to recognise a particular species of monkey in outdoor Amazon forest scenes. The approach uses a Colour constancy algorithm to compensate for changes in ambient lighting. An unsupervised segmentation algorithm using graph theoretical clustering based on image chromaticity is applied post colour constancy. Subsequently 6 Colour-texture angle-indices are calculated and the Frobenius norm is applied to measure the distance between candidate and prototype colour images. Experimental results are provided to illustrate the performance of the segmentation algorithm and classification results using an image database comprising 77 Amazon forest scenes show that a "Red Face" Monkey can be detected with 74% success.

Key words. colour recognition, colour-texture angles, colour constancy, real images

I. Colour Constancy

One of the most useful pieces of information used by the human eye in order to recognise a known object or being is colour, however, the machine vision community has largely ignored colour. This is due partly to the increased complexity and cost of colour vision solutions but it is also due to the fundamental problem of colour constancy. Unlike the human visual system, whose perception of colour is largely unaffected by the colour of illuminating light, a machine vision system relies on data from CCD cameras, and these values are illumination dependent. The term colour constancy has been coined to describe the ability to make statements about surface colour properties which are invariant to surface illumination. For machine vision, there are at least two major approaches to this issue:

- the colour of a surface is regarded as its colour under a fixed canonical light.
- the colour of a surface is regarded as the result of a reflectance function.

Both ways of dealing with the problem of colour constancy rely on von Kries's 19th century theory of coefficient adaptation [1]. The first method involves estimating the canonical image illuminant by assuming that its value is constrained only by the colours observed in the image. The method is invariant to surface reflectance [2]. The second solution places only weak constraints on the illuminant but uses a set of illumination invariant vectors computed from a large database of images captured under different lighting conditions.

When working with natural outdoor scenes illuminated by sunlight, it is impractical to construct a comprehensive database of illumination invariant vectors from a relatively small image database. Therefore we have chosen to perform an estimation of the illuminant based on the constraint that surfaces can reflect no more light than is cast on them [2]. We also constrain the illumination parameters further by using our knowledge that the images were acquired from the Amazon region. The illumination characteristics are calculated using some features from the Coefficient Rule (CRULE) algorithm and encoded as a simple diagonal transform matrix, which is applied to each pixel in the image to yield its expected value under canonical light.

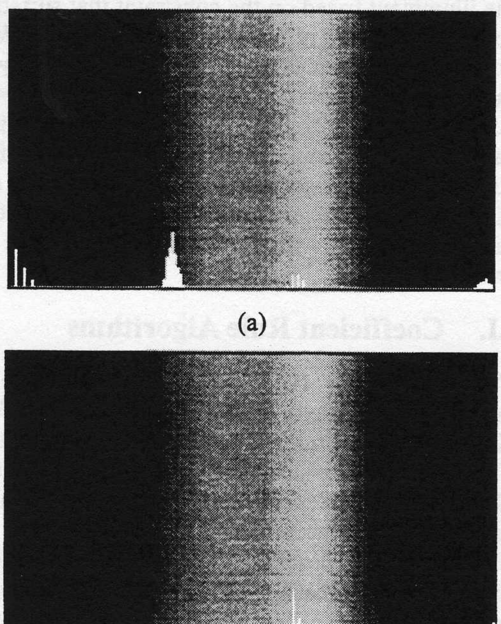
I.I. Coefficient Rule Algorithms

Coefficient Rule (CRULE) algorithms have been developed by Forsyth [2] in order to improve the von Kries coefficient rule algorithm for colour constancy. This algorithm adjusts the gain of each class of photoreceptor independently (for example the red, green and blue channels of a colour camera) to obtain surface colour descriptors which are invariant to changes of illuminant. The factors, by which the gains are adjusted, are called coefficients. Several ways of determining the factors are known; for example, Brill and West [3] divide the output of each photoreceptor class by its output response to a surface patch known as being white, and Land [4] chooses coefficients such that the geometric average of photoreceptor responses is constant for each class.

For CRULE, Forsyth [2] uses the appearance of a patch under a canonical illuminant as a descriptor. According to this approach, colour constancy involves predicting what an image would have looked like had it been captured under a canonical illuminant. This, in turn, requires knowledge of the illumination used for the original.

I.II. Illuminant Estimation

In this case, the illuminant is assumed to be equatorial sun light, as tabulated by D55, D65 or D75 [5], modulated by natural colour filters encountered in an Amazon forest environment, due to green leaves or tree branches, for example. As the images contain no known colour surfaces, we assume that the illuminant is mostly influenced by the dominant colour. In order to improve the speed of the computation only the strongest colours in the image are considered so, dark colours, which do not reflect much light, do not influence the lighting model. In this way, using only the most significant two bits of each pixel colour, valuable illumination information can be extracted. For example, the histogram of Figure 1a exhibits a significant amount of blue light (due to the sky backgrounding the image) whereas the histogram of Figure 1b contains a peak in the green band due to background foliage.



(a)

(b)

Figure 1 - (a) Colour histogram with a blue zone peak
(b) Colour histogram with a green zone peak

I.III. The Diagonal Transform

As a first step in our approach, we use a simple diagonal constancy transform of the form

$$d = Dp \quad (1)$$

where D is a diagonal transform matrix and p and d are input and output colour vectors [6]. Following Finlayson, we have found out that it is possible to neglect the sensor transformation T in Eq. 2 since the images were captured using a video camera with narrow band sensors.

$$Td = DTp \quad (2)$$

Thus, we calculate D by assuming the diagonal matrix performs a linear transform of the r, g and b values such as to reduce the effect of the illuminant on colours in the scene.

The r, g and b coefficients for matrix D are calculated using Eq. 3 and Eq. 4.

$$C_i = 1 - 0.15\frac{\sum i}{\sum (r+g+b)}, \text{ where } i \in \{r, g, b\} \quad (3)$$

In the equation above, each normalised band is weighted by 0.15. This value was chosen empirically to account for naturally occurring variations in the illuminant. The coefficients are weighted such that the geometrical mean of the colours in the image before and after the linear transformation remains the same.

$$C'_i = \frac{C_i}{\sqrt[3]{C_r C_g C_b}}, \text{ where } i \in \{r, g, b\} \quad (4)$$

II. The Segmentation Algorithm

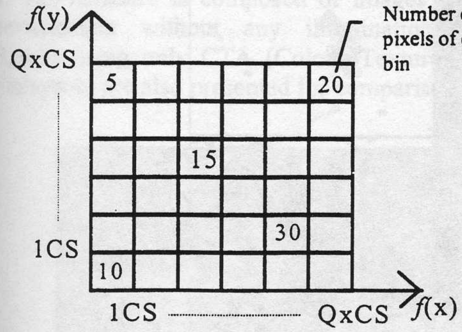
Segmentation based on image chromaticity is carried out by using a Graph Theoretical (GT) clustering algorithm [7] to group pixels together on the basis of their colour affinity. A threshold is used to ensure only strong chromaticity clusters of the image are passed to the recognition algorithm. The following steps describe the process:

- The chromaticity values for each pixel are calculated.
- Then they are divided into Q quantization levels each representing a chromaticity increment of

$$CS_i = \frac{chr_{i(max)} - chr_{i(min)}}{Q}, \text{ where } i \in \{r, g, b\} \quad (5)$$

Consequently, the histogram is divided into QxQ bins and each pixel belongs to one of the bins.

Histogram C[x][y]



- The N-neighbours of each bin are examined and the biggest one is chosen in order to establish a link. The neighbour needs to be bigger than the present bin as can be seen in Figure 2. If the current bin has the same value as the biggest neighbour, then one of them is chosen to be the father and a link is established.

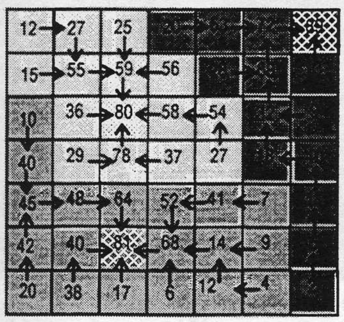


Figure 2 - 4-neighbourhood unimodal GT 2-D clustering

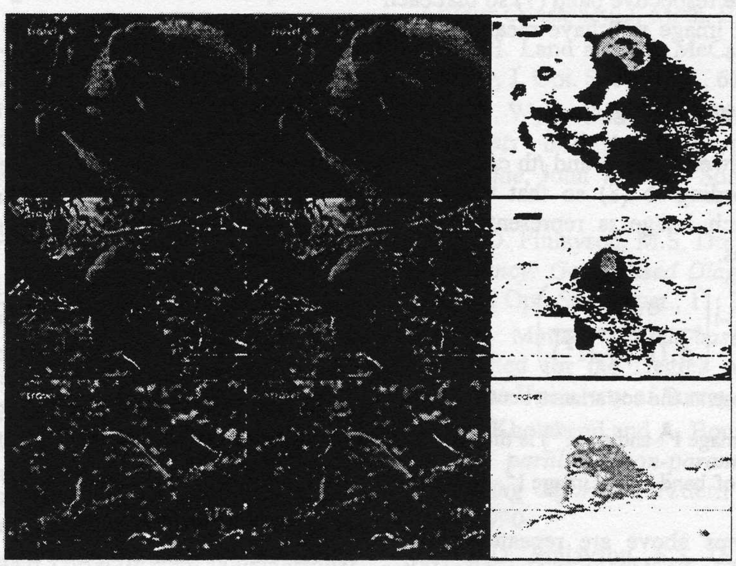
For each pixel the indices of its bin is:

$$x = \text{INT}\left\{\frac{f(x) - f_{min}}{\text{CS}} + 1\right\}$$
$$y = \text{INT}\left\{\frac{f(y) - f_{min}}{\text{CS}} + 1\right\} \quad (6)$$

Where $f(x) = chr_r$ and $f(y) = chr_g$ and C[x][y]=C[x][y]+1

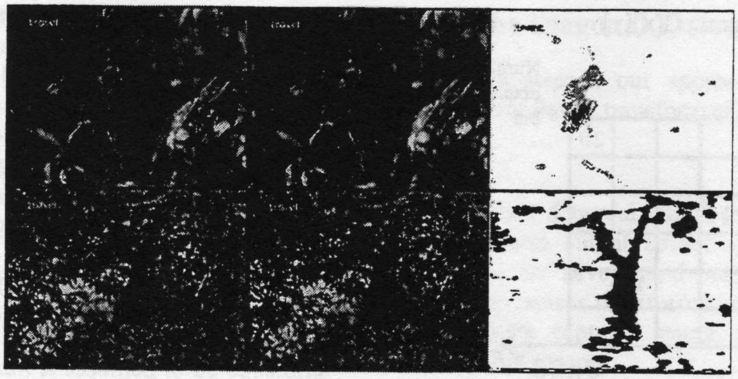
An optimisation method was used to determine the best chromaticity quantization. This involves repeating the clustering process for 8-NN and 4-NN and choosing the quantization according to a measure of clustering compactness [8]. Two methods of automatic removal of the background were used. First, a reliable chromaticity threshold, image independent, is applied before segmentation in order to eliminate the basic background structure, which is composed of leaves, branches and sky. After segmentation, small clusters, most of them belonging to the background, are also eliminated. The list of pixels belonging to a cluster needs to be larger than a threshold, which is a percentage of the maximum cluster, otherwise the cluster is eliminated.

Applying the clustering algorithm to the Colour constant image (Figure 3b and 4b) produces a segmented image containing only the principal chromatic image features (Figure 3c and 4c).



(a) (b) (c)

Figure 3 - (a) Image of "Red Face" monkey in the Amazon Forest, (b) Result after Colour Constancy, (c) Result after GT clustering



(a) (b) (c)

Figure 4 - (a) Image of "others" monkey in the Amazon Forest, (b) Result after Colour Constancy, (c) Result after GT clustering

III. Recognition algorithm

The purpose of the recognition stage is to analyse the segmented image, and extract colour features which can be used to match candidate objects in the scene. The task of the recognition system is made much simpler since most of the background artefacts have been removed in the segmented image. This significantly improves the efficiency of the recognition algorithm. The recognition approach is based on a comparison of Colour Texture Angles [9] but, in this case, the colour angles are calculated from just the above segmented clusters and not the whole image. The algorithm is described by the following steps:

Extraction of Colour Texture Angles

- First the mean from each band is calculated and subtracted from the respective band (7) so that each band of the result image will have mean equal to zero.

$$I_i^2 = I_i^1 - \mu_i(I^1) \text{, where } i \in \{r, g, b\} \tag{7}$$

- Then the angle between the *i*th and *j*th colour band is computed according to (8) so that the colour distribution of each image is represented by its three Colour angles.

$$\Phi_{ij}(I^2{}_i, I^2{}_j) = \cos^{-1}\left(\frac{\sigma_{ij}(I^2)}{\sigma_i(I^2)\sigma_j(I^2)}\right) \tag{8}$$

where $\sigma_{ij}(I^2)$ represents the cov ariance between bands i and j from image I^2 and $\sigma_i(I^2)$ is the s tan dard deviation of band i from image I^2

- After that the steps above are repeated for the image convolved with a Laplacian of Gaussian Filter (9). This filter is often used to recognise local surface texture [10] so these angles are called Colour-texture angles.

$$I_i^1 = \nabla^2 G * I_i \text{, where I is the original image post color cons tan cy and } i \in \{r, g, b\} \tag{9}$$

Therefore, each image is represented by its 6 angles which will be used for indexing.
Finlayson [9][11] has proved that (8) represents the angle between two colour bands and also that this measure is illuminant independent.

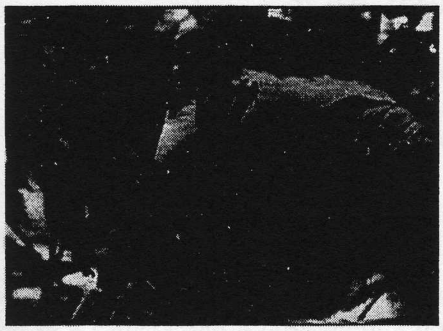
Index Matching

A similarity indice for each pair of images (one from the model database and the other from scene database) is computed as the root-mean square error between the respective Colour-texture angle vectors. If the best index for each scene is below a threshold it means that the model matches the scene and the algorithm is terminated.

IV. Results

Regarding the Colour Constancy algorithm, there is an improvement in terms of colours in most of the processed images as the influence of the reflected Colours from the background onto the animal is reduced. The transform does not ensure overall colour constancy, but adjusts the gains of the camera photoreceptors in a way similar to that of the human eye (i.e. by decreasing the gain of the most intensive stimulus).
An original image and the result after transformation can be seen in Figure 5 as an example.
In this case, the influence of green light on the colour of the monkey fur is reduced.
Results applying the segmentation and recognition algorithm before Colour constancy (STA) and after Colour constancy (CCSTA) to a model database composed of 1 monkey (red face) in three different conditions and a test database comprising 43 images of the same species of monkey (red face) and 31 images

of three different monkeys (others) are given in Table 1. The database is composed of images in a natural environment without any illumination constraint. Results using only CTA (Colour Texture Angles by Finlayson) are also presented for comparison.



(a)

(b)

Figure 5 - (a) Image without Colour constancy
(b) Image with Colour constancy

Table 1 - Experimental Result

		Match	Don't	Match(%)
CCSTA	red face	32	11	74.4
	Others	0	31	0
STA	red face	23	20	53.4
	Others	5	26	19.2
Finlayson CTA	red face	13	30	30.2
	Others	13	18	41.9

It is evident that approaches based solely on CTA performed poorly. More than half were indexed incorrectly and there were many false positives. Matching using a Colour constancy pre-processing and also GT clustering for segmentation and then applying (CTA) to represent each segmented image yields a better result with no false positives and 74.4% of the test images are matched correctly. We conclude that for recognition of animals in a natural scene it is important to reduce the influence of illuminant change (Colour constancy) and background (segmentation) and also to use indices that exploits the 3 bands of the image.

V. Future Work

Currently we are not being able to recognise the whole database, as just 6 Colour-texture angle indices are not sufficient to represent large database. Therefore, we seek to extend the approach by incorporating other invariant features. We also plan to investigate other segmentation methods such as Gaussian Markov Random Field (GMRF) models [12] and then apply the CTA for each cluster in the segmented image so that we could match not just the whole animal but also important clusters that can better represent the animal. We are also researching other colour constant methods in order to make the recognition of colour pictures more robust.

Acknowledgement

This work is partly supported by Brazilian Government (CAPES) and TEMPUS S-JEP 08012-94.

References

[1] J. von Kries, *Beitragzur Physiologie der Gesichtsempfindung*, Arch. Anat. Physiol., 2: 505-524, 1878.

[2] D.A. Forsyth, *A Novel Algorithm for Colour Constancy*, International Journal Computer Vision, 5: 5-36, 1990.

[3] G. West and M.H. Brill, *Necessary and Sufficient Conditions for von Kries Colour Adaptation to Give Colour Constancy*, J. Math. Biol., 15:249-258, 1982.

[4] E.H. Land and J.J. McCan, *Lightness and Retinex Theory*, J. Opt. Soc. Amer., 61: 1-11, 1971.

[5] G. Wyszecki and W.S. Stiles, *Colour Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley & Sons editor, 2° ed., pg. 164, 1982.

[6] G.D. Finlayson, M.S. Drew and B.V. Funt, *Colour Constancy: Generalised Diagonal Transforms Suffice*, Journal Opt. Soc. Amer., 11: 3011-3019, 1994.

[7] J. Mattas, *Colour-based Object Recognition*, Submitted for the Degree of Doctor of Philosophy from the University of Surrey, September 1995.

[8] A. Khotanzad and A. Bouarfa, *Image segmentation by a parallel, non-parametric histogram based clustering algorithm*, Pattern Recognition, pages 961-973, 1990.

[9] G.D. Finlayson, S.S. Chatterjee and B.V. Funt, *Colour-texture indexing*, Technical Report from Department of Computer Science, University of York, 1996.

[10] B.V. Funt and G.D. Finlayson, *Colour constant Colour indexing*, IEEE transactions on Pattern analysis and Machine Intelligence, 1995.

[11] G.D. Finlayson, S.S. Chatterjee and B.V. Funt, *Colour angular indexing*, In The Fourth European Conference on Computer Vision (Vol. II), pages 16-27. European Vision Society, 1996.

[12] D.K. Panjwani and G. Healey, *Markov Random Field Models for Unsupervised Segmentation of Texture Colour Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 10, October 1995.