Published in the Proceedings of the 10th European Signal Processing Conference, Tampere, Finland, Sept. 4-8, 2000

ILLUMINATION INVARIANT OBJECT RECOGNITION USING THE MNS METHOD*

D. Koubaroulis[1] *J.Matas*[1,2] *J.Kittler*[1]

[1] Centre for Vision Speech and Signal Proc.
University of Surrey, Guildford, UK

[2] Center for Machine Perception
CTU Prague, 121 35 Czech Republic

e-mail: {d.koubaroulis, g.matas}@ee.surrey.ac.uk

ABSTRACT

The suitability of the Multimodal Neighbourhood Signature (MNS) method for illumination invariant recognition is investigated. The MNS algorithm directly formulates the problem of extracting illumination invariants from local colour appearance of an object. The invariants are the channel-wise ratio and the cross-ratio computed from modes (pairs of modes respectively) of colour density function in neighbourhoods with multimodal density function.
The MNS algorithm is tested on a colour object recognition task designed to test the effectiveness of algorithms claiming illumination invariance properties. The image set used is publicly available from the Simon Fraser University. Results previously reported using colour constancy and histogram matching were comparable to the performance of the presented method that achieved recognition rate of 60%. When the pose of the objects was fixed recognition performance was 84%.

1 Introduction

An interesting aspect of research in retrieval from image databases is the development of methods based on the visual content i.e. the appearance of the objects in the images. In particular, colour-based image retrieval and object recognition systems have demonstrated acceptable results in the past few years (see [18] for a survey). Many applications require retrieval of images, where the query object (or region) covers only a fraction of the database image, a task essentially identical to colour-based object recognition with unconstrained background (also referred to as region-based image retrieval).

Retrieving images of objects based on their surface reflectances must take into account the factors that influence colour image formation i.e. the viewing geometry, the illumination conditions and the image acquisition system used to record the surface appearance. Depending on the application, illumination colour and intensity as well as viewpoint and background may change. Moreover, partial occlusion and/or deformation of non-rigid objects must also be taken into consideration. Other not so well modeled or not readily measured effects like mutual illumination and individual camera characteristics (e.g. gamma correction settings) also result in variations of object appearance. Consequently, invariance (or at least robustness) to a number of different factors is highly desirable.

Most current colour-based retrieval systems use various versions of the colour histogram [19] which has proven useful as a descriptor of the global colour structure of an image. However, histogram matching cannot be directly applied to the problem of region-based retrieval. A recent effort to extend the colour histogram so that it can handle region-based queries was described in [8]. Despite their overall simplicity and flexibility, histograms are not invariant to varying illumination and not generally robust to background changes. Applying colour constancy methods to achieve illumination invariance for histogram-based methods has been a challenging subject of research for many years [4]. Other methods have also been used to retrieve scenes with similar appearance e.g. using wavelets [9] and moments of the image colour distribution [7, 14]. Finally, graph representations of colour content (like the colour adjacency graph [12] and its extension to a hybrid graph [17]) have provided good recognition for scenes with fairly simple colour structure.

For region-based retrieval and recognition, a description of the local object colour structure seems more appropriate than full-image (global) approaches. The main advantages of describing images with localized invariant features are robustness up to considerable background changes and partial occlusion. Moreover, illumination invariant features can be computed from compact image regions even for the case of spatially varying illumination conditions within the same image (thus a requirement for global illumination change across the images is not necessary). It is well known [11, 5], that (under the assumption of locally constant illumination and the diagonal model of illumination change) the channel-wise ratio of colours between two locally neighboring differently coloured surfaces is invariant to lighting changes. In the past, histograms of colour ratios computed locally from pairs of neighbouring pixels [5] and across detected edges

*The research was supported EC fifth framework project IST-13082 ASSAVID. The second author was supported by the Czech Ministry of Education under the grant VS96049.

have been used [6]. In order to avoid the limitations of the global nature of the histogram representation invariant ratios were computed across the boundaries of segmented image regions [16, 15]. However, reliable image segmentation for image retrieval has proven difficult to obtain [18, 15].

As an alternative to segmentation, the image is covered by compact regions where local colour features are computed. In our previous work [13], the *multimodal neighbourhood signature* (MNS) has been proposed to address the colour-based retrieval task. Local invariant features are computed from image neighbourhoods with multimodal colour density function. MNS signatures of compact size (typically a few hundred bytes) are constructed from robust measurements, the modes of the density function. According to the model of illumination change in the application environment, a number of illumination invariants has been proposed. Acceptable results both for region-based image retrieval and object recognition have been reported [13].

In this paper, the suitability of the MNS method for illumination invariant recognition is investigated. Using simple invariant ratios computed from multimodal image neighbourhoods, promising results were obtained on a publicly available dataset. The motivation of this work stems from a recent experiment by Funt et al. [4] on the same image set where the authors concluded that colour constancy methods as a preprocessing step for colour-based applications were not good enough even for a carefully designed image database. We present a method to recognise and retrieve objects regardless of the illumination conditions and without a need for colour constancy normalisation.

A brief outline of the computation of the MNS signature and the invariant features used is given in the next section. Section 3 describes the object recognition experiment and the results obtained are presented and discussed in section 4. Section 5 concludes the paper.

2 Computing Illumination Invariant Features from Multimodal Neighbourhoods

As detailed in our previous work [13], the image is covered with small neighbourhoods of compact size over a grid with randomly perturbed rectangular spacing. For each neighbourhood, the modes (local maxima) of the local colour density function are efficiently located using the mean shift algorithm [3]. Based on the number of the local maxima each neighbourhood is characterised as unimodal, bimodal, trimodal etc. Clearly, for unimodal neighbourhoods no illumination invariant features can be computed. Therefore, we focus on multimodal neighbourhoods. The proposed MNS signature consists of a number of colour pairs originating from the mode values of detected multimodal neighbourhoods. From the MNS signature, a number of invariant features can be computed. For the ease of exposition we will describe feature extraction from bimodal neighbourhoods which are the simplest multimodal ones.

Consider a local image patch with two adjacent surfaces i and j. According to the monochromatic model of surface reflectance [12, 6] the two estimated mode colours will be given by

$$\begin{aligned} r_i &= (R_i, G_i, B_i) &= s_i^k g_i c_i^k \\ r_j &= (R_j, G_j, B_j) &= s_j^k g_j c_j^k \quad , \quad k = R, G, B \end{aligned}$$

where s_i^k is the illumination factor, g_i is the geometric factor and c_i^ks the k-th sensor response to the surface reflectance of patch i under white light (the surface colour). Besides modelling the effects of change in viewpoint and object pose, the geometric factor g of the monochromatic model encompasses all factors that have the same effect on each colour channel, e.g. change of aperture or camera gain and change in illumination intensity. Coefficients s_i^k represent factors that effect individual colour channels, e.g. the change of illumination colour in the diagonal colour constancy model described below.

According to the diagonal model (also called Von Kries model or coefficient rule), illumination change is modeled by an independent scaling of the colour channels by a different constant i.e. $s_c'^k = d_k s_c^k,\ c = i, j\ ,\ d_k \in \mathbb{R}^+$. The diagonal model has been shown plausible when camera sensors are sufficiently narrow-band filters [2]. Under the diagonal model assumption, it is easy to show that the ratio of colours of two neighbouring surfaces with different surface reflectance is invariant to lighting changes [11, 5]. Nevertheless, for the assumption to hold, the two neighbouring surfaces must have the same orientation i.e. $g_i = g_j$. Invariant features can be computed from the 3 colour channel ratios of the mode RGB values

$$f_{cg} = \left(\frac{R_i}{R_j}, \frac{G_i}{G_j}, \frac{B_i}{B_j} \right)$$

In the most general situation, where orientation is different for the two surfaces $g_i \neq g_j$ and $\frac{g_i}{g_i'} \neq \frac{g_j}{g_j'}$, the 2-dimensional cross-ratio vectors

$$f_{cgd} = \left(\frac{R_i G_j}{G_i R_j}, \frac{G_i B_j}{G_j B_i} \right)$$

are invariant under the diagonal model as shown in [6].

For matching relative features, a simple formula was devised

$$d_{fr}(p, q) = \frac{|a * d - b * c|}{\sqrt{a + b + c + d}} \tag{1}$$

where $p = \frac{a}{b}$ and $q = \frac{c}{d}$ are 1-dimensional fractions. The distance between the colour ratio between two RGB values p_i, p_j defined as

$$r_1 = \left(r_R^1, r_G^1, r_B^1\right) = \left(\frac{R_i^1}{R_j^1}, \frac{G_i^1}{G_j^1}, \frac{B_i^1}{B_j^1} \right)$$

and another ratio $r_2 = \left(r_R^2, r_G^2, r_B^2\right)$ between two other colours q_i and q_j is then

$$d_{rat}(r_1, r_2) = \frac{1}{3} \left(d_{fr}(r_R^1, r_R^2) + d_{fr}(r_G^1, r_G^2) + d_{fr}(r_B^1, r_B^2) \right) \tag{2}$$

The modification of d_{rat} to measure the distance between 2-dimensional cross-ratio features is trivial, ignoring one colour channel in (2).

3 Experimental Setup

We repeated the experiment described by Funt et al. [4] that tested color constancy algorithms in the context of object recognition. In [4], matching was performed using histogram intersection after application of a color constancy algorithm. The tested algorithms were grey world, white-patch retinex, neural net, 2D gamut mapping and 3D gamut mapping.

The database used is publicly available from Simon Fraser University [1] and consists of images of 11 household objects, imaged on black background under 5 different lights. Extreme care was taken in image acquisition (linear high quality camera, calibrated, temporal averaging to effectively increase gray resolution, spatial filtering). Due to the careful preparation of the database there were no clipped (saturated) pixel values which strongly influence colour ratios. In total, there are 55 registered images comprising the *model* database. Another *test* image set of 55 images is available, similar to the model set except that the objects were slightly displaced, rotated or occluded. In total, 110 images with resolution 465×635 are available.

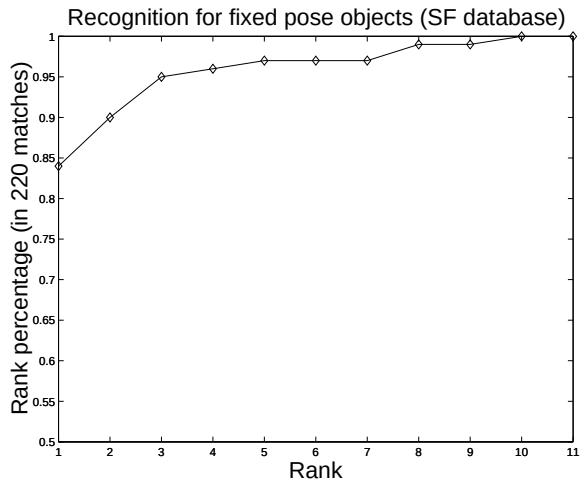
Despite the "ideal" imaging conditions, the recognition rate (i.e. percentage of correct matches) reported in [4] was only 67% for the best color constancy algorithm. If the color transformation was defined using knowledge about the light source, 92% of rank 1 matches were obtained. The authors therefore concluded that neither of the tested color constancy methods is good enough for object recognition.

We repeated the experiment using 3-dimensional f_{cg} ratio features originating from bimodal neighbourhoods. The neighbourhood used was a rectangular window of dimensions 10×10.

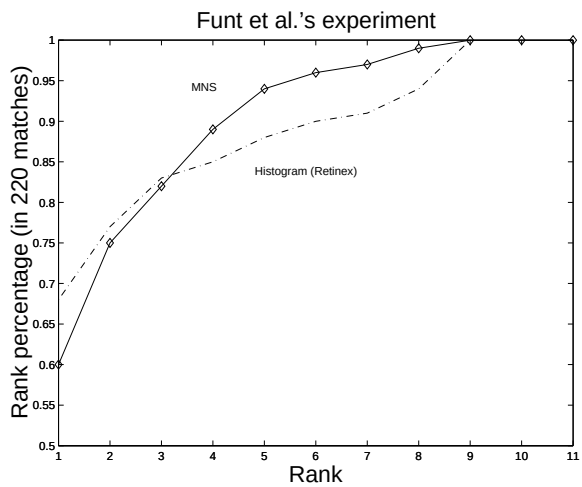
4 Results

Recognition performance for the MNS was comparable to previously reported results (Fig. 1(b)). The distribution of ranks of the correct model is presented in table 1. First rank percentage was 60% and slightly lower than the best performance of reported by Funt et al. However, MNS results are better for ranks two and three. The MNS has 83% models ranked 3 or lower, Funt et al. report 84%. When the pose of all the objects was fixed (i.e. matching models against models under different illumination) the MNS algorithm performed much better achieving a first rank recognition rate of 84% (see Fig. 1(a)).

Note that Simon Fraser database favours full-image (i.e. global) methods like the colour histogram since objects are imaged on a uniform background and they dominate the scene. In the presence of background clutter or other effects that influence the relative areas of the surfaces in the image will negatively affect performance of any method relying on area (or relative area) size. Another factor limiting



(a)



(b)

Figure 1: Recognition results for the Simon Fraser image database. Results for objects with (a) a fixed pose and (b) different pose and light source position.

MNS performance stem from the following. The illumination invariant features of MNS assume *local* change of illumination colour, reducing the six measurements available at a bi-modal neighbourhood to three. Adopting a global model would only reduce the *total* number of measurements by three.

5 Conclusions

The Multimodal Neighbourhood Signature algorithm directly formulates the problem of extracting illumination invariants from local colour appearance of an object. The channel-wise ratio (and cross-ratio) between pairs of modes originating from image neighbourhoods with multimodal density function is stable under the diagonal model of illumination change.

We tested the MNS algorithm on a colour object recog-

Rank	1	2	3	4	5	6	7	8	9	10	11
Histogram (Col. Constancy)	0.68	0.09	0.06	0.02	0.03	0.02	0.01	0.03	0.06	0.01	0.01
MNS (fixed vs. fixed)	0.84	0.07	0.05	0.01	0.01	0.00	0.00	0.02	0.00	0.01	0.00
MNS (moved vs. fixed)	0.60	0.15	0.07	0.07	0.05	0.02	0.01	0.02	0.01	0.00	0.00
MNS (fixed vs. moved)	0.61	0.13	0.10	0.07	0.03	0.04	0.02	0.01	0.00	0.00	0.00

Table 1: Recognition results for Simon Fraser database

nition task designed to test the effectiveness of algorithms claiming illumination invariance properties. The image set used is publicly available from the Simon Fraser University. Results previously reported using colour constancy and histogram matching were comparable to the performance of the presented method that achieved recognition rate of 60%. When the pose of the objects was fixed recognition performance was 84% (percentage of correct model ranked first).

Improvements to the current algorithm include selection of a more appropriate distance function for colour ratios (determined in a learning stage) and use of features suitable for global change to illumination colour. Finally, use of more complex illumination invariants from multimodal neighbourhoods with more than 2 modes is being studied.

References

[1] http://www.cs.sfu.ca/~colour/image_db/.

[2] G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In *Proceedings of the International Conference on Computer Vision, Berlin*, pages 164–171, 1993.

[3] K. Fukunaga and L. Hostetler. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. In *IEEE Transactions in Information Theory*, pages 32–40, 1975.

[4] B. Funt, K. Barnard, and L. Martin. Is Machine Colour Constancy Good Enough? In *Proceedings of the 5th European Conference on Computer Vision*, pages 445–459, 1998.

[5] B. Funt and G. Finlayson. Color Constant Color Indexing . *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

[6] T. Gevers and W. M. Smeulders. Color-based Object Recognition. *Pattern Recognition*, 32(3):453–464, 1999.

[7] G. Healey and D. Slater. Global Color Constancy - Recognition Of Objects By Use Of Illumination Invariant Properties Of Color Distributions . *Journal Of The Optical Society Of America A-optics Image Science And Vision*, 11(11):3003–3010, 1994.

[8] J. Huang, R.S. Kumar, M. Mitra, J.-W. Zhu, and R. Zabih. Image Indexing Using Color Correlograms . In *IEEE Computer Vision and Pattern Recognition, San Juan, Puerto Rico*, pages 762–768, 1997.

[9] D. Jacobs, P. Belhumeur, and R. Basri. Comparing Images Under Variable Illumination. In *IEEE Proceedings in Computer Vision and Pattern Recognition*, pages 610–616, 1998.

[10] D. Koubaroulis, J. Matas, and J. Kittler. Colour-based Image Retrieval from Video Sequences. In *The Challenge of Image Retrieval, Brighton, UK (to appear)*, 2000.

[11] E. Land and J. McCann. Lightness and the retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

[12] J. Matas. *Colour Object Recognition*. PhD thesis, University Of Surrey, 1995.

[13] J. Matas, D. Koubaroulis, and J. Kittler. Colour Image Retrieval and Object Recognition Using the Multimodal Neighbourhood Signature . In *Proceedings of the 6th European Conference in Computer Vision, Dublin, Ireland (to appear)*, 2000.

[14] F. Mindru, T. Moons, and L. Van Gool. Recognizing Color Patterns Irrespective of Viewpoint and Illumination. In *Proceedings of the Computer Vision and Pattern Recognition,Fort Collins, Colorado*, pages 368–373, 1999.

[15] K. Nagao and W. Grimson. Using Photometric Invariants for 3D Object Recognition. *Computer Vision and Image Understanding*, 71(1):74–93, July 1998.

[16] S. Nayar and R. Bolle. Reflectance Based Object Recognition. *International Journal of Computer Vision*, 17(3):219–240, 1996.

[17] K. Park, Il-Dong Yun, and Sang Uk Lee. Color Image Retrieval Using a Hybrid Graph Representation. *Journal of Image and Vision Computing*, 17(7):465–474, 1999.

[18] Y. Rui, T.S. Huang, and S.-F. Chang. Image retrieval: Current techniques, promising directions, and open issues . *Journal Of Visual Communication And Image Representation*, 10(1):39–62, 1999.

[19] M. J. Swain and D. H. Ballard. Color Indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.