

# Colour-based object recognition under spectrally non-uniform illumination\*

J Matas, R Marik and J Kittler

---

This paper describes a colour-based recognition system with three novel features. Firstly, the proposed system can operate in environments where spectral characteristics of illumination change in both space and time. Secondly, benefits in terms of speed and quality of output are gained by focusing processing to areas of salient colour. Finally, an automatic model acquisition procedure allows rapid creation of the model database.

**Keywords:** computer vision, recognition, colour

---

In this paper we present a colour-based recognition system that aims to demonstrate the advantages of selective processing. We do not attempt to analyse the whole image in the spirit of traditional segmentation methods<sup>1,2</sup>; instead we try to find areas where distinctive colour provides least ambiguous information about the presence of objects from the model database. Using this approach, standard recognition tasks (e.g. *What is in the scene? Where is object X?*) can be accomplished without wasting computational resources in parts of the image where pixel colour analysis is complex, e.g. where mutual illumination effects or specularities must be taken into account.

Pixel colour depends upon a number of factors – spectral reflectance of the viewed object, spectral distribution, intensity and relative position (photometric angles) of illumination sources. We show how the effects of changing illumination and geometry can be predicted, allowing recognition in environments with spectrally variable illumination (in both time and space). Moreover, we do not impose any restriction on the spectral reflectance of objects.

---

Department of Electronic and Electrical Engineering, University of Surrey, Guildford, Surrey GU2 5XH, UK (e-mail: g.matas@ee.surrey.ac.uk)

\*An earlier version of this paper was presented to the *5th British Machine Vision Conference*, York, UK, 13–16 September 1994.

Paper received: 28 July 1994; revised paper received: 1 November 1994

Objects are modelled as sets of coloured patches. A description of each patch for a number of ‘canonical’ illuminants is stored in the database. Automatic model acquisition is generally desirable for any model-based system. Considering the number of objects (> 50), patches and canonical illuminants (4–6), it is clearly necessary. We used a modified MODESP clustering algorithm<sup>3</sup> to accomplish the task.

The present paper makes a contribution to the state-of-the-art in colour processing, as well as our earlier work<sup>4</sup>, by:

- operating under illumination with spectral distribution varying in both space and time;
- adopting a recognition strategy with focus of attention; and
- implementing automatic acquisition of models (learning).

The rest of the paper is structured as follows. The attention mechanism, the overall structure of the recognition system and the region growing method that expands interest points into object hypotheses are described in the next section. We then describe the automatic model acquisition procedure. Experiments on two test images are presented, and our results are summarized.

## SURFACE REFLECTANCE, GEOMETRY AND ILLUMINATION

Sensor response of a standard imaging device is well modelled by a spectral integration process:

$$p_k^x = \int_{\lambda_1}^{\lambda_2} \rho_k(\lambda) L(\lambda, \cdot) d\lambda, \quad k = 1, \dots, n \quad (1)$$

where  $p_k^x$  is the response of the  $k$ th sensor at location  $X$  of the sensor array,  $L(\lambda, \cdot)$  is the light emitted from the surface patch that is projected on pixel  $X$ , and  $\rho_k(\lambda)$  is

the responsivity function of the  $k$ th sensor. The  $n$ -dimensional vector  $\mathbf{p}^x$  will be referred to as pixel value, pixel colour or object colour, assuming that light from a surface patch belonging to a single object falls on pixel  $X$  (the results of this section apply to any pixel; we therefore simplify expressions by dropping  $X$  from  $p_k^X$ ). Besides the optical properties of the patch, the spectral power distribution (SPD) of  $L(\lambda, \cdot)$  depends upon the relative position of the patch, illumination source(s) and the viewpoint (defined by photometric angles), as well as on the spectral power distribution of the illumination source(s). Clearly, any system making use of pixel colour for recognition must separate the dependence of  $\mathbf{p}$  on object material from the effects due to changes in illumination or geometry.

The relationship between illumination SPD and pixel colour has been studied by researchers interested in colour constancy<sup>5-7</sup>. Unfortunately, the theory of colour constancy is developed mainly in the context of the Mondriaan world, i.e. a world consisting of a single planar surface composed of a number of matte (Lambertian) patches. Light striking a Mondriaan world is assumed to be spectrally unchanging and of uniform intensity<sup>8</sup>. Under such conditions, light reflected from a patch is independent of viewing geometry, and can be expressed as:

$$L(\lambda) = S(\lambda) E(\lambda) \quad (2)$$

$S(\lambda)$  is a surface reflectance function of the patch and  $E(\lambda)$  is the (global) illumination SPD. Furthermore, surface reflectances and illuminant SPDs are approximated by a weighted sum of basis functions  $S_j(\lambda)$  and  $E_i(\lambda)$ , respectively:

$$S(\lambda) = \sum_{j=1}^{d_S} S_j(\lambda) \sigma_j \quad E(\lambda) = \sum_{i=1}^{d_E} E_i(\lambda) \varepsilon_i \quad (3)$$

Substituting equations (3) and (2) into equation (1), we obtain:

$$\mathbf{p}_k = \sum_{j=1}^{d_S} \sigma_j \sum_{i=1}^{d_E} \varepsilon_i \int_{\lambda_1}^{\lambda_2} S_j(\lambda) E_i(\lambda) \rho_k(\lambda) d\lambda \quad (4)$$

where the expression inside the integral depends only upon the sensor responsivity  $\rho_k(\lambda)$  and the choice of basis functions for illumination and reflectance. Equation (4) lies at the heart of most colour constancy algorithms; variations exist in the assumed number of sensors at each pixel, and the dimensionality of the illuminant and reflectance spaces ( $d_E$  and  $d_S$ , respectively). From equation (4), it can be seen that if a representation of a spectral reflectance in terms of the vector of mixing weights  $\sigma$  is known, then object colour can be computed for any illuminant described by  $\varepsilon$ .

Unfortunately, comparatively little work has been carried out to establish the applicability of the low dimensionality assumptions. Surface reflectance were studied by Maloney<sup>9</sup>, who analysed 337 spectral reflectances of natural formations collected by Krinov. Maloney concludes that five to seven basis functions provide an almost perfect fit. In our opinion, Maloney's

results are difficult to interpret, yet the quality of the fit of the first three basis function seems sufficient for computer vision applications. On the other hand, it is unlikely that the same basis functions are applicable to a larger set of natural and man-made objects. In contrast, SPDs of a number of artificial illuminants are known. Furthermore, three basis functions providing practically a perfect fit to all phases of daylight have been found<sup>10</sup>.

The effects of geometry on SPD of reflected light have been extensively studied<sup>1,11</sup>. The dichromatic reflection model<sup>12</sup> is generally regarded to be accurate for a large class of materials<sup>13</sup>. The dichromatic model states that reflected light  $L$  consists of two independent components: light reflected on the interface and light due to sub-surface (body) reflection. Furthermore, it is assumed that the SPD of neither of the two components depends upon geometry. Therefore:

$$L(\lambda, g) = m_i(g) L'_i(\lambda) + m_b(g) L'_b(\lambda) \quad (5)$$

where  $g$  denotes the geometry (expressed, for example, in terms of photometric angles),  $m_i(g)$  and  $m_b(g)$  are scaling factors, and  $L'_i(\lambda)$  and  $L'_b(\lambda)$  are the relative spectral distributions of light reflected by interface and body reflection, respectively. The quantities  $L'_i(\lambda)$ ,  $L'_b(\lambda)$  depend only upon the surface reflectance and relative illuminant SPD. Besides geometry, the scaling factors model absolute changes of illumination intensity. The dichromatic model does not specify *how* quantities  $L'$  depend upon illumination and spectral reflection, therefore its application always requires the assumption of spectrally unchanging illumination SPD.

For a colour-based recognition system operating in the real world, the assumptions imposed by colour constancy algorithms are too restrictive. We adopt a weaker set of assumptions by modelling surface reflection by a unichromatic reflection model:

$$L(\lambda, g) = m(g) L'(\lambda) \quad (6)$$

In our opinion, ignoring specularities is justifiable for a number of reasons. In case of metals, the model is equivalent to the dichromatic reflection model. For dielectrics we neglect the specular component. Moreover, specularities almost always cover only a fractional part of an image. Very often the high intensity of specular part saturates the sensor, making colour analysis meaningless. Finally, the attention mechanism will skip over specularities as their colour is inherently more ambiguous than areas of diffuse reflection.

To predict the effects of changing illumination, we substitute for  $L'$  from equation (2) into equation (5):

$$L(\lambda) = m(g) S(\lambda) E'(\lambda) \quad (7)$$

where  $E'(\lambda)$  is the relative SPD of illumination. Assuming low dimensionality of illuminant SPD but a number of sources  $j = 1 \dots N_s$  (with different SPDs), we obtain, after substituting in (1):

$$p_k = \sum_{j=1}^{N_s} m_j(g) \sum_{i=1}^{d_E} \varepsilon_{ij} \int_{\lambda_1}^{\lambda_2} S(\lambda) E_i(\lambda) \rho_k(\lambda) d\lambda \quad (8)$$

where  $m_j(g)$  is a scaling factor covering the effects of change in the intensity and geometry of the  $j$ th source,  $\varepsilon_{ij}$  defines the  $j$ th source SPD in terms of the basis function  $E_i(\lambda)$ . Replacing the integral (which is independent of geometry and illumination) we obtain:

$$p_k = \sum_{j=1}^{N_s} m_j(g) \sum_{i=1}^{d_E} \varepsilon_{ij} p_k^{E_i} \quad (9)$$

where  $p_k^{E_i}$  is the colour of the object with spectral reflectance  $S(\lambda)$  as seen under illuminant  $E_i$ . Because both  $m_j(g)$  and  $\varepsilon_{ij}$  are non-negative, pixel colour  $\mathbf{p}$  will lie inside a convex polyhedron – a pyramid with an apex at  $\mathbf{0}$  and edges coincident with vectors  $p_k^{E_i}$ . In the case of artificial illumination, we choose illuminant SPD as the basis function  $E_i$ . Values  $p_k^{E_i}$  are obtained in a straightforward manner as they are identical to object colour under the given artificial illumination. The three  $p_k^{E_i}$  corresponding to pixel colour under daylight are approximated by points on the convex envelope of  $\mathbf{p}$ 's from images taken under different daylight conditions.

Equation (9) shows that the absolute value of  $p_k$  arbitrarily changes with  $m_j(g)$  and  $\varepsilon_{ij}$ , and therefore carries no information about the object. Therefore, we project  $\mathbf{p}$  onto the chromatic plane (projection on a unit sphere or parametrization by hue and saturation would achieve the same objective). The pyramid is projected on a convex polygon whose vertices correspond to canonical illuminants. Such polygons serve as our model of object colours – pixel colour of a surface patch with arbitrary surface reflectance will lie inside a convex polygon (regardless of geometry) if illuminated by a light source well approximated by the set of basis functions we adopted.

With the presented approach we are able to recognize object colour under any mixture of canonical illuminants. Moreover, we do not have to assume that the whole scene is illuminated by a single source (imagine a common indoor environment with windows and lit lamps). Information about illumination source can be easily incorporated into the scheme; only vectors  $\mathbf{p}$  corresponding to the illuminants known to be present in a given environment are taken into account when computing the convex hull.

## MODEL ACQUISITION

In the analysis above, we made no assumptions about spectral reflectances of object materials. As a necessary consequence, object colour under a given illumination can only be predicted as linear combination of its colour under canonical illuminants. Therefore, for each object a colour model must be acquired for every canonical illumination. Taking into account the number of objects stored in the database ( $> 50$ ), the fact that an object can have a number of colour patches, and the number of illuminants, automatic acquisition is called for.

At present, every object is defined as a collection of coloured patches (rather than a graph). The colour of each patch is characterized by the convex polygon whose

vertices correspond to patch chromaticities under canonical illuminations. Recall that any artificial illuminant is by definition a canonical illuminant; for daylight a number of samples must be taken; points on the convex hull represent the respective canonical illuminants.

The automatic model acquisition proceeds as follows. First a background image is taken. Next, an object is presented to the system (Figure 4a). Pixels that significantly differ in chromaticity from the background image are assumed to belong to the object (Figure 4b). A chromaticity histogram is computed for the subimage (Figure 4d). A procedure based on the MODESP clustering algorithm<sup>3</sup> builds the model by first looking for the chromatic bin with the highest count. Next the immediate neighbourhood of the maximum bin is searched, and the bin with highest count is added to the list of bins. The search continues until a local minimum is reached. The counts in the set of resulting bins provide an approximation of the probability  $P(\text{chroma}|\text{patch})$ . The same steps are carried out for the unprocessed part of the histogram until no more bins above a threshold can be found. The number of maxima found corresponds to the number of coloured patches in the model of the current object.

In the image depicted in Figure 4a, three significant peaks were found. Figure 4c shows pixels labelled by the patch number; the labelling is needed only for verification. The clustering procedure is repeated for every object and every canonical illumination.

## OVERALL STRUCTURE

The overall structure of the colour recognition system is depicted in Figure 1. We will first focus on the most complex part of the system – the attention mechanism. The pixel location on which the module focuses its attention depends, besides the input data, on the definition of the recognition task, the contents of colour database, model of the environment and the current state

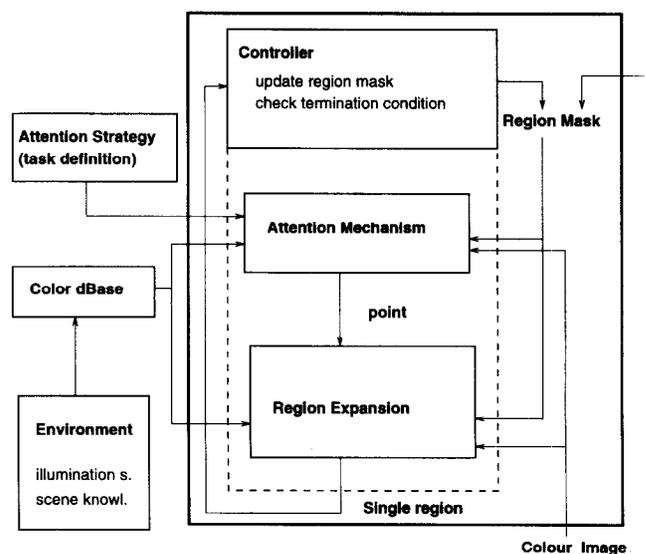


Figure 1 System structure

of the region mask. The region mask serves two purposes. Pixels already processed are marked in the mask and are ignored in further processing. The mask can be set up prior to processing to control in a natural way the image area on which colour processing is performed.

The model of the environment allows the system to exploit specific information about the current scene and illumination. The environment defines a set of objects that can appear in the scene and a set of permissible illuminants. The colour database manager builds a chromatic model for every patch of every specified object. Each patch is characterized by a convex polygon in the chromatic plane with vertices corresponding to patch colour under canonical illumination. At present, the system can cope with scenes with up to five illuminants, for which patch chromaticities are stored in the colour database (tungsten filament lamp, fluorescent lamp and three illuminants modelling daylight). So far we have tested the approach in scenes which were lit by natural light and one artificial illuminant.

The recognition process adopted in our system represents a significant departure from our earlier work<sup>4</sup>, where a Bayesian decision rule was invoked at every pixel in the image. Here, prior to recognition, the colour database is transformed into a look-up table. The table representing a discretized chromatic plane stores at each cell a list of patch labels that can assume the chromaticity. The conditional probability  $P(\text{patch-label}|\text{chroma})$  is computed for each chromatic cell and the labels sorted according to it. Briefly, the process can be described as follows. A set of points where  $P(\text{chroma}|\text{patch-label})$  is non-zero is obtained by raster-

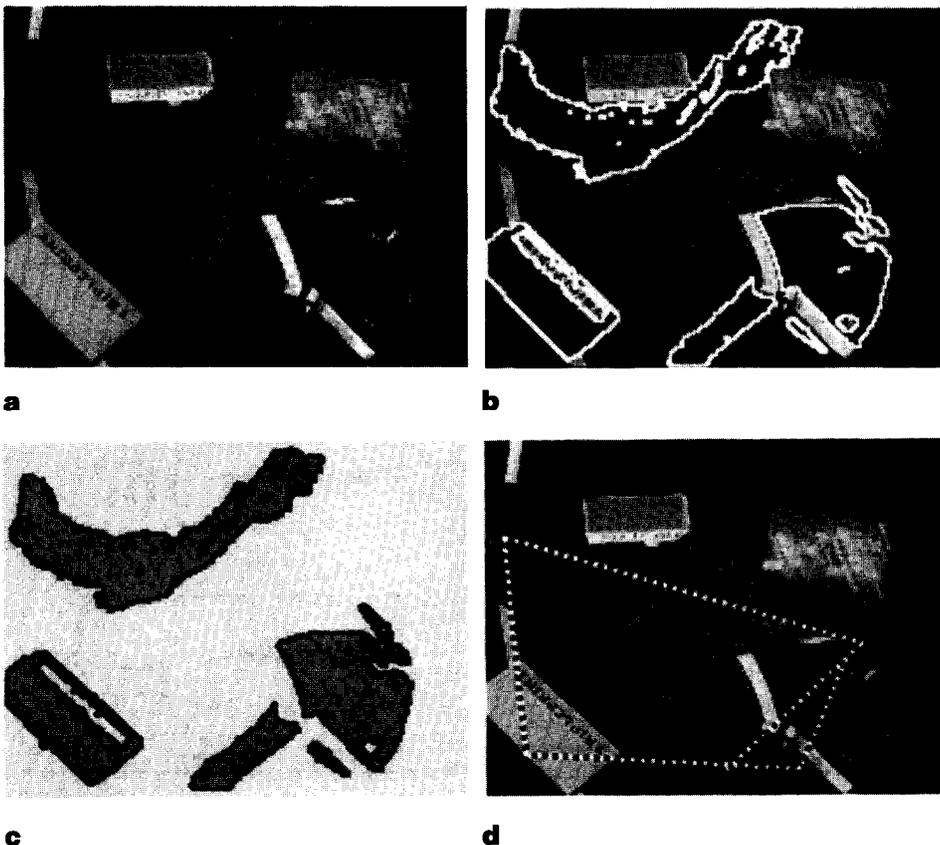
izing the chromatic polygon of the patch. The rasterized polygon is convolved with a smoothing filter to model effects of noise. The shape of the smoothing kernel depends upon the shape of the distribution detected in the colour histogram during model acquisition. Values in the raster are taken as approximations of  $P(\text{chroma}|\text{patch-label})$ . The conditional probability  $P(\text{patch-label}|\text{chroma})$  is finally computed from probabilities  $P(\text{chroma}|\text{patch-label}_i)$  of all patches (with non-zero  $P(\text{chroma}|\text{patch-label}_i)$ ) and the probability  $P(\text{chroma}|\text{background})$ .

The chromatic look-up table enables the attention mechanism to implement efficiently recognition strategies best described as *Where* (is object X) and *What* (is in the scene). In the case of *Where* the attention selects a pixel with chromaticity that maximizes  $P(\text{patch-label}_x|\text{chroma})$ . Note that, although a *list* of labels is stored at every chromatic cell, only the first item of a suitably sorted list need be accessed.

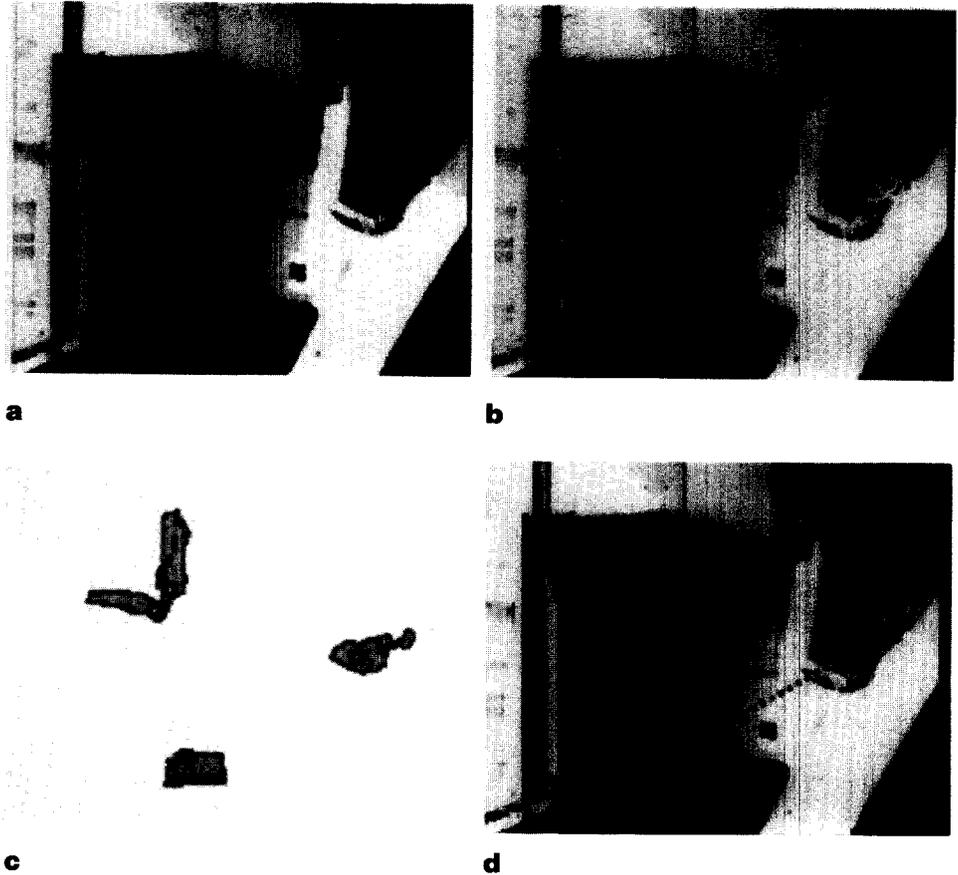
Once the point of interest is selected a standard region growing algorithm<sup>14</sup> is employed to detect a region. The region, together with a list of labels with probabilities, is passed to the controller. The controller updates the region mask and decides whether to terminate processing or run another iteration of the attention-region expansion loop.

## EXPERIMENTS

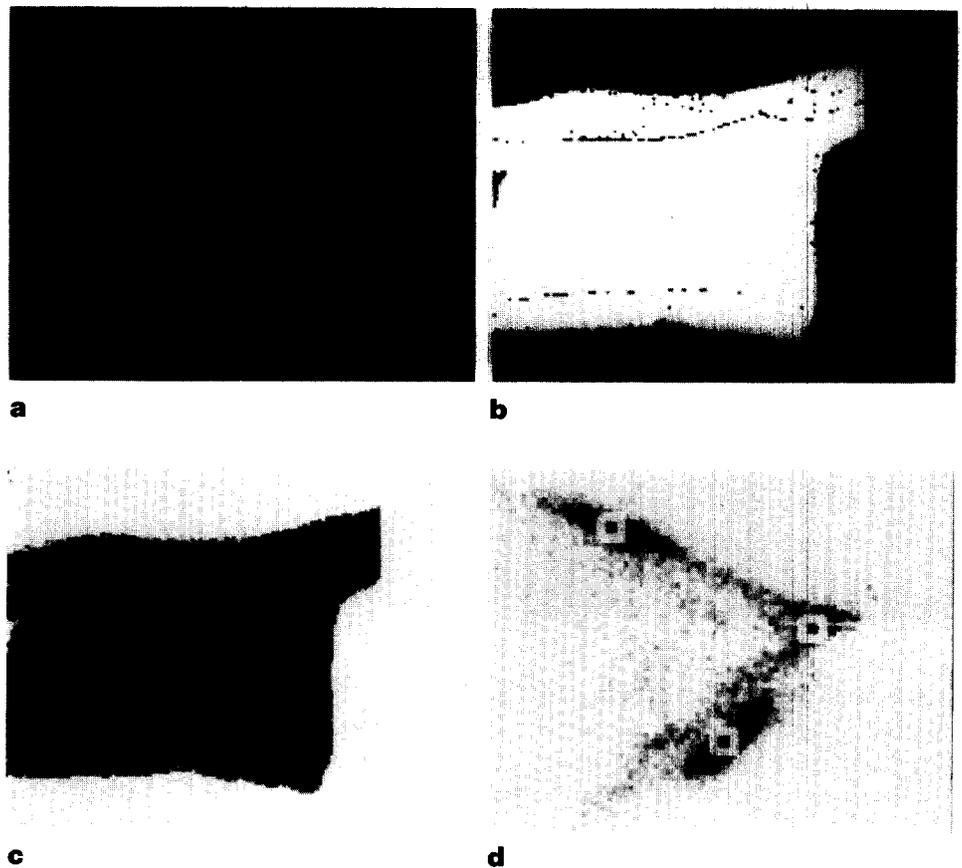
We demonstrate performance of the colour recognition system on images shown in *Figures 2a* and *3a*. The



**Figure 2** Experiment 1. (a) Image of the scene; (b) regions expanded from the six first attention points superimposed on the image; (c) regions with attention points; (d) movement of the point of attention. The sequence starts on the trunks in the bottom-right part of the image



**Figure 3** Experiment 2. (a) Image of the scene; (b) regions expanded from the four first attention points superimposed on the image; (c) regions with attention points; (d) movement of the point of attention. The sequence starts on the trunks on the PUMA robot arm



**Figure 4** Model acquisition. (a) Image of the object; (b) part of the image belonging to the object detected from chromatic differences; (c) labelled patches; (d) chromaticity histogram with markers highlighting local maxima. Each maximum corresponds to one colour patch

colour database contained 51 models depicted in *Figure 5*. The environment model specified that daylight and fluorescent light could be present in the

scene. The termination strategy was set as 'find  $n$ -best regions'.

The first experiment (*Figure 2*) shows the perfor-

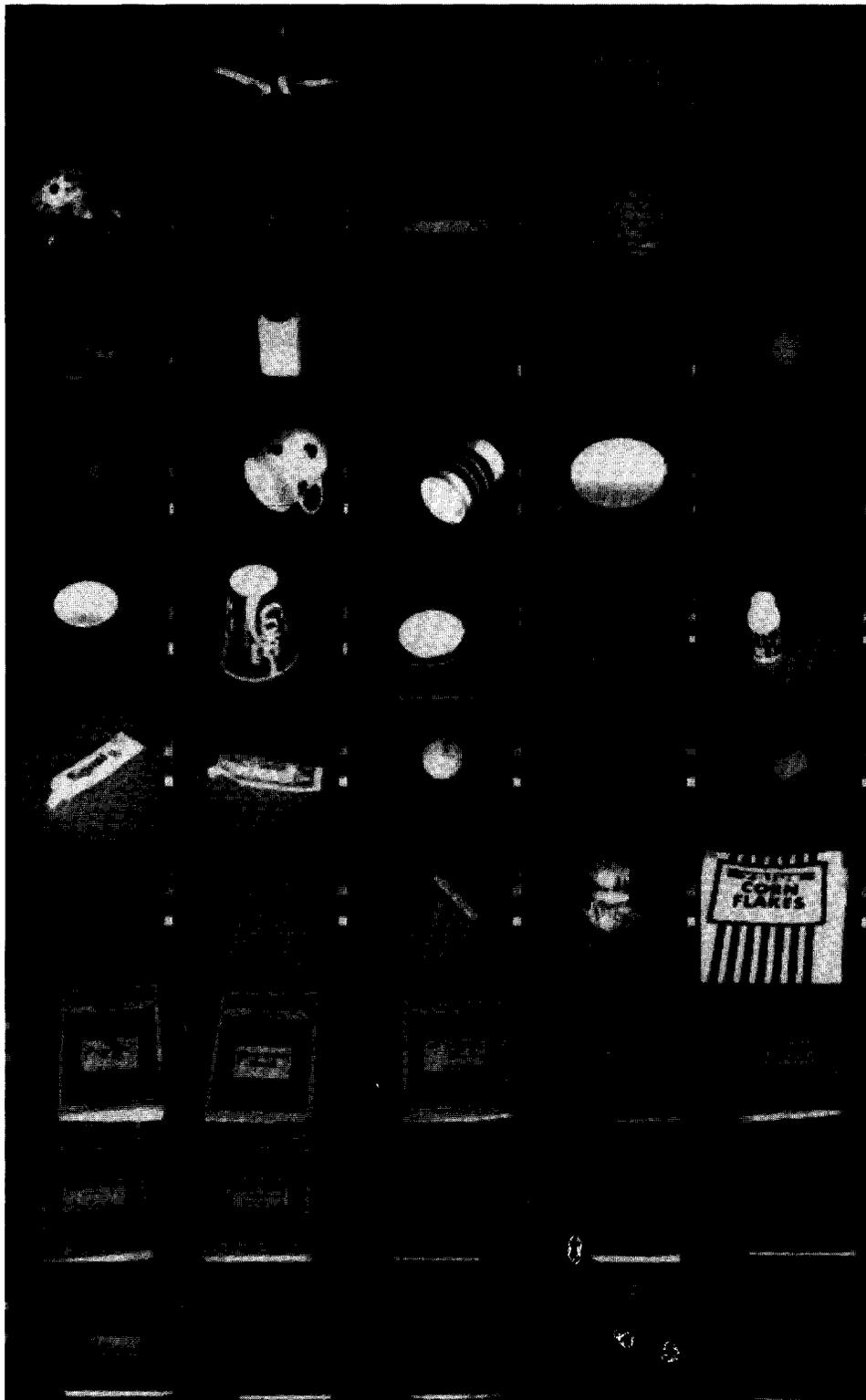


Figure 5 Fifty selected objects from the database

mance of the recognition system under most favourable conditions reminiscent of the Mondriaan world. The objects are placed close to each other on a single plane; it is therefore likely that they are illuminated by light of unchanging spectral distribution. The sequence of points of attention is shown in *Figure 2d*. The points, together with regions into which they were expanded, are depicted in *Figure 2c*. The sequence starts on the red trunks. Next another part on the inside of the trunks is picked (this part is separated from the main part of the

trunks by a white stripe). The attention point then moves clockwise to the yellow envelope and the purple sleeve. Next a small triangular patch to the left of the trunks is detected. This patch is a part of the back side of the pyjama top and has the same colour. The process was terminated after focusing on the bottom purple part of the pyjama top (the small triangular patch). *Figure 2c* shows the position of the attention points in the expanded regions. It might appear counter-intuitive, but there is no reason why the points should lie near

the region centres as the colour near the centre (of a sufficiently large object) is, on average, equally likely to provide the point with least ambiguous chromaticity (in case of a tie, the top-left pixel is chosen). With the exception of the group of holes on the top part of the pyjama top, the expanded regions fit well the image data. The holes correspond to almost black points on the dark side of the creases.

The second experiment was carried out using the same objects as in experiment one (Figure 3). However, the complexity of the analysis is increased by two factors. Firstly, the proportion of daylight to fluorescent light is higher for objects closer to the window (on the left edge of Figure 3a). Secondly, objects are not placed on a planar surface. Results of the second experiment are presented in the same way in Figure 3. The point of attention shifted from right to left, i.e. from the trunks to the pyjama top.

## CONCLUSIONS

We have presented a colour-based recognition system with three novel features. Firstly, the proposed system can operate in environments with spectrally uneven and changing illumination. Secondly, benefits in terms of speed and quality of output are gained by focusing processing to areas of salient colour. Finally, an automatic model acquisition procedure has been implemented. The results presented show that the approach is viable.

## ACKNOWLEDGEMENTS

The work was carried under ESPRIT project 7108, *Vision as process*, and the BRITE project BREU-CT91-

0496, *Automated Visual Inspection Applied to Stone Process*.

## REFERENCES

- 1 Klinker, G J, Shafer, S A and Kanade, T 'Using a color reflection model to separate highlights from object color', *Proc. 1st Int. Conf. Comput. Vision*, London, UK (1987) pp 145-150
- 2 Gershon, R, Jepson, A D and Tsotsos, J K 'Highlight identification using chromatic information', *Proc. 1st Int. Conf. Comput. Vision*, London, UK (1987) pp 161-170
- 3 Kittler, J 'A locally sensitive method for cluster analysis', *Patt. Recogn.*, Vol 8 (1976) pp 23-33
- 4 Matas, J, Marik, R and Kittler, J 'Generation, verification and localisation of object hypotheses based on colour' in J Illingworth (ed), *Br. Machine Vision Conf.*, BMVA Press (1993) pp 539-548
- 5 Forsyth, D A 'A novel approach to colour constancy', *Proc. 2nd Int. Conf. Comput. Vision*, Tampa, FL (December 5-8 1988) pp 9-18
- 6 Maloney, L T 'Evaluation of linear models of surface spectral reflectance with small number of parameters', *J. Opt. Soc. Am.*, Vol 3 (October 1986) pp 1673-1686
- 7 Tsukada, M and Ohta, Y 'An approach to color constancy using multiple images', *Proc. 3rd Int. Conf. Comput. Vision*, Osaka, Japan (December 4-7 1990) pp 385-389
- 8 Finlayson, G D, Drew, M S and Funt, B V 'Diagonal transforms suffice for color constancy', *Proc. 4th Int. Conf. Comput. Vision*, Berlin, Germany (May 11-14 1993)
- 9 Maloney, L T and Wandell, B A 'Color constancy: a method for recovering surface spectral reflectance', *J. Opt. Soc. Am.*, Vol 3 (January 1986) pp 29-33
- 10 Wyszecki, G and Stiles, W S *Color Scene: Concepts and Methods, Quantitative Data and Formulae*, John Wiley, Chichester (1982)
- 11 Healey, G and Binford, T O 'Local shape from specularity', *Proc. 1st Int. Conf. Comput. Vision*, London, UK (1987) pp 151-160
- 12 Shafer, S A *Using color to separate reflection components*, Technical report, University of Rochester (1984)
- 13 Tominaga, S 'Surface identification using the dichromatic reflection model', *IEEE Trans. PAMI*, Vol 13 (June 1991) pp 658-670
- 14 Marik, R *Regions and image segmentation*, Technical report (PhD, thesis), Department of Control Engineering Czech Technical University, Prague, Czechoslovakia (1992)