Generation, Verification and Localization of Object Hypotheses based on Colour

J. Matas, R. Mařík, and J. Kittler
Dept. of Electronic and Electrical Engineering,
University of Surrey,
Guildford, Surrey GU2 5XH, United Kingdom

Abstract

This paper presents a model-based method for colour-based recognition of objects from a large database. The algorithm is based on the assumption that surface reflectances of objects in the model database follow the extended dichromatic model proposed by Shafer [Sha84]. Adoption of the dichromatic model allows recovery of *body colour* - the component of sensor responses (*RGB-values*) that is independent of scene geometry and illumination intensity. Both theoretical studies [Hea89b] and experiments [LB90][KSK88] confirm that Shafer's model gives a suitable approximation for reflectances of a wide range of materials.

Instead of using traditional techniques (eg. clustering, split-and-merge) to obtain regions of 'similarly' coloured pixels followed by classification a novel approach is argued for. First, for each pixel a list of models with nonzero aposteriori probabilities $P(model_i|body\ colour)$ is computed using Bayes formula. Next, regions are formed by grouping pixels with identical most probable hypothesis. Probabilities $P(model_i|region)$ are obtained trough a standard group decision rule [FT80].

We show that the proposed scheme can be used for a number of visual tasks - localization of objects, generation and verification of object hypotheses. Experiments on images of complex indoor scenes confirm that the proposed method can provide reliable information about the surrounding environment.

1 Introduction

> This system is based on the philosophy that there normally exist easy ways of finding objects. *T.D. Garvey, SRI, 1976*

Today, with almost twenty more years of experience, vision researchers may or may not share Garvey's optimism [Gar76]. In this paper we attempt to show that colour can provide, in many complex natural scenes, an efficient if not 'easy' way of localising objects. The proposed model-based algorithm can be outlined as follows. Assuming that surface reflectance follows the dichromatic reflection model [Sha84], information that is invariant to geometry and illumination intensity can be extracted from sensor responses at every pixel location. This information, the *body colour*, is then compared with body colour of prestored models and object

BMVC 1993 doi:10.5244/C.7.54

hypotheses are generated for every pixel. Finally, pixels with identical most likely hypotheses are grouped to form regions.

The reversal of the classical segmentation paradigm of group-and-classify [Hea89a], [Fat92] offers a number of advantages. A pixel with body colour not matching any model will be immediately discarded from further processing. Regions formed from pixels indexing a single model are output directly after grouping. Multiple hypothesis are kept at regions containing pixels with ambiguous colour; wider context (neighbouring regions), or possibly a different visual cue (eg. shape, texture), can be used to resolve the ambiguity. A single object can be represented by a number of colour models; therefore the proposed approach , unlike the recently published colour-based indexing methods [SB90], [Wix90], can successfully locate objects with surfaces with single and multiple colours.

The Bayesian approach employed for hypothesis generation can be easily primed with apriori information. It is therefore possible to use the proposed framework in goal-directed, verification mode by setting all but one a priori model probabilities to 0 ('Where is X?') or in a data-driven, invocation mode with apriori probabilities equal ('What is in the image?').

The rest of the paper is organised as follows. Section 2 discusses the advantages and limitations of the dichromatic reflection model. The details of the colour-based matching scheme are presented in Section 3 . Section 4 reports the results of experiments carried out to assess the proposed approach. Finally Section 5 draws some conclusions about the advocated method.

2 Analysis of Surface Reflection

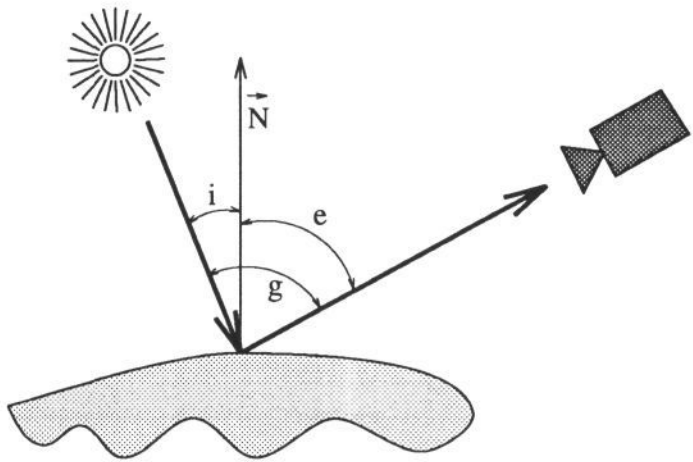


Figure 1: Reflection of light from a surface patch. In the dichromatic reflection model, the dependence of surface reflection L on scene geometry is defined in terms of three angles. The incidence angle i between the surface normal N and the light source direction, the emittance angle e between the surface normal and the viewing direction and the phase angle g between the light source and viewing directions.

If sensor response is assumed to be additive over wavelength then the output q_k of the k-th sensor (eg. red, green, blue) can be expressed as

$$q_k = \int_0^{+\infty} \rho_k(\lambda) L(i, e, g, \lambda) d\lambda, \qquad k = 1, \cdots, n \tag{1}$$

where $\rho_k(\lambda)$ is the spectral responsibility of the k-th sensor and $L(i, e, g, \lambda)$ denotes surface radiance in viewpoint direction which is defined in terms of standard photometric angles i, e and g (fig. 1). The standard symbol λ refers to the wavelength of light, n denotes the number of sensors.

In Shafer's dichromatic model of surface reflection [Sha84], total radiance L of reflected light is defined as the sum of two independent parts: the radiance L_i of the light reflected at the interface (surface reflectance) and the radiance L_b of the body (sub-surface reflectance).

$$L(i, e, g, \lambda) = L_i(i, e, g, \lambda) + L_b(i, e, g, \lambda) \tag{2}$$

Furthermore, the dichromatic model states that each of the radiance components can be expressed as a product:

$$L_i(\lambda, i, e, g) = m_i(i, e, g) c_i(\lambda), \qquad L_b(\lambda, i, e, g) = m_b(i, e, g) c_b(\lambda) \tag{3}$$

where $c_b(\lambda)$, resp. $c_i(\lambda)$ depends solely on the relative spectral power distribution of the light source and spectral characteristic of surface and $m_{\cdot}(i, e, g)$ represent the influence of scene geometry on radiance $L_{\cdot}$.

Substituting (2) and (3) into (1) we obtain after a simple manipulation

$$\begin{aligned} q_k &= \int_0^{+\infty} \rho_k(\lambda)(m_i(i, e, g) c_i(\lambda) + m_b(i, e, g) c_b(\lambda)) d\lambda \\ &= m_i(i, e, g) \int_0^{+\infty} \rho_k(\lambda) c_i(\lambda) d\lambda + m_b(i, e, g) \int_0^{+\infty} \rho_k(\lambda) c_b(\lambda) d\lambda \\ &= m_i(i, e, g) q_k^i + m_b(i, e, g) q_k^b \end{aligned} \tag{4}$$

where sensor responses q_k^i, q_k^b due to interface reflection, and body reflection respectively are defined as

$$q_k^i = \int_0^{+\infty} \rho_k(\lambda) c_i(\lambda) d\lambda, \qquad q_k^b = \int_0^{+\infty} \rho_k(\lambda) c_b(\lambda) d\lambda \tag{5}$$

The usefulness of the dichromatic model ensues from the fact that the n-dimensional vectors $\underline{q^b}$, $\underline{q^i}$ contain valuable information about the viewed object; it is apparent from equation (4) that $\underline{q^b}$ is geometry-independent and proportional to illumination intensity. The dichromatic model is widely applicable: a number of experiments [KSK88], [KSK87], [GJT87] as well as physical analysis of refraction-reflection phenomena [LB90], [HB87] suggest the surface reflectance approximation expressed by equations (2) and (3) is accurate for a wide range of materials. On the other hand we found the assumption of a single light source unrealistic for experiments carried out in daylight. To cater for effects of the diffuse component of daylight we introduced constant term to equation (5). As a final modification of the original dichromatic model we decided to drop the term corresponding to (specular) interface reflection. Firstly, specularities are very bright and the intensity

of image irradiance saturates colour channels. Secondly, specularities cover only a small proportion of surfaces; in our opinion giving up on correct classification of these patches is worth the simplification and increased efficiency of processing of colour. The reflection model thus becomes:

$$q_k = m_b(i, e, g)q_k^b + q^{ambient} \tag{6}$$

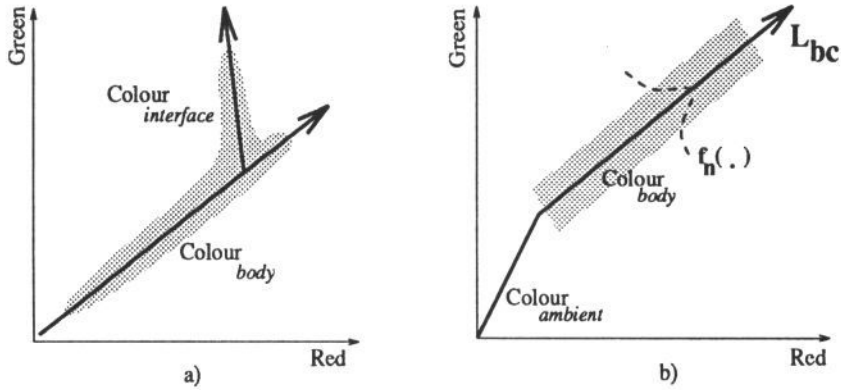


Figure 2: Clusters of RGB values corresponding to surface patches of identical material (body colour) as predicted by Shafer's dichromatic model (a) and the model adopted in the paper (b) .

The difference between the original dichromatic model and the model defined by equation (6) used for colour matching is depicted in figure 2.

It can be seen from equation (5) that sensor response $\underline{q}$ depends through c on the spectral power distribution (SPD) of light incident on the imaged surface. One possible approach to eliminate the dependence, estimation of illuminant colour and recovery of illuminant-independent colour description (the colour-constancy problem) [For88], [MB86] has been avoided because of its complexity. Instead we adopted the approach of [KSK88] and [LB90] and assumed that all surfaces in the scene are lit by light of the same SPD (and that body colour models were acquired under the same illumination).

3 From Pixels to Object Hypotheses

Equation (6) predicts that RGB values corresponding to surface patches of the same body colour will lie on a broken line L_{bc} depicted in fig. 2b. Assuming additive nature of the noise in the acquisition chain, the probability of an RGB-value being a noisy measurement of body colour can be expressed as

$$P(RGB|M_{bc}^i) = f_n(\|L_{bc}^i, RGB\|) \tag{7}$$

where $\|.\|$ denotes Euclidean distance of the RGB triplet from line L_{bc} and f_n represents properties of noise. Combining conditional probabilities of equation (7) with apriori model probabilities using Bayes formula gives

$$P(M_{bc}^i|RGB) = \frac{P(RGB|M_{bc}^i)P(M_{bc}^i)}{\sum_j P(RGB|M_{bc}^j)P(M_{bc}^j) + P(RGB|bg) * P(bg)} \tag{8}$$

where $P(bg)$ refers to the apriori probability that a pixel does not correspond to any model and $P(RGB|bg)$ denotes the colour distribution of non-model (background) pixels.

The object hypotheses are generated by straightforward application of equation (8):

1. At every pixel location, generate a list of models with aposteriori probabilities $P(M^i_{bc}|RGB)$ greater then 0 (or a small threshold)

2. Form regions by grouping pixels with identical i, such that $P(M^i_{bc}|RGB) = \max_j P(M^j_{bc}|RGB)$. Apply the standard group decision rule [FT80] to compute $P(M^i_{bc}|region)$ (index p iterates over all pixels of the region):

$$P(M^i_{bc}|\forall_p RGB_p) = \frac{\{\prod_p P(RGB_p|M^i_{bc})\}P(M^i_{bc})}{\sum_j\{\prod_p P(RGB_p|M^j_{bc})\}P(M^j_{bc})} \tag{9}$$

Complexity of our implementation of the algorithm is proportional to the number of pixels multiplied by the number of models. However, a more sophisticated implementation could use techniques akin to Delaunay triangulation to preselect candidate models to keep complexity almost independent of the number of models. Most properties of the algorithm can be deduced from equation (8), eg.:

- Discernibility grows with brightness. All lines L^i_{bc} meet at (0,0,0) in the RGB space and, consequently, $P(M^i_{bc}|RGB)$ are similar for dark patches.

- The influence of $P(bg)$ on classification results is weak because of the small value of $P(RGB|bg)$ (it integrates to 1 over the whole RGB space)

- If a model has a highly conspicuous colour (with respect to other models) it will be recognised even if the body colour is strongly influenced by noise.

- The method can be easily modified to accept non-stationary $P(M^i)$, ie. a function of pixel location, allowing incorporation of additional knowledge (from previous processing or different visual modules).

4 Experiments

The algorithm described in section 3 was tested on a sequence of images (approx. 70) taken in a large office. Thirty-five test objects (fig. 5, table 1) were arranged randomly in the room. During acquisition of the sequence the viewpoint and zoom of the hand-held camera were changed and objects moved. Some of the non-rigid objects (doll mc301, puppet mc291, bag mc321) changed their shape. The acquisition chain was linearized by a method of varying aperture [AM92] (linear camera was tacitly assumed in the development of surface reflection models in section 2).

The potential of the proposed method is demonstrated on three experiments. In the first experiment, localization of model mc071 (Minsky and Papert: Perceptrons) is attempted in the cluttered bookshelf scene depicted in the top-left corner of fig. 4. Except for mc071, all model apriori probabilities were set to 0. The aposteriori probability $P(mc071|RGB)$ for every pixel location is plotted in the bottom-left corner of fig. 4. Pixels with $P(mc071|RGB)$ above 0.5 are shown

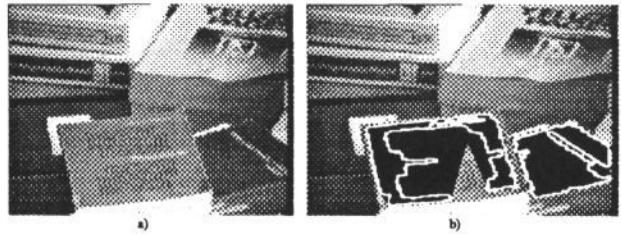


Figure 3: Generating hypotheses about books. (a) Original image. (b) Exemplary regions (white contour) and their convex hull (black).

in the middle-left image. The most likely pixels indeed correspond to the sought book. Because of interreflection, shadows and blurring (mixing of colours) *not all* parts of the book were assigned high $P(mc071|RGB)$. But does it matter?

The second experiment is documented in the right hand column of fig. 4. The set up was identical to the first experiment with mc321 (ochre plastic bag) replacing Minsky's book. $P(mc321|RGB)$ is high for pixels corresponding to the bag. The brownish colour matches comparatively well the colour of the table (especially in shadows) as well as the colour of puppet mc291 in the foreground; the background 'noise' is therefore much stronger.

In the third experiment (fig. 3) the system generated hypotheses about all books. In this case, the output of the colour-based algorithm is voluminous; the two books in the foreground were selected for presentation of the results. The region to the right, corresponding to model mc081 (Arbib's book) was classified in the following way:

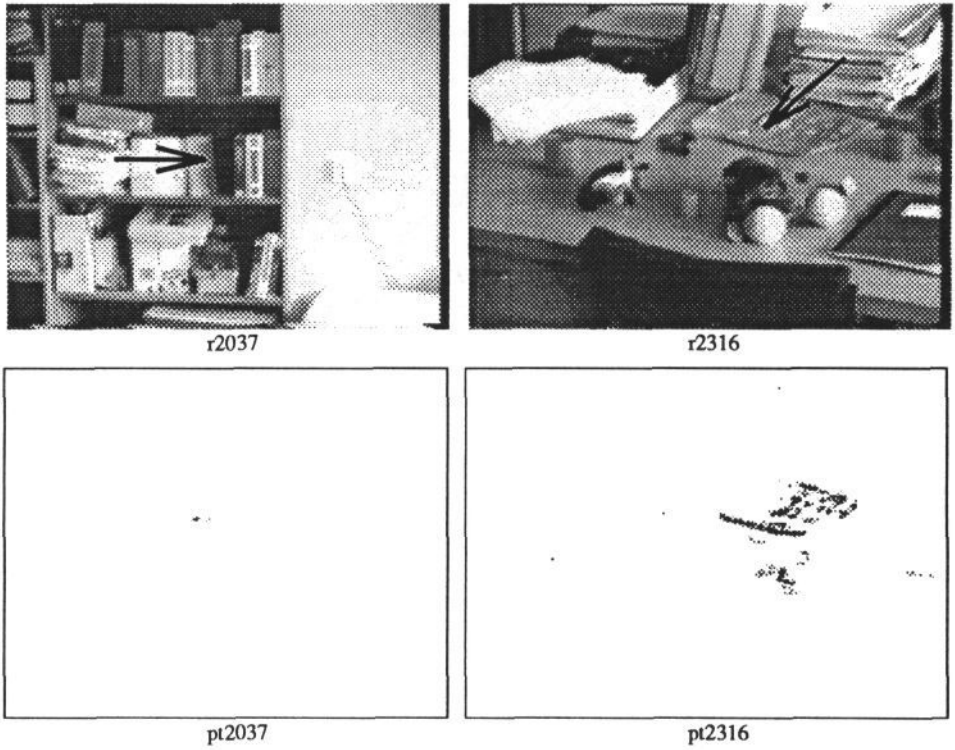
Prob.	model	description	Prob.	model	description
0.812	mc081	red book (Arbib)	0.051	mc151	red wooden block
0.114	mc201	red plastic pot	0.019	mc231	orange block

Pixels corresponding to Minsky's book were split into two regions (the split was caused by blurring of the title). The hypotheses generated for the two regions were almost identical :

P(left region)	P(right region)	model	description
0.977	0.979	mc071	(Minsky's book)
0.023	0.020	mc181	(light green plastic pot)

5 Conclusions

We have presented a new algorithm for generating, verifying and localising of object hypotheses that is based on two sound building blocks: modified Shafer's dichromatic reflection model and Bayesian decision theory. Introduction of colour models greatly simplifies the problem of grouping pixels into regions corresponding to a single object. The algorithm can be easily primed with additional information (coming from previous experience or other visual modules) which makes it useful in the context of complex cooperative or continuously operating vision system as a tool for determining focus of attention, object disambiguation and object tracking. The speed of the algorithm, especially in the single object localisation mode, make it a very attractive component for top down scene interpretation strategies. Our



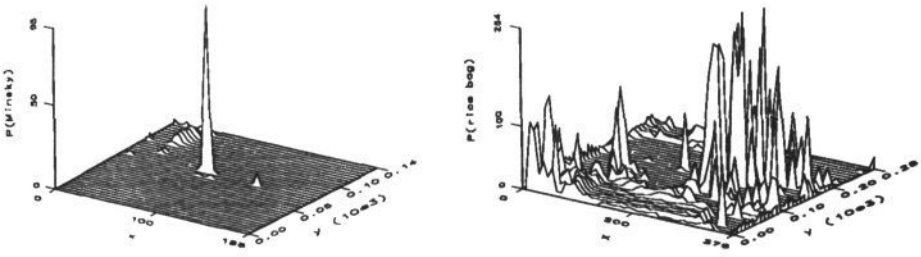


Figure 4: Localization of objects: top - original images; center - pixels with aposteriori probability $P(Model\_of\_interest|RGB)$ greater than 0.5; bottom - $P(Model\_of\_interest|RGB)$ for all image points

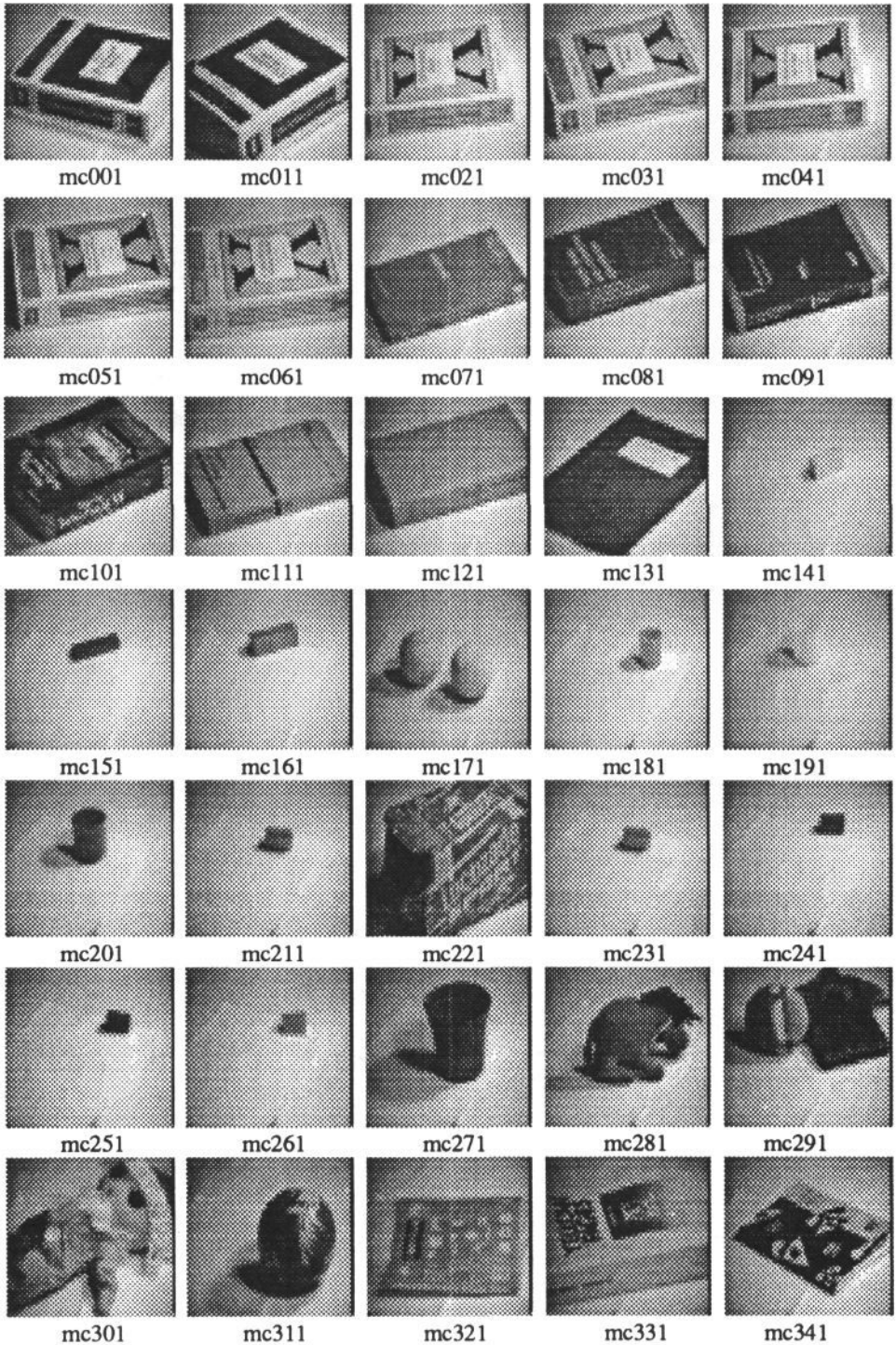


Figure 5: Thirty-five objects used in experiments of section 4.

experiment shows that the method is viable on its own and can provide valuable information about object at various scales even in complex scene.

References

[AM92] N. Asada and T. Matsuyama. Color image analysis by varying camera aperture. In *11th IAPR International Conference on Pattern Recognition (The Hague, The Netherlands, August 30–September 3, 1992)*, volume A, pages 466–469, Washington, DC, 1992. IEEE Computer Society Press.

[Fat92] S.T. Fathima. Data and model-driven selection using color regions. In *Image Understanding Workshop (San Diego, CA, Jan 26-29, 1992)*, pages 705–716, San Mateo, CA, 1992. Defense Advanced Research Projects Agency, Morgan Kaufmann.

[For88] D.A. Forsyth. A novel approach to colour constancy. In *Second International Conference on Computer Vision (Tampa, FL, December 5–8, 1988)*, pages 9–18, Washington, DC,, 1988. Computer Society Press.

[FT80] K.S. Fu and T.S.Yu. *Statistical Pattern Classification Using Contextual Information.* John Wiley, 1980.

[Gar76] T.D. Garvey. An experiment with a system for locating objects in multisensory images. pages 319–325, 1976?

[GJT87] R. Gershon, A.D. Jepson, and J.K. Tsotsos. Highlight identification using chromatic information. In *First International Conference on Computer Vision, (London, England, June 8–11, 1987)*, pages 161–170, Washington, DC., 1987. IEEE Computer Society Press.

[HB87] G. Healey and T.O. Binford. The role and use of color in a general vision system. In IUW1987 [IUW87], pages 599–613.

[Hea89a] G. Healey. A parallel color algorithm for segmenting images of 3-d scenes. In *Image Understanding Workshop (Palo Alto, CA, May 23-26, 1989)*, pages 1038–1042, San Mateo, CA, 1989. Defense Advanced Research Projects Agency, Morgan Kaufmann.

[Hea89b] G. Healey. Using color for geometry-insensitive segmentation. *Journal of the Optical Society of America*, 6:86–103, June 1989.

[IUW87] Defense Advanced Research Projects Agency. *Image Understanding Workshop (Los Angeles, CA, Feb. 23-25, 1987)*, San Mateo, CA, 1987. Morgan Kaufmann.

[KSK87] G.J. Klinker, S.A. Shafer, and T. Kanade. Using a color reflection model to separate highlights from object color. In IUW1987 [IUW87], pages 614–615.

[KSK88] G.J. Klinker, S.A. Shafer, and T. Kanade. Image segmentation and reflection analysis through color. In *Image Understanding Workshop (Cambridge MA, April 6-8, 1988)*, pages 838–853, San Mateo, CA, 1988. Defense Advanced Research Projects Agency, Morgan Kaufmann.

[LB90] S.W. Lee and R. Bajcsy. Detection of specularity using color and multiple views. In *Second European Conference on Computer Vision (Santa Margherita Ligure, Italy, May 19-22, 1992)*, pages 99–114. Springer-Verlag, 1990.

[MB86] L.T. Maloney and B.A.Wandell. Color constancy: a method for recovering surface spectral reflectance. *J. Opt. Soc. Am.*, 3:29–33, January 1986.

Model	*Red*	*Green*	*Blue*	*Object Description*
mc001	0.235	0.075	0.969	Xlib Reference Man. 1 (blue)
mc011	0.421	0.613	0.669	Xlib Prog. Man. 2 (light blue)
mc021	0.654	0.394	0.646	Xlib Prog. Man. 1 (purple)
mc031	0.673	0.333	0.661	Xlib Reference Man. 2 (lilac)
mc041	0.929	0.223	0.296	X Windows System User's Guide 3 (pink)
mc051	0.964	0.260	0.063	X Toolkit Intrinsics Prog. Man. 4 (orange)
mc061	0.884	0.464	0.061	X Toolkit Intrinsics Reference Man. 5 (ocher)
mc071	0.264	0.959	0.107	Minsky and Papert: Perceptrons (green)
mc081	0.999	0.044	0.000	Arbib ed.: Vision, ... (carmin)
mc091	0.164	0.059	0.985	Wang: An Introduction to Berkeley Unix (dark blue)
mc101	0.434	0.086	0.897	Mastering Turbo Pascal 5.5 (violet)
mc111	0.587	0.570	0.575	Andel: Matematicka statistika (light blue)
mc121	0.775	0.569	0.273	Rektorys: Prehled Uzite Matematiky (light brown)
mc131	0.985	0.138	0.107	VSSP technical report (dark red)
mc141	0.805	0.588	0.076	yellow wooden block (approx. 2x2x4cm)
mc151	0.992	0.104	0.066	flat red wooden block (approx. 1x2x6cm)
mc161	0.430	0.877	0.213	green wooden block (aprox. 2x2x6 cm)
mc171	0.678	0.733	0.049	light green Penn tennis balls (well in use)
mc181	0.387	0.908	0.161	light green plastic pot
mc191	0.692	0.720	0.038	yellow plastic pot
mc201	0.994	0.087	0.065	red plastic pot
mc211	0.962	0.076	0.261	pink plastic pot
mc221	0.094	0.415	0.905	Lucozade 6-pack paper cover
mc231	0.990	0.137	0.043	orange block with cross cross-section
mc241	0.135	0.162	0.978	blue cylinder
mc251	0.333	0.174	0.927	purple clock with a star-of-David cross-section
mc261	0.461	0.877	0.138	green block with triangular cross section
mc271	0.955	0.265	0.136	brown cup-like flower pot
mc281	0.593	0.527	0.609	grey stuffed donkey
mc291	0.940	0.290	0.182	brown monkey puppet
mc301	0.515	0.461	0.722	red and light blue doll ('Clown')
mc311	0.891	0.454	0.014	red, blue and yellow rubber ball
mc321	0.970	0.245	0.003	long american grain rice bag (orange)
mc331	0.806	0.592	0.011	yellow pages (Guildford 89), well in use
mc341	0.806	0.547	0.025	UoS telephone directory, December 1992
mc351	0.653	0.535	0.535	white xerox paper

Table 1: Description of objects used for experiments.

[SB90] M.J. Swain and D.H. Ballard. Indexing via color histograms. In *Third International Conference on Computer Vision (Osaka, JAPAN, December 4–7, 1990)*, pages 390–393, Washington, DC,, 1990. IEEE Computer Society Press.

[Sha84] S.A. Shafer. Using color to separate reflection components. Technical report, University of Rochester, 1984.

[Wix90] L.E. Wixson. Real-time qualitative detection of multi-colored objects for object search. In *Image Understanding Workshop (Pittsburgh, Penn., Sep 11-13 1990)*, pages 631–638, San Mateo, CA, 1990. Defense Advanced Research Projects Agency, Morgan Kaufmann.