Object Recognition Using Colour, Shape and Affine Invariant Ratios

Paul A. Walcott
Centre for Information Engineering
City University, London EC1V 0HB, England
P.A.Walcott@city.ac.uk

Abstract

This paper describes a spectral-spatial model (for colour object recognition) which exploits the shape, colour and position of regions on the surface of a rigid object in describing it. Given a model and test image (with colour constancy pre-processing) suitably segmented into colour regions, model and test regions with similar shape and colour are identified. If at least three model regions have matching test regions then the consistency of these matches are verified using distance/area affine invariant ratios. Subsequently, model regions are affine transformed into image space for matching, from which a match probability is determined. Experimental results demonstrate that this model is significantly invariant to illumination changes, affine deformity and partial occlusion.

1 Introduction

In [12] a spectral-spatial model (which was used for colour object recognition) called the colour landmark model (CLM) was presented — which was significant invariant to illumination changes, affine deformity and partial occlusion. This paper extends that model by: improving robustness in shape matching; more adequately describing the object match probability; and introducing affine invariant area/distance ratios to verify matches between model and image regions. To demonstrate the performance of the extended model, a set of test results are presented.

The CLM represents a significant deviation from the "state of the art" spectral-spatial object models [7], [8], and [10] which are based on the colour region adjacency graph (CRAG); however, it bears resemblance (the spectral part of the model) to the spectral-based (colour histogram based) models presented in [3], [9] and [13]. Quite strikingly, the CRAG model is capable of representing non-rigid objects, however the sub-graph matching process — resulting from a search for a model CRAG sub-graph in an image CRAG — is computationally expense. Even more importantly, the CRAG is incapable of representing non-adjacent regions which often result when considering only prominent colours. In none of these CRAG implementations has the problem of colour constancy been seriously addressed, thus ignoring the effects of biased colour values!!

BMVC 1996 doi:10.5244/C.10.64

The CLM was formulated to overcome the before-mentioned problems of the CRAG — although the current implementation is restricted to rigid object modelling — and to provide a match probability measure. It describes the n (generally n>=3; however n=2 is also allowed and treated separately) regions on the surface of a rigid object by selecting n_l of these regions as references and describing the positions of the remaining $n-n_l$ regions from these references. These reference regions are formally known as colour landmarks.

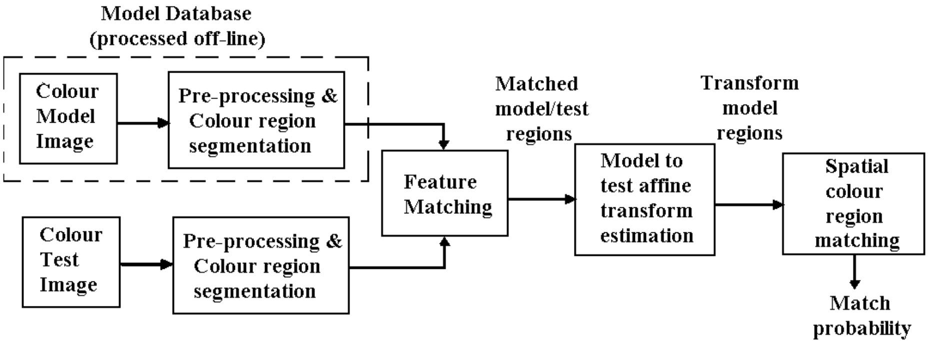

Model Database
(processed off-line)
Colour Model Image
Pre-processing & Colour region segmentation
Colour Test Image
Pre-processing & Colour region segmentation
Feature Matching
Matched model/test regions
Model to test affine transform estimation
Transform model regions
Spatial colour region matching
Match probability

Figure 1. An Overview of the CLM Based Object Recognition System

As illustrated in Figure 1, a colour test image is introduced into the system and is pre-processed by a colour constancy algorithm [5] (CCA) which transforms model and image colours into the same canonical illumination colour space for matching. Since colour constancy is a pre-processing step, any CCA could have been used. The test image is then suitably segmented into coloured regions (using a modification of the colour software filter [11] described in earlier work) and the shape, colour, centroid and area of these regions recorded.

The second stage of the system is concerned with the matching of model and image regions based on their colour, shape and affine invariant distance/area ratios. When the number of model regions is greater than or equal to three, it is necessary for at least three of these regions to be matched with image regions (for the minimum requirement of six points needed for affine transformation parameter estimation). By comparing model and image affine invariant distance ratios (between region centroids) and area ratios, the consistency of these matched regions is determined; this step removes model/image region mismatches. Subsequently, n_l pairs of these matching model/image regions are used as colour landmark. In the special case of two region objects, shape, colour, and the affine invariant area ratio are used to determine model/image matches.

In the final stage, the centroids of the colour landmarks are used in an affine transformation estimation process. These parameters are used to transform the centroids of each non-landmark model region into image space. If there are image regions close to (in a Euclidean sense) these transformed model regions — which are of the same colour — then a match is recorded. The overall number of these matches is used to determine the match probability.

2 The CLM Parameterisation

The CLM is characterised by the parameters <L, R, A, D> where: L $=\{l_1, l_2, \ldots, l_{n_l}\}$ the set of n_l landmark regions; R = $\{r_1, r_2, \ldots, r_{n_r}\}$ the set of n_r non-landmark regions; A, the matrix of area pair ratios; and D, the matrix of distance (between region centroids) ratios. Each region (landmark and non-landmark) is parameterised by: (x_c,y_c) the centroid of the region; I_1^i, an affine invariant moment for different region resolutions; a_r, the absolute area of the region; and C, the region colour where C = $\{(a_1,b_1), (a_2,b_2), \ldots, (a_t,b_t)\}$ where (a_i,b_i) are the set of histogram bin co-ordinates of a cluster in a 2D opponent colour histogram space.

3 Colour Segmentation using the Software Colour Filter

The software colour filter (SCF) [11] is used to segment multiple images into regions of similar colour. It performs this segmentation in two steps; first, it identifies distinct colours in the images to be segmented; then it determines which of these colours are similar. The SCF uses the opponent colour space [1] which transforms RGB into two chroma (rg and by) and an intensity channel (wb):

$$rg = R - G \qquad [1]$$

$$by = -R - G + 2B \qquad [2]$$

$$wb = R + G + B \qquad [3]$$

By discarding the wb and quantising the rg and wb channels (into 16 x 16 bins) and recording the frequency of colours in an RGB image, a 2D opponent colour histogram is generated.

Clusters corresponding to distinct colours are identified in these histograms using a one-pass peak climbing clustering algorithm [6]. The implementation of this algorithm used proceeds by determining, for each histogram bin, whether another histogram bin in its 8-neighbourhood has a greater bin count. If such a bin is found, then the current bin points to it. This process is repeated for all the bins in the colour histogram, after which any bin that does not point to another bin is a parent bin (local maxima). This parent bin and all the other bins that point to it form a cluster in histogram space representing a distinct colour.

Given two images (pre-processed by a CCA[5]), 2D opponent colour histogram are generated, and corresponding model/test histogram clusters identified. Histogram cluster correspondence requires a measure of cluster overlap and closeness of the parent bins of these clusters. The selected measure is a cross correlation, of the model and image colour histogram bin clusters, defined by:

$$\boldsymbol{O(a,b) = \sum_{y}\sum_{x} H_t(x+a, y+b) \times H_m(x,y)} \qquad [4]$$

where H_m and H_t are given by:

$$H_m(x,y) = \begin{cases} 0 & (x,y) \notin C_a \\ H(x,y) & (x,y) \in C_a \end{cases}$$

where H(x,y) is the 2D opponent colour histogram of the model image and C_a is the set of histogram bins corresponding to a distinct colour (cluster C_a) — H_t is created using the opponent colour histogram of the test image; and a, b = 0, ±1, ..., ± d. The resulting bins in the correlated space O(a,b) are clustered (using the above-mentioned clustering algorithm). If a parent bin exist within a chessboard distance d from O(0,0) then these clusters are corresponding clusters. The correlation used is isotropic since the relative position of model/image clusters is difficult to predict.

4 Shape Matching and Verification

In order to identify colour landmarks, the shape of model and image regions of the same colour are compared. It is important that all test regions with minor shape distortions be considered since shape errors due to occlusion, colour segmentation and image noise may have occurred; mismatched regions are removed by comparing model and test affine invariant area and distance ratios. Since the shape descriptor needed to be tolerant of noisy borders, a region-based descriptor (affine invariant moments) was chosen.

Given the discrete form of the (p+q) order moment of a *binary* image function f(x,y), the general moment m_{pq} and the central moment μ_{pq} are defined by:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \qquad [5]$$

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q f(x,y) \qquad [6]$$

where $x_c = m_{10}/m_{00}$ and $y_c = m_{01}/m_{00}$. Further, a second order affine invariant moment I_1 [4] can be defined:

$$I_1 = \frac{(\mu_{20}\mu_{02} - \mu_{11}^2)}{\mu_{00}^4} \qquad [7]$$

I_1 is less sensitive to digitalisation errors, minor shape deformations, camera non-linearity and non-ideal camera positions and is less expensive computationally than higher order moments.

To model digitalisation errors in I_1 due to scale reduction, moment values are calculated for all regions at a number of different resolutions. Given a model region, I_1 is calculated for successively halved resolutions. For a model/image shape match, the calculated value of the moment I_1 for the test region must fall within the range defined by the model. If only a single model I_1 is calculated (due to the region's small area) then this is the minimum allowed value of the image region moment. This simple method of modelling the variation in I_1 was sufficient since it was required only to distinguish between dissimilar region shapes. In the experiments performed, a maximum of three moment values (I_1) were used.

Assume that after feature (shape/colour) matching, model regions m1, m2, m3 and m4 have matching image regions r1, r2, ..., r10:

m_1: r_1, r_2
m_2: r_3, r_4, r_5
m_3: r_6, r_7, r_8, r_9
m_4: r_{10}

Now, by comparing area ratios of model region pairs (m_1m_2, m_1m_3, ..., m_3m_4) with corresponding image pairs (e.g. for model pair m_1m_2, image pairs r_1r_3, r_1r_4, r_1r_5, r_2r_3, r_2r_4 and r_2r_5) the consistency of the image pairs is determined. For any model pair m_im_j (i>j), the consistency of image pair r_cr_d is determined from:

$$\left| A_{ij} - A'_{cd} \right| < T_{area} \qquad [8]$$

given $$A_{ij} = \frac{a_i}{a_j} \quad \text{and} \quad A'_{cd} = \frac{a'_c}{a'_d}$$

where a_i, a_j and a'_c, a'_d are absolute areas of model and image pairs, respectively; and T_{area} is a pre-defined threshold. If it is further assumed that two occurrences of the model exist, $r_1r_3r_6r_{10}$ and $r_2r_5r_9$, in the test image, therefore the resulting match pair lists are:

m_1m_2: r_1r_3, r_2r_5
m_1m_3: r_1r_6, r_2r_9, r_1r_7
m_1m_4: r_1r_{10}
m_2m_3: r_3r_6, r_5r_9
m_2m_4: r_3r_{10}
m_3m_4: r_6r_{10}

where the image pair r_1r_7 is purposely mismatched. It is then required to determine those regions which are related by consistent area ratios. In the above example these are {r_1, r_3, r_6, r_{10}}, {r_1, r_3, r_6, r_7, r_{10}} and {r_2, r_5, r_9}. Distance ratios are used to further determine the consistency of a given set of regions; for three regions r, r’ and r’’ (which have been matched with model regions, say m_1, m_2 and m_3 respectively), three (however, only two are independent) affine invariant distance ratios between centroids of these regions can be calculated:

$$\frac{d_{rr'}}{d_{r'r''}}, \quad \frac{d_{rr''}}{d_{r'r''}} \quad \text{and} \quad \frac{d_{rr'}}{d_{rr''}} \qquad [9]$$

where $d_{rr'}$ is the Euclidean distance between the centroid of region r and r’ (similarly for $d_{rr''}$ and $d_{r'r''}$).

The consistency of the three image regions can be determined by comparing with corresponding model distance ratios using a root mean square measure:

$$\sqrt{\frac{1}{n}\sum_i (a_i - b_i)^2} < T_{dist} \qquad [10]$$

where a_i is the list of n model distance ratios and b_i the corresponding list of n image distance ratios; and T_{dist} a pre-defined threshold. For three regions, n=2 for independent ratios. By selecting groups of three regions from a given list, the consistency of these region is determined. Subsequently, a minimum of three consistent regions (which will be used as colour landmarks) are identified. Affine parameters are estimated from these three pairs of matched model/image regions and a match probability is calculated (Section 5). In the experiments presented three regions were used for colour landmarks (n_l = 3) since the models had a relatively small number of coloured regions.

5 Model Region Transformation and Match Probability

Given the centroids of the n_l landmark regions (X_i', Y_i') and the corresponding image regions (X_i, Y_i), affine transformation parameters a, b are estimated using [11] if n_l =3 or [12] if n_l>3:

$$X^T = [X_1, X_2, \ldots, X_{n_l}],\ Y^T = [Y_1, Y_2, \ldots, Y_{n_l}],$$
$$a^T = [a_0, a_1, a_2],\ \ b^T = [b_0, b_1, b_2] \text{ and}$$

$$A = \begin{bmatrix} 1 & X_1^{'} & Y_1^{'} \\ 1 & X_2^{'} & Y_2^{'} \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & X_{n_l}^{'} & Y_{n_l}^{'} \end{bmatrix}$$

$$a = A^{-1}X\ ,\ b = A^{-1}Y \qquad [11]$$
$$a = (A^T A)^{-1}(A^T X)\ ,\ b = (A^T A)^{-1}(A^T Y) \qquad [12]$$

The centroid (x',y') of each non-landmark model region is affine transformed into image space where the new centroid (x,y) is given by x = **Aa** and y = **Ab**, where **A** = [1 x' y']. Now, if any image region with centroid (x_r,y_r) satisfies:

$$\sqrt{(x - x_r)^2 + (y - y_r)^2} < T_d \qquad [13]$$

(where T_d is a pre-defined threshold) and the two regions are the same colour then a match is recorded. The total number of transformed region matches plus the number of landmarks is the overall number of region matches n'. The probability for n' matches is approximated by the function (suitably normalised):

$$\boldsymbol{P(n', n, k) \approx 1 - \frac{1}{\left(\frac{k \cdot n'}{n}\right)^{3} + 1}} \qquad [14]$$

where n is the total number of model regions and k is a constant (k=3 is used in these experiments). Clearly however, any function with similar characteristics could have been used.

6 Implementation and Results

The database illustrated in Figure 2 was used in the experiments described. CLMs were generated for each model image by creating a 16x16 2D opponent colour histogram and applying the clustering algorithm described in Section 3 to identify distinct histogram clusters. For each identified cluster a binary image was generated by filtering all pixels whose colour did not fall within the histogram cluster and copying all remaining pixels to the binary image (which was the same size as the model image).

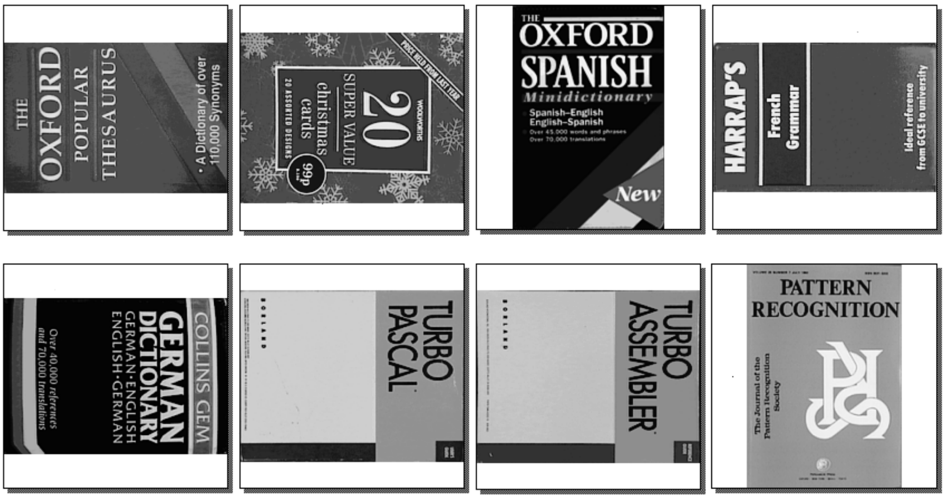
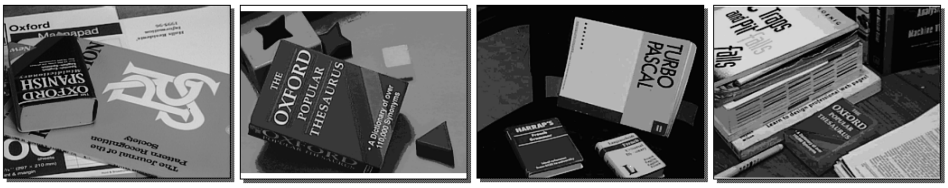

THE OXFORD POPULAR THESAURUS
• A Dictionary of over 110,000 Synonyms
20 SUPER VALUE christmas cards
99p
THE OXFORD SPANISH Minidictionary
Spanish–English English–Spanish
New
HARRAP'S French Grammar
Ideal reference from GCSE to university
COLLINS GEM GERMAN DICTIONARY
GERMAN·ENGLISH ENGLISH·GERMAN
Over 40,000 references and 70,000 translations
TURBO PASCAL
TURBO ASSEMBLER
PATTERN RECOGNITION
The Journal of the Pattern Recognition Society

Figure 2. A reduced model database of coloured objects.

The boundaries of the image regions in these binary images were located using a boundary follow algorithm, each region boundary filled (all of the pixels within the boundary were turned on) and the area, centroid, and moment I_1^i calculated for a maximum of three successively halved region resolutions. These parameters, as well as the histogram bin co-ordinates and values, for each histogram cluster, were written to the landmark files and stored in the model database.

Given a test image, its CLM was generated and compared with each model in the database. For a given comparison, model and image regions of similar colour were identified using correlation matching (the SCF in Section 3). Affine invariant moments were used to determine matching model and test regions. The number of mismatched regions were reduced using affine invariant area and distance ratios (Section 4.1). For each set of hypothesis model regions, affine transformation parameters were determined and the regions transformed. The match probability was then calculated.



(a) (b) (c) (d)

Figure 3. A data set of test images.

The test images illustrated in Figure 3 were used to test the CLM. These images were captured under random illumination and illustrate problems of reduced scale, affine deformity, shadows, and partial occlusion.

Region	Colour	x	y	Area	I_1^0	I_1^1	I_1^2
0	0	209	219	1.6e+05	6.0e-13	6.4e-08	6.7e-05
1	0	542	81	1.7e+04	1.4e-08		
2	1	495	153	2.2e+04	6.3e-09	2.2e-04	
3	3	445	196	3.0e+04	1.1e-09	4.9e-06	

Table 1. The region parameters for the Oxford thesarus model.

The region parameters calculated for the Oxford thesarus model image and the test image in Figure 3(b) show the values of the affine invariant moment of the original region resolution (I_1^0) and succesively halved resoltions (I_1^1, I_1^2). Table 1 illustrates the Oxford thesarus model region parameters: colour number (representing a colour histogram bin cluster), the centroid (x,y) of the region, its area and as previously describing the variation of I_1 with reduced scale. The blank entries in the table represent region resolutions below a threshold of about 1% of the total image area, therefore I_1 was not calculated.

The region parameters calculated for the test image illustrated in Figure 3(b) are illustrated in Table 2. The corresponding model and image regions are model region numbers 0, 1, 2 and 3 and image region numbers 1, 2, 5 and 9, respectively. The moment value for image region 0 (7.3e-11) is bounded by the calculated model moment values (6.0e-13 and 6.7e-05); and single moment value (1.4e-08), calculated for model region 1, represents the minimum allowed moment value for image region 2 (3.8e-06).

Region	Colour	x	y	Area	I_1
0	0	324	230	2.1e+05	6.0e-13
1	1	182	242	5.9e+04	7.3e-11
2	1	412	219	3.9e+03	3.8e-06
3	1	525	340	3.9e+03	5.7e-06
4	2	88	70	1.8e+04	3.8e-09
5	2	366	260	7.5e+03	3.7e-07
6	3	90	56	3.7e+03	2.8e-06
7	3	250	71	2.6e+03	6.5e-04
8	3	333	294	3.9e+03	1.9e-05
9	4	328	271	9.1e+03	1.6e-07

Table 2. Region parameters for the test image illustrated in Figure 3(b).

The results of matching the database against the test image set is illustrated in Figure 4. For test image 3(a), the model yielding the largest match probability was the Pattern Recognition Journal, 0.97. All models tested against test image 3(b) yield a match probability of 0.0 except the Oxford Spanish dictionary (0.87) and the Oxford thesarus (1.0). For 3(c) the Turbo Paascal model was matched correctly (0.92), however the Harrap's model was not detected because only two prominent colour regions exist at reduced resolution. If not occluded, as in this case, the single area ratio is consistent. Finally the Oxford thesarus was again recognised correctly (1.0) with all other model probabilities of 0.0 except the Oxford Spanish dictionary (0.71).

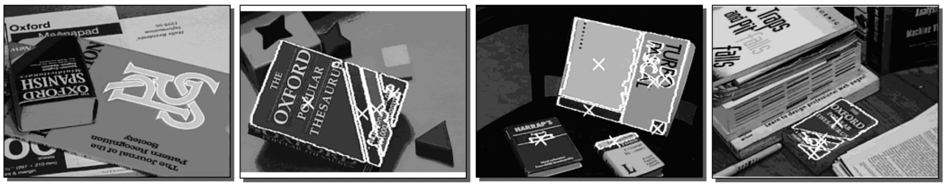


Figure 4. The localised models for the test image set illustrated in Figure 3.

7 Conclusion

A spectral-spatial model for colour object recognition has been presented, which describes rigid objects. It has been shown through experiments that this model is viable and provides solutions to some of the common problems of the CRAGs, namely computational expense due to sub-graph matching and the modelling of non-adjacent regions. Since the affine invariant moments used in the CLM are some what expense computationally, alternative affine invariant descriptors will be examined, specifically the contour descriptor in [2]. Also, a CRAG will be added to the CLM so that it will be able to exploit the benefits of region adjacency and non-rigidity while maintaining the features of the CLM.

References

[1] Ballard, D, Brown, C, Computer Vision, Prentice Hall, 1982.

[2] Cyganski, D, Vaz, R, “A Linear Signal Decomposition Approach to Affine Invariant Contour Identification”, Pattern Recognition, **28**, 12, 1995, pp 1845-1853.

[3] Finlayson, G, “Colour Object Recognition”, Master’s thesis, Simon Fraser University, 1992.

[4] Flusser, J, Suk, T, “Pattern Recognition By Affine Moment Invariants”, Pattern Recognition, **26**, 1, 1993, pp 167-174.

[5] Hung, T, W, R, Ellis, T, “Spectral Adaptation with Uncertainty using Matching”, IEE Proc. of the Fifth Int. Conf. on Image Processing and its applications, Scotland, 1995, pp 786-790.

[6] Khotanzad, A, Bouarfa, A, “Image Segmentation By a Parallel, Non-Parametric Histogram Based Clustering Algorithm”, Pattern Recognition, **23**, 9, 1990, pp. 961-973.

[7] Matas, J, Marik, R, Kittler, J, “On Representation and Matching of Multi-coloured Objects”, Proc. of ICCV, Boston, 1995.

[8] Olatunbosun, S, Dowling G, Ellis, T, “Topological Representation For Matching Coloured Surfaces”, ICIP 96, Switzerland, September 1996.

[9] Swain, M, “Colour Indexing”, Ph. D. thesis, University of Rochester, 1990.

[10] Syeda-Mahmood, T, “Data and Model-Driven Selection Using Color Regions”, Image and Understanding Workshop 92, Morgan Kaufmann, 1992, pp 705-716.

[11] Walcott, P, Ellis, T, “The Localisation of Objects in Real World Scenes Using Colour”, proc. of the 2nd ACCV conf., Singapore, December 1995, pp 243-247.

[12] Walcott, P, Ellis, T, “Modelling Colour Surfaces Using Colour Landmarks”, proc. of the ninth IMDSP conf., Belize, March 1996, pp 100-101.

[13] Wixson, L, Ballard, D, “Real-time Detection of Multi-coloured Objects”, SPIE Sensor Fusion II: Human and Machine Strategies, **1198**, November 1989, pp 435-446.