ACCV2002: The 5th Asian Conference on Computer Vision, 23–25 January 2002,Melbourne, Australia

Evaluating Colour-Based Object Recognition Algorithms Using the SOIL-47 Database

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

[1]Centre for Vision Speech and Signal Processing
University of Surrey, Guildford, GU2 7XH , UK
{d.koubaroulis,j.kittler}@ee.surrey.ac.uk

[2]CMP, CTU Prague
Karlovo nám. 13, 121 35 CZ
matas@cmp.felk.cvut.cz

Abstract

In this paper, a new image set, called the Surrey Object Image Library (SOIL-47) is introduced, on which the performance of two colour-based object recognition methods is evaluated. The data was collected specifically for testing colour-based recognition algorithm and is publicly available.

In the conducted experiments on SOIL-47, we evaluate two recognition algoritms; the Multimodal Neighbourhood Signature (MNS) approach and a method based on a Attributed Relational Graph (ARG). The MNS approach represents object appearance by measurements computed from image neighbourhoods with a multimodal colour density function. The ARG approach computes a graph of affine invariant measurements of the colour and shape of segmented image regions.

Using only a single model image of each of the 47 objects, MNS performed well even for extreme test views close to ± 90 *degrees. The ARG method assumes a locally planar surface, therefore a second experiment was conducted on a subset of box-like objects of SOIL-47. MNS performance was fairly stable, outperforming ARG for most viewing angles.. Note, that this is the first systematic test of MNS with controlled 3D viewpoint change.*

1 Introduction

Reproducible and meaningful comparison of colour-based object recognition and image retrieval algorithms is not simple. The publicly available set of colour images collected by Swain and Ballard and the related testing protocol adopted in their seminal paper [12] is a popular benchmark on which a number of methods have been tested e.g. [12, 7, 9]. However, that dataset has some limitations. Firstly, only two views of each object are available. Secondly, the transformation between the two views is very close to rotation, so only invariance to Euclidean transformation is needed for successful recognition. Swain's recognition task has become fairly simple by current standards, so it is not very discriminative - most methods achieve almost perfect results.

Another dataset of colour images is a publicly available database collected at Simon Fraser University which has been used in a number of colour recognition experiments [6, 4]. The database is designed for studying colour constancy, i.e. the effects of change in spectral power distribution on recognition performance. Twenty objects are viewed from a single viewpoint under 11 different illuminants. 3D variability in appearance is present in another publicly available database, the so called COIL-100, collected at Columbia University [11] which was the testbed for several methods (e.g. [10]). Most experiments using this database assume multi-view object recognition i.e. more than one images available to learn object appearance. The methods tested often compute shape and/or grey-scale object descriptors which are out of the scope of this paper. In addition, many objects in the database are single-coloured and with the same or very similar colour. This poses difficulties to testing recognition algorithms that are designed to recognise multi-coloured objects. Finally, very few objects in COIL-100 consist of planar surfaces which is a fundamental requirement of algorithms like [2].

It is often important to understand the behaviour of recognition methods when full projective or affine effects are present, in particular when only a single view is available to learn the appearance of a 3D object. With this objective in mind, a database of coloured household objects, many of the same shape, viewed over a significant portion of the viewing sphere was collected at the University of Surrey. The images show mainly multicoloured objects, many of which consisting of planar surfaces (boxes) and with generally complex colour structure. The so called SOIL-47 (Surrey Object Image Library) database contains 987 images. Altogether, 21 images of 47 objects were acquired by a robot arm moving around the object at intervals of approximately 9 degrees spanning a range up to ± 90 degrees. Appearance variations are mainly due to 3D viewpoint change and self-occlusion. Another set of 987 images was acquired under different geometry pf light sources, to test algorithms claiming invariance to changing illumination intensity. In order to separate between the two sets, we distinguish SOIL-47A and SOIL-47B, where A and B correspond to the two different lighting settings. Some objects of the database are shown in Fig. 4. All views of the 47 objects can be found on-line. [1].

In the experimental section, we evaluate the performance of the MNS method, introduced by Matas et al. in [9], on images from SOIL-47A. We assume that only a single image is available for learning the object representation. In a first experiment, MNS is shown to perform well for a fairly wide range of viewpoints. Reasonable recognition rate is achieved even for objects whose views differ significantly from the frontal view. MNS performance is next compared with a relational attributed graph-based approach, proposed by Ahmadyfard and Kittler [2]. In a repeated experiment, MNS outperformed ARG matching.

An outline of both MNS and ARG is given in the next section. Details about the experimental setup can be found in sections 3 and 5. Section 6 discusses the results obtained and section 7 concludes the paper.

2 The MNS and ARG methods

The Multimodal Neighbourhood Signature (MNS) is a colour-based object recognition and retrieval method for multi-coloured objects. *Local* colour structure is represented by illumination invariant features computed from image neighbourhoods with a multimodal colour density function. The positions of the modes used for the computation of the invariants are robustly filtered, stable values, efficiently established in the RGB colour space with the mean shift algorithm [5]. Each MNS signature consists of a number of invariants and representative locations allowing for localisation of the object in another image. MNS matching is posed as an assignment problem, i.e. a problem of uniquely associating each model feature (pair of modes) to a test feature. In the literature, MNS matching is implemented a model-oriented stable matching problem [8]. A stable match is established between a model and the *closest* (in terms of the specified distance function) test feature to it that is not closer to any other model feature and within a maximum allowed distance. A final score is computed from all matched features and the models are ranked by score (for details see [9]).

In the Attributed Relational Graph (ARG) method, the authors proposed a representation of colour structure based on a relational attributed graph (ARG), built from selected regions of a segmented image. Affine invariants, based on the pair-wise properties of segmented regions, like their colour and normalised relative area, are computed and stored in a graph node. Correspondence is established by comparing both graph node information and supporting evidence from adjacent regions, using a modified probabilistic relaxation method [3]. From all matched regions, the probability of the test image showing a specific model is computed. The model objects are ranked according to the number of matched regions. Advantages of the method include effective use of region shape and topological properties to infer correspondence of regions and finally to establish the transformation between the test and the corresponding model object. In addition, the method can recognise objects also in intensity images (for details see [2]).

3 The SOIL-47 Database

The SOIL-47 database comprises 47 household objects imaged on black background using a 2/3-inch CCD JVC TK-1070 camera mounted on a robot arm. Database capturing has been carried out for two illumination conditions, giving the SOIL-47A and the SOIL-47B data sets respectively. The illumination settings during database acquisition were as follows. Two 1500 Watt tungsten bulbs were used. To make the lighting more diffuse, a white paper shield was placed between the light(s) and the object in the scene. The approximate distance between the light(s) and the object was approximately 2.5 meters and the distance between the lights and the shield was about 0.4 meters. For the acquisition of SOIL-47A, one light was used, placed in front of the object. For SOIL-47B, two light sources were used in approximate distance of 0.75 meters from each other. In the latter case, the line connecting the lights was perpendicular to the optical axis.

4 The Database Protocol

The frontal views of all 47 objects of SOIL-47 comprise the model database. The test images depict the same objects imaged systematically using a robot arm rotated around the vertical axis up to ± 90 degrees in intervals of approximately 9 degrees. This results in 20 test images for every object. The test images corresponding to one of the objects are shown in Fig. 1. The grabbed images are in portable network graphics (PNG) format and their size is 576×720. Due to storage space limitations, a scaled version of the original images is available on the WWW [1] with size 288×360. In addition, frontal views of the objects in the original size (576×720) were included in the on-line version of the database to allow evaluation of algorithms claiming scale invariance. No image processing was applied during or after capturing of the database. For the scaling the unix utility `pnmscale` was used. The full database is publicly available from the CVSSP lab upon request.

5 The Testing Protocol

5.1 Experiment 1: 47 model objects

For the reported experiments, illumination is assumed to be constant, therefore only SOIL-47A is used. A single image of each object, its frontal view (view 0), is inserted in the model database. The resolution of the model images is 576×720. All other images (views 1 to 20) are used as test images (resolution 288×360).

For each viewing angle considered, performance is calculated after matching all test objects viewed under that angle to the model database. For each test image, the model objects are ranked by the similarity value computed during matching. We are interested in the rank (position) of the corresponding (correct) model object in the sorted list of model objects. Finally, recognition performance for this test angle is measured as the percentage of rank 1 recognitions.

The thresholds involved with MNS signature computation and matching are identical to those used in other published experiments and correspond to default settings of the MNS algorithm.

5.2 Experiment 2: 24 model objects

Using the experimental setup described above, we repeated the experiment using a preselected set of 24 SOIL-47A objects, for which reported results by the ARG method were provided by the authors of [2]. In their experiment, a subset of the original database was selected including objects with planar surfaces, to avoid inaccuracies of image segmentation due to shading of curved objects. The selected objects (shown in Fig. 4) correspond to database objects from 1 to 19 and to objects from 24 to 29 except 26. For efficiency, in the ARG experiment, the test images were scaled down to 231 $\times$ 288 and the model images to 461 $\times$ 576. Although the scale change between test and models was preserved in the scaled images, for consistency we used the image sizes described in Experiment 1.

6 Results

The results obtained from Experiment 1 and 2 are listed in Table 1 and graphically presented in Fig. 2(a)-(b). In Experiment 1, average correct (rank 1) recognition rate was 55.3% for viewing angles in ± 20 degrees and 51.6% for angles in the range ± 60 degrees. The slightly lower average for very small angles was probably due to the fact that the near frontal views of some complex objects produced a large number of similar colour pairs which confused the matching algorithm. Note, that approximately 56% of the objects were correctly recognised even when viewed by a viewpoint different by 90 degrees than the frontal view (Fig. 3). Adding rank 2 and rank 3 percentages, the above figures become 78.7% for viewing angles in ± 20 degrees and 78.7% for ± 60 degrees. Moreover, 75% of the objects were recognised in the first 3 ranks, viewed from 90 degrees different angle. The same figure for a viewing angle close to the frontal view (-27 degrees) was 96%. In both cases, the averages denote stable (robust) performance for most viewing angles considered.

In the last two columns, comparative results from Experiment 2 are presented. A graphic display of the result can be found in 2(c)-(d). The performance of the MNS method was fairly stable over a wide range of angles, achieving an average correct recognition rate of 75% for a viewing angle of ± 20 degrees and 71% for a angles in the range of ± 60 degrees. The corresponding figures for ARG were 87.5% and 67.8%. Although MNS performance is slightly lower than ARG for viewing angles close to the frontal view, it outperforms ARG for most viewing angles tested. The difference in performance is explained since many objects' frontal views are very similarly coloured and the spatial arrangement of regions is important for the discrimination. Nonetheless, objects with similar colours were almost always in the first 3 ranks. The percentage of objects recognised in the first 3 ranks was 90.6% for a viewing angle of ± 20 degrees and 88.2% for angles in the range of ± 60 degrees.

Amongst the objects that were not recognised with first rank by MNS are objects with almost identical colour appearance like those shown in Fig. 3. These images differ slightly only in the image areas occupied by each colour. MNS does not use region shape or area properties. This, on the other hand, is advantageous in many situations where region area and/or shape is not preserved. For example, compare the colour regions in Fig. 3(a) with Fig. 3(b) showing the same object rotated by 90 degrees. In principle, MNS is not as discriminative as methods that use the shape and/or topology of regions, but it is very stable (robust) with respect to viewpoint change. A possible extension of MNS to incorporate spatial information is a priority in our future work.

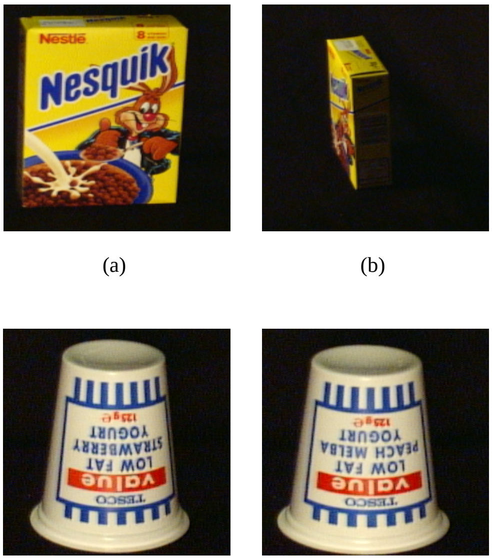
7 Conclusions

In this paper, a new image dataset, the Surrey Object Image Library (SOIL-47), was introduced as a testbed for colour-based object recognition algorithms. Images of 47 household objects acquired by a camera mounted on a robot arm from a wide range of angles (in the range $\pm$ 90 de-

Table 1. Recognition results for MNS and ARG

Viewing angle (approx. degrees)	% rank 1 recognition rate		
	MNS Experiment 1	MNS Experiment 2	ARG Experiment 2
-90	36.2	41.7	4.5
-81	51.0	50.0	0.0
-72	48.9	78.3	0.0
-63	55.3	75.0	0.0
-54	59.6	79.2	18.2
-45	57.5	79.2	54.5
-36	42.6	79.2	81.8
-27	55.3	83.3	77.3
-18	70.2	87.5	81.8
-9	57.4	66.7	90.9
9	38.3	78.3	90.9
18	44.7	66.7	86.4
27	40.4	66.7	63.6
36	44.7	66.7	63.6
45	57.4	75.0	50.0
54	55.3	79.2	54.5
63	61.7	75.0	13.6
72	53.2	83.3	9.1
81	63.8	79.2	4.5
90	55.3	45.8	0.0
Total Average (rank 1)	52.5	71.0	42.3
Average ±60 deg.	51.6	71.0	67.8
Average ±20 deg.	55.3	74.6	87.5
Total Average (ranks 1-3)	78.0	85.6	–
Average ±60 deg.	78.7	88.2	–
Average ±20 deg.	78.7	90.6	–

Figure 1. All views of one object from SOIL-47A viewed by a robot arm from various angles between -90 and 90 degrees



(a) (b)

(c) (d)

Figure 3. (a)-(b) Appearance variation of a Corn Flakes box due to a 90 degree rotation (c)-(d) Two objects with identical colour structure

grees), under two different illumination geometries. The database is publicly available. The images exhibit full 3D appearance variability due to 3D viewpoint and illumination change and significant resolution change between the test and model images.

In the conducted experiments, the performance of the Multimodal Neighbourhood Signature (MNS) method and the Attributed Relational Graph (ARG) approach was evaluated and compared, using a single view to learn object appearance. Performance was measured as the percentage of correct recognitions per viewing angle. Using all 47 model objects in the model database, 80% of all the objects were recognised in the first 3 ranks for test views up to $\pm$ 90 degrees. To satisfy ARG's requirement of a locally planar surface, a second experiment was conducted, using a subset of 24 box-like objects. MNS performance was stable, higher than ARG for most viewing angles. Note, that this was the first systematic evaluation of MNS under controlled 3D viewpoint change.

Acknowledgements

This authors were supported by the EC IST-13082 ASSAVID project, the I. Latsis Foundation and the Czech Ministry of Education grant VS96049. We are grateful to A. Ahmadyfard from CVSSP for providing the ARG results for the recognition experiment, and to J. Burianek and A. Ahmadyfard for capturing the database images.

References

[1] http://www.ee.surrey.ac.uk/EE/VSSP/demos/colour/-soil47/.

[2] A. Ahmadyfard and J. Kittler. Region-based Object Recognition: Pruning Multiple Hypotheses and Representations. In M. Mirmehdi and B. Tomas, editors, *Proceedings of the British Machine vision Conference*, pages 745–754. BMVA, September 2000.

[3] W. Christmas, J. Kittler, and M. Petrou. Structural Matching in Computer Vision Using Probabilistic Rrelaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):749–764, 1995.

[4] D.Finlayson, Bernt Schiele, and James L. Crowley. Comprehensive Colour Image Normalization . In *ECCV'98*, pages 475–descriptions490, 1998.

[5] 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.

[6] 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.

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

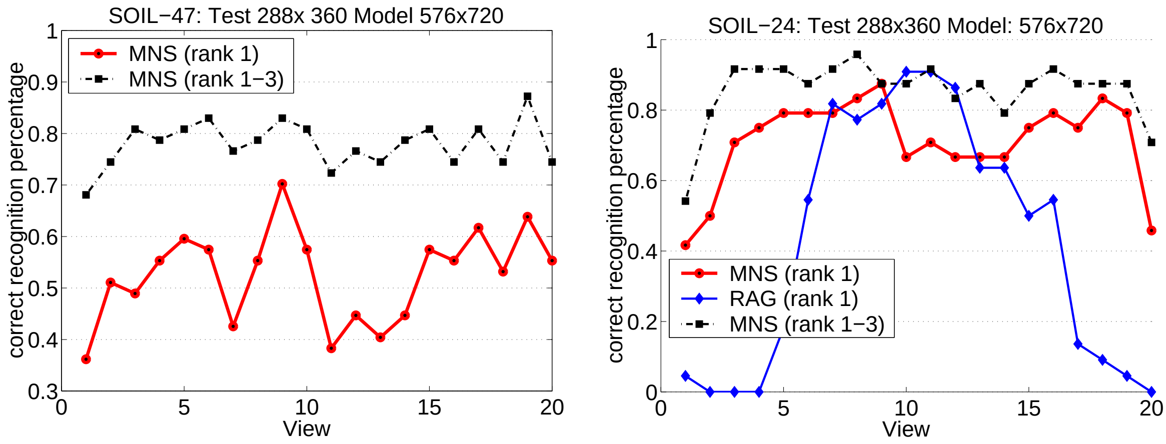
[8] D. Gusfield. *The Stable Marriage Problem:Structure and Algorithms*. MIT Press, 1989.

[9] J. Matas, D. Koubaroulis, and J. Kittler. Colour Image Retrieval and Object Recognition Using the Multimodal Neighbourhood Signature. In *Proceedings of the ECCV*, LNCS vol. 1842, pages 48–64, Berlin, Germany, June 2000. Springer.

[10] S. K. Nayar, S. A. Nene, and H. Murase. Real-Time 100 Object Recognition System. Technical Report CUCS-019-95, Columbia University, 1994.

[11] S. A. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-100). Technical Report CUCS-006-96, Columbia University, 1996.

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



(a) (b)

Figure 2. Recognition performance for (a)-(b) Experiment 1 (c)-(d) Experiment 2

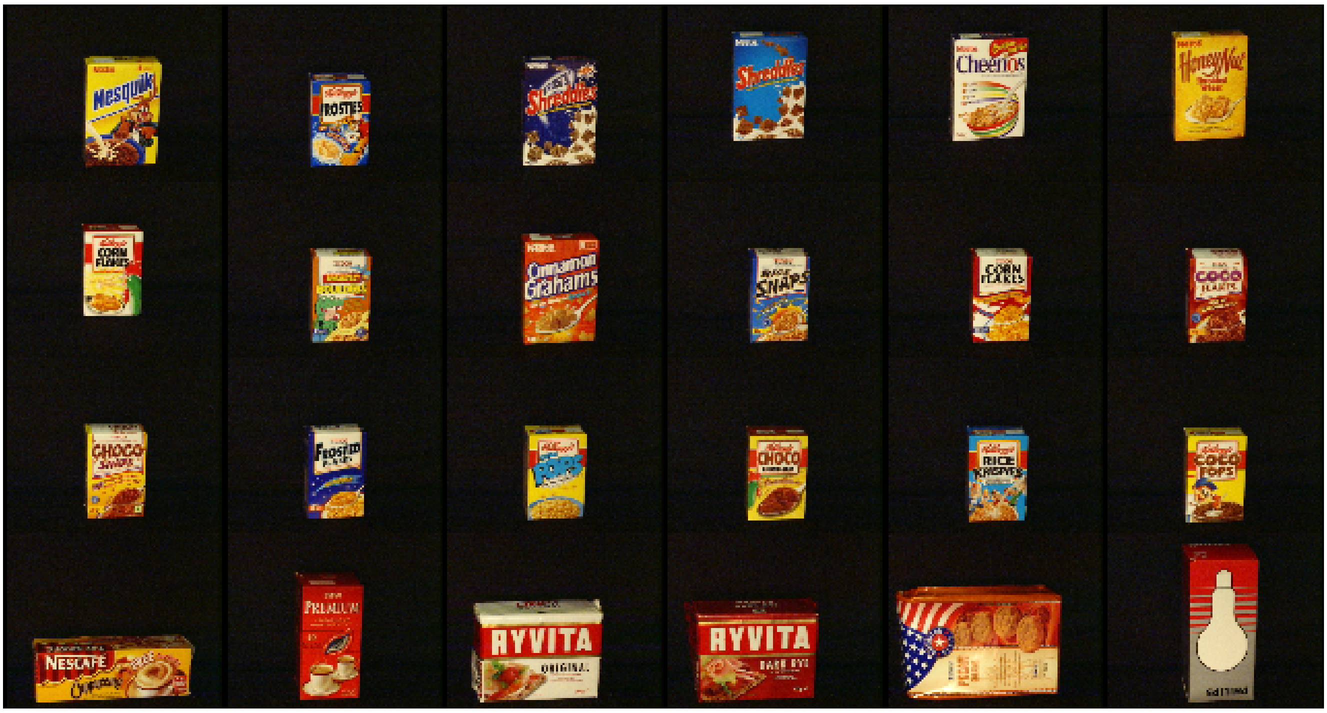


Figure 4. 24 selected images of boxes from the SOIL-47A database