



# Chapter 11

## Conclusions

### 11.1 Summary

In the first chapter of the thesis, object recognition was defined in a broad sense as the task of determining the presence of objects or object categories in images. Motivated by a wide range of applications, a method was sought for representing the (possibly complex) colour structure of multicoloured objects. In an example-based framework, the object representation had to be derived from one or more image (or region) examples. The approach should be general enough to be useful to applications such as image retrieval, object recognition and video annotation.

In Chapter 2, a number of colour-based methods described in the literature were reviewed and categorised depending on assumptions of appearance and the image formation process. To judge the performance of different methods in the presence of changing imaging conditions, an overview of the main factors influencing image formation, e.g. the viewing geometry, the characteristics of the acquisition device and the illumination, was given. In a critical review, region-based (as opposed to full-image based) methods were argued to be more accurate in recognising objects in images with cluttered background. Finally, the main ideas underpinning our approach were outlined and compared to related previous research.

A detailed description of the proposed Multimodal Neighbourhood Signature (MNS) approach was given in Chapter 3. An algorithm for computing and matching MNS signatures was implemented. Object colour structure was represented by stable measurements, computed from

robustly filtered values, the modes of a local colour density function, estimated from sample colour values of image neighbourhoods. A subset of all image regions, compact neighbourhoods with a multimodal density function, were considered for image processing. A number of illumination invariants was possible to compute from pairs of modes of the density function. After measurement computation, identical measurements were suppressed and a representative was selected. Note, that the actual density of measurements was not exploited and neither was their spatial arrangement. Finally, a signature matching algorithm was introduced, posing signature comparison as a stable matching problem.

Below, the main advantages of the proposed MNS method are listed:

- Our recognition system, based on local appearance modelling and partial matching can cope with many appearance variations present in a variety of applications.
- The method was extensively tested in its default configuration and shown to perform well for object recognition, image retrieval and video annotation.
- The method is example-based; automatic computation of the object representation is desirable in many applications. Even a single image (region) example can be used.
- The selection of the internal parameters of the MNS algorithm was shown to be non-critical. Good performance was achieved for many applications/data using the same default settings.
- The method was experimentally shown to be robust to significant scale and 3D viewpoint change, as well as partial occlusion.
- The algorithm is computationally simple and has low storage requirements.
- Reasonable speed was reported for signature computation, matching and object localisation.
- Changing image resolution and/or the acquisition device (e.g. camera, scanner, DTP application) had only a marginal effect on performance.

The limitations of the implemented MNS method are discussed in Section 11.3 and possible extensions are proposed.

---

In Chapter 4, both the performance and the efficiency of the proposed MNS method were evaluated on a well known baseline colour object recognition experiment first performed by M. Swain. In that experiment, the MNS performance was almost perfect, outperforming a number of other methods and comparing favourably to algorithms that exploit colour region area and structure. The comparative study shown in Table 4.1 is –to our knowledge– comprehensive, and is interesting in its own right. From that study, we concluded that MNS has good potential for 3D object recognition. Furthermore, an efficiency evaluation study highlighted attractive properties of MNS such as fast signature computation and matching as well as low storage requirements. Finally, the suitability of the proposed method for illumination invariant object recognition was tested. Using a selected set of MNS parameters, the results obtained were comparable to those previously published in the literature.

In the comparative experiment of Chapter 4, most methods achieved high performance on Swain’s data, probably due to the limited appearance variations of the imaged objects. For this reason, another experiment was repeated and described in Chapter 5. For that experiment, a novel database of 47 objects (called SOIL-47) was introduced, which was collected at the University of Surrey. Our method was compared with a graph-based approach which exploits the spatial arrangement of segmented colour regions. As expected, the graph-based method provided better discrimination for viewpoints close to the frontal view, where the size of similarly coloured regions of different objects was approximately equal, but its performance degraded dramatically with changing viewpoint. On the contrary, MNS performance was generally stable over all viewpoints tested, even for extreme views, very different than the single frontal view example inserted in the database.

The suitability of the MNS method for image retrieval was investigated in three experiments described in Chapters 6, 7 and 8. The data used in each experiment demonstrated different aspects of appearance variation present in typical applications such as trademark catalog searching, colour advertisement retrieval and finally object-based retrieval of video frames. For the last experiment, a database of 1300 images grabbed from various TV broadcast sequences was introduced. Using a single example image (region) to represent the sought object, MNS was compared to two other retrieval methods and shown to compare favourably with published results. In general, good performance was achieved by MNS for image retrieval. A large number of retrieval examples can be found in the appendix.

In some applications, the location at which an object appears in the image, is important. In general applications, accurate localisation and pose estimation may be hard to achieve, therefore an algorithm for approximate localisation was described in Chapter 9. In the conducted experiments, the algorithm was shown to perform well at localising compact objects viewed from changing viewpoint in images with significant background clutter. In the same experiments, MNS outperformed a well known method called histogram backprojection.

Another application of MNS, described in Chapter 10, is video annotation. An object-based approach was taken; the image labels were assigned depending on the presence of specific objects in the images. Our experiments were conducted on sport video sequences for which ground truth, in the form of sport labels, was available. Unlike previous experiments, object appearance was learnt from more than one example regions and a training set comprised of example images showing each object of interest. A novel matching algorithm was proposed, based on measurements computed through training. In addition, another localisation algorithm was designed to localise possibly non-compact objects. Video annotation was viewed as a classification problem using binary features. In the conducted experiment, low error rate was achieved using MNS to annotate video frames from sequences of 4 sports.

## 11.2 Contributions

The main contributions of this work are summarised below:

- Region-based recognition methods often require image segmentation or edge detection which may be unreliable in the presence of appearance variations. In this work, a novel representation of local object colour structure was proposed. Local appearance was described by invariant features computed from robustly filtered colour values computed from image neighbourhoods.
- The proposed object model was derived from one or more example images; a realistic assumption in many applications. The method was shown to perform well even with a single example view of the object, therefore a large set of object views was not required.
- Most published recognition methods have been demonstrated to work well with images carefully selected for their experiments. In this work, the implemented algorithm, with

---

its default algorithmic settings, was evaluated using a number of different data sets, representing a variety of applications.

- Good results were presented for recognition of objects in cluttered scenes, in the presence of partial occlusion and appearance variations due to changes in viewpoint, illumination, scale and image resolution.
- In contrast with previous approaches, the type of invariant features used for recognition is not fixed at run-time. Instead, it depends on the illumination model which in turn is defined by the application.
- A number of colour-based object recognition and image retrieval methods were reviewed and compared. These comparative studies are a contribution in their own right.
- A new image set, the SOIL-47 database, was introduced. The data is designed for evaluating colour-based recognition algorithms. All experiments in the thesis, except one, were performed on publicly available data sets which enables further comparison and improvement.

### 11.3 Possible extensions

In the implemented MNS algorithm, measurements are computed from pairs of robustly filtered colour values, computed from multimodal neighbourhoods. It would be interesting to consider measurements computed from triplets or other combinations of neighbourhood colours as they are expected to provide more discriminative features.

The frequency of similar measurements, a measure of the area covered by each colour patch in the image, or even within an image neighbourhood, was not exploited for representing the objects. However, the area of the object's surface patches may be a discriminative property in some applications (e.g. those assuming controlled viewpoint change) and its use within the MNS framework should be investigated.

The spatial arrangement of the computed measurements in the image may be another useful measurement for representing object appearance. In our experiments, the spatial arrangement

of measurements was often not preserved, and therefore it was not exploited. Nevertheless, in some applications it is likely to improve discrimination.

In some applications, the assumption of globally constant and uniform illumination is realistic. For instance, this was the case in some of the experiments carried out in this thesis. For other applications, other than 6-dimensional illumination invariants should be used. The experiment described in section 4.4 confirmed the view that more work is needed to achieve illumination invariance for general-purpose recognition.

A method for the automatic selection of some algorithmic settings like the neighbourhood width, the mode seeking kernel and the matching threshold would be useful for achieving performance optimised for a specific data set.

Although good results were obtained using RGB colour measurements, a colour space such as  $YC_bC_r$  in which the colour components are non-correlated would be interesting to test and compare.

Regarding image retrieval, a significant gain in matching speed could be obtained if an indexing data structure such as a tree index or a hash table were used. The efficiency of the implemented algorithm was sufficient for the reported experiments, therefore minimising search time was not attempted in this work.

In generic applications like web-based image retrieval, the image collection to be searched may not be available prior to the search. When it is, properties of the images in the collection can be exploited in order to optimise performance. Although in our experiments the database was available off-line, no preprocessing was performed.

We view our colour-based system as part of an integrated application that recognises objects based on more than one type of visual (colour, texture, motion) and/or non-visual information (e.g. text, speech). We are currently working on the development of such a system. Some preliminary results are presented in [47].