A Demonstration of Color-Based Object Recognition on a Grid[*]

F.J. Seinstra and J.M. Geusebroek
ISLA, Informatics Institute, University of Amsterdam,
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
{fjseins, mark}@science.uva.nl

Abstract

Multimedia data is rapidly gaining importance along with recent developments such as the increasing deployment of surveillance cameras in public locations. In a few years time, analyzing the content of multimedia data will be a problem of phenomenal proportions, as digital video may produce data at rates beyond 100 Mb/s, and multimedia archives steadily run into Petabytes of storage space. Consequently, for urgent problems in multimedia content analysis, Grid computing is rapidly becoming indispensable.

This demonstration shows the viability of wide-area Grid systems in adhering to the heavy demands of a real-time task in multimedia content analysis. Specifically, we show the application of a Sony Aibo robot dog, capable of recognizing objects from a set of learned objects, while connected to a large-scale Grid system comprising of cluster computers located in Europe, the United States, and Australia. As such, we demonstrate the effective integration of state-of-the-art results from two largely distinct research fields: multimedia content analysis and Grid computing.

1. Introduction

Irrespective of the application of a robot, the problem of object recognition is to determine which, if any, of a given repository of objects appears in an image or video stream. It is a computationally demanding problem that involves a non-trivial tradeoff between specificity of recognition (e.g., discriminating between different faces) and invariance (e.g., to different lighting conditions). Due to the rapid increase in the size of multimedia repositories consisting of 'known' objects [3], state-of-the-art sequential computers no longer can adhere to the computational demands, making high-performance distributed computing indispensible.

This demonstration shows a real-time object recognition task performed by a Sony Aibo robot dog, connected to a Grid system spanning our entire globe (see Figure 1). In the first 'learning' phase of the demonstration, we present a set of objects under a single visual setting. In the second 'recognition' phase, we validate the learning step by showing the objects again, under varying lighting conditions, lighting color, and viewing position. In earlier experiments we have shown that our dog is capable of accurately recognizing more than 300 objects from a set of 1,000 learned objects, under a diversity of imaging conditions. Interestingly, this recognition rate is higher than the recognition rate of around 200 objects reported for a real dog [4].

2. Color-Based Object Recognition

Color is a powerful cue in the recognition of objects. Recognition based on color, rather than just intensity, provides a broader class of discrimination between objects. The use of RGB values, however, does not directly increase recognition performance, certainly not when variations in imaging conditions are encountered. Differences in intensity, direction, and color of the illumination, as well as shading and cast-shadow significantly effect the appearance of an object. Therefore, it is meaningful to transform the RGB

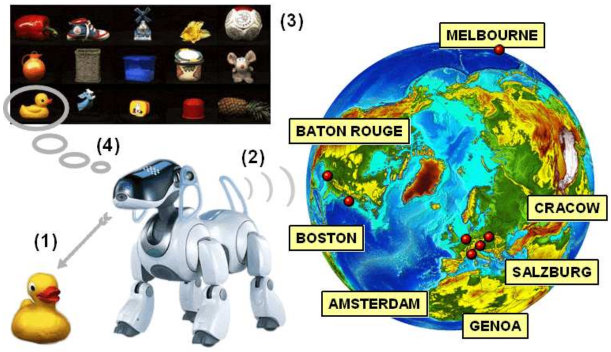

Figure 1. *Object recognition by our robot dog: (1) an object is held in front of the dog's camera; (2) video frames are processed on a per-cluster basis; (3) given the resulting feature vectors describing the scene, a database of known objects is searched; (4) in case of recognition, the dog reacts accordingly (see also: www.science.uva.nl/~fjseins/aibo.html).*

---

[*] This work is sponsored by the Netherlands Organization for Scientific Research (NWO) and by the DELOS NoE on Digital Libraries (EU FP6).

Figure 2. *In the recognition phase of our demonstration, the dog will make a 360 degrees turn, recognizing the objects it has learned previously (see also: www.science.uva.nl/˜fjseins/aibo.html).*

values to invariant properties, which relate to surface properties rather than to object appearance. We previously derived a broad class of invariants [1, 2], which are shown to be robust under noisy conditions. Furthermore, these invariants can be scaled to the size of the object structure. Recognition with these invariants boils down to learning an invariant representation of the object, rather than learning every possible appearance of a single view of the object.

We decompose the recognition of object appearance into two schemes. First, we have different views or aspects of an object, each of which has to be learned. Secondly, there is the illumination, drastically influencing object appearance. For this class of appearance effects, we demonstrate invariants to be very effective. Object recognition may be based on a weak description of the important features in the scene, as long as mutual correspondence between observation and objects in the world is maintained. Therefore, we learn local histograms of invariant features for each aspect of an object.

3. Services-based Multimedia Grid Computing

From the user's perspective, Grid computing is still far from being more than just an academic concept. Essentially, this is because Grids do not yet have the full basic functionality needed for extensive use. Consequently, as long as programming and usage is hard, most researchers in multimedia computing will not regard Grids as a viable alternative to more traditional computer systems.

The Parallel-Horus project [5, 6, 7] attempts to overcome this problem by providing a software architecture that allows multimedia researchers to implement *fully sequential* applications for efficient execution on cluster systems. To further stimulate the use of clustered wide-area Grid systems in the multimedia community, the user transparent programming model is supported by an easy-to-use execution model based on *Wide-Area Multimedia Services*, i.e. high-performance multimedia functionality that can be invoked from sequential applications running on a desktop machine. With Parallel-Horus, dynamic systems of distributed multimedia services, in which clients and servers can participate at will, can be created without any parallelization and distribution effort from the user. Our robot dog application indeed constitutes a dynamic system of this kind.

4. Demonstration

The color-based object recognition is performed using the video data obtained from the camera hidden in the dog's nose. Each video frame is processed on one of the available clusters (see Figure 1). Using this world-wide computing capacity, we first demonstrate the learning phase, by presenting our dog with new objects, potentially obtained from people in the audience. Next, we demonstrate the recognition phase, by letting our dog walk around in a circle and indicate the objects it has previously learned (see Figure 2).

Acknowledgements

The authors would like to thank all researchers and system maintainers who have granted us access to their cluster computers. We are also grateful to Michiel van Liempt for his excellent efforts in implementing the original sequential object recognition code. Thanks go out to Edwin Steffens and Arnoud Visser for providing us with a robot dog.

References

[1] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and H. Geerts. Color invariance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(12):1338–1350, 2001.

[2] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and T. Gevers. Color constancy from physical principles. *Pat. Rec. Let.*, 24(11):1653–1662, 2003.

[3] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam library of object images. *Int. J. Comput. Vis.*, 61(1):103–112, January 2005.

[4] J. Kaminski, J. Call, and J. Fischer. Word Learning in a Domestic Dog: Evidence for ”Fast Mapping”. *Science*, 304, June 2004.

[5] F. J. Seinstra and D. Koelma. User Transparency: A Fully Sequential Programming Model for Efficient Data Parallel Image Processing. *Concurr. Comput.-Pract. Exp.*, 16(6):611–644, May 2004.

[6] F. J. Seinstra, D. Koelma, and A. D. Bagdanov. Finite State Machine Based Optimization of Data Parallel Regular Domain Problems Applied in Low Level Image Processing. *IEEE Trans. Par. Dist. Syst.*, 15(10):865–877, Oct. 2004.

[7] F. J. Seinstra, D. Koelma, and J. M. Geusebroek. A Software Architecture for User Transparent Parallel Image Processing. *Parallel Comput.*, 28(7–8):967–993, Aug. 2002.