LAVA

Title: Learning for Adaptable Visual Assistants
Proposal Number: IST-2001-34405
Acronym: LAVA
Action Line: IST-2001 IV.2.1

ABSTRACT

The key problem that must be solved in order to build cognitive vision systems is the robust, efficient and learnable categorisation and interpretation of large numbers of objects, scenes and events, in real settings. LAVA will create technologies enabling such systems and an understanding of the systems - and user-level aspects of their applications, via a novel alliance between statistical learning theory, computer vision and cognitive science experts. For practical computational efficiency and robustness, we shall devise methods for goal-directed visual attention and the integration of multiple asynchronous visual cues. These results will be embodied in two integrated systems: one will employ vision for information retrieval in a mobile setting; the other will derive symbolic representations from video sequences, enabling a wide range of ambient intelligence scenarios.

OBJECTIVES

Our goal is to create fundamental enabling technologies for cognitive vision systems and to understand the systems - and user-level aspects of their applications. Technologically, the objectives are the robust and efficient categorisation and interpretation of large numbers of objects, scenes and events, in real settings, and automatic online acquisition of knowledge of categories, for convenient construction of applications. Categorisation is fundamentally a generalisation problem, which we shall solve using measures of distance between visual descriptors known as "kernels". We aim to dramatically improve generalisation performance by incorporating prior knowledge about the behaviour of descriptors within kernels, and by exploiting the large amounts of unlabelled data available to vision systems. Finally we aim to exploit this technology in integrated systems that employ vision for information retrieval in a mobile setting, and systems that derive symbolic representations from video.

WORK DESCRIPTION

Five of the project's seven workpackages are devoted to LAVA's core technologies. The first two concentrate on learning and visual descriptors. This will involve close collaboration around kernel design and the incorporation of models of the behaviour of descriptors. The learning work will emphasise improvement of the generalisation properties of classifiers, for example by exploiting the vast amounts of unlabelled data available to vision systems. Visual descriptors will be designed to enable efficient learnability and discrimination between categories in the face of much extraneous information such as lighting, viewpoint, occlusion and natural within-class variation. The next two workpackages focus on higher-level issues of generic categorisation, interpretation and cue integration. These will collaborate closely in the study of attention mechanisms, but each has its specific focus: one is principally concerned with the practice of goal-directed search with mainly static cues of specific types, and the

other aims at a unifying theory of attention for integrating arbitrary asynchronous cues. One workpackage is devoted to building and evaluating the two integrated demonstrators, which contain contributions from all partners. It will also conduct the important task of data gathering. Both of these operations will be conducted in two phases, which are reflected by the task divisions of the other workpackages. The first phase will integrate early versions of components from the other workpackages. User- and systems-level evaluation will identify areas for improvement in the application scenarios and the system architecture, while maximising the opportunity for feedback on issues regarding the components. The second phase will rectify such issues and incorporate more advanced components as necessary.

MILESTONES

YR1 Initial data gathering. Baseline descriptors, learning methods and static cue integration methods leading to initial evaluation of association assistant.
YR2 Learning with unlabelled data, temporal cue integration and dynamic attention mechanisms leading to initial evaluation of event interpreter.
YR3 Final data gathering. Optimised sparse and online learning, high-level descriptors, learning and attention for interpretation leading to final evaluation of integrated demonstrators.

PROJECT DATA

Start date: 1 May 2002   **Duration:** 36 months
Total cost: 4,330,530 €   **EC Contribution:** 2,463,170 €

CONTACT DETAILS

Project Coordinator
Christopher DANCE
Xerox – Research Centre Europe
61 Regent Street
CB2 1AB Cambridge
UNITED KINGDOM
Tel: +44 1223 431521
Fax: +44 1223 431510
e-mail: chris.dance@xrce.xerox.com

PROJECT PARTICIPANTS

Xerox Research Centre Europe	F
Institut National de Recherche en Informatique et en Automatique	F
Royal Holloway and Bedford New College	UK
Lund University	S
Technische Universität Graz	A
Institut Dalle Molle d'Intelligence Artificielle Perceptive	CH
Australian National University	AU