Understanding and Interpreting the Activities of Experts: Towards a Cognitive Vision Methodology

by Stelios Orphanoudakis, Antonis Argyros and Markus Vincze

The ActIPret project, funded by the European IST programme, aims to build advanced systems, able to recognize and interpret the activities of experts in the context of a cognitive framework.

One of the fundamental abilities possessed by humans is that of acquiring skills through observation. Teaching by demonstration is consequently a powerful way to provide training. Despite the advantages of teaching, it is not always possible because of distance and time barriers; experts cannot teach groups of trainees, at a certain location and for a limited time period. These barriers could be overcome by realising a computational vision system that is capable of understanding, interpreting and indexing the activities of experts. The combination of such a system with recent advancements in virtual and augmented reality (VR/AR) could be used to effectively search, retrieve and reproduce the activities of experts anywhere and anytime.

Humans have a remarkable ability to visually interpret the activities of other humans, transforming these interpretations into knowledge and subsequently exploiting this knowledge in acquiring related skills. Teaching by demonstration therefore constitutes a powerful training technique. Currently, teaching by demonstration entails an expert demonstrating her/his expertise to small groups of trainees on specific occasions. Time and distance barriers hinder the observation of experts in action. Recording activities on video provides a partial solution to this problem and allows repeated viewing but is subject to other important limitations. Although video records a dynamic sequence of events, in some sense it is quite a 'static' source of information. The fixed viewpoint restricts visibility and may lead to ambiguous interpretations. Moreover, activities cannot be indexed and effectively searched, as is the case with a manual or user's guide.

With recent developments in virtual and augmented reality (VR/AR), it is now possible to produce high quality representations of a reconstructed scene and a realistic replay of activities therein. Such capabilities could prove invaluable in developing tools for teaching by observation, provided that the interpretation of the activities of an expert is also possible. The coupling of these capabilities could result in the removal of most of the important barriers in teaching through observation. The expert's demonstration can be replayed anywhere, anytime and from any viewpoint. Moreover, the activities can be indexed efficiently and effectively, and retrieved by the trainee based on her/his needs.

The overall objective of ActIPret is to develop a cognitive vision methodology that permits the recording and interpretation of the activities of people handling tools. Interpreted activities can be stored in an activity plan that can be referenced later by the user. The activity plan is an indexed manual in the form of 3D reconstructed scenes, which can be replayed at any time and location to many users using VR/AR equipment. Research and development is focused on the active observation and interpretation of the activities, on the extraction of the essential activities and their functional dependence, and on organising them into constituent behaviour elements. The approach is active in the sense that the user obtains views that facilitate the interpretation of the observed activities. Moreover, the activities are interpreted in an efficient and effective manner, and retrieved by the trainee based on her/his needs.
The act of interpreting the activities of experts involves using a conceptual language to observe and record expert activities during use cases. During replay, the trainee/user searches for specific activities of interest using a conceptual language. The user is then able to choose between two replay of the sequence from arbitrary viewpoints, depending on the training level or (2) use of the ActIPret system in the form of a personal teaching assistant. In this case, for a selected task, the activities performed by the user are automatically compared with the activities of the expert and improvements are suggested. This results in more effective training, compared to repetition without feedback.

ActIPret has two main technical objectives: the design and evaluation of a cognitive vision framework that extracts and interprets activities, and the development of purposive visual processing techniques to provide the required perceptual capabilities.

To achieve a robust interpretation of activities, the interaction of visual attention, active recognition, understanding, and knowledge from models, tasks, and context are being investigated. The cognitive vision framework makes it possible to discriminate between activities that are essential to the task at hand (and should therefore be maintained) and those that are irrelevant (and should therefore be eliminated from the training sessions). The final outcome of the cognitive approach is the activity plan, which contains an index into activities and behaviours for access in user-driven training and for feedback while the trainee is rehearsing.

To achieve the cognitive ability of the framework, vision techniques must provide the required functionality in the form of self-contained, cooperating components. The framework consists of both top-down (behaviour-/context-driven) and bottom-up (data-driven, self-evaluating) interacting components. There are four types of visual processing components:

- extraction of cues and features
- detection of context-dependent relationships between cues/features
- recognition of activities and objects handled, taking into account potential occlusion
- synthesis of behaviours and tasks that modify the context of the other components.

All four types of components report visual evidence with confidence measures. These interpretation are interlaced with the attentive and investigative behaviours that provide feedback to purposively focus processing. Robust interpretation results are achieved with methods that actively seek desirable viewpoints and obtain elucidative information for detection, recognition and synthesis. Robustness is also enhanced using context-dependent information integration between the components.

The ActIPret consortium consists of the following partners: Institute of Automation and University of Technology (Project Coordinator), Center for Machine Perception at the Czech Technical University, School of Cognitive and Computing Sciences at the University of Sussex, Vision and Robotics Laboratory, ICS-FORTH and PROFACTOR - Produktionsforschung.

Link:
Project web site: http://actipret.infa.tuwien.ac.at/

Please contact:
Stelios Orphanoudakis, ICS-FORTH
Tel: +30 2810 391605,
E-mail: orphanou@ics.forth.gr

Markus Vincze, Project Coordinator, Technical University Vienna
Tel: +43 1 5041446 11
E-mail: vincze@acin.tuwien.ac.at