ACTIPRET

Title: Interpreting and Understanding Activities of Expert Operators for Teaching and Education
Proposal Number: IST-2001-32184
Acronym: ACTIPRET
Action Line: IST-2001-IV.2.1

ABSTRACT

Rapid technical development generates the need to train many people in expert operations. To teach many users a system to interpret the expert’s activities is required. The user can replay activities at any time and from any viewpoint. Due to the cognitive framework of the vision system envisioned it is possible to index activities and objects involved. The index is based on natural language terms and allows user-driven retrieval. The system provides feedback to motivate the trainee and to enhance the training effect. The cognitive vision framework builds on purposive and reactive vision techniques which focus processing to obtain real-time performance. Integration and active selection of techniques realises robust interpretation. The final presentation will interpret activities in an assembly scenario, e.g. changing a car wheel.

OBJECTIVES

The objective of ActIPret is to develop a vision methodology that interprets and records the activities of people handling tools. The tasks considered are observable by video streams. Focus is on active observation and interpretation of activities, on parsing the sequences into constituent behaviour elements, and on extracting the essential activities and their functional dependence. By providing this functionality ActIPret will enable observation of experts executing intricate tasks such as repairing machines and maintaining plants. The expert activities are interpreted and stored using natural language expressions in an activity plan. 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 Augmented Reality equipment. Due to the interpretative ability of the system, ActIPret can provide the trainee with feedback when repeating the operation (in simulation or reality), which results in a superior training effect compared to repetition without feedback.

WORK DESCRIPTION

The project is organised into eight interlaced technical work packages to build the cognitive vision framework and its purposive and reactive processing components. In the first year the framework and its constituent parts are designed and a first prototype is implemented. Every six months all components will be tested, evaluated and the functionality extended. Iterative progress of work is required to thoroughly study the interactivity of the components. The approach involves associating attention pragmatic interpretation with specific phases of tasks and context to zoom in on the relevant objects and activities. The components of visual processing are all task and context-driven and report visual evidence with confidence measures. These components are the extraction of cues and features, the detection of context-dependent relationships between cues/features, the recognition of the objects handled taking into account potential occlusion, and the recognition of activities and the synthesis of behaviours and
tasks that bias the context at the other components. These levels of visual interpretation are interlaced with the attentive and investigative behaviours that provide the feedback to purposively focus processing to obtain real-time performance. Robust interpretation results are obtained with methods of learning and methods to actively seek good viewpoints and to obtain disambiguating information for detection, recognition and synthesis.

Robustness is further enhanced using context-dependent information integration between the components and with integration of cues, features and complementary recognition methods.

The framework is tested with two scenarios where various people handle objects: placing a CD in a player and an industrial assembly task. From interpreting observed activities the activity plan will be extracted. Using the stored reconstruction of the scene and activities, users will be able to index the expert operations for replay using AR equipment.

**MILESTONES**

Year 1: Prototype framework implemented; Recognition of single activities and learning of object representations; Conceptual language defining activities;
Year 2: Interpretation of one-handed activities with objects; Qualitative description of spatial relations between objects and activities; Conceptual activity description for activity plans;
Year 3: Interpretation of two-handed activity sequences with objects; Temporal relations between objects and activities; Activity plan synthesis and replay.

**PROJECT DATA**

- **Start date:** 1 November 2001
- **Duration:** 36 months
- **Total cost:** € 2,569,877
- **EC Contribution:** € 1,894,000

**CONTACT DETAILS**

**Project Coordinator**
Markus VINCZE
Technische Universitaet Wien
Institut für Flexible Automation
Karlsplatz 13
1040 Wien
AUSTRIA
Tel: +43-150-4144611
Fax: +43-158-8013619/9
Email: vm@infa.tuwien.ac.at

**PROJECT PARTICIPANTS**

- Technische Universitaet Wien
- Foundation for Research and Technology - Hellas
- Czech Technical University in Prague
- The University of Sussex
- Profactor Produktionsforschungs GmbH