Communicating Agents Architecture with Applications in Multimodal Human Computer Interaction

Maximilian Krüger, Achim Schäfer, Andreas Tewes, Rolf P. Würtz
Institut für Neuroinformatik, Ruhr-Universität Bochum

Abstract: We present our idea of solving parts of the vision task with an organic computing approach. We have designed a multiagent system (MAS) of many different modules working on different tasks towards the same goal, the recognition of a human gesture. We will introduce our point of view of an agent and how a group of agents can cooperate with other organically inspired techniques like the democratic integration for information merging and bunch graph matching for face/object recognition. In this article we want to explain the need, the concept and the software for our work. Two applications currently limited to the tracking problem, demonstrate that the concept is fruitful and in use.

1 Introduction

Human computer interaction (HCI) is one of the future's main tasks, and its improvement is the goal of our work. Although there are many ways of communicating with a machine like speech and touch-screen, we have focused on computer vision, a field in science that has shown many advances, but still is under construction [vdM04]. We split the task of human gesture recognition in two areas. The first part is the analysis of visual data, the identification and tracking of human extremities. To solve this problem, multiple modalities like color, texture, motion, etc have to be interpreted. In the second part, gesture recognition based on the collected information is performed. Both areas shall form a system that can handle varying environmental conditions. Each work step has to solve the same kind of problem: imperfect information sources. We believe that organic computing, with its self-x properties is leading into the right direction. One of the key concepts of organic computing is sub-system integration. The technique of democratic integration [TvdM01] implements a self-organizing integration of cues with self-healing properties. It has proven to work quite well for simple tracking problems. As we started to extend this method, we realized that the existing implementation was too rigid. The need for multiple independent instances of trackers working on their individual set of cues and the integration of top down information led us to the construction of a multiagent system (MAS). Our architecture is built in a modular fashion, agents are working on different tasks in different layers. These agents show the following self-x properties. They are autonomous, have a perception of their surrounding to which they can adapt and they are capable of communicating with other entities. The breakdown of sensor information and a changing environment is treated in a robust and flexible manner. Our multiagent network can handle agents with tasks ranging from coordination of sub processes, the tracking of an image point up to the recognition of human extremities. To accomplish our aim of human gesture recognition, teamwork between agents and a hierarchically ordered structure are essential. The following sections present our approach of a MAS and sketch the software architecture. Because the system is still under development, the two applications cover only the tracking part of the task.

2 Framework

As a realization of interacting entities, a MAS was chosen. The design is modular with well defined interfaces. The components are agents (with a cue integrator and different sensors as sub-modules), a blackboard and an environment.

The environment supplies information about the world. Based on the desired functionality of the whole system, this can simply be an image sequence. The environment may, however, also supply preprocessed images like e.g. gray value images and difference images treated as different modalities. An agent is an autonomous entity, which gets information from the environment and processes it to reach a decision about further actions. Each agent can decide which parts of the information provided by the environment it requests and makes available to its sensors. They in turn individually process their information and represent the agent's perception of the environment. Sensors can be flexibly attached or detached during run-time. In each agent, integration of the sensors' output is done by a cue integrator, which is implemented regarding the agent's general goal and according to the specific sensors it uses. Alternatively, an agent can consult knowledge sources which provide information about the world stored in the system. We call these agents top down agents. For example, the face recognition agent presented in the next section is a top down agent that uses general face knowledge. Training the top down agents, i.e. learning the world knowledge from examples is also a crucial task requiring organic computing methodology, see [Wü04].

Communication within the system is done via a blackboard. After registering with the blackboard, each agent can write messages onto the blackboard and read messages other agents have posted. All messages can be quite complex, have a defined lifetime and are posted to a group of receivers. The message handling allows the creation of new agents with specified properties, the change of properties of agents and also the elimination of agents during run-time, provided they were given this functionality. This architecture offers the possibility to establish a hierarchy of agents, where higher agents manage lower ones and collect, process and evaluate the information from them.

Altogether, the software framework enables to develop autonomous, communicating and possibly hierarchically ordered multiagent systems. So far, the framework is not restricted to vision, but can incorporate non visual modalities as well.

3 Applications

As we are working on HCI, our applications are designed to solve problems like pointing gesture and sign language recognition. They depend on a reliable and robust tracking of human body extremities. Since human motion has many degrees of freedom, the tracking is faced with the problem of cluttered scenes and the possibility of objects leaving the image borders. Based on the organic computing concept and the presented framework, we developed a multiagent network with different entities working on local or global scales. We designed tracking agents whose task is to follow an object. These agents merge different visual cues like color, texture etc. Cue fusion is done by the scheme of democratic integration, which offers a flexible and robust way of tracking. For a detailed view refer to [TvdM01]. The tracking agents possess the ability to rate their action. They offer a confidence value, which assesses the tracking quality that becomes lower if the individual agent is loosing its target. Due to the fact that they are only working on a small part of the image they are called local agents. The network is hierarchically organized, information from the tracking agent (like position and movement of the object) are sent to higher global working stages. These are currently monitoring the tracking agents, instantiating and deleting them. Further high level agents will be implemented in the course of the project.

3.1 Hand Gesture Recognition

For hand gesture recognition a special agent is performing a global search for skin colored and moving objects. If successful, it can assign a tracking agent to the object. Tracking agents can be seen in figure 1. Their region of interest is marked by a gray rectangle. The white circle encloses the estimated target position. Because tracking is a difficult task, we need a self-controlling mechanism. If the tracking agent is not convinced of its tracking result, if two agents are tracking the same object or if the object being tracked is of no use for the gesture recognition, the agent has the possibility of self-destruction, a self-healing mechanism of the system. The tracking agent receives information from top down agents like a face detecting agent. This agent checks whether an agent is following a face. Face recognition is performed by bunchgraph matching [WFKvdM97] see also [Wü04]. In figure 1 the dark gray rectangle indicates that this tracking agent was informed by the face detecting agent that the tracked object is a face. This and other top down knowledge that is brought to the system supports the classifying of objects and gestures. The first results are encouraging, some more work will be put into the dynamic to make the tracking more robust and reliable.

3.2 Arm Gesture Analysis

To identify arm or body gestures, motion of points other than the hand and the face need to be analyzed. This task induces additional difficulties because e.g. the color and the

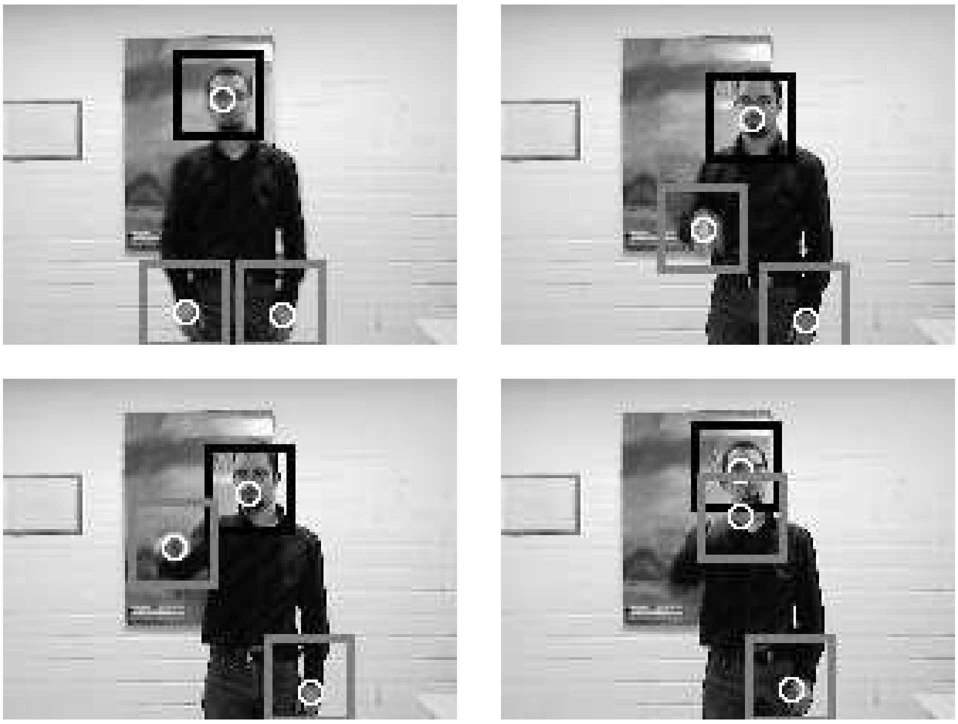

Figure 1: Parts of an image sequence of a hand tracking process. Tracking agents' region of interest is marked by a gray rectangle and their target position by a circle.

texture of clothes have a greater variety than skin color. In the current state, the application supports user assisted learning of arm gesture models. With one mouse click per agent, the user needs to position tracking agents at pre-defined points at the body of the person performing the arm gesture. The tracking agents' sensors and parameters can be chosen individually. Another agent is designed to control the birth of new tracking agents at the given positions with the given parameters and to monitor their activity. It collects and saves all data, which can then be analyzed.

The short-term objective is the analysis of the tracking process. These results will help to increase the tracking performance in other applications. On the one hand, we want to understand better which sensor configuration at which position contributes to the performance of a tracking agent. On the other hand, we want to learn what is important to successfully discriminate one gesture against the others. The analysis should give information about the essential features of gestures in general and of the specific gesture in particular. Learned properties of different arm gestures can then be used for efficient gesture recognition. In future versions, successful tracking parameter sets shall be determined automatically by the process of self-organization and self-monitoring. The learning itself and the gesture recognition based on the learning still need to be realized.

At the current state of development, the system must be bootstrapped with manual selection of points to be tracked. This burden will be removed from the user by passing over decisions gradually to specialized agents.

4 Discussion

We have set up a flexible framework for communicating agents and we have shown the usefulness and flexibility with two applications. Tracking of points – especially hands and head in the hand gesture recognition application – already works quite well. Furthermore, the application for arm gesture analysis provides an interface for learning parameters, where tracking agents can easily be positioned on image sequences and the data can be analyzed. Positioning other types of agents can easily be built in.

In the future, the communication between the agents will be used to produce a positive feedback that will further stabilize the system. Learning relative positions of human body extremities from examples is another goal in order to achieve gesture recognition. Also, more top down knowledge like a hand recognition agent will be attached to the system. Up to now, a module for the recognition of different gestures is in the conceptual state.

We have presented an approach to gesture recognition by organic computing methodology. A multiagent system has been chosen because it meets the flexibility and self-healing requirements, which are crucial for the task of, ultimately, understanding the movements a user performs in front of a camera and make this understanding useful for human-friendly human computer interaction.

Acknowledgements: Funding by the European Commission in the Research and Training Network MUHCI (HPRN-CT-2000-00111) is gratefully acknowledged.

References

[TvdM01] Triesch, J. und von der Malsburg, C.: Democratic integration: Self-organized integration of adaptive cues. *Neural Computation*. 13(9):2049–2074. September 2001.

[vdM04] von der Malsburg, C.: Vision as an exercise in Organic Computing. In: *This workshop*. 2004.

[WFKvdM97] Wiskott, L., Fellous, J.-M., Krüger, N., und von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine, Intelligence*. 19(7):775–779. 1997.

[Wü04] Würtz, R. P.: Organic Computing for face and object recognition. In: *This workshop*. 2004.