
KUNGL
TEKNISKA
HÖGSKOLAN


CID-148 • ISSN 1403-0721 • Department of Numerical Analysis and Computer Science • KTH

# Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction

**Sören Lenman, Lars Bretzner, Björn Eiderbäck**

**CID, CENTRE FOR USER ORIENTED IT DESIGN**

**Sören Lenman, Lars Bretzner, Björn Eiderbäck**

Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction
**Report number:** CID-148
**ISSN number:** ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)
**Publication date:** January 2002
**E-mail of author:** lenman@nada.kth.se

**Reports can be ordered from:**

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE- 100 44 Stockhom, Sweden
Telephone: + 46 (0)8 790 91 00
Fax: + 46 (0)8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

# Computer Vision Based Recognition of Hand Gestures for Human-Computer Interaction

**Sören Lenman**
Centre for User-Oriented IT-Design, CID
Royal Institute of Technology
100 44 Stockholm, Sweden
lenman@nada.kth.se

**Lars Bretzner**
Computer Vision and Perception, CVAP
Royal Institute of Technology
100 44 Stockholm, Sweden
bretzner@nada.kth.se

**Björn Eiderbäck**
Centre for User-Oriented IT-Design, CID
Royal Institute of Technology
100 44 Stockholm, Sweden
bjorne@nada.kth.se

**ABSTRACT**
This paper describes a project that explores computer vision based analysis of hand gestures for developing new forms of HCI. A prototype for remote control of home electronics, such as TV and CD-player, has been developed. Pie- and marking menus are used to create the gesture command set. Initial user tests points to problems of fatigue and difficulties with using free gestures for marking menus. These problems are currently being addressed in the project.

**Keywords**
Hand gesture, computer vision, HCI, marking menu.

**INTRODUCTION**
Ubiquitous, embedded computing, e.g., in domestic environments requires new human-computer interaction styles that are natural, convenient and efficient. They should be less like the current desktop paradigm and more like the way we interact in the real world, using speech and gesture [1][8]. This paper describes a project that explores computer vision based analysis of hand gestures for developing new forms of HCI. For embedded computing in home environments computer vision based solutions have the advantage over other techniques for recognition of hand gestures, e.g., the data-glove, that users do not have to carry any equipment.

**MARKING MENUS FOR GESTURE CONTROL**
A fundamental concern in all kinds of gestural control is what the command set should be. Exactly what hand postures and movements should be used? A possible strategy is to base the command set on a menu system. The "language" is determined by menu layout and organization, and can be made culturally neutral and self-explanatory. Gestures can be kept relatively simple. The assumption here is that Pie- and marking menus are especially well suited for the purpose, because they offer a possibility for users to develop the skill to work with no feedback from the menus.

**Pie- and Marking Menus**
Pie menus were first described in [3]. They are pop-up menus with the alternatives arranged radially, often used in pen-based interfaces. Because the gestures ("marks") are directional users can learn to make selections without looking at the menu items. With expert users, menus need not even be popped up. Hierarchic marking menus [6] are a development of pie menus that allow more complex choices. The shape of the path, rather than the series of distinct menu choices, can be recognized as a selection. If the user, e.g., a novice, works slowly, or hesitates, the underlying menus can be popped up to provide feedback.

**THE PROTOTYPE**
Following [4] we chose a scenario for the first prototype that is well known to most: remote control of appliances in a domestic environment. A hierarchic menu system for controlling the functions of a TV, a CD player, a VCR, and a lamp is under development and some initial user trials have been performed. The prototype has been set up similar to a home environment in an open lab /demo space at CID. In order to maximize speed and accuracy, gesture recognition is currently tuned to work only against a uniform background within a limited area, approximately 0,5 by 0,65 m in size, at a distance of approximately 3 m and under relatively fixed lighting conditions.

**The Recognition Algorithms**
The computer vision system for tracking and recognizing the hand postures that control the menus is based on a combination of multi-scale color feature detection, view-based hierarchical hand models and particle filtering. The hand postures or states are represented in terms of hierarchies of multi-scale color image features at different scales, with qualitative inter-relations in terms of scale, position and orientation. In each image, detection of multi-scale color features is performed. The hand postures are then simultaneously detected and tracked using particle filtering, with an extension of layered sampling referred to as hierarchical layered sampling. To improve the performance of the system, a prior on skin color is included in the particle filtering. In Figure 1, white ellipses show detected multi-scale features in a complex scene and the correctly recognized hand posture is superimposed in gray. A detailed description of the algorithms is given in [2].

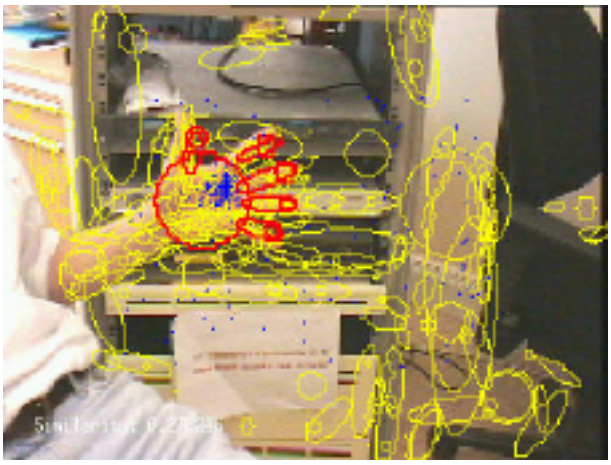

**Fig. 1** Detection of multi-scale features in a video stream of a complex scene.

### Equipment

A Dell Workstation 530 with dual 1,7 ghz Intel Xeon P4 processors, running Linux. Mvdelta 2 framegrabber, IRdeo remote IR control, and a DI-01 Data interface (X10).

### Menus

The exact design, the arrangement, and the organization of the menus are still under development. A rudimentary version with three hierarchical levels, but with only a few active choices (TV on/ off, channel +/ -, CD play/ stop/ forward/ back) is currently available. A hand posture with the index finger and thumb outstretched is used for activating the menus (mouse-down). A hand with five fingers outstretched is used for making a selection (mouse-up). Evidently, any two postures could be used.

Menus are activated when the mouse-down posture is detected in the recognition area. The hand is tracked as long as the mouse-down posture is held. If the hand is moved over the periphery of a sector that has a submenu, the parent menu disappears, and the submenu is shown. Assuming the mouse-up posture in an active field, e.g., "TV on", makes a selection. All other ways of ending the interaction are ignored. The menus are currently shown on a computer screen, placed side-by-side with the TV, but in a future version menus will be presented in an overlay on the TV screen.

## EARLY RESULTS

Only a limited number of informal user trials have been performed so far. From these it is obvious that the menu-based system requires more instruction than an earlier prototype based on hand postures only. Direct control via hand postures is immediate, but limited in the number of choices. Menu-based systems are indirect and more complex, and there is simply more to learn. However, learning to use the system was not a major issue. The current setup, with subjects seated facing the TV, making gestures with the right arm and hand held out by the side of the body with no support, is inconvenient and fatigue quickly sets in. This problem must be addressed, e.g., by providing support, miniaturizing gestures, and making the recognition system more position independent.

Hierarchic marking menus have not yet been completely implemented. However, initial observations indicate that pie menus (or single-level marks) are feasible, but that hierarchic marks might not be. It is difficult to make spatio-temporal gestures for rapid multiple-level selection that are sufficiently distinct. Thus, algorithms for recognition of marks might be required. Also, with a deep hierarchy the gesture has to cover a relatively large area, and the risk is high to end up outside the recognition area.

## FUTURE WORK

There is ongoing work to make recognition of gestures more position independent, to increase the tolerance for varying lighting conditions, and to increase recognition performance with complex backgrounds.

We are currently working on the design and organization of menus as well as the display of idealized marks to indicate selections. We also intend to test Flow Menus, a variant of hierarchical marking menus, where successive levels of the hierarchy are shown in the same position [5] for reducing the area gestures have to cover with deep hierarchies. An additional concern is that not all kinds of functions, e.g., increasing sound volume, are suitable for standard pie menus. Thus, we are developing a variant of control menus |7] where repeated control signals are sent as long as the hand is kept within the menu item in a mouse-up posture.

## REFERENCES


1. Abowd, G.D. & Mynatt, E.D. (2000) Charting Past, Present, and Future Research in Ubiquitous Computing. ACM ToCHI, Vol. 7, No. 1, pp. 29-58
2. Bretzner, L., Laptev, I. & Lindeberg, T. (2002) Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering. Submitted to 5th Intl. Conf. on Automatic Face and Gesture Recognition.
3. Callahan, J., Hopkins, D, Weiser, M. & Shneiderman, B. (1988) An Empirical Comparision of Pie vs. Linear Menus. Proceedings of CHI'88, pp. 95-100.
4. Freeman, W.T. & Weissman, C.D. (1994) Television Control by Hand Gestures. In 1st Intl. Conf. on Automatic Face and Gesture Recognition.
5. Guimbretière, F. & Winograd, T. (2000) FlowMenu: combining Command, Text and Data Entry. Proceedings of UIST'2000, pp. 213-216.
6. Kurtenbach, G. & Buxton, W. The Limits of Expert Performance Using Hierarchic Marking Menus. Proceedings of CHI'94, pp. 482-487.
7. Pook, S., Lecolinet, E., Vaysseix, G. & Barillot, E. (2000) Control Menus: Execution and Control in a Single Interactor. CHI'2000 Extended Abstracts, pp. 263-264.
8. Segen, J. & Kumar, S. Look, Ma, No Mouse! Communications of the ACM, July 2000/ vol 43, 7.