Automatic adaptation of Human-Computer Interaction system

Jaromir Przybyło

AGH University of Science and Technology, Department of Automatics

Abstract: Human-computer interaction (HCI) is an emerging field of science aimed at providing natural ways for humans to use computers. Among many systems are vision-based approaches, utilizing face tracking and facial action recognition. Many sources of variation in facial appearance exists, which make recognition a challenging task. Often, uncomfortable manual calibration of the system is required. Therefore, in our opinion the key issue is the automatic adaptation of the machine to human rather than vice versa.

Keywords: human-computer interaction; image and video processing; adaptation and calibration

1. Introduction

Typical input devices, such as QWERTY keyboard and a mouse, remains the dominant way of human-computer interaction. Unfortunately they are not designed for people with severe physical disabilities. People who are quadriplegic – as a result of cerebral palsy, traumatic brain injury, or stroke – often have great difficulties or even cannot use mouse or keyboard. In such situations facial actions and small head movements can be viewed as alternative way of communication with machines (computer, electric wheel chair, etc.).

There are various approaches to build mouse alternatives. Some of them use infrared emitters that are attached to the user's glasses, head band, or cap [7]. Other try to measure eyeball position changes and determine gaze direction [1, 14]. Also, there are attempts to leverage electroencephalograms for human-machine interaction [5]. Typically these systems use specialized equipment such as: electrodes, goggles and helmets which are uncomfortable to wear and also may cause hygienic problems.

On the other hand, optical observation gives possibility of easy adaptation to serve the special needs of people with various disabilities and does not require uncomfortable accessories. Therefore, vision based systems are promising solution for HCI interface. The potential of other applications of such systems exists as well (i.e. entertainment and game systems, AV content analysis, etc).

2. The need of machine adaptation

There are many approaches to vision-based human-computer interaction. One of them is the „Camera Mouse" system [2], which provides computer access for people with severe disabilities. User's movements are translated into the movements of the mouse pointer on the screen. System tracks facial features such as the tip of the user's nose or finger. The visual tracking algorithm is based on correlation technique and requires manual selection of tracked features.

Another, widely known, example is CAMSHIFT algorithm [4] used as a computer interface for controlling commercial computer games and for exploring immersive 3D graphic worlds. It operates on color probability distributions derived from video frame sequences. The efficiency of the algorithm depends on created skin-color model. Many other face localization and tracking algorithms use a skin color model as well.

There are many sources of variation in facial appearance. They can be categorized [9] into two groups – intrinsic factors related to the physical nature of the face (identity, age, sex, facial expression) and extrinsic factors due to interaction of the light with the face and observer (illumination, viewing geometry, imaging process, occlusion, shadowing).

Above examples, show typical problems in designing robust vision-based HCI system. Careful manual intervention during initialization of the algorithm and recalibration during execution are required.

On the other hand, human object recognition seems effortless. From generic knowledge of an object, one can easily recognize novel instances of the object (for example: smile, sadness, etc). The main question is how this knowledge is encoded and used?

In this paper, we propose a vision-based automatic calibration method designed to initialize HCI system or adapt its parameters during execution. It uses generic model of human face to detect facial features. After that system can learn other visual cues, important to recognize facial actions.

3. Algorithm outline

In order to minimize manual intervention during initialization of the HCI system, the blink gesture detection is performed. Using only blink signal, allows people with severe physical disabilities to perform initialization without help or with minimum assistance of caretakers. Te user has to stay still in front of the camera and blink.

Proposed algorithm makes use of both – static image analysis and motion detection performed simultaneously. Algorithm outline is given in Fig. 1 (details are in [15]).

After preprocessing stage (color conversion, median filtering), motion detection is performed in luminance component by creating Motion Energy Images – MEI [3]. MEI representation shows where in the image the motion occurs. The delay parameter defines the temporal extent of a movement and is set to match the typical eye blink period. Sum of the MEI representation gives information about how big motion is and it is used to detect head movements or significant scene lighting changes.

The aim of object analysis stage is to remove objects from binary MEI image which do not fulfill assumed criteria (object size, etc.). This step removes artifacts generated by image noise or small head movements. All remaining objects are considered as possible eye candidates. To select proper pair of objects which are most likely located in eye regions of the image, additional analysis based on face anatomy is performed (the distance and the angle of line passing through eye candidates must lie within assumed limits).

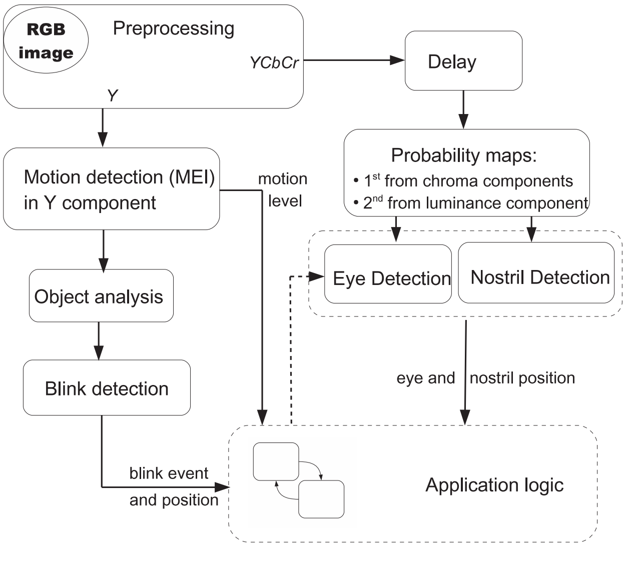

Fig. 1. Overview of the algorithm
Rys. 1. Schemat ogólny algorytmu

To detect blink event the remaining pair of eye candidates must exists on the same image region on N consecutive frames. Approximate eye positions obtained are used to define region of interests (ROIs) for accurate eye and nostril localization. The blink event triggers feature (eye, nostril) localization algorithm to execute until head movement or significant scene lighting change is detected. In such situation (for example when user shakes head or move) feature detectors are suspended until next blink.

Eye and nostril detection is performed on computed ROIs using both luminance and chrominance components. Two probability maps are created, each of them based on different image representation. The probability map from the chroma is based on the observation that high Cb and low Cr values can be found around the eyes [8]. Since nostrils, eyes (also other significant face features) usually form dark or light blobs on the face image, the second probability map is constructed from luminance component using blob detection algorithm based on a scale-space representation [11] of the image signal.

After eye and nostril localization stage the following information is available for calibration and initialization HCI application: blink event and location, eye and nostril position, as well as face approximate boundary; head movement (or significant lighting change) signal.

Gathered information can be used to learn visual cues and initialize essential parameters of the facial action recognition and feature tracking algorithms – for example: skin-color model used in CAMSHIFT tracker or correlation tracker templates. Also, blinks can be used to recalibrate system during its execution.

4. Preliminary experimental results

Proposed algorithm has been modeled and tested on MATLAB/Simulink platform running on Windows-based PC. Because, MATLAB is not suitable to perform tests in real-time, part of the algorithm also has been integrated with C++ framework using Simulink Coder which offers code generation capabilities. This enables possibility of additional evaluation of algorithm performance with user interaction.

To our knowledge, most blink detection algorithms [6, 10, 13, 16] are evaluated only on image sequences from one camera. Our main requirement was that algorithm designed for system calibration should work well under various conditions. without the need of adjusting its parameters. Therefore, test video sequences have been taken by different cameras in various lighting conditions and background complexity.

Signals obtained from algorithm (blink events, eye position, direction of head movement, etc.) are suitable to control applications. For example – blinks can emulate mouse clicks, and head movement can be used to control cursor.

5. Conclusion

Preliminary experimental results show that proposed algorithm can be used to automatic adaptation and calibration of HCI system. Adaptation of the machine to human rather than vice versa is essential shift toward a human-centered interaction architecture, which is observed in Human-Computer-Interaction domain [12]. Although, results are promising there are several issues that should be addressed in future work. More tests and evaluation in various scenarios (especially use by disabled people) can help improve algorithm efficiency.

References

1. Augustyniak P., Mikrut Z.: *Complete scanpaths analysis toolbox.*, EMBS'06, 2006, 5137–5140.
2. Betke M., Gips J., Fleming P.: *The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities.* IEEE Trans.

on Neural Systems and Rehabilitation Eng., Mar 2002, 10(1):1–10.

3. Bobick A.F., Davis J.W.: *Action Recogn. Using Temporal Templates*, 1997.
4. Bradski G.R.: *Computer Vision Face Tracking For Use in a Perceptual User Interface.*, „Intel Technology Journal", (Q2), 1998.
5. Broniec A.: *Motion and motion intention representation in eeg signal and possibilities of application in man machine interface.*, AGH Univ. of Science and Technology Krakow, Master's thesis, 2007.
6. Chau M., Betke M.: *Real time eye tracking and blink detection with USB cameras.*, Technical report, Boston University Computer Science, Apr 2005.
7. Evans D.G., Drew R., Blenkhorn P.: *Controlling mouse pointer position using an infrared head-operated joystick.* IEEE Trans, Dec 2000, 8(1): 107–117.
8. Hsu R-L., Abdel-Mottaleb M., Jain A.K.: *Face detection in color images.* „IEEE Trans. Pattern Anal. Mach. Intell.", 2002, 24(5):696–706.
9. Gong S., McKenna S.J., Psarrou A.: *Dynamic Vision: From Images to Face Recognition.* Imperial College Press, London, UK, 2000.
10. Lalonde M., Byrns D., Gagnon L., Teasdale N., Laurendeau D.: *Real-time eye blink detection with gpu-based sift tracking.*, In CRV, 2007, 481–487.
11. Lindeberg T.: *Scale-space theory: A basic tool for analysing structures at different scales*, „Journal of Applied Statistics", 1994, 21(2):224–270.
12. Lisetti Ch.L., Schiano D.J.: *Automatic facial expression interpretation: Where human-computer interaction, artiffcial intelligence and cognitive science intersect.*, „Pragmatics and Cognition (Special Issue), 2000, 8(1):185–235.
13. Lombardi J., Betke M.: *A self-initializing eyebrow tracker for binary switch emulation*, 2002.
14. Ober J., Hajda J., Loska J.: *Applications of eye movement measuring system ober2 to medicine and technology.* SPIE, Apr 1997, 3061:327–336.
15. Przybylo J.: *Eye and nostril localization for automatic calibration of facial action recognition system.* Computer Recognition Systems 3, Springer-Verlag, May 2009, vol. 57/2009, 53–61.
16. Zhiwei Zhu: *Real-time eye detection and tracking under various light conditions.* ACM Press, 2002, 139–144. ■

Automatyczna adaptacja interfejsu człowiek-komputer

Streszczenie: Interakcja człowiek-komputer (ang. human-computer interaction - HCI) to nowa dziedzina nauki ukierunkowana na rozwój naturalnych i intuicyjnych sposobów korzystania z komputerów i rządzeń. Wśród wielu stosowanych rozwiązań znajdują się systemy potrafiące śledzić cechy i rozpoznawać mimikę twarzy, wykorzystujące algorytmy przetwarzania i analizy obrazów. Rozpoznawanie mimiki jest trudnym zadaniem ze względu na istnienie wielu źródeł zmienności w wyglądzie twarzy. Bardzo często skutkuje to koniecznością ręcznej i niewygodnej kalibracji systemu. Dlatego, też zdaniem autora, kluczową sprawą jest takie konstruowanie algorytmów, aby system adaptował się automatycznie do człowieka, a nie odwrotnie.

Słowa kluczowe: interakcja człowiek-komputer; przetwarzanie i analiza obrazów, adaptacja i kalibracja

Jaromir Przybyło, PhD

Jaromir Przybylo, born in 1975 in Krakow, Poland. He graduated with an outstanding proficiency, with MSc, Eng. degree in 2000 from the Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, AGH-UST. PhD degree (with honours) received in 2008 in Computer Science: Biocybernetics and Biomedical Engineering. Since 1999 he works at the AGH-UST Institute of Automatics as a research and teaching assistant, since 2009 as an assistant-professor.

e-mail: przybylo@agh.edu.pl