Towards Building User Seeing Computers

Bessens, G.; Gorodnichy, Dimitry

Publisher’s version / Version de l'éditeur:

Second Canadian Conference on Computer and Robot Vision Workshopon Face Processing in Video (FPIV'05) [Proceedings], 2005

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

Towards Building User Seeing Computers *

Bessens, G., and Gorodnichy, D.
May 2005

* published at Second Canadian Conference on Computer and Robot Vision Workshop on Face Processing in Video (FPIV'05). May 9-11, 2005. Victoria, British Columbia, Canada. NRC 48210.

Canada

Towards building user seeing computers*

Gilles Bessens and Dmitry O. Gorodnichy
Institute for Information Technology (IIT-ITI)
National Research Council of Canada (NRC-CNRC)
Montreal Rd, M-50, Ottawa, Canada K1A 0R6
http://iit-iti.nrc-cnrc.gc.ca

Abstract

Of all human senses, vision is the primary source of information about the outside world. When humans lose a sense of touch or a sense of hearing, they can still communicate with each other using vision. The same is for computers. When information cannot be entered into computer using hands or speech, vision could provide a solution, if only computers can see. Based on our recent work on designing a vision-based software to enable people with disabilities to work with a computer, we summarize the main problems related to the task and indicate the directions for the design of computers which can see with a specific task in hand of seeing a computer user.

1. Seeing tasks

When humans lose a sense of touch or hearing, they can still communicate using vision. The same is for computers: When information cannot be entered into a computer using hands or speech, vision could provide a solution. Therefore it is important to develop computers which can see.

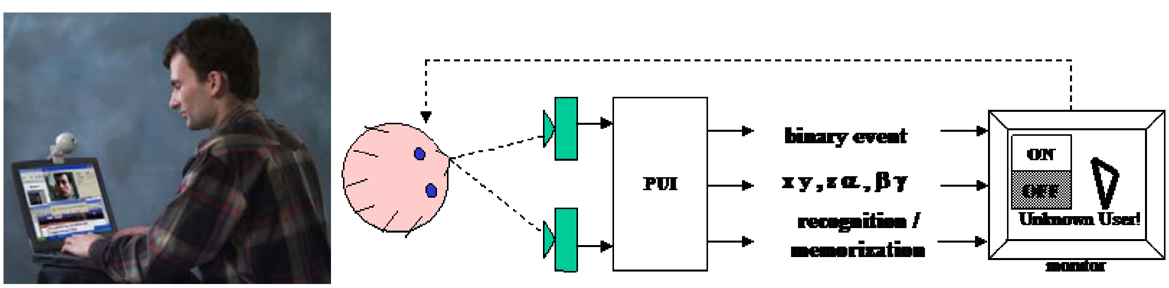


Figure 1: Seeing a user involves three tasks.

Below we define the three primary tasks involved in seeing a user (see Figure 1):

1) "Where" task - to see where is a user, i.e. to know the position of the user's face (or facial feature) in 3D. This, in general case, is represented as a tuple of six numbers: $\{x,y,z,\alpha,\beta,\gamma\}$, of which often only two can be extracted and/or used.
2) "What" task - to see what is a user doing, e.g. opening the mouth, the eyes, lifting eye-brows, turning or nodding the head etc.

*Presented at the Second Workshop on Face Processing in Video (FPiV'05) held jointly with AI/GI/CRV'05, 9-11 May 2005, Victoria, BC, Canada, 2005.

3) "Who" task - to see who is the user, which would usually imply associating a nametag to the face observed.

The first of these tasks has been already widely approached by computer vision researchers, and as a result a number of face tracking technologies have been recently developed. The main goal is therefore seen in developing a framework to combine these techniques with those to be developed for the other two tasks, and to tailor the entire thus built vision-based user perception systems for the users' needs.

2. Main results

Figure 2 shows typical results of detecting a user's face position with web-cameras using three video modalities: motion, colour and intensities (See also Table 1 which summarizes milestone results related to all three computer seeing tasks). These results show that now it is possible to quite reliably to detect an approximate position of a user's face.

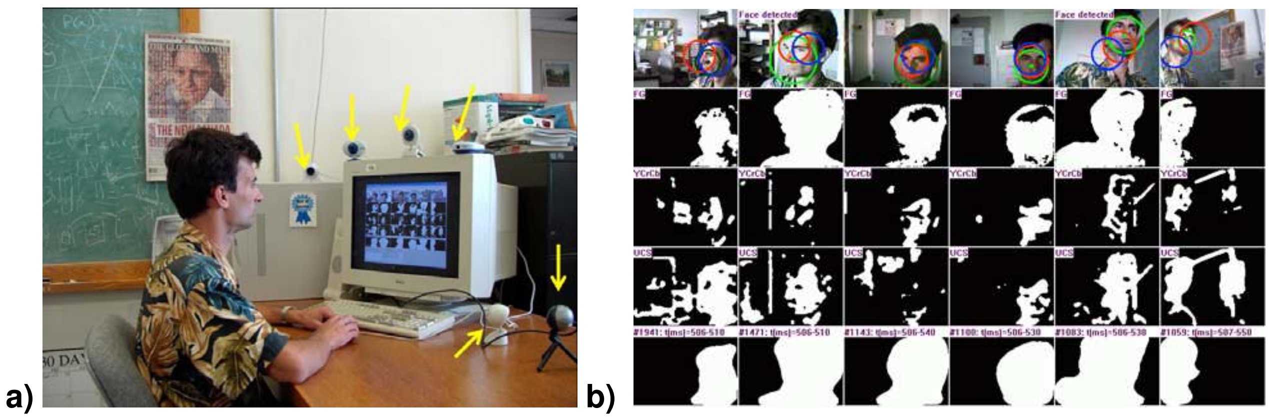


Figure 2: Typical "where" task results achieved without convex-shape nose tracking: a) computer-user setup showing six different web-cameras, b) face detection results obtained using these cameras.

The invention on the *Nouse* convex-shape nose tracking technique (see Figure 3) made it possible to detect and track the user's face, represented by the tip of the nose, with the precision sufficient for pointing the cursor and controlling the computer programs. Table 2 from the *Planeta Digital* magazine (Brazil, August 2003) lists other computer-vision based face tracking technologies developed for the application of moving a cursor.

- 1998. Proof-of-concept colour-based skin tracking [Bradski'98] – not precise
- 2001. Motion-based segmentation & localization – not precise
- 1999-2002. Several skin colour models developed – reached the limits
- 2001. Rapid face detection using rectangular wavelets of intensities [fg02]
- 2002. Subpixel-accuracy convex-shape nose tracking: NRC [*Nouse*™, fg02, ivc04]
- 2002. Stereo face tracking using projective vision: NRC [ivc04]
- 2003. Second-order change detection: NRC [*Double-blink,* ivc04]
- 2003-05. Neuro-biological recognition of low-res faces: NRC [avbpa05,fpiv04,fpiv05]

Table 1: Milestone results related to developing seeing computers.

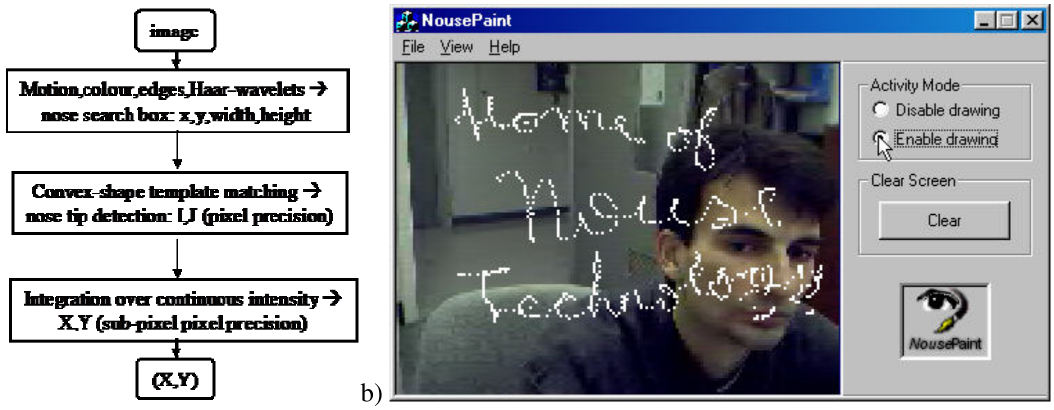
a) 

Figure 3: The *Nouse* convex-shape nose tracking: a) the flow-chart, and b) a typical result ("*Home of Nouse Technology*" phrase is written hands-free using the motion of the user's face).

ON-LINE — *Confira endereços com informações, depoimentos e softwares para deficientes físicos na rede:*

NOUSE	www.cv.iit.nrc.ca/research/Nouse/download.html	MICROSOFT	www.microsoft.com/enable
DOSVOX	http://intervox.nce.ufrj.br/dosvox	HP	www.hp.com.br/acessibilidade/soft-def.html
MOTRIX	http://intervox.nce.ufrj.br/motrix	APPLE	www.apple.com/disability
REDE SACI	www.saci.org.br	TRACE CENTER	http://trace.wisc.edu/world/computer-access/multi/sharewar.htm
DEFNET	www.defnet.org.br		
ABRAS	www.acessibilidade.org.br/softwares.htm	KOLLER	www.koller.com.br/produtos.html
VIRTUAL VISION	www.micropower.com.br	LENIRA LUNA	http://intervox.nce. ufrj.br/=lenira
JAWS	www.hj.com	INSTITUTO BENJAMIN CONSTANT	www.ibcnet.org.br
IBM	www-3.ibm.com/able	ELIZABET SÁ	www.lerparaver.com/bancodeescola

Table 2: List of vision-based face tracking technologies developed for moving a cursor (from *Planeta Digital* magazine, Brazil, August 2003)

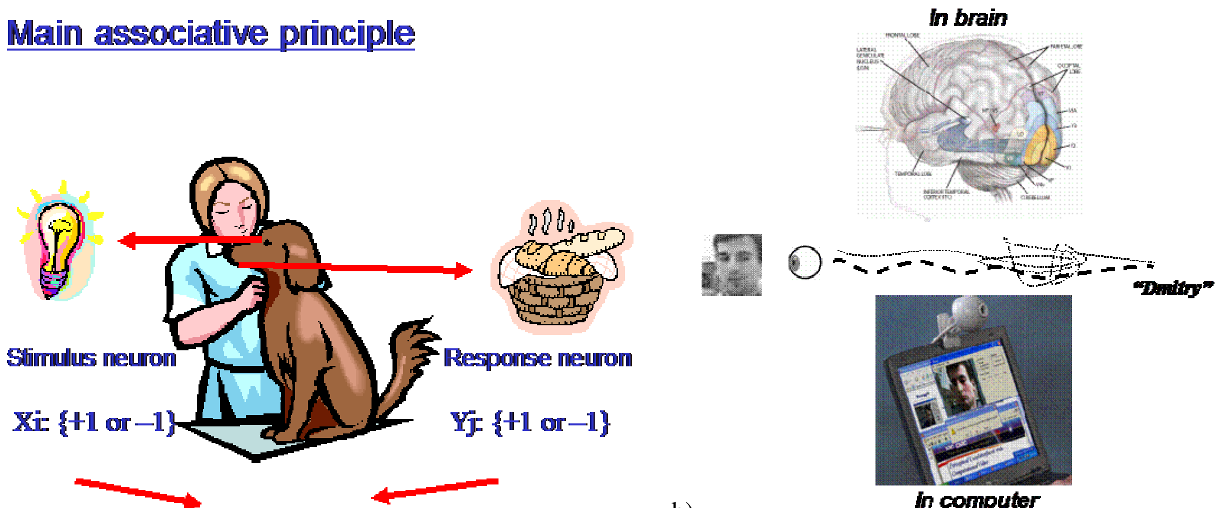
Finally, recently (see Table 1) it has also become possible to memorize and recognize face from low-resolution low-quality video, such as the one commonly obtained with off-the-shelf web-cameras shown in Figure 2.a, using the techniques developed in [1,2], the idea of which is shown in Figure 4.

3. Wish list

Based on the communication with the potential users of the technology, such as the residents of the SCO Health Center in Ottawa with accessibility needs, who would benefit the most from the development of these systems,
we identify the following tasks the users wish the seeing computer would do.

1. Detecting a user and automatically extracting the user's parameters and settings. This includes a) loading personal windows settings, such as the font size, application window layout etc, which is a very tedious work for users with disabilities. This also includes b) finding the range of user's motion it can be mapped to the computer control coordinates.
2. Enabling written communication, an example of which is typing a message in an email browser or internet.

3. Enabling navigation in Window environment, which includes selecting items from windows menus and pushing buttons of Windows applications
4. Detecting visual cues from users, such as intentional blinks, mouth opening, repetitive or predefined motion patterns, for hands-free remote control, a simple example of which is generation of mouse-type "clicks" and a more complex example of which includes generation of vision-based lexicon or computer control commands such as "go to next/last window", "copy/cut/paste", "start Editor", "save and quite".



a) b)

Figure 4: The idea behind face memorization and recognition using low-quality video data (b) is the same as behind establishing the light-saliva association in Pavlov dog's experiments (a): tuning inter-neuron connections to represent the association.

4. Constraints and problems

From the computer vision prospective, the above tasks have be to approached with the understanding of the constraints, of which the following three are identified as being of great importance:

a) computer limitations - the system should run in real time,
b) user mobility limitations - users have limited range of motion; besides, camera field of view and resolution are limited,
c) environmental limitations - changing lighting conditions may effect significantly the vision-based results.

We analyze the applicability for the problem of each of the video modalities: colour information, motion information, shape information and corners-edges information.

We also identify two very important problems related to the vision-based computer control, without resolving which a viable vision-based control can not be achieved. The first problem is the need for the missing feedback, which is what the feeling of touch provides to the users who operate with the mouse. The second problem is the need for limited-motion-based cursor control and key entry.

5. Perceptual Vision Interface *Nouse*

As a result of our deliberations, which included analyzing the above described tasks and problems and proposing the solutions to these problems, the prototype of the Perceptual Vision Interface *Nouse* system, which evolved from a single demo program [3] to a hands-free perceptual operating system combining all techniques presented, has been developed. The main interface of the PVI *Nouse* is shown in Figure 5. It runs in background as part of Windows OS and self-activates when it detects a face of a user, upon which it automatically recognizes a user or enrolls him or her as a new user. In doing that the PVI Nouse retrieves the user's motion range, the zero (0,0) Nouse position, needed for the hands-free cursor control and typing.

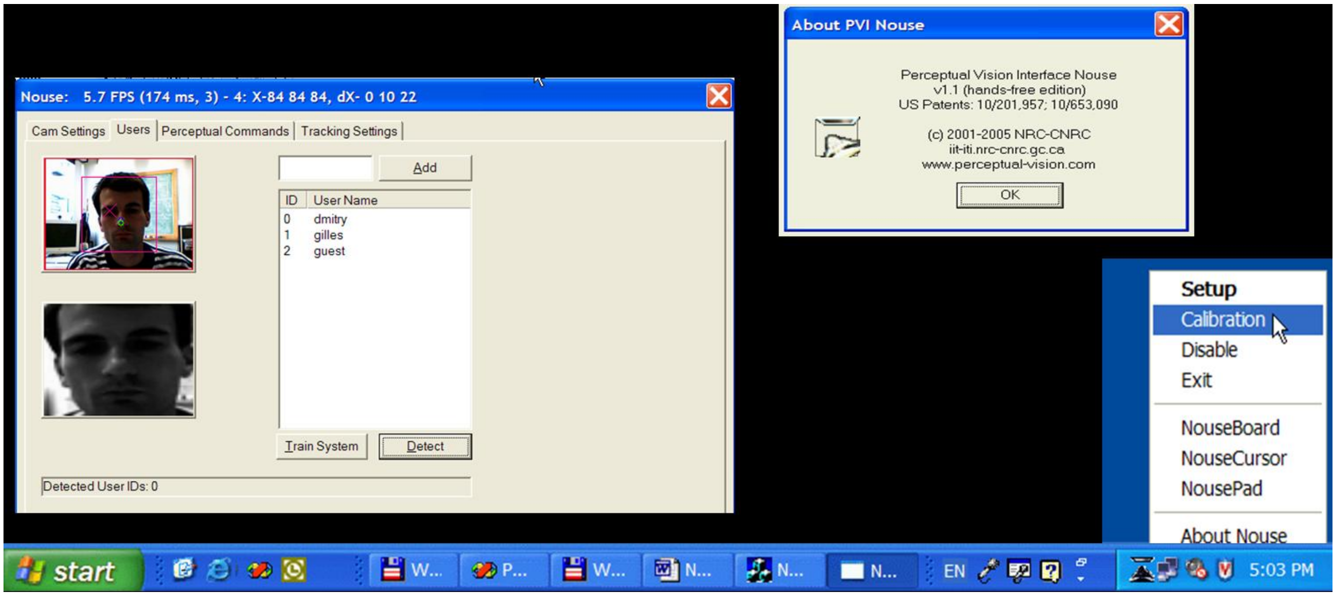


Figure 5: The snapshot to the Perceptual Vision Interface *Nouse* running within Windows OS.

While further tuning and software design improvement is required, the Perceptual Vision Interface Nouse system, has provided a clear vision on how to design a truly seeing computer capable of performing all seeing tasks.

Acknowledgement

The communication with the SCO Health Center in Ottawa, who expressed the interest to test the NRC Perceptual Vision Interfaces, is acknowledged.

References

[1] Dmitry O. Gorodnichy, "Video-based framework for face recognition in video", Second Workshop on Face Processing in Video (FPiV'05) in Proceedings of Second Canadian Conference on Computer and Robot Vision (CRV'05), pp. 330-338, Victoria, BC, Canada, 9-11 May, 2005. ISBN 0-7695-2319-6. NRC 48216.

[2] D.O. Gorodnichy, "Projection learning vs correlation learning: from Pavlov dogs to face recognition", AI'05 Workshop on Correlation learning, Victoria, BC, Canada, May 8-12, 2005. NRC 48209.

[3] Dmitry O. Gorodnichy and Gerhard Roth. Nouse 'Use your nose as a mouse' perceptual vision technology for hands-free games and interfaces. Image and Vision Computing, Volume 22, Issue 12, Pages 931-942, 2004. NRC 47140.