From Recognition in Brain to Recognition in Perceptual Vision Systems

Gorodnichy, Dimitry; Bessens, G.

National Research Council Canada   Conseil national de recherches Canada

Canada

National Research
Council Canada

Conseil national
de recherches Canada

Institute for
Information Technology

Institut de technologie
de l'information

NRC·CNRC

*From Recognition in Brain to Recognition in Perceptual Vision Systems. Case Study: Face in Video. Example: Identifying Computer Users with Low-Resolution Webcams ***

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

Canada

From recognition in brain to recognition in perceptual vision systems.
Case study: face in video.
Example: identifying computer users with low-resolution webcams. *

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

Abstract

This presentation summarizes the effort of our group in designing a non Von-Neumann, biologically motivated approach to video processing and recognition. The need for such an approach is seen from the very fact that, while for humans recognition in video is easy, most recognition approaches developed to date still perform very poorly when applied to video data. Instead of focusing the effort on making the video data of better quality, as suggested by the Face Recognition Vendor Test Grand Challenge, we start from the premise that video data is *inherently* of bad quality, and hence the effort should be directed towards developing approaches which can deal with such low-quality data. We build our approach using the neuro-associative mechanism which is known to be of prime importance for biological vision recognition systems in enabling the accumulation of learning data in time and which is implemented by means of tuning the synaptic connections of a multi-connected neural network. As a result, we build a memorization-recognition system which can memorize a face from a *video sequence* and then identify this face in another *video sequence*. The testing of the system is deliberately done on video of very low quality, i.e. such that is just sufficient for humans to identify the faces. The application of the system for identifying computer users using a low-resolution webcam is shown. Other applications include video-annotation of TV programs and tele-conferences, and fusion of hard and soft biometrics for homeland security.
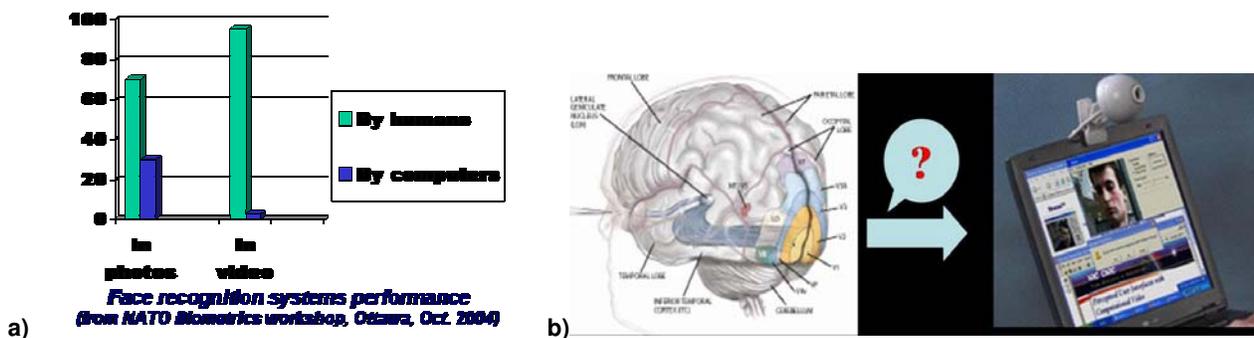
Figure 1: The solution to face recognition in video problem is seen in the way biological systems recognize objects.

1. Presentation of the problem

As the title indicates and illustrated in Figure 1, our goal is to design computer-based recognition systems using the mechanisms employed by the biological recognition systems. The need for the development of biologically driven recognition systems is seen from the very fact that, despite many resources invested recently into the development of the systems capable of recognizing faces in video, there has not been practically any success story related to the Face Recognition in Video (FRiV) problem yet.

At first glance, such a poor state of things may appear puzzling, knowing how much success has been already achieved by the face recognition vendors in designing the systems capable of recognizing faces in documents. However, should we examine closer the situation, we may find such an outcome not surprising at all, because, as summarized in Table 1 and illustrated in Figure 2 (see also [2,6]), *the facial data taken by a photographer's photo-camera and the facial data captured by a surveillance camera belong to two different modalities.* What is surprising however is that, despite the above stated fact, the direction for the future development of face recognition technologies, as presented by the face recognition vendors test website [1], is seen "in making the video data of better quality" instead of developing approaches which can deal with low-quality data, such as video data is.

One way to finding the approaches that can deal with memorization and recognition from low-quality data is to study the way memorization and recognition is done in human brain, since humans do not have problem memorizing and recognizing faces from TV which shows faces of very modest resolution (close to 10-20 pixels between the eyes) and quality.
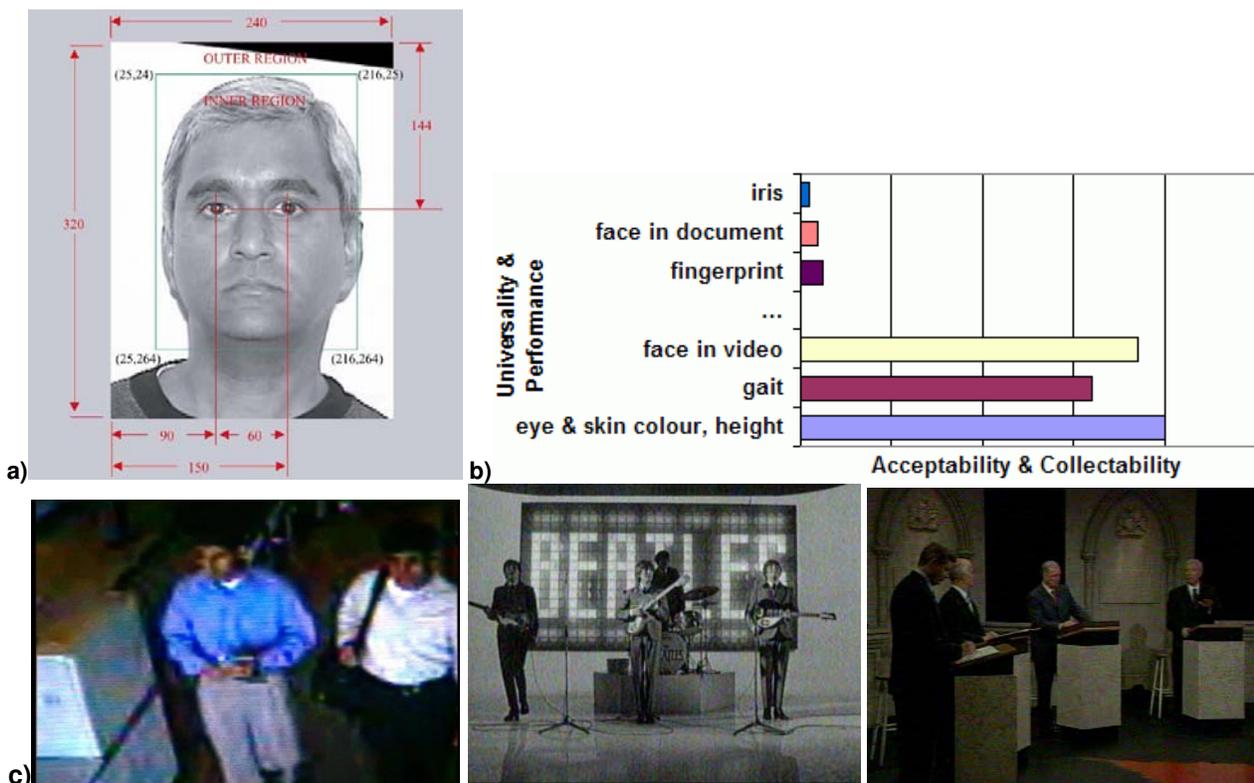

Figure 2: Facial data on photographs and video belong to two different modalities.

Photographic facial data and video-based facial data are two different modalities:	
• different nature of data , different biometrics , different approaches , different testing benchmarks	
In video: faces *are meant to be* of low quality and resolution.	
Humans: recognize faces on TV of 12 pixel between the eyes all the time	
Photos:	**Video:**
Taken in controlled environment *(similar to fingerprint registration)*	Taken in unconstrained environment, *in a "hidden" camera setup (similar to gait recognition or height estimation)*
• high quality • face "nicely" positioned • high resolution Nominal face resolution: 60 pixels between the eyes	• People don't look into camera • Poor illumination • Blurriness, bad focus, interlacing • Individual frames of poor quality • *Resolution: 320x240 is the top in most cases* • *Video can also be MPEG compressed (for storage & transmitting)* Nominal face resolution: 12 pixels between the eyes

Table 1: Facial data in photographs vs facial data in video.

1.1 Recognizing computer users

Beside the already hinted applications of FRiV to security and TV annotation, there is another very important application of this problem – designing Perceptual Vision Systems (PVS), which are computer systems that can see a user (see Figure 3). There are three primary seeing tasks we identify for this application [5]:

- the "Where" task - to see where is a user, i.e. to know the position of the user's face (or a facial feature) in 3D, so that it can be used to point or control the computer,

- the "What" task - to see what is a user doing, e.g. opening the mouth or the eyes so that it can be used to sent the binary "click"-type commands to the computer, and

- the "Who" task - to see who is the user so that the user settings and defaults can be retrieved *hands-free*.
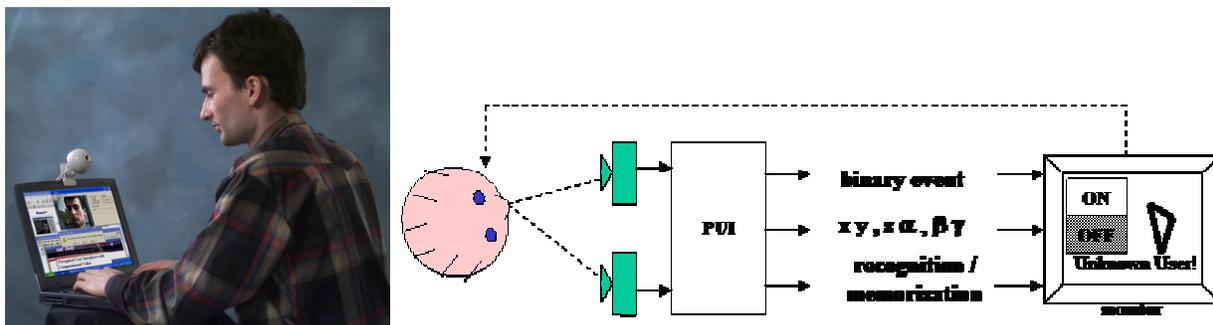
Figure 3: Seeing of a user involves three tasks.

While the first two of these tasks have been already widely approached by computer vision researchers, and as a result a number of face tracking technologies have been recently developed, the third task seems to be not well addressed nor has it a good solution. The reason for this is the same - lack of recognition approaches capable of recognizing faces in real-time, under unconstrained conditions, with low quality web-cameras.

This work is therefore aimed to developing FRiV techniques to be used and tested in Perceptual Vision Systems. In a formalized way our task is stated as follows: *given a 160x120 video showing a*

face of a user occupying 1/8 – 1/4 of the image (i.e. having a resolution of about 12-20 pixels between the eyes), identify who of the already registered by the system users he or she is.
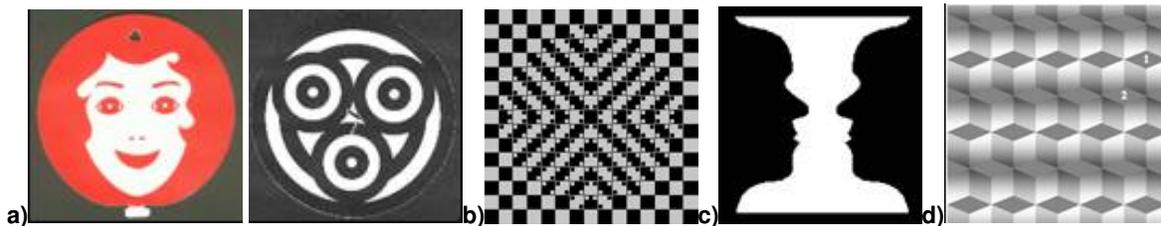

Figure 4: Pictures illustrating several phenomena related to the visual perception by humans, namely: saliency–based attention (a), accumulation over time and space (b), time-filtered context-based decision (c), fovea vision and local brightness adjustment (d).

2. How brain does it: the model of the process

Of the mechanisms that contribute to the excellent ability of humans to recognize faces of low resolution (of which some are illustrated in Figure 4) we emphasize the following. *We use an efficient neuro-associative mechanism which allows us to accumulate learning data in time and in space.* This neuro-associative principle (illustrated in Figure 6) is accomplished in living organisms by using the specially designed by the mother Nature cells located in visual cortex of brain called neurons, which are connected among each other through the system of axons and synapses (see Figure 5). It is the strength of these synaptic connections which makes associative accumulation and recognition of information possible. The number of neurons in the human brain is estimated to be in the range of 10^{10} to 10^{13}, and the number of synaptic interconnections is even larger number, which makes it possible for humans to memorize practically unlimited quantity of information.
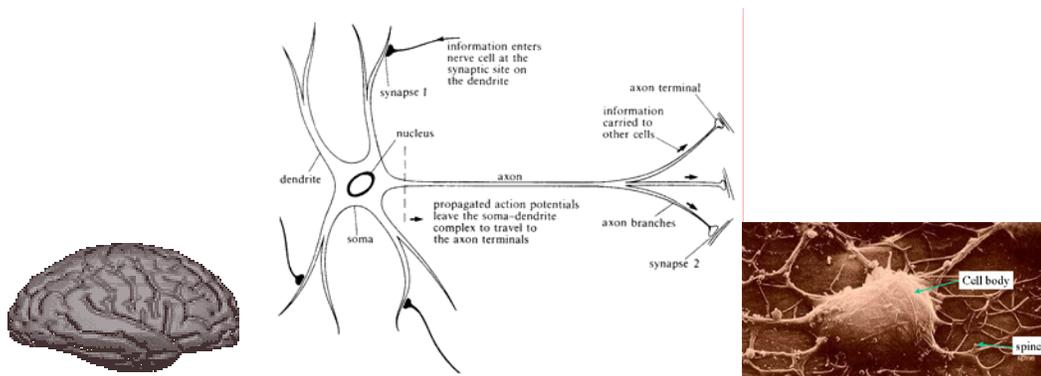
Figure 5: Brain memorizes by means of adjusting inter-neural synaptic connections, which is formalized as computing weights Cij based on the current stimulus V and the current memory content C.

What makes neuro-associative memorization and then the retrieval efficient are the following three properties related to the neural processing: *1) non-linear processing, 2) massively distributed collective decision making, and 3) synaptic plasticity.* These properties can be models as follows.

- Synapses which, for simplicity and because we do not know exactly what is connected in the brain to what, are assumed to connect all neurons, are represented using real value numbers Cij.
- Neurons, which are either in rest or activated depending on values of other neurons Yj and the strength of synaptic connections (Cij), are modelled as binary values Yi={+1,-1} .
- Brain is thus a network of "binary" neurons evolving in time from the initial state (e.g. stimulus coming from retina) until it reaches a stable state – attractor, according to the update rule:

$$Y_i(t+1) = \text{sign}(\sum_{j=1}^{N} C_{ij}Y_j(t))$$

- Attractors are what the brain retrieves as a result of an associative recall.

The main question in this model is how to tune synaptic values Cij so that the attractors are the desired memories and so that the best associative recall is achieved. This synaptic tuning should be done so that the computation of the synaptic values

- does not require the system to go through the already presented stimuli, i.e. there are no iterations involved, and
- would update the synapses based on the currently presented stimuli pair only, without knowing which stimuli will follow, i.e. no batch mode is involved.

These two conditions represent the idea of incremental learning, one of the best implementations of which is the projection leaning rule described in [4,8] which can be written as follows.

$$C_{ij}^m = C_{ij}^{m-1} + \frac{(v_i^m - s_i^m)(v_j^m - s_j^m)}{E^2}$$ where $$E^2 = \|\vec{V}^m - C^{m-1}\vec{V}^m\|^2$$
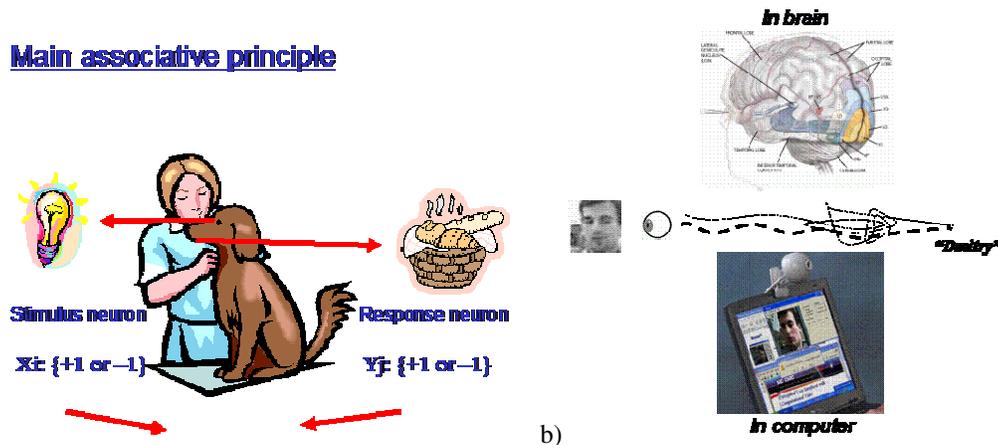

Figure 6: The idea behind face memorization and recognition in webcam video (b) is the same as in establishing the light-saliva association in Pavlov dog's experiments (a): tuning inter-neuron connections to represent the association.

3. The demo programs

The method of incremental learning described above is implemented in the program which is made available for testing from our web-site. This program, two snapshots of which are shown in Figure 7, allows one memorize a face either from an avi video file or a webcam connected to the computer. In the recognition mode, it shows the output of the neurons responsible for retrieval of the user's name. It also shows the state of the synaptic matrix, represented as a grey-scale image.

The presented ideas are also implemented in the Perceptual Vision Interface system *Nouse* [5,7]. The Nouse PVI system runs in background as part of Windows OS and self-activates when it detects a face of a user, upon which it automatically recognizes the user. The snapshot of the *Nouse* interface is shown in Figure 8. The details on how the facial image is extracted from video and converted to a binary feature vector Y={Yi} can be found in [2,3].
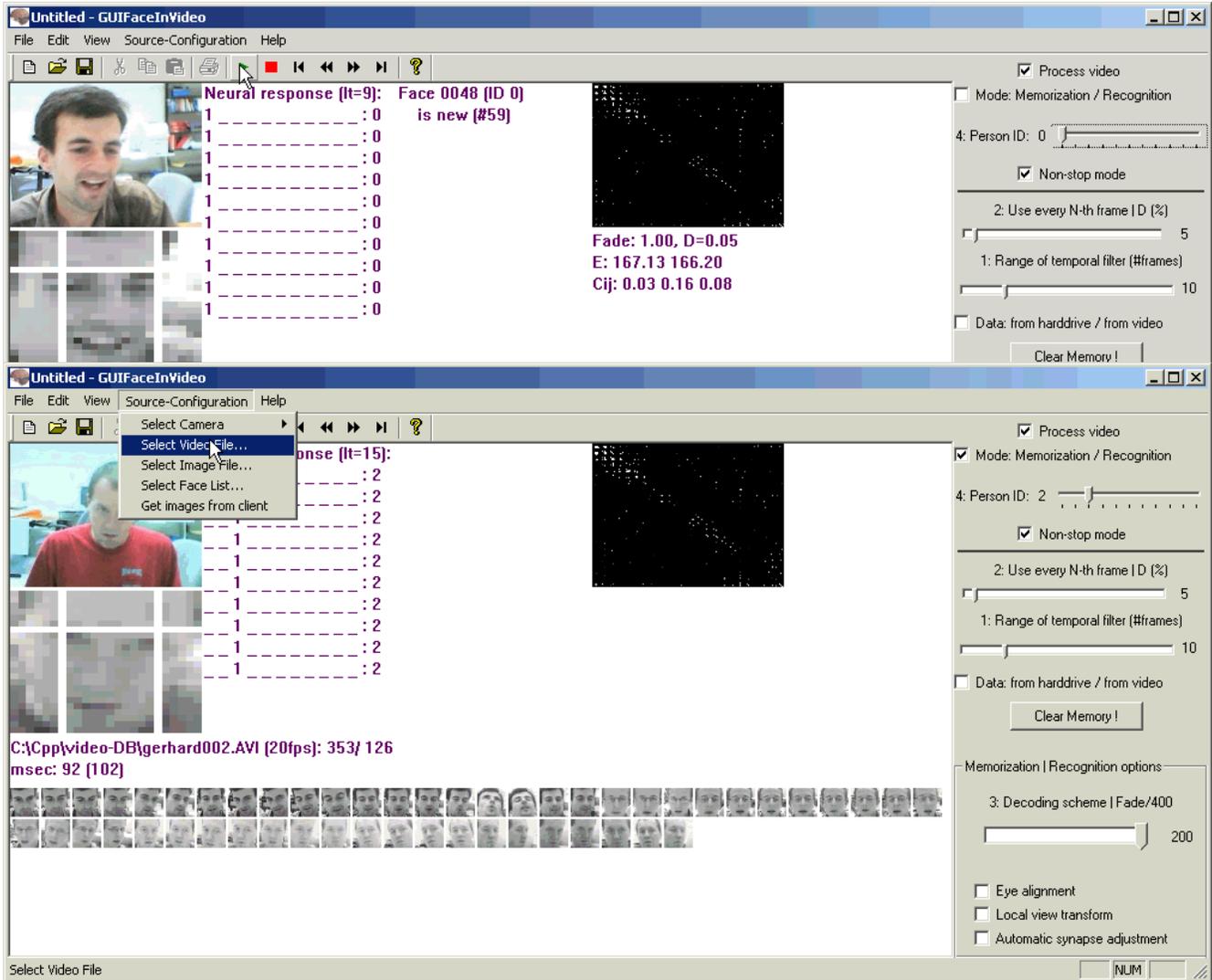

Figure 7: The snapshots to the FRiV testing program. Facial images shown in the bottom are used to tune synaptic connections corresponding to the video-name association

4. Conclusion

By looking at the facial images extracted from a web-cam shown in Figures 7 and 8, one can notice that, in fact, the resolution, the quality and the variation of orientation and expression of these facial images is such that they could be considered practically useless for the traditional face recognition techniques. For the neuro-associative techniques we have developed however these images provide enough information both for learning a face and for its memorization.

To meet the real-time constraint, we have limited the size of the face recognizing brain model to N=1739 neurons. This size allows us to encode on the adjacent pixels relationships (vertical and hori-

zontal gradient images of the face), which seems sufficient of the application where only a few faces are stored. For the scenarios with the larger number of faces to be memorized, one has to wait until computers become faster or perform the processing off-line so that other inter-pixel relationship can be used to increase the memorization capacity of the system and to improve it recognition rates.
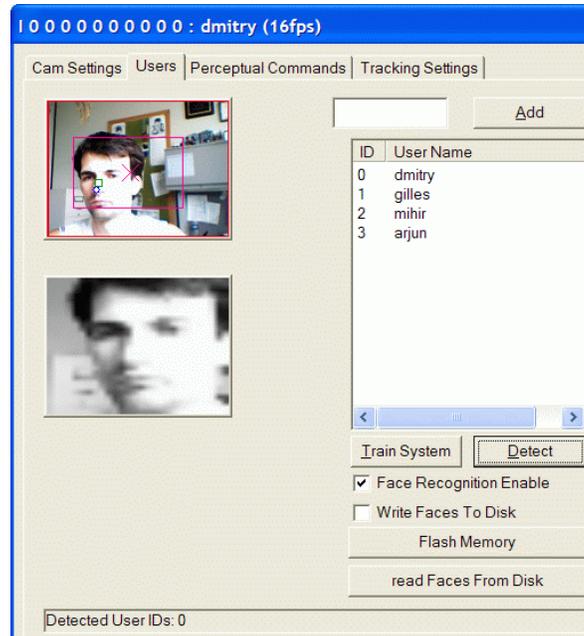

Figure 8: The snapshot to the Perceptual Vision Interface *Nouse* running within Windows OS.

References

[1] P. J. Philips, P. Grotherand R. J. Michealsand D. M. Blackburnand, E. Tabassi, and J. M. Bone, "Face recognition vendor test 2002 results: Overview and summary," in http://www.frvt.org, 2003.

[2] Gorodnichy, D. **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. NRC 48216.

[3] Gorodnichy, D. **Associative neural networks as means for low-resolution video-based recognition**. International Joint Conference on Neural Networks (IJCNN'05). Montreal, Quebec, Canada. July 31-August 4, 2005. NRC 48217.

[4] Gorodnichy, D. **Projection Learning vs. Correlation Learning: From Pavlov Dogs to Face Recognition**. In Correlation Learnings AI'05 Workshop. May 8, 2005. Victoria, B.C.. NRC 48209.

[5] Bessens, G., Gorodnichy, D. **Towards Building User Seeing Computers**. Second Canadian Conference on Computer and Robot Vision Workshop on Face Processing in Video (FPiV'05). May 9-11, 2005. Victoria, B.C. NRC 48210.

[6] Gorodnichy, D. **Recognizing Faces in Video Requires Approaches Different from Those Developed for Face Recognition in Photographs**, NATO IST - 044 Workshop on "Enhancing Information Systems Security through Biometrics". Ottawa, Ontario, Canada. October 18-20, 2004. NRC 47149.

[7] 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 , 1 October 2004, Pages 931-942, 2004. NRC 47140.

[8] D.O. Gorodnichy, A.M. Reznik. **Increasing Attraction of Pseudo-Inverse Autoassociative Networks**, Neural Processing Letters, volume 5, issue 2, pp. 123-127, 1997.