Recognizing Faces in Video Requires Approaches Different from Those Developed for Face Recognition in Photographs
Gorodnichy, Dimitry

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

*Recognizing Faces in Video Requires Approaches Different from Those Developed for Face Recognition in Photographs ***

Gorodnichy, D.
October 2004

Canada

Recognizing faces in video requires approaches different from those developed for face recognition in photographs

Dmitry O. Gorodnichy
Computational Video Group
Institute for Information Technology
National Research Council of Canada
M-50 Montreal Rd, Ottawa, Ontario, K1A 0R6, Canada

Email: Dmitry.Gorodnichy@nrc-cnrc.gc.ca

ABSTRACT

With the increasingly growing demand for security and biometrics systems the problem of detecting and recognizing faces using video cameras has been given a lot of attention by many groups and individuals: from the media to the government to the military. At the same time, there seems to exist a misconception in that the problem of recognizing faces in video is often treated as an extension of another problem – that of recognizing faces in photographs. From the computer vision research point of view, however, this is an erroneous assumption, and it may lead to wrong understanding on what can be done and what not using video data.

On one hand, due to the nature of video, video processing has to deal with much lower resolution and much lower quality of images, compared to that in photograph processing. Even assuming that the lighting conditions are perfect when taking a snapshot, which is rarely true, the object of interest may be located too far from the camera or at angle which makes recognition very difficult. On the other hand, video images can be easily acquired and they can capture the motion of a person. This makes it possible to track people until they are in a position convenient for recognition. However, this requires techniques for automatically tracking and detecting people in video, because it is impossible in practice to examine large amounts of video data manually. Therefore, it is important to understand the limitations and the advantages of video so that unnecessary expenses can be avoided and proper allocation of resources to the problems can be done.

As a scientist actively working in the area, I will introduce the audience to the state of mind and the state of art in the area of recognizing faces in video. In my presentation, I will cover the main scenarios within which face recognition in video has been already proved to be successful. I will also describe the main limitations of the video-based recognition technologies.

1.   LEGACY OF FACE RECOGNITION

Research in face recognition has started from the problem of face recognition in photo-images and as such it has a long history and has produced a number of well established results and standards. One of the best known and most used results is the development of the canonical face model deployed by the International Civil Aviation Organization (ICAO) for the Machine Readable Travel Documents [1]. This is the model used to store face images in databases and many passport and immigration offices in many countries presently use it.

Following the Government of Canada Forum on Biometrics held last year, this canonical face model might be chosen to for store faces in the national identity cards [2]. However, will this face model, shown in Figure 1, be appropriate for recognizing faces in video?
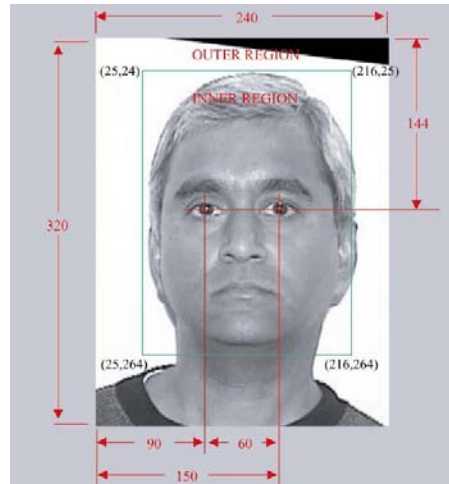

Figure 1: Canonical face model adopted by ICAO for recognizing faces in travel documents.

2. VIDEO IMAGES VS PHOTO IMAGES

The problem is not only that in video a person will very unlikely look straight into the camera. The problem is also in that the quality of face images in video is very much different from their quality in photo-images (see Figure 2).

Photograph images are taken under good illumination, with sufficiently fast exposure, wide aperture and good focus to avoid any blurriness in the picture, and as a result they can capture human faces in very fine detail. It makes sense therefore to store photo-images using high resolution, as this is done in the canonical face model of Figure 1 which stores faces using 60 pixels between the eyes, so that the details of the face can be explored later if needed. And in fact, most face recognition approaches developed to date, such as those based on extracting local features, rely upon the high resolution of the images.

The images captured by a video camera, on the other hand, may not and, in fact, should not be assumed to be of high resolution and good quality. Some of such images captured by the Washington Dulles International Airport surveillance camera showing the hijacker of 9/11 became very well-known around the world [3]. From a human prospective, the video captured by the surveillance camera is good enough to recognize the persons, provided that we saw these persons before. For computerized facial recognition technology developed to date, however, this video is practically useless. Thus it is important to realize that it is not the resolution that matters the most for the problem of recognizing people in video, but the approaches used.
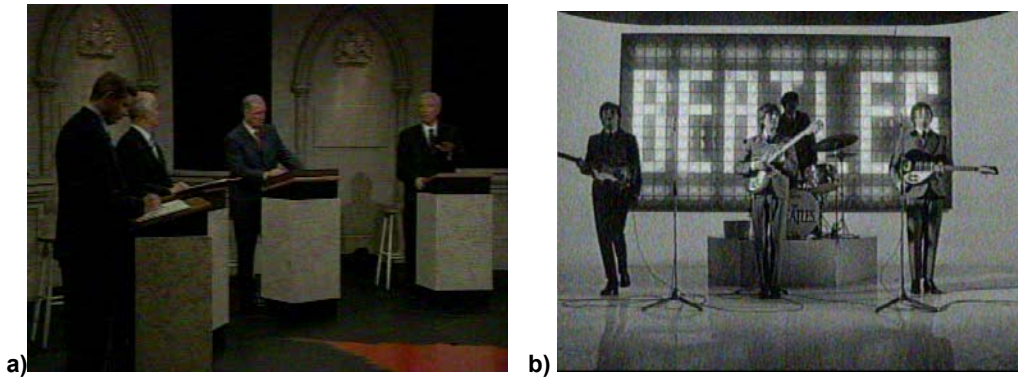
Figure 2: Video sequences which can be used for testing facial recognition systems.

3.   VIDEO-DRIVEN SOLUTIONS

For humans it is easier to recognize faces when seeing them live rather than seeing them in photographs. For computer-based facial recognition however, the way it is approached now, this is the opposite. Computer technologies, which have been already shown to be quite capable of recognizing faces in still images (e.g. [4,5]), exhibit very poor performance when dealing with video (including the above mentioned [4,5]). It seems therefore appropriate to assume that the development of successful vision-based face recognition systems will benefit the most from the expertise in computer and biological vision rather than from the conventionally used expertise in image processing and pattern recognition.

The main reason why humans are so successful in recognition of faces in video, while computers are not, is that, first, humans track an object prior to its recognition, and second, humans use the collection of data over time, rather than an individual video image, when both memorize and recognize an the object. As outlined below, the recent results in face processing in video [6], allow us to build computer-based face recognition systems based on the same principle.

3.1 From tracking and capturing a face to its memorization and recognition

Humans are known to be able to recognize faces in very low resolution images, as long as there are at least 10 pixels between the eyes. A few computerized face detectors are now available which can automatically and in real time detect faces in video of almost the same small size (10 pixels between the eyes), provided that they are facing the camera. The range of facial orientation and expression which is tolerated by these face detectors is shown in Figure 3. For a regular 640x480 video camera this means that we can detect and *track* a person in video within a large range of motion even when his/her faces occupies only one thirtieth of the image. This is close to the scenario with a surveillance camera in the airport. If a face is too small to be detected by a face detector or if it faces away from the camera, then motion detection techniques allow us to keep it all the time in the view of the camera.

Figure 3: A variety of facial orientations and expressions tolerated by real-time face detectors.
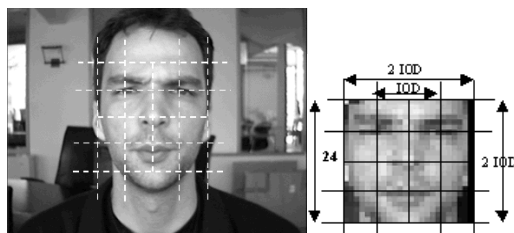

Figure 4: Canonical face model suitable for face recognition in video

It has been also realized that only the inner part of the face is important for recognition. This finding led many computer vision scientists to use the canonical face model similar to the one shown in Figure 4. When decimated to 24 x 24 size, this face model is shown to be good for memorization and recognition of faces in video on fly in real time.

Finally, a recently NRC-CNRC developed associative framework for memorizing and recognizing faces [7] offers a way to deal with a flow of video data, allowing us to combine several images, showing different facial orientations and/or expressions of the same person, to yield a robust facial memory.

CONCLUSION

Despite the fact that present facial recognition technologies have not yet been shown to be successful in recognizing faces in video, the situation may soon change for the better. However, this will happen only if the problem of recognizing faces in video will not be considered as a mere extension of the face recognition from photographs problem, which is unfortunately the case now. Also, a proper testing benchmark should be used for all existing and new video-oriented face recognition technologies. In particular, these technologies should always be tested on video sequences rather than on individual snapshots. That is, a system should be able, after having seen a *video sequence* of a person, to recognize this person in another *video sequence*. Examples of such tests could be recognizing people on TV, starting from low-complexity scenarios such recognizing guests of a talk show (as in Figure 2.a) to middle-complexity scenarios such recognizing the same musicians on different stages (as in Figure 2.b) to high-complexity scenarios such as recognizing actors in different movies. Unless this is done, in a few years we may find ourselves in same cul-de-sac situation of having spent a lot of resources for building computerized facial recognition systems and then having an army of personnel watching video identifying persons.

ACKNOWLEDGEMENT

A demo of the face recognition program developed at NRC-CNRC which can memorize and recognize faces, such as those seen in video of Figure 2, will be shown. The program runs on a 2.1GHz laptop equipped with

an off-the-shelf web camera.

REFERENCES

[1] International Civil Aviation Organization (ICAO): http://www.icao.org

[2] Citizenship and Immigration Canada Forum on Biometrics. October 7 & 8, 2003, Ottawa, Canada. http://www.cic.gc.ca/english/pub/biometrics

[3] "Video shows 9/11 hijackers at airport" (CNN report) http://www.cnn.com/2004/US/07/21/attacks.surveillance.video.ap

[4] Face recognition companies: http://www.cognitec-systems.de, www.identix.com, http://www.viisage.com, http://www.bioid.com

[5] Canadian face recognition companies: http://www.visionspheretech.com, http://www.findbiometrics.com, http://www.biodentity.com

[6] The First IEEE Workshop on Face Processing in Video, June 28, 2004, Washington, D.C., USA www.visioninterface.net/fpiv04

[7] "Seeing Faces in Video: By Humans and Machines. Recent Advances in Perceptual Vision Technology" (D.O. Gorodnichy), NRC-IIT 2004-2005 Colloquium Series Talk. http://iit-iti.nrc-cnrc.gc.ca/colloq/0405/04-09-16_e.html