Seeing Faces in Video by Computers: editorial for special issue on Face Processing in Video Sequences

Gorodnichy, Dimitry

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.imavis.2005.11.001
Image and Video Computing, 24, 6, pp. 551-556, 2006-06

National Research Council Canada
Institute for Information Technology

Conseil national de recherches Canada
Institut de technologie de l'information

NRC · CNRC

*Seeing faces in video by computers **

Gorodnichy, D.
May 2006

* published in Image and Video Computing, Vol. 24, No. 6 (Special Issue on Face Processing in Video Sequences, Editor: D.O. Gorodnichy). May 2006. pp. 551-556. NRC 48295. Editor: D.O. Gorodnichy.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Image and Vision Computing 24 (2006) 551–556
www.elsevier.com/locate/imavis

Editorial

Seeing faces in video by computers. Editorial for Special Issue on Face Processing in Video Sequences

1. Video-processing phenomenon

The beginning of the 20^{th} century was met with the invention that changed forever the way people saw the world. The invention, named the motion picture, was made using a technological breakthrough that enabled displaying pictures so fast (faster than 16 frames a second) that the human eyes perceived them as a single object—a movie.

Now, in the beginning of the 21^{st} century, it is fascinating to see that we are experiencing the arrival of another phenomenon related to video. This time it is due to the recent technological advances in computer and video hardware, which enable us to process video in real-time (faster than 16 frames a second), opening the way for new vision-based applications and technologies.

Many of the newly arrived vision-based applications such as computer–human interfaces, multi-media, immersive and collaborative environments, video conferencing, video coding and annotation, video computer games, entertainment, public surveillance, information security, biometrics at a distance, to name a few, have a human face as the object of prime interest. This is why a major research effort has been recently put into face processing in video, which is an area of research dedicated to the extraction and manipulation of information about human faces in video sequences.

The most important tasks of face processing in video are face detection, face tracking, and, of course, face recognition. The problem of recognizing faces from video, however, should not be considered as a mere extension of the problem of recognizing faces in photographs, since there are critical differences between the two, both in terms of the nature of processed data and approaches used.

On one hand, because of real-time, bandwidth, and environmental constraints, video-processing has to deal with much lower resolution and image quality, when compared to photograph processing. Even assuming that the lighting conditions are perfect when taking a video snapshot, which is rarely true, faces in video 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 a person over a period of time. This makes it possible to track people until they are in a position convenient for recognition and also to accumulate evidence over time.

Besides, image-based face recognition traditionally belongs to the field of pattern recognition, and as such is mainly driven by mathematical principles. By contrast, video-based face recognition may also be approached by using neurobiological principles, the study of which may hopefully result in making the performance of computer vision systems closer to that of biological vision systems.

2. Face processing we do

Another part of the video-processing phenomenon is that it is part of everybody's every day life. We, humans, do it all the time and, in fact, we even know how well the performance of our biological face processing system works. This provides both an inspiration and the ultimate bench-mark for the computer-made video-processing systems [10]. For example, consider the TV programs, which we watch on a daily basis. The video resolution of TV is only 352×288 (PAL/SECAM) and 352×240 (NTSC) pixels. In a common scenario when a camera shows several people, the faces often occupy 1/16th of the image width, meaning that resolution of the faces in video, measured by the intra-ocular distance (i.o.d.), is only about 12 pixels. And yet we do not have a problem recognizing faces on TV!

Here are a few other known observations about the properties of the face processing performed in our brain [11]. People recognize faces in grey-scale images just as well as they do in colour images. Motion and colour, while not assisting the recognition itself, are used to effectively focus our attention to the area of interest. Eyes are the most salient points on a facial image and are attended first by humans and animals; intra-ocular distance serves as a natural measurement unit used to relate facial parts to one another. People are believed to memorize several representative images of a face, rather than build a 3D face representation based on several images. Finally, for people it is often easier to memorize and recognize a person's face from video than from a photograph, given the variety of facial poses and expressions the video can capture.

doi:10.1016/j.imavis.2005.11.001

Some of these biological vision performance results have been already matched by their computerized counterparts, while some, such as the last mentioned, have still not [25]. The debate, however, also exists on whether computer systems should try to imitate the biological exemplars altogether, rather than to use the power that the technology endows us to invent new solutions which would be just as good if not better than that of the mother Nature.

3. Precursory workshop

The state of mind presented above was the inspiration for the organization of the *First Workshop on Face Processing in Video* (FPiV'04), which was held on 18 July, 2004 in Washington, DC, USA, in conjunction with the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'04), and which was the precursor of this Special Issue on Face Processing in Video Sequences.

The FPiV'04 workshop was aimed at providing a forum for scientists from different backgrounds: biological vision, computer vision, pattern recognition, machine learning, computer–human interaction, etc.—to share their experiences and to develop approaches which would make use of the advantages of video for face processing. A total of 43 papers from 11 countries were submitted, of which 30 were accepted to appear in the workshop proceedings. Based on reviewers' recommendations, authors of 11 papers were invited to revise and extend their workshop papers for consideration of publication in this special issue. The revised papers then went through a separate and independent review process managed by the IVC editors to assure that the quality of the special issue meets the IVC's publication standard. The nine papers that appear in this special issue represent the result of that review process and the final selection.

4. State of the art: perusing the papers

The papers brought to the readership of the IIVC show a good balance of face processing in video research areas. They are organized according to the order by which biological vision systems are believed to process faces: starting from the papers that address the problem of face detection and 2D tracking [1–3] to the papers that focus on 3D face tracking [4,5], to the papers that analyze facial expression [6,7] and then to the papers that deal with the recognition of faces in video [8,9].

The paper by Bonaiuto and Itti commences the issue by presenting a biologically inspired approach to narrowing the field of search for face recognition by using the bottom-up saliency-based attention mechanisms. The approach computes saliency maps for each video frame using (i) intensity, (ii) colour, (iii) orientation, (iv) flicker, and (v) motion information of the video sequence. The fixation points are then computed using a winner-take-all neural network with lateral and local inhibition, which insures that most salient parts of the video are attended in sequence. The advantage of this approach is demonstrated using the facial landmark based face recognition, in which faces are matched to one another using the actively used in computer vision local scale invariant SIFT features developed by David Lowe. The experiments are carried out using the colour video clips taken from television programs, with i.o.d. ranging from 10 to 40 pixels. The proposed visual attention guidance technique runs in real-time and is shown to speed up the recognition process.

One of biggest milestones for the area of face processing in video was the work of Paul Viola and Michael Jones [15] followed by its implementation by Reiner Lienhart [16] for the Intel OpenCV public domain library [17,18]. It not only demonstrated that it is possible to detect face in video in real-time, but also provided many researchers with the tool to do that. The Viola–Jones face detector, used and enhanced by several authors of this issue [3,7,9], detects close-to-frontal horizontally aligned faces using a cascade of pretrained classifiers of increasing complexity applied to binary rectangular Haar-like binary wavelet features efficiently computed from video frames using the 'integral image' preprocessing technique.

It is worth mentioning that for the Viola–Jones face detector the canonical face size, which is the size used to store and detect faces and which is also the smallest size that can be detected, is 20×20 (or 24×24) pixels, with i.o.d. equal to 12 pixels, which matches exactly the resolution humans are able to deal with.

In parallel with Viola and Jones, several other research groups around the world have been developing their own versions of rapid face detectors, one of the most successful of which is presented in this issue. Christian Küblbeck and Andreas Ernst continue the work of their colleague Bernhard Fröba and describe their method for rapid face detection. As in the approach of Viola–Jones, the cascade of classifiers of increasing complexity is used. The analysis is done on 22×22 image patches. The binary face features used in the work are obtained by using 3×3 local structure kernels originally proposed for detection of edges, corners and interest points. The thrust of the paper is to show that with the existence of rapid face detectors, the face tracking problem can be replaced by performing face detection in each frame. This resolves the 'initialization' and 'lost tracking' problems. In order to smooth the jitter in the obtained face positions, the authors propose to use the Kalman filter. The demo of the proposed technique, which runs with off-the-shelf USB cameras, is made publicly available at the authors' website.

The authors of the next paper, Zhengrong Yao and Haibo Li, note that one of the Viola–Jones face detection fails to detect rotated faces. In order to circumvent the problem, the authors propose to track faces from one frame to another by performing the deformable graph matching, where deformable graphs of faces are made using the same Haar-like facial features that were used in face detection and where the Viterbi algorithm, a popular Dynamic Programming technique, is used to perform the match. The proposed technique is shown to be superior to histogram-based face colour tracking on the basis of its better encoding of the structural knowledge of the object. The work is motivated by the Model-Based Coding application, which is a coding technique aimed for low bit rate video transmission and for which real-time facial information extraction is required to achieve very high video compression. The approach is tested using the publicly available database of facial motions [21] containing 320×240 video sequences with i.o.d. ranging from 30 to 40 pixels.

As higher resolution of the faces becomes available, it becomes possible to track face not only in 2D, i.e. with respect to image coordinates, but also in 3D, i.e. with six degrees of freedom. This is addressed by the next two papers of the issues.

In their paper, Le Lu, Xiang Tian Dai and Gregory Hager propose an improvement to the Particle Filtering (PF) technique. This technique also known as Condensation is frequently used for sequential state estimation, and in particular for vision-based face tracking, where particles to be tracked are generated using the frequently used in computer vision Harris corner detectors. The authors tackle one of the main deficiencies of the technique, which deals with a very large number of points needed for tracking. By embedding the Random Sampling (RANSAC) into the particle selection process, the authors significantly reduce the number of particles, make the approach much less dependable on the model of state dynamics, and, as a result, make tracking much more robust. Several variations of their combined technique (RANSAC-PF) are described, including the one which makes use of temporal coherency of the propagating particles to improve the overall tracking performance. The authors present results obtained on simulated data with known ground truth and on facial data captured by a video camera. The approach is applied to the problem of 3D face pose tracking using 3D face models and is shown to deal well with facial occlusions and large variations of facial orientation and expression. The recorded experiments, conducted on 640×480 video sequences, are provided at the authors' website. At this video resolution, the faces occupying 1/8 in the image have i.o.d. of 60 pixels, which ensures that Harris corner detector, which works on patches of 11×11 pixels, can be applied on facial image.

Ralph Gross, Iain Matthews, and Simon Baker focus their work on another technique intensively used in the tracking of non-rigid objects such as faces—active appearance models (AAMs). This technique fits the AAM of a face, constructed from training facial images using a mesh of hand-marked feature points, to input images. Previously, AAM fitting was known to work well only if all feature points are visible. Since in video sequences faces are often partially occluded, robust methods are necessary for model construction as well as model fitting, both of which are introduced in the paper. The authors show how to use Principal Component Analysis with missing data to construct the shape and appearance components of an AAM from training data containing occlusions. The paper then proposes a new fitting method, the Robust Normalization algorithm, which allows one to efficiently track faces showing significant amounts of occlusion. Experiments are carried out on artificially occluded facial data as well as using real 640×480 video sequences showing self-occlusions of the face due to head rotation and occlusion by other objects; facial resolution (measured by i.o.d.) ranges from 35 to 60 pixels. The video files showing the results are provided at the authors' website.

It is one of the distinctive features of video that it is able to capture live emotions and facial expressions of people. Automatic detection of facial expression, however, remains to be a very challenging video-processing task. The next two papers of the issue are contributions to this area of study.

To facilitate facial expression analysis in video, Y. Chang, C. Hu, R. Feris and M. Turk develop a six-dimensional embedding of the facial expression space, corresponding to six basic facial actions: happiness, sadness, fear, anger, disgust, and surprise. To construct this embedding, Active Shape Models, similar to those used in [5], are manually initialized and trained on two individuals who performed six basic facial expressions many times. The obtained reduced dimension of the face state allows the authors to efficiently track the facial deformations caused by the change in facial expression. This is done by constructing a Mixture of Gaussians Model to describe the distribution of data, followed by applying the ICondensation probability distribution propagation technique, similar to the one used in [4]. The experiments are done using 320×240 video recordings (showing faces with i.o.d. of about 60 pixels). The speed of execution is close to real-time. The video sequences showing the learned manifold of facial expressions, as well as the results of facial expression recognition, are made available from the authors' website.

There are many processing steps leading from seeing a video sequence to recognizing a facial expression in it, each one contributing to the success of the process. The state of the art techniques to many of these steps are summarized and comparatively analyzed in the paper by Gwen Littlewort, Marian Stewart Bartlett, Ian Fasel, Joshua Susskind, Javier Movellan, who by conducting experiments on two popular databases of facial expressions find such techniques and parameters that allow them to achieve the best reported to date recognition rates for the problem. The authors use their own real-time face detection system, which extends the work of Viola–Jones to include better feature search and novel training procedure, to detect faces in video. Faces are resized to a canonical face model with i.o.d of 16 pixels, on which Gabor filters, known to model the spatial frequency and orientation saliency detection mechanism employed in biological vision systems, are used to generate features. Based on these features, a face is classified as belonging to one of the seven basic facial expression classes: six as used in [6] plus one corresponding to a normal state. Among the classifiers tested by the authors are such popular machine learning techniques as support vector machines (SVM), linear discriminant analysis (LDA), and Adaboost (also used in [2] and Viola–Jones face detector). The results obtained, including a real-time animation of computer-generated character, as well as the source code of the techniques presented are made available at authors' website.

The last two papers of the issue are dedicated to the problem which by many is considered as one of the most difficult and demanded—recognizing faces in video.

Research in face recognition has started long ago from the problem of recognizing faces in still images. As such it has a long history and produced a number of well established results and approaches. Many of these approaches, such as appearance-based techniques based on principal component analysis (PCA) (also tested in [7]), independent component analysis (ICA), and linear discriminant analysis (LDA), have become classical for the area. Now, as we move on into the area of recognizing faces in video sequences, a natural question arises.—How to apply the already established approaches, such as the ones mentioned above, to a sequence of images showing a face? This question is the focus of the paper by Yongbin Zhang and Aleix M. Martinez, who propose a probabilistic approach to propagate recognition results over several images. The authors model the errors caused by imprecise face localization and occlusion by dividing a detected face into set of several overlapping ellipsoidal-shaped areas, localization error of each of which is estimated using the mixture of Gaussians. In addition, the authors introduce a weighting scheme which factors the similarity of test images to those used in the training, in terms of their pose and facial expression. This all makes recognition robust to occlusions, orientation and expression changes. The experiments are conducted on 640×480 video sequences. The technique available from [20] is used to detect person's eyes, nose and mouth in each video frame. The detected facial features are used to transform a face to a canonical 120×170 pixels face model with i.o.d. = 60, which is similar to the canonical face model deployed by the International Civil Aviation Organization (ICAO) for the machine readable travel documents presently used by passport offices and the police in many countries. which is presently used by many passport offices as well as by police for criminal identification. The training is performed on still images of a person. The recognition, which is performed off-line on a video sequence of the person, is shown to be much better than that done on any single video frame.

One of the main advantages of video over still images is that video provides a set of images showing a face. It is critical therefore for the success of face recognition in video systems, say the authors of the last paper of the issue, Ognjen Arandjelovic and Roberto Cipolla, to develop recognition approaches that can match sets of images, rather than the images themselves, to one another. These sets, which, for the case of set of images showing the motion of a face, are called the face motion manifolds by the authors and which are similar to the manifolds shown in [6] obtained for the case of facial expression motion, are intrinsically low-dimensional and non-linear. Modeling them compactly and finding an efficient procedure to compare them to one another are the main two keys for the success of the video-based face recognition system, argue the authors. By using the Resistor-Average Distance for measuring the dissimilarities between the two probability density functions preprocessed by the RANSAC-based Kernel Principle Component Analysis, the authors achieve both expressive modeling of nonlinear face motion manifolds and an efficient procedure to compare them, and a significant overall improvement of the recognition system performance, as a result. To achieve higher manifold sample density and better generalization, the authors synthetically populate facial data using an affine model of the face transformations. Faces are detected using the Viola–Jones face detector. Colour histograms inside the detected face area and outside of it are used to detect and filter out the pixels belonging to the background. Theoretical study of the paper is supported by the experiments conducted on the 320×240 video database maintained by the authors. The results are obtained on facial data transformed to a canonical grey-scale 20×20 pixel presentation (i.o.d. = 10), which allows the authors to perform recognition in real-time.

5. Implications

It is hard to believe that only 5 five years ago video-cameras were so expensive and computers were so slow that doing research in video-processing was a prerogative of a very few, whereas today it is within reach of everybody. This is due to the fact that affordable camera solutions are now widespread: from very low cost USB web-cams, to cell-phones with build-in cameras, to surveillance cameras mounted in public places. This favourable state is also due to the effort of researchers who have developed and made publicly available many useful video-processing libraries. These libraries significantly facilitate the development of the technology in the area and are usable for building commercial applications. Of the ones used by the authors of this issue [18–20], the Open Source OpenCV project [17,18], initiated by Intel researchers and now widely supported by contributions from all around the globe, deserves a special acknowledgement.

Another important resource for the development of the face-in-video applications are publicly available video-based facial datasets, which can be used for testing and tuning the algorithms. On one hand, there are datasets collected and maintained by the research community, as used by the authors of this issue [19–23]. On the other hand, it is also proposed to use publicly available recordings of popular TV shows, concerts and movies for the purpose [10].

All these factors have brought about this rapid development of face-in-video technologies that we are witnessing now. And this journal issue is a proof of that. As the papers of this issue show, as well as other papers presented at the First and Second Workshops on Face Processing in Video, the second of which (FPiV'05) was held in 2005 jointly with the Canadian conference on Computer & Robot Vision (CRV'05) in Victoria, Canada, many solutions to the face-processing-in-video problems have been published. This is especially true for the face-camera setup, within which a person sits in front of a camera. There is unlikely to be the same type of breakthrough for this type of scenario as there has been in the past [24].

At the same time, the situation where human faces are observed by hidden surveillance cameras still has many open problems. The issues of low resolution and lack of frontal views makes processing such video a difficult task. Therefore, more research effort is needed to deal with this type of situation. This need is particularly very motivated by the mandate from the security industry which sees video capturing technology as a very important tool to combat terrorism.

This is why the next international workshop, now called the *First International Workshop on Video Processing for Security* (VP4S'06), will be again organized under the umbrella of Canadian conference Computer & Robot Vision (CRV'06) sponsored by the Canadian Image Processing and Pattern Recognition Society (CIPPRS) and the National Research Council of Canada (NRC-CNRC). It will be held in June 7-9, 2006 in Quebec City, Canada. Its focus has been extended from face detection, tracking and recognition to people, objects, scene and action detection, tracking and recognition. The workshop website is http://computer-vision.org/4security.

6. Future trends

In the following we list the research directions that we believe to be very promising and important in the context of face processing in video.

1. *Recognition-assisted tracking*. Face recognition in surveillance-type video deserves thorough consideration or, even perhaps, reconsideration. While facial image quality in surveillance video may not be sufficient for identifying a person, it should be quite sufficient for finding the same person in an entire video sequence or several video sequences. This leads to such an important application of face recognition in video as tracking and backtracking of people in video sequences, where video recordings obtained from surveillance cameras need to be searched through to find out where a person captured as committing a crime came from or where he is going to.
2. *Video and biometrics*. Even though the low resolution and quality of facial images obtained from video may not be sufficient to allow its use as hard biometrics needed for person identification, it provides data that can be treated as soft biometrics, which, when combined with other biometric modalities, may improve the overall performance of the biometrics-based person identification. Just as other examples of soft biometrics such as a person's height, weight, skin and eye colour, face-in-video biometrics can be acquired at a distance and without person's involvement, which makes it very attractive for security applications.
3. *Video for best-shot selection*. Since video allows one to track a person's face and detect instances when it is conveniently positioned for recognition, it appears very promising to combine regular-resolution video cameras with high-resolution still image photo-cameras in such a way that the face detection performed in video automatically triggers the capture button of the photo-camera.
4. *Data accumulation over time*. One way of improving face recognition in video is to develop techniques for combining visual evidence over time both in face memorization and recognition. While each piece of evidence may be of low quality; the quality of the aggregated evidence may be high enough to be used for person identification. Such techniques can be implemented on a hardware level as in [26], or on a software level as in [8–10].
5. *3D head model assisted recognition*. Another way of improving face recognition in video is to use the precomputed 3D head models. It has been recently shown that using 3D head models for face recognition improves significantly the recognition, as it provides a way to deal with large variations in illumination and head pose commonly observed in video [25].
6. *Recognizing facial gestures*. A person's gestures are by nature dynamic and video allows computers to capture them. Making computers recognize them would have a great impact on the development of computer–human interaction systems as well as on the security industry.
7. *Other visual cues for face recognition*. For humans, gestures and gait are known to significantly facilitate person recognition, especially in low resolution when fine facial features cannot be seen. The same approach should improve the recognition performance of computer systems.
8. *Video-based bench-marks*. Video-based technology, whether it is face recognition, facial expression and gesture recognition, or face and person tracking using single or multiple video sources, should be tested using video-based datasets. Until now there are not many such datasets available. This is partially due to the fact that video data consumes a lot of computer memory and requires considerable effort to annotate. This is also partially due to the fact that the area of face processing in video is relatively young compared to other areas of computer vision science. However, the belief is that with fast growing awareness of the potential of face processing in video technologies, which this special IVC issue is hoped to contribute to, the solution to this problem as well as to the other mentioned problems will soon be found.

In conclusion, I want to thank all authors of this issue for bringing the work they started a while ago to their best level of presentation, the editorial team of IVC for agreeing to dedicate a special issue of their journal to the topic, and the program committees of both workshops on Face Processing in Video, many of whom were the reviewers for this issue.

7. Dedication to Bernhard Fröba (1969–2004)

It is with the great sadness that I am writing these last lines of the editorial. Bernard Fröba started working with me as a Co-Editor on this Special Issue soon after the workshop was held in July. Bernhard, best known for the face detection techniques he developed for the Fraunhofer Institute in Erlangen, Germany [12–14], was very excited to serve as an editor for this special issue. Together we selected the papers to be invited for the issue. But as I was sending the invitations to the authors and communicated with Bernhard by email during the next two months, I did not know that the recipient never received my letters. — Bernhard died of a heart attack, while jogging with a cold, in August 2004 at the age of 35. His colleagues Christian Küblbeck and Andreas Ernst finished the paper which Bernhard was hoping to see in the issue, and I was left to finish the editorial work, feeling an enormous loss.

Bernhard graduated from Friedrich–Alexander University in Erlangen and from 1997 to 2004 worked at the Fraunhofer Institute for Integrated Circuits in the same city. The face detection demo which he made for the institute was one of the first which showed the public that the problem of real-time face detection in video is now practically resolved. He supervised 31 bachelor and master thesis students who respected him very much for being not only an expert in the area but also a person always open for questions. Besides his scientific and teaching achievements, Berhard will also be remembered as a very friendly and active person, who liked mountaineering and travelling, and whose positive attitude towards many of his friends, colleagues and students will always make them remember him as a very genial "gemütlich" person.

References

[1] J. Bonaiuto, L. Itti, The use of attention and spatial information for rapid facial recognition in video, Image and Vision Computing (this issue).

[2] Christian Kublbeck, Andreas Ernst, Face detection and tracking in video sequences using the modified census transformation, Image and Vision Computing (this issue).

[3] Zhengrong Yao, Haibo Li, Tracking a detected face with dynamic programming, Image and Vision Computing (this issue).

[4] Le Lu, Xiangtian Dai, Gregory Hager, Efficient particle filtering using RANSAC with application to 3D face tracking, Image and Vision Computing (this issue).

[5] Ralph Gross, Iain Matthews, Simon Baker, Active appearance models with occlusion, Image and Vision Computing (this issue).

[6] Ya Chang, Changbo Hu, Rogerio Feris, Matthew Turk, Manifold based analysis of facial expression, Image and Vision Computing (this issue).

[7] Gwen Littlewort, Marian Stewart Bartlett, Ian Fasel, Joshua Susskind, Javier Movellan, An automated system for recognizing facial expression in video, Image and Vision Computing (this issue).

[8] Yongbin Zhang, Aleix M. Martinez, A weighted probabilistic approach to face recognition from multiple images and video sequences, Image and Vision Computing (this issue).

[9] Ognjen Arandjelovic, Roberto Cipolla, An information-theoretic approach to face recognition from face motion manifolds, Image and Vision Computing (this issue).

[10] 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), 2005, pp. 330–338, NRC 48216.

[11] Dmitry O. Gorodnichy, Facial recognition in video, in: Proceedings of IAPR International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'03), 2003, pp. 505–514, NRC 47150.

[12] Bernhard Fröba, Christian Küblbeck, Orientation template matching for face localization in complex visual scenes, International Conference on Image Processing (ICIP 2000), 2000, pp. 251–254.

[13] Bernhard Fröba, Christian Küblbeck, Real-time face detection using edge-orientation matching, Audio- and Video-based Biometric Person Authentication (AVBPA 2001), 2001 pp. 78–83.

[14] Bernhard Fröba, Christian Küblbeck, Face tracking by means of continuous detection, Proceedings of the First IEEE CVPR Workshop on Face Processing in Video (FPiV'04), Washington DC, 2004.

[15] Paul Viola, Michael Jones, Rapid object detection using a boosted cascade of simple features, CVPR 2001/Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.

[16] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, ICIP 2002. 1 (2002) 900–903.

[17] Gary Bradski, OpenCV: examples of use and new applications in stereo, recognition and tracking, Proceedings of the International Conference on Vision Interface (VI 2002), Calgary, Canada, 2002, p. 347, online at http://www.cipprs.org/vi2002.

[18] OpenCV library: http://sourceforge.net/projects/opencvlibrary.

[19] UCSD MPLab Dataset and Software: http://mplab.ucsd.edu.

[20] MIT CBCL Dataset and Software: http://cbcl.mit.edu/cbcl/software-datasets/index.html.

[21] BU ground truth annotated dataset of facial motions: http://www.cs.bu.edu/groups/ivc/HeadTracking.

[22] Head, gaze, hands, feet and shoulder ground-truth annotated video dataset http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/.

[23] NRC IIT Low resolution video-based face dataset http://synapse.vit.iit.nrc.ca/db/video/faces/.

[24] NewScientist, Breaking News Report. Nose-steered mouse could save aching arms, September 16, 2004: http://www.newscientist.com/article.ns?id=dn6400 (Also CNN, Special Report Follow your nose, September 30, 2004: http://edition.com.cnn.com/2004/TECH/09/29/explorers.nouse/index.html, and Discovery Channel, The Nose Knows, December 7, 2004 http://www.exn.ca/dailyplanet/story.asp?id=2004120751).

[25] Face recognition grand challenge website:http://www.frvt.org/frgc/

[26] M. Ben-Ezra, A. Zomet, Shree K. Nayar, Jitter camera: high resolution video from a low resolution detector, Conference on Computer Vision and Pattern Recognition (CVPR'04), 2004 pp. 135–142.

Dmitry O. Gorodnichy

Institute for Information Technology, National Research Council of Canada, Ottawa, Ont., Canada K1A 0R6

E-mail address: dmitry.gorodnichy@nrc-cnrc.gc.ca