FGNet - 1st Foresight Report

Date of preparation: 23 Aug 2002~~1 Jan 2002~~

Content List

1. Introduction

One of the major objectives of the FGNet Network of Excellence in Face & Gesture Recognition is the organization of foresight workshops, where the FGNet members and invited experts get together in order to define visions of possible future scenarios enabled by intelligent methods in face and gesture recognition.

The first FGNet foresight workshop was hosted by J. L. Crowley at INRIA in Grenoble, from 1.-2. November, 2001. This document reports about the content and outcome of this workshop. The outline of this report is as follows: First, a brief introduction into the workshop topic "smart spaces" is given, then the following section contains summaries of the talks presented by the invited speakers. The next section describes the foresight visions that were defined by the different working groups at the last afternoon of the workshop and a final section summarizes the major conclusions. The appendix contains the workshop agenda and the list of participants.

2. Introduction into Workshop Topic "Smart Spaces"

It was decided by the workshop organizers that each foresight workshop should be associated with a specific topic from the field of face & gesture recognition. Talks and discussions at the workshop should be mainly centered around, but not strictly limited to this workshop topic.

This year's workshop topic was selected to be "Smart Spaces". This decision has been taken because this scenario is relatively wide and includes areas such as

- Smart Homes
- Smart Offices
- Smart Streets
- Smart Parking Lots
- Smart Meeting Rooms
- .....

while still focussing on some well-defined scientific fields with strong relation to the areas of ubiquitous and pervasive computing. It has also strong connections to several themes that were preliminarily selected as potentially interesting for database collections, such as "FG behavior in the home", "pointing & command/attention gestures" as well as "FG for access control" and covers several focal themes within the F&G community, such as gesture and face recognition, action recognition and video surveillance.

3. Summary of the Invited Talks

Four invited talks were presented at the workshop, which are briefly summarized in the following paragraphs.

The first invited talk was presented by Dr. Clive Norris from the University of Hull, UK. In his talk, entitled “Social Issues about Surveillance”, Dr. Norris explained how the application of advanced technologies in criminology has pushed the development of surveillance techniques especially in the United Kingdom. Actual events such as the latest terrorist attacks in the US are expected to accelerate this development. He pointed out that surveillance has always been misused by totalitarian governments and states. Today, there is a danger that democratic states follow a similar direction by introducing surveillance methods all over and with an almost complete coverage of all individuals in a country. A good example is England, where the first video surveillance systems have been deployed in 1984, an almost complete coverage of the country has been reached in 1989 and today around 1.5 million surveillance cameras are installed in public and private places. Two factors have accompanied these developments: The “time-space-distantiation” has led to the fact that there is no ~~more~~ longer necessity to perform a certain action and be physically present for its performance. All tele-services, including telephone and internet services, are good examples for that. Consequently, also video surveillance can be performed from remote locations, and most often is handled that way. Furthermore, the fact of “disembodiment” expresses that e.g. the credit card number of a person might become more important that the physical appearance of this person in the future.

Video surveillance in the UK has led to overall high installing and processing costs, and to an information flooding due to the generation of too much image data with storage problems. The original motivation was the reduction of crime and several studies were conducted in order to investigate if video surveillance has indeed reduced the crime rate. The results show a strong variation for different cities in the UK and are often contradictory and thus still difficult to interpret. Newest developments show that one should distinguish between passive surveillance (which is simply recording from mostly fixed cameras) and active surveillance, where interaction is taking place by zooming and tracking operations. A similar distinction can be made between online- and offline-surveillance, where again the latter variant is mainly concerned with storage of recorded video data.

Besides the problem of over-examination and intrusion into people’s private spheres, the massive introduction of video surveillance techniques can also lead to other problems, e.g. the “disembodiment” of natural surveillance accomplished so far by people. An example is the vision of an automatic lifeguard system in pool areas, that replaces human lifeguards and thus would also abolish their additional “human capabilities”, such as helping people, enforcing pool rules, etc....

In summary, this talk demonstrated very impressively that “smart spaces” and “surveillance” are closely related to each other and that it is necessary to reflect about the possible social impacts of this technology, if smart spaces will become a more and more accepted F&G application scenario in the future.

The second invited talk, entitled “Smart Spaces and the Aging Population” was presented by Dr. Andrew Sixsmith from the University of Liverpool in the UK. At the beginning of his presentation, Dr. Sixsmith pointed out that Europe is the largest potential market for elderly people who often suffer from physical and mental handicaps, especially when they reach ages over 85. Generally, there are many technical aids that can be deployed in order to facilitate the lives of these people, such as home support, remote care, transport, restoration, communication, medical care and access from distance. The area of smart spaces including video surveillance is

particularly important in the field of home monitoring using intelligent sensors, where of course smart and tiny cameras would play an important role in the future.

A special issue within these home monitoring scenarios is the area of "fall detectors", where currently mostly simple sensors such as e.g. infrared detectors or simple motion detectors are employed. Here, the potential for more sophisticated methods, including video-based sensors, is very high. This would also encourage the usage of more interactive surveillance methods, where the person could give some feedback to the surveillance system (e.g. if he does need help or not). It is not surprising, that quite a lot of EU projects have been and currently are concerned with service for the elderly people, especially within the TIDE programme (from 94-98) as well as in the current 5th framework programme within the action lines that deal with services for the citizens.

By showing the importance of surveillance techniques for elderly people, this talk demonstrated how large and rich the scenario spectrum for smart spaces will be in the future and that here again non-technical issues will become very important.

The third invited talk was presented by Paco Gomez from Visual Tools in Spain and was titled "Smart Spaces in Buildings and Commercial Applications". In this talk, especially the market situation and marketing opportunities for video surveillance applications have been presented. The author pointed out that several conditions should be fulfilled in order to realize a market success. Among them is the necessity to focus on rather simple applications and to use standard distributed architectures for installation of these systems. As already hinted by the title of the talk, the author concentrated on the application of surveillance techniques in buildings, thus focussing on the "smart building" scenario as one of the possible scenarios of smart spaces. Besides the classical application of video surveillance in buildings and shops, the author suggested the new application area of tele-observation as one of most promising future applications in that area. In this case, a number of shops are monitored in parallel, and the outcome of the surveillance process is not only the transmission and storage of the video data, but additionally some intelligent evaluation, containing e.g. information on how many people have entered the shop, what their reaction was on different products (e.g. by examining their facial expressions), how many people have been attracted by the shop window, passed the entrance or stopped in front of the shop.

In summary, this talk was a very market and application oriented presentation and highlighted the chances as well as the problems in this business as seen from the perspective of a person who really is involved in these issues in his daily activities.

The fourth and final invited talk was presented by Prof. Niels Ole Bernsen from the University of Southern Denmark in Odense, entitled "Natural Interactivity Resources and Tools". Its goal was the view of the topic "smart spaces" from the viewpoint of multimodality, also with strong emphasis on natural language issues. The author presented several research challenges in multimodal human-computer interaction, such as domain-oriented speech recognition, multilingual speech understanding, graphical animation, prosody generation, interface agents, tracking & identification, audio-visual speech recognition, emotion interpretation, and mobile communication. He listed several important projects in that area, including he projects CLASS,

MATE, ISLE and NITE and later on presented more details on the ISLE project. This project is especially concerned with multimodal data resources and tools and thus closely related to he title of this invited talk. The project was also very much concerned about the availability of already collected databases in the area of face, gesture & speech technology and it turned out that in total 64 different databases could be made available within this project. In many domains, the so-called coding scheme of the databases became very important, especially in cases where the meaning of the data is quite ambiguous. The coding scheme is basically identical to the alphabet that was used in order to label the collected data. This issue is e.g. less crucial for speech databases, which are mostly labeled according to phones or words. It is much more important in the area of gesture databases, because there is obviously no standard alphabet for gestures available. There were however several alternative coding schemes defined by different organizations. This is also the case for facial databases, where similar transcription problems arise. The author concluded his presentation with a review on popular tools that have been mostly developed by research organizations for handling and processing of multimodal databases.

This talk demonstrated the necessity for viewing the area of smart spaces in the framework of multimodality and to consider a smart space as an environment that requires multimodal communication capabilities, because smart spaces are not isolated but always involve humans as users of its facilities.

4. Foresight Visions

In order to increase the productivity and creativity of the workshop members, they were divided into four groups which each defined a set of possible foresight visions in separate sessions. Roughly, the following procedure has been agreed on:

- definition of possible scenarios and classification into short (S), medium (M) or long term (L) range
- definition of possible terrain map axes
- identification of enabling background technologies for these scenarios
- identification of the generic capabilities required for those scenarios
- definition of specific database collection issues for selected scenarios

The following paragraphs contain a summarization of the results obtained from all groups, merged into one unified result.

4.1 Scenarios

Home Scenarios

- home personal assistant (L)
  - possibly embodied (talking head)
  - endowed with personality and voice
  - integrated view of your life
  - faithful with respect to privacy to the occupants or to individuals?
- home medical advisor (L)
  - possibly embodied (talking head)

- endowed with personality and voice
- integrated view of your health and lifestyle
- faithful with respect to privacy
- monitor and assess your physiological state
- monitor and assess your life style
- access to medical history including nonverbal cues
- able to react to medical emergencies

· entertainment, educational or expert assistant (M)
- possibly embodied (talking head)
- endowed with personality and voice
- faithful with respect to privacy
- access nonverbal cues
- assess ones competence
- Examples:
  - celebrity chef
  - expert tutor
  - chess tutor
  - music star
  - exotic dancer

· smart home managing unit (speech & gesture for home device control) (M)
· smart home care system for elderly people (M)

Portable Personal Assistant Scenarios

· general issues:
- is the personal assistant the master controller for spaces?
- all persons have "their" personal assistant who negotiates with other electronic servants
- personal assistants as servants
- certified identity (backed with biometrics - for example for theft protection)

· portable assistants for the disabled (M)
- people recognizer
- informational translation

· other portable personal assistant examples:
- celebrity chef
- expert tutor
- child's imaginary friend

Office Scenarios

· office personal assistant (M)
- possibly embodied (talking head)
- endowed with personality and voice
- integrated view of your life
- faithful with respect to privacy to the occupants or to individuals?
- is the personal assistant the master controller for spaces?
- does it negotiate with other assistants?

· gestures for turn-taking in video conferences (S)
· smart meeting room (complete facial-gesture-verbal meeting transcription) (M)
· cooperative workspace using AR technologies (command gestures) (M)

Sports Scenarios

- personal referee (L)
- body skills training (sports, dancing, gymnastics) (M)
- smart sports referee (L)
- coach (L)
- golf instructor (M)
- tennis assistant (M)

Commercial Scenarios
- shop customer monitoring: purchasing, queuing, shop lifting, abnormal behavior (M)
- shop staff monitoring (shop assistants, cleaning personnel, security guards): cash transactions, tidying up, assisting customers (M)
- person identification for smart space access control (M)

Street Scenarios
- intelligent revolving door (S)
- smart pedestrian management system (M)
- total surveillance (continuous, constant monitoring of individuals) (L)
- public monitoring with individual data protection methods (L)
- detection of threatening behavior (in highly sensitive areas) (L)

Criminal Justice Scenarios
- detective - investigative agent (L)
- crowd observation (prediction of panic…), public safety / public order (L)
- investigative interrogation assistant (using non-verbal behavior during interrogation for lie detection etc.) (L)
- safety of people in cells (M)
- generalized people tracking in public space (widespread, in space and time) (M)

4.2 Terrain Map / Axes

- environmental
  - indoor / outdoor
  - controlled vs uncontrolled lighting and background
  - observability
- spatial
  - ego space (single user is primary focus of attention, e.g. cave)
  - work space (space populated by several users, e.g. AR environment)
  - collective space (group of users, e.g. meeting room, video conference)
  - communication space (e.g. smart home, intelligent agent)
- social
  - public / private
  - potential for abuse
  - open access / restricted access
  - voluntary / non-voluntary
  - cooperative vs unaware
  - fairness, misappropriation, false knowledge

- functional
  - number of involved persons
  - degree of hazard in area
  - active / passive sensing
  - degree of autonomy
  - association with space or with person
- technical
  - modality rich vs modality poor
  - image acquisition noise
  - control of acquisition conditions (imaging, sound)
  - adaptive (immediate vs. long term)
  - occlusions and occluding targets
  - number of targets
  - resolution and target speed
  - adaptive control vs fixed control
  - supervised vs unsupervised learning
  - 2D vs 3D vs 4D
- data collection issues
  - size
  - resolution
  - frame rate
  - lighting conditions
  - noise conditions
  - color vs. gray scale
  - sound track inclusion
  - format
  - compression
  - data protection issues
  - camera parameters
- end use
  - accuracy requirements
  - business potential

4.3 Enabling Background Technologies Landscape

- wireless networks
- digital video cameras
- embedded technologies (sensors/computing/…
- inexpensive flat displays
- wearable computers
- holographic displays (by 2020?)
- free tera-flop computing

4.4 Generic capabilities

- face tracking

- face configuration / posture tracking
- face recognition
- face verification
- facial expression recognition
- hand configuration tracking and recognition
- hand position tracking
- person type classification (age, sex)
- body tracking
- body posture measurement
- multi-person tracking
- gesture recognition (dynamic & static)
- pointing gestures
- intention recognition
- activity recognition (individuals & groups)
- emotion recognition
- modeling face to face interaction
- gaze recognition
- deictic command recognition and measuring
- speech recognition
- speaker recognition
- speech and face/lips dynamics integration

4.5 Selected scenarios for database collection

Home Scenario

· smart home
  - data capture: room with several persons coming in and out, sitting, pointing
  - analysis: test tracking, test gesture recognition, person recognition, intruder detection, abnormal behavior detection, gaze detection, speech & gesture & pointing

Outdoor Scenario

· counting and tracking
· groups activities
· detecting facial images for person identification in a crowd
· shop (external)
  - data capture: single camera observing passing people
  - analysis: attention to shop window, person classification, body tracking, person counting, duration measures, speed, facial expressions

Office Scenario

· modeling interaction of two or more people in a meeting
· modeling interaction of two or more people in a video conference

Commercial Scenario

· counting people in a queue

- counting people on a bus or train
- shop (internal)
  - data capture: multiple indoor cameras, recording customer activities
  - analysis: test body tracking algorithms, test person counting algorithms, duration measures, queuing, abnormal behavior, attention to products, consumer behavior, person classification, facial expressions, interaction between people
  - privacy issues: anonymization of data, data corruption

Human interface scenario

- interacting people home scenarios
- interacting people in meetings

5. Summary and Conclusion

During this 2-day long workshop, the state-of-the-art in smart spaces and the non-technical issues resulting from the deployment of this technology have been demonstrated by the invited speakers and discussed by the workshop participants.

Finally, a large variety of foresight visions have been defined and elaborated by the different groups formed at the final afternoon of the workshop. These should be helpful in order to indicate development roadmaps and opportunities for the technology in the medium (5-7 years) and long term (>10 years).

Appendix I: Final Workshop Programme

FGNet Technology Foresight Workshop
INRIA Rhône Alpes
Programme

Thursday, 1. November 2001

9:00-9:30	Arrival - Discussion
9:30-9:45	Welcome from the local host (J. Crowley) General Information on workshop venue and local environment
9:45-10:00	Message from the coordinator (P. Courtney)
10:00-10:15	Message from the workshop organizer (G. Rigoll) Introduction to major workshop topic "Smart Spaces", schedule, background, goals
10:15-10:30	Coffee Break
10:30–11:15	Invited Speaker I: Clive Norris, University of Hull, "Social issues about surveillance"
12:00-13:30	Lunch
13:30 -14:45	Invited Speaker II: Andrew Sixsmith, University of Liverpool, "Smart spaces & the aging population"
15:00-15:30	Coffee Break
15:30-16:15	Invited Speaker III: Paco Gomez, Visual Tools, "Smart spaces in buildings and commercial applications"
16:15-16:30	Introduction to Theme 1: "FG behaviors in the home" (T. Cootes) Database & scenario issues for Theme 1 (P. Courtney) Discussion (J. Crowley)
16:30-18:00	Small group discussions
18:00-18:30	Group reports & wrap-up of Day I (J. Crowley)
18:30	Return to Hotel
20:00	Dinner

Friday, 2. November 2001

9:00	Arrival - Discussion
9:30-9:45	Introduction (E. Granum)
9:45-10:30	Invited Speaker IV: Niels Ole Bernsen, University of Southern Denmark, Odense, “Natural interactivity resources and tools”
10:30-11:00	Coffee Break
11:00-11:15	Introduction into Theme II: “Pointing & command/attention gestures” (E. Granum)
11:15- 11:30	Database & scenario issues for Theme II (M. Stoerring)
11:30- 11:45	Database & scenario issues for Theme IV: “Gestures to support verbal communication” (S. Marcel)
11:45-12:00	Define Groups
12:00-13:30	Lunch
13:30-15:00	Group Discussions : 1) Define Brief Scenarios 2) Define Axes of Terrain map
15:00-15:30	Groups Report Back
15:30-15:45	Coffee Break
15:45-17:30	Group Discussions: 3) Identify Generic Capabilities 4) Define Data Collection Scenarios
17:30-18:00	Group reports and summaries (J. Crowley)
18:00-18:30	Foresight-visions for Day II themes, workshop summary & wrap-up (G. Rigoll)
18:30	Return to Hotel
20:00	Dinner

Appendix II: List of Participants

NAME	INSTITUTION	COUNTRY	ROLE
Olivier Barnier	France Telecom R&D	F	Invited Participant
Niels Ole Bernsen	University of Southern Denmark	DK	Invited Speaker
Tim Cootes	University of Manchester	UK	FGNet Partner
Patrick Courtney	University of Manchester	UK	FGNet Coordinator
James L. Crowley	INRIA Rhône Alpes	F	FGNet Partner
Paco Gomez	Visual Tools	ES	Invited Speaker
Erik Granum	University of Aalborg	DK	FGNet Partner
Sebastien Marcel	IDIAP	CH	FGNet Partner
Kimberly Moravec	Xerox Research, Cambridge	UK	Invited Participant
Clive Norris	University of Hull	UK	Invited Speaker
Justus Piater	INRIA Rhône Alpes	F	FGNet Partner
Olivier Riff	INRIA Rhône Alpes	F	FGNet Partner
Gerhard Rigoll	Duisburg University	D	FGNet Partner
Andrew Sixsmith	University of Liverpool	UK	Invited Speaker
Jean-Bernard Stefani	INRIA Rhône Alpes	F	Invited Participant~~FGNet Partner~~
Moritz Störring	University of Aalborg	DK	FGNet Partner
Chris Taylor	University of Manchester	UK	FGNet Partner