

IM2 Newsletter

Contents

COVER STORY

- IM2, third phase (2010-2013) 1

FOCUS

- IM2 Review Panel Report 2
- A new collaboration between IM2 and Affective Sciences 3
- Completed PhD Thesis, IM2.VP 3

INSIDE IM2

- News 4
- Selected publications 4

News

Picture of the month: Information management

January 2009

**KLEWEL: An innovative conference
archiving solution**

IM2, first steps towards its third phase (2010-2013)

Pre-proposal submitted to SNSF

End of September 2008, jointly with the 7th annual report, the IM2 Management submitted to SNSF a pre-proposal for its third phase (2010-2013).

During the last few months, the SNSF Research Council thoroughly discussed each of the 14 NCCR pre-proposals, and took the final funding decisions for phase III at its December meeting.

A few weeks ago, the Research Council was pleased to inform the IM2 Management that after 7 years of operation, IM2 had turned out to be a success story, and that, therefore, IM2 will be funded for a third phase.

Next September, IM2 will have to provide the SNSF with a full proposal which will take into account the allotted budget and the comments of the last Review Panel report. The pre-proposal summary available below presents the research topics, their focus and objectives, and the added value resulting from IM2. This description of work will be further revisited in the full proposal:

Executive summary of the pre-proposal:

In the preface of the initial IM2 Full Proposal dated March 15, 2000, we state that:

By studying, inventing and developing technologies that support multimodal interaction with computers, this NCCR will become a focal point of excellence in the development of related technologies such as human-computer dialogue, computer vision, machine learning, and automatic speech and speaker recognition. By doing so it will carry out the highest quality computer science research on problems of importance to business and society.

At the end of the 7th year of IM2, we actually believe that many of the stated goals (in terms of research, technology transfer, education and training) set in the IM2 Full Proposal are being successfully met, and even often surpassed. As of this writing, IM2 is clearly perceived as a Center of Excellence in the area of multimodal processing, while also pioneering new approaches towards the analysis, modeling and understanding of (multimodal) human-to-human interaction.

In IM2-Phase 1 (2002–2005), we (1) initiated the IM2 common vision (human-to-human interaction in multimodal meeting recordings, and indexing and retrieval of relevant multimedia information), (2) developed the required hardware (instrumented meeting rooms at Idiap and the University of Fribourg) and software infrastructure, and (3) collected and annotated a large common multimodal corpus.

On this basis, we initiated and performed fundamental research in mono-modal and multimodal processing, encouraging the IM2 partners to test their research on the IM2 corpus, while adapting their software to the IM2 vision.

In IM2-Phase 2 (2006–2009), we mainly aimed at consolidating the IM2 research and development efforts on our common vision, while at the same time putting more emphasis on multimodal processing, system integration, HCI, and system evaluation (at the component level, as well as at the application level).

At the beginning of Phase 2, and in addition to the consolidation of the common annotated corpus (now often used as a standard in multiple international multimodal projects), key software (the Hub) was also made available to allow easy integration of multiple modalities and software tools, allowing multimodal 'data producers' (audio, video, different levels of annotator outputs, etc.) to easily exchange data with 'data consumers' in an interoperable environment (e.g., with applications running on cell phones, PDA, meeting browsers, etc). In addition to normalized data annotations (used by more than 20 institutions worldwide) and exchange formats, IM2 was also enriched with a flexible user interface development tool (JFeret), allowing easy adaptation and customization of HCI interfaces.

Towards the middle of IM2-Phase 2, these efforts, coupled with leading edge research, resulted in multiple end-to-end systems, demonstrating the information and processing flow from user requests to meeting data.

To be continued on page 2

Cover Story

IM2 third phase (continued)

Finally, significant structural impact started to be highly visible during the 3rd year of IM2-Phase 2, not only at the Leading House but also in some of the partner institutions, including joint Idiap/EPFL tenure track professor positions, one new chair at EPFL in BCI (in addition to several tenure track, leading to tenured, professor positions at EPFL), one SNSF professor position at ETHZ (and an upcoming new chair for a full professor to be announced soon), a permanent professor position in the IM2 domain at the University of Geneva (in addition to a new assistant professorship and a SNSF professorship), and the potential to start a new institute at the University of Fribourg.

Starting from this solid basis, which enabled the partners to successfully implement a major restructuring during IM2-Phase 2 due to a mid-phase budget cut, IM2-Phase 3 will work towards:

1. Integrated multimodal processing: Core multimodal technologies geared towards integration into end-to-end applications and consolidating all IM2 activities developed in Phase 1 and Phase 2 through a single "Individual Project" (IP) involving:

- 3rd Phase Pre-Proposal (Jan. 2010 - Dec. 2013)
- All speech processing activities, including: natural speech recognition, speaker tracking, segmentation, and recognition.
- All visual processing input, including shape tracking, face and gesture recognition, visual focus of attention.
- Integration of modalities and coordination among modalities, including (asynchronous) multi-channel processing, integration of multimodal, and multimedia indexing and retrieval.
- Further development and evaluation of meeting browsers and assistants

2. Human centered design and evaluation: In order to better evaluate the generalization capabilities of the IM2 technologies, and as often questioned by the Review Panel, we will develop and evaluate new educational applications, including formal ergonomics, usability, and user studies, through the involvement of two new partners.

3. Social signal processing: As acknowledged by the Review Panel, an important by-product of IM2 was to pioneer a new research area, often referred to as 'social signal processing', analyzing and modeling non verbal communication and human behavior. Given our leading edge in this area, and based on recent successes, we believe it is appropriate to devote a small part of the budget (15% of the IM2-Phase 2 budget) to perform further research in this area, but still in the context of the IM2 vision.

4. Community building: One of the goals of the NCCR is to build a sustainable network of researchers and research institutions in the targeted area. During the 3rd phase of IM2, it will thus be important to create all types of incentives towards the retention of the network, collaboration between its affiliated members (with or without direct IM2 funding), and fostering the growth of the community. This will be achieved through multiple actions spanning white papers (competitive grant attribution, also opened to newcomers), organization of summer institutes and international workshop series, wide distribution of corpora and software libraries, and a provision of minimum funding to encourage newcomers (typically new professors) to take part in IM2 activities. In addition to the above activities, IM2 will continue its efforts towards active technology transfer and internal/external training, also increasing its integration into large EU projects and exchange programs (e.g., with ICSI/Berkeley).

IM2 Technical Committee

Excerpt from the 7th NCCR IM2 Review Panel Report

GENERAL IMPRESSION AND REACTION TO THE OUTCOME OF THE LAST REVIEW

The NCCR focuses on multimodal research in the context of human-to-human interactions in meetings. This is a very challenging task in terms of speech and language processing, speaker recognition, video analysis with person/object/gesture recognition. The component technologies need to be fused in order to support subsequent context extraction from the recordings of the meetings. Meeting transcription is also a common task that is pursued by several leading re-search groups under the NIST Automatic Meeting Transcription Project. The task demands much infrastructure building in order to acquire appropriate meeting data. It also requires much research into adapting existing technology to support recordings of highly interactive and spontaneous human activities in meetings. In this regard, the leading house of this NCCR, Idiap, has collaborated closely with ICSI, which is one of the US participants in the NIST Automatic Meeting Transcription Project. The meeting room domain offers great opportunity for the group to focus and target both basic and applied research in tackling the real-world problem. Through this application, the group has collected large corpus of data, which offer them great advantage to be leading in research on large real-life data in this domain. Thus the meeting room scenario has been fruitful for



Annual review meeting at Idiap

gaining and maintaining focus across teams and for advancing key enabling technologies. Yet the practical impact seems limited. The NCCR has delivered strong scientific contributions in speech and computer vision. The research conducted in the IPs AP and VP seems to have the greatest impact in terms of systems research and quality publications. In Phase II, the NCCR has made significant progress in sharing data and tools. This was a major problem in the past. There has been a significant accomplishment to integrate the various system components.

The major added value produced by the NCCR compared to individual funding has been the large amount of funding which

in turn has allowed to assemble a very large team that has been able to take a leadership position on the international scene. Therefore the international standing of IM2 is excellent, mostly due to its size and early start. The NCCR is also intertwined with AMI and other European projects, which makes it even more visible. Generally, the group seems to be very successful in leveraging the NCCR funding in acquiring more support from other sources. The NCCR has to a large extent been responsive to the panel's recommendations of last year as well as to the concerns expressed by the SNSF Research Council.

IM2 Review Panel

A new collaboration between IM2 and Affective Sciences:

"CROSS-CULTURAL PERSONALITY PERCEPTION"

Did you ever notice that the voice of talking machines is always female, warm and kind? This applies to personal devices like GPS suggesting car drivers the direction to take, as well as to synthetic voices diffusing announcements in large public spaces like airports and stations. In both cases, it is much more natural to attribute the voice to a woman of beautiful appearance and kind manners than to its actual source, a cold and unattractive piece of hardware. This is not just a coincidence, the designers of such machines know that people tend to attribute personalities to voices and know that devices with a nice personality are perceived to work better by the users. Hence, it is important to understand how voices influence the perception of personality to build human-centered interfaces more enjoyable for the users.

Alessandro Vinciarelli (Idiap Research Institute), active in IM2.MCA, has just been granted a new project, called Cross-Cultural Personality Perception (CCPP), aimed at investigating this interesting phenomenon. CCPP takes place in the framework of the Indo-Swiss Joint Research Program (<http://indo-swiss.epfl.ch>) and involves two other partners, namely Bayya Yegnanarayana (Indian Institute of Information Technology) and Marcello Mortillaro (University of Geneva), a psychologist involved in the NCCR Affective Sciences (<http://www.affective-sciences.org>). The presence of Marcello further improves the collaboration record between IM2 and the NCCR led by Prof. Klaus Scherer in Geneva. Furthermore, CCPP benefits from the expertise provided by SSPNet, the European Network of Excellence on Social Signal Processing coordinated by Alessandro (<http://www.sspnet.eu>). This strong academic context, providing links with some of the best groups at both national and international level, gives CCPP the best chances to achieve the expected scientific results.

The project will investigate how nonverbal vocal behavior (prosody,

hesitations, vocalizations, silences, pauses, etc.) influences the perception of personality in humans. This includes three major stages: The first is the identification of voice characteristics (in particular those that can be detected with machines) that elicit the attribution of certain personality traits rather than others. The second is the development of an approach that uses such characteristics to infer automatically the way humans perceive personality from voice. The third aims at synthesizing voices capable of eliciting the attribution of specific personality traits. In simple terms, the goal of the project is to 1) build machines that perceive the personality of a given speaker in the same way as humans do, and 2) build machines that synthesize voices perceived by human listeners as belonging to people with desired personality characteristics.

As the Indo-Swiss Joint Research Program offers the unique opportunity of working across different cultures, CCPP will dedicate particular attention to the effect of culture on personality perception. Personality assessments provided in India and Switzerland about the same voice will be compared to estimate what are the characteristics that are perceived in the most different way. This will help to build synthesizers capable of adapting to different cultures, i.e. capable of changing the voices so that they elicit the same personality perception in different cultures.

With its exciting mix of human sciences and technology, CCPP provides an excellent opportunity to address a totally new challenge (the personality perception problem is almost unexplored in the technical literature) with a fully multidisciplinary approach. The multi-cultural dimension of the Indo-Swiss framework will further enrich the project. All the best to the lucky participants!

*Alessandro Vinciarelli
Idiap Research Institute*

Alessandro.Vinciarelli@idiap.ch



Alessandro Vinciarelli, Idiap

Effrosyni Kokiopoulou, EPFL, IM2.VP

GEOMETRY-AWARE ANALYSIS OF HIGH-DIMENSIONAL VISUAL INFORMATION SETS

Effrosyni started her PhD studies in September 2005 in the Signal Processing Laboratory (LTS4) at EPFL under the supervision of Prof. Pascal Frossard. She successfully defended her PhD thesis on December 19th, 2008.

Her PhD thesis has been performed within the framework of IM2.VP, led by Prof. Thiran from EPFL. It proposes new methods for the multi-view analysis of visual pattern manifolds based on sparse geometric expansions and graph models. The thesis' contribution is three-fold; (i) it leverages the use of sparse representations towards invariance with respect to geometric transformations of pattern, as well as supervised dimensionality reduction, (ii) it proposes graph-based methods for pattern classification with multiple observations in both centralized and distributed environments and (iii) it designs fast consensus methods for distributed data analysis in multimedia sensor networks. Effrosyni's thesis work is the first one to solve optimally the image registration problem

under transformations consisting of a synthesis of translation, rotation and scaling.

Effrosyni has received several awards for her work. For example, she has been selected as one of 8 top students worldwide for the 'Emerging Leaders in Multimedia Research' workshop organized in 2007 by IBM TJ Watson Research Center, NY, USA. She has also been a finalist of the Google Europe Anita Borg Scholarship 2007.

Effrosyni has joined the Seminar for Applied Mathematics at ETHZ in January 2009.

Contact:
Effrosyni Kokiopoulou
Seminar für Angewandte Mathematik
HG J 49
Rämistrasse 101
8092 Zürich



E. Kokiopoulou, EPFL

News

Picture of the month: Information management

KLEWEL: An innovative conference archiving solution

The Swiss National Science Foundation (SNSF) places utmost importance on co-operation with the media. SNF's Press and Information Office sends out regular press releases in German and French, organises media conferences and provides the media with information in other ways as well. Pictures of the month are research projects presented in images with a press release on methods and results.

The picture of January 2009 is about Klewel, a start-up from Idiap Research Institute:

"Under the auspices of National Centre of Competence in Research IM2, researchers at the Idiap Research Institute have developed an innovative conference archiving solution. By combining several technologies derived from the smart meeting room concept with the extraction and analysis of information from video sequences, they have been able to come up with a system that is both simple to use and reliable. The researchers have set up a company called Klewel to market their innovative technology. Prestigious customers such as Nestlé, Unicef or ACM are already using their services."

Klewel provides leading edge solutions for effectively capturing, sharing and searching the information contained in multimedia digital recordings of presentations and conferences. Klewel ensures the continued existence and visibility of events. Thanks to a fully innovative digital solution, the content is always available and easily promoted online. The solution automatically references the full content of presentations (including audio, video and slides). Organizations such as UNICEF, Nestlé, Conseil communal of Montreux, and ACM (Association for Computing Machinery) one of the largest scientific conference organizers in the world trust Klewel.

Maël Guillemot, Klewel



Alessandro Vinciarelli, Maël Guillemot and Jean-Marc Odobez demonstrate their software.



IMD Start-up Competition

KeyLemon: Winner in 2009

The IMD Startup Competition provides a unique opportunity for early-stage companies to benefit from the support and insights of MBA and Executive MBA participants. The objectives of this unique collaboration are:

- To provide hands-on learning experience in entrepreneurship and new ventures for the IMD MBA and EMBA classes
- To provide input and support to early stage companies, to improve their chances of funding, product development and market success.

Twenty-two young companies have been chosen to work with IMD's MBA and Executive MBA students in 2009. They were selected from among many worthy competitors based not only on the value of their ideas and quality of their teams, but also on their fit with the learning objectives of our programs.

KeyLemon, among the full time MBA program winners, is a software company which offers simple, fast and convenient computer access solutions for internet users, based on face and speech recognition.

KeyLemon provides a new experience to access your computer. Instead of typing your password, you just need to smile to the webcam and KeyLemon will recognize you. It is fast, simple and convenient. Download LemonScreen for FREE: <http://www.keylemon.com>

Valérie Devanthery, Idiap

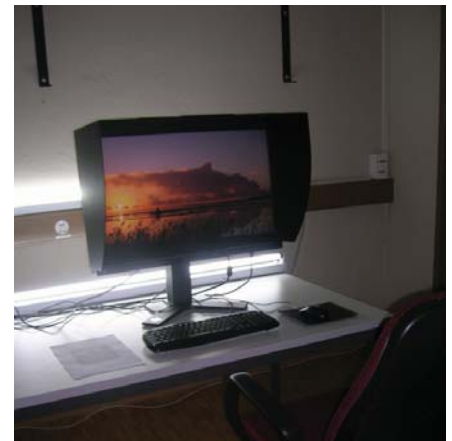
Test space available at MMSPG/EPFL to perform subjective visual quality evaluations

Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics

The MMSPG laboratory at EPFL has now a test space to perform formal subjective visual quality test. The space includes an High Resolution monitor (2560x1600 pixels) with calibration tool, a powerful server able to display up to high resolution uncompressed video sequences, and a lighting system with color temperature of 6500 K. This equipment is being shown in the photographs below.

Related to the subject, the paper 'A subjective study of the influence of color information on visual quality assessment of high resolution pictures', authored by Francesca De Simone, Frederic Dufaux, Touradj Ebrahimi, Cristina Delogu and Vittorio Baroncini, has been accepted and presented by Professor Ebrahimi at the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-09), Scottsdale, Arizona, U.S.A, Jan. 15-16, 2009 (www.vpqm.org).

This paper presents the design and the results of a psycho-visual experiment which aims at understanding how the color information affects the perceived quality of a high resolution still picture. The results of this experiment help to shed light into the importance of color for human observers and could be used to improve the performance of objective quality metrics.



Equipment for the test space at the MMSPG laboratory (EPFL): an high Resolution monitor (2560x1600 pixels)

Francesca de Simone, EPFL

Selected publications

A hardware-independent fast logarithm approximation with adjustable accuracy

O.Vinyals, G.Friedland

Proceedings of the 10th IEEE International Symposium on sMultimedia, pp. 61-65, Berkeley, CA, USA, December 2008.

Integration of n-gram language models in multiple classifier systems for offline handwritten text line recognition.

R.Bertolami and H.Bunke

Int. Journal of Pattern Recognition and Artificial Intelligence, 22(7):301 – 321, November 2008.

Impact of feature correlations on separation between bivariate normal distributions.

K.Kryszczuk, A.Drygajlo

19th International Conference on Pattern Recognition, Tampa, Florida, USA, December 8-11, 2008.