

Techware: www.sspnet.eu: A Web Portal for Social Signal Processing

Please send suggestions for Web resources of interest to our readers, proposals for columns, as well as general feedback, by e-mail to Dong Yu ("Best of the Web" associate editor) at dongyu@microsoft.com.

In this issue, "Best of the Web" focuses on introducing the social signal processing network (SSPNet), a large European collaboration aimed at establishing a research community in social signal processing (SSP), the new, emerging domain aimed at bringing social intelligence in computers.

One of the most exciting challenges that a researcher can face is to pioneer a new domain and to foster its acceptance and recognition in the scientific community. The success in such an endeavor depends on how promising the new domain is in terms of scientific results but also on a second factor that should not be neglected, namely how difficult it is to enter the new domain for an institution, group, or even an individual researcher. As a matter of fact, entry barriers can prevent even the most interested researchers from entering a fully or largely unexplored domain no matter how interesting and promising the domain is. This is the consideration that drives the efforts of the SSPNet. The SSPNet involves some of the earliest SSP researchers and its ultimate goal is to smooth, if not to eliminate, the three main entry barriers that people face when starting to work on SSP, the lack of knowledge, data, and tools.

The strategy of the SSPNet is reflected on a Web portal (www.sspnet.eu)

that aims not only at diffusing information about SSP but also at providing the most important and yet difficult to obtain resources for working in SSP, i.e., the knowledge, data, and tools corresponding to the above-mentioned barriers. The SSPNet portal has been online since August 2009 and, thanks to a collaborative effort involving both SSPNet members (11 institutions scattered across Europe) and contributors from the rest of the scientific community, provides a large amount of SSP-related resources. Both data and tools (see below for more details) are freely available to the scientific community. To share resources through the SSPNet portal is not only an important contribution but also an excellent opportunity for achieving high visibility in the emergent and dynamic community growing around SSP.

In addition to the above-mentioned resources, the portal offers an up-to-date view of the SSP state of the art through an extensive (often updated) bibliography and the Virtual Learning Centre (VLC), a repository of lecture and presentation recordings collected at scientific and training events revolving around SSP. This contributes to the elimination of the last important entry barrier, i.e., the lack of knowledge.

SOCIAL SIGNAL PROCESSING

Social intelligence is the ability of dealing effectively with the complex web of social interactions we cope with in our everyday life and, at its core, it consists of effectively perceiving, correctly understanding, and appropriately reacting to social signals, the complex constellations of nonverbal behavioral cues (e.g., facial expressions, gestures, or vocalizations) through which we express our relational

attitudes (e.g., empathy, disagreement, or hostility) with respect to others and social situations.

In this respect, SSP aims to answering the following three main questions:

- Is it possible to detect nonverbal behavioral cues from signals captured through microphones, cameras, or any other suitable sensor?
- Is it possible to automatically infer and understand social signals from nonverbal behavioral cues detected in possibly multimodal signals?
- Is it possible to synthesize appropriate social signals (as a set of synthesized nonverbal behavioral cues) via different forms of embodiment?

In correspondence to the above questions, two main kinds of technologies are involved in SSP: approaches for analysis and synthesis of nonverbal behavioral cues like, e.g., facial expression analysis and synthesis, prosody extraction and synthesis, gesture and posture recognition, and synthesis, as well as techniques for inferring social signals from behavioral cues like, e.g., machine learning and pattern recognition. Equally important for both kinds of technology is the investigation of psychological, anthropological, and social laws underlying human-human interactions. These laws and principles identify the predictable behavioral patterns that actually allow technology to be effective with social signals.

Synthesis and understanding of social signals are mostly data driven, large corpora of data annotated in terms of social signals that become a fundamental resource, hence the data barrier. Furthermore, as nonverbal behavioral cues are typically captured with different sensors (e.g., facial expressions with cameras and vocalizations with microphones), tools addressing a wide

spectrum of diverse needs become crucial, hence the tool barrier. Finally, the multidisciplinary nature of SSP, spanning across multiple technical competences (speech processing and synthesis, computer vision) and human sciences, makes it difficult for a group or even for an institution to have all necessary knowledge at disposition, hence the knowledge barrier.

The three barriers shape the structure of the SSPNet portal and drive the selection of the resources being accumulated in its different sections. The material on the portal is at disposition of the scientific community for research purposes (in some cases upon signing an end user license agreement) and any contribution is welcome as long as it is annotated rigorously (in the case of the data) and relevant to the SSP research. The portal guarantees storage of the material and, most importantly, high visibility in the emergent SSP community.

RESOURCES TO BREAK THE KNOWLEDGE BARRIER

As SSP is a young domain, its state of the art is still relatively limited, but it is rapidly growing, and it is fragmented across a large number of disciplines and research areas. Thus, it can be difficult for people entering the domain to identify the relevant literature and to access the latest developments in the field. To this end, the SSPNet portal hosts two important sections. The first is an exhaustive bibliography including not only the most important SSP works published so far, but also a large number of works providing the necessary background to enter the field (http://sspnet.eu/category/sspnet_resource_categories/resource_type_classes/publication/). The repository is fully searchable in terms of meta data (title and authors) as well as in terms of tags defined by SSPNet researchers and corresponding to the most important aspects of SSP (<http://sspnet.eu/resources/search/>) like the behavioral cues being investigated in the article (e.g., “facial analysis” and “speech synthesis”), or the modeling classes (“linguistic modeling” and “psychological



A slightly modified version of this cartoon appeared in *IEEE Antennas and Propagation Magazine*, vol. 52, no. 1, p. 201, 2010.

modeling”). At the time this column was written, the bibliography included around 300 titles, but it is constantly

increasing with contributions from both SSPNet researchers and the rest of the scientific community.

Senior Research Scientist For Signal and Image Processing

The RDECOM CERDEC Night Vision and Electronic Sensors Directorate, Ft Belvoir, VA, is looking for an Engineer or Scientist to serve in a Scientific or Professional (ST) Position as the Senior Research Scientist for Signal and Image Processing. NVESD is located approximately 30 minutes south of Washington, D.C., and is the Army's premier laboratory for the development of next generation electro-optical, infrared and countermine sensor technology. Over 400 engineers, scientists and technicians work together in a collaborative environment with co-located customers to field the latest EO/IR and countermine technology to the Soldier. Position is responsible for the development of new signal and image processing techniques that extract and optimize information from advanced sensors, optimizing and identifying the signal(s) associated with targets/threats while separating them from signals associated with background clutter and compression of sensor data for transmission over tactical sensor networks. The ST position reports directly to the Director of NVESD and is expected to identify and solve signal/image processing problems at the strategic level that have far ranging impacts to the Army.

The incumbent of this position must have specialized experience in sensor signal and image processing technology. ST positions represent the highest level of technical accomplishment and are of very limited number (approximately fifty ST positions within the Army). Typically, applicants for ST positions are expected to have a graduate degree, significant research experience, and a national or international reputation in his/her field.

How to apply: U.S. Citizenship and ability to obtain a TOP SECRET security clearance is required. Refer to www.opm.gov, job announcement number DA-ST-01-10 for application requirements/process and additional information. Questions should be directed to Mrs. Genie Shires, 703-704-1140, or by email: genie.shires@us.army.mil

The second important section of the portal dedicated to the knowledge barrier is the VLC (<http://sspnet.eu/virtual-learning-centre/>), a repository of lecture and presentation recordings collected at scientific events (workshops, special sessions) and training initiatives (summer schools, courses) dedicated to SSP. When this column was being written, the VLC included around 40 presentations collected at the first events organized by the SSPNet. An extensive recording campaign is taking place to further improve depth and breath of the material (the inclusion of 80–100 more presentations is planned by the end of 2010). The VLC is fully searchable in terms of keywords appearing in the presentations slides, a simple textual query (like those submitted to Web search engines) returns presentation intervals corresponding to those slides that are detected as relevant to the query itself.

These two sections of the portal make it possible for any interested researcher to acquire the necessary knowledge about the current state of the art in SSP, including most recent trends, as well as about the background necessary to deal with SSP problems.

RESOURCES TO BREAK THE DATA BARRIER

In SSP, data typically consist of large corpora of video and audio recordings portraying social interactions. The collection of this kind of data is one of the most expensive and time-consuming aspects of SSP. On one hand, recording social interactions often requires large experimental apparatuses like smart meeting rooms, or devices capable of synchronizing multiple sensors. On the other hand, data must be annotated, i.e., trained observers must identify the social phenomena taking place in the recordings, at the exact time when they appear and following a rigorous methodology that allows repeatability of the experiments and a sufficient degree of objectivity (in terms of agreement between multiple annotators). Such a process can take significant amount of time, especially when the corpus is large and the annotation is fine grained (e.g., the annotation of facial expressions

requires to track every facial muscle during the time a face is portrayed).

The SSPNet portal hosts a data repository that, at the time this column was being written, contained around 240 h of annotated material (http://sspnet.eu/category/sspnet_resource_categories/resource_type_classes). The data repository includes, among others, the Augmented Multiparty Interaction (AMI) Meeting Corpus (150 meeting recordings annotated in terms of roles, dominance, and subjectivity), the Canal9 Database (75 television debates annotated in terms of conflict, roles, agreement and disagreement), the Belfast Naturalistic Database (298 clips showing 125 speakers in both neutral and emotional states annotated in terms of acoustic features), the Human Communication Research Centre (HCRC) Map Task Corpus (128 task oriented dialogues annotated in terms of discourse phenomena), the IDIAP Head-Pose Database (eight meetings annotated in terms of participant head pose), the Green Persuasive Database (eight dialogues annotated in terms of persuasive behavior), the ICSI Meeting Corpus (75 meeting recordings), the Man-Machine Interaction (MMI) Facial Expression Database (2,894 video clips annotated in terms of facial expressions), and the FreeTalk Corpus (a collection of Japanese phone calls annotated in terms of interactional phenomena).

RESOURCES TO BREAK THE TOOL BARRIER

Once the interested researchers know what SSP is about and have the relevant data at disposition, the last barrier is the lack of suitable tools to perform actual research work. This applies, for instance, to researchers who work on the automatic understanding of social signals but do not have the competencies for the extraction of nonverbal behavioral cues or to researchers who know how to process only one modality (e.g., speech) but would like to develop approaches involving other modalities as well (e.g., gestures).

For this reason, the SSPNet portal provides a repository of tools that addresses diverse needs in SSP work (<http://sspnet.eu/2009/12/gabor-facial-point-detector/>).

While this column was being written, the repository contained the Nite XML tool kit (an open source cross-platform framework for handling multimodal annotations that are related both temporally and structurally), a salient point detector for human gestures (it finds spatio-temporal salient points in an image sequence), a real-time gaze and head-pose estimation system (it can use a plain Web camera mounted on top of the user's screen and produces two-dimensional yaw/pitch vectors for the user's eye gaze and head pose, including roll), PRTtools (a toolbox for pattern recognition algorithms), the SEMAINE research platform (an open-source software package containing state-of-the-art software tools for audio-visual behavior analysis and synthesis), and the Gabor facial point detector (combining face detection with facial point detection).

ACKNOWLEDGMENTS

This work was funded in part by the European Community's Seventh Framework Programme (FP7/2007-2013) under the grant agreement 231287 (SSPNet). The work of Alessandro Vinciarelli was funded in part by the Swiss National Science Foundation through the National Center of Competence in Research Interactive Multimodal Information Management. The work of Maja Pantic was funded in part by the European Research Council under the ERC Starting Grant agreement ERC-2007-StG-203143 (MANHOB).

AUTHORS

Alessandro Vinciarelli (vincia@dcs.gla.ac.uk) is a lecturer with the Department of Computing Sciences, University of Glasgow, United Kingdom. He is also a senior researcher with Idiap Research Institute, Switzerland.

Maja Pantic (m.pantic@imperial.ac.uk) is a reader in multimodal HCI in the Computing Department with Imperial College London, United Kingdom. She is also a professor of affective and behavioral computing in the Department of Computer Science at the University of Twente, The Netherlands.

