Department of
Information Engineering
and Computer Science
DISI

UNIVERSITY
OF TRENTO - Italy

DISI - Via Sommarive 5 - 38123 Povo - Trento (Italy)
http://disi.unitn.it

EVENTMASK

Andrea Rosani, Duc-Tien Dang-Nguyen,
Giulia Boato, and Francesco G.B. De Natale

May 2014

Technical Report # DISI-14-009
Version 1.0

Published also: ICASSP 2014 – Show and Tell.

EventMask

Andrea Rosani, Duc-Tien Dang-Nguyen, Giulia Boato, and Francesco G.B. De Natale

***Abstract*—Events emerged in the last years as a driving concept to efficiently index and retrieve media. Several approaches have been proposed to analyse the relationship between events and related media, to enable event discovery, event-based tagging, or event-based media retrieval. Notwithstanding the outstanding work done by several researchers in this area, a major still unsolved problem is how to understand the inherent link between visual concepts and events, and in particular, which are the key elements that allow a human being to perceive from a media collection what is the underlying event. In this demo, we address this problem in an original way by exploiting human knowledge using a game with a hidden purpose. Users are engaged in a competition, where they can alternatively try to mask a photo collection to prevent other users to recognize the related event or they can see a photo partially masked and try to guess the event connected. A set of rules and an adequate score system force players to focus on real important details, thus allowing the detection of most event-related salient parts of the media. The final result is a saliency map that, differently from the typical concept of saliency, highlights the visual concepts that are considered fundamental to perceive the nature of an event.**

***Index Terms*—Event Detection, Gaming, Saliency, Photo Galleries**

I. Introduction and motivations

RECENT studies demonstrate that users find it easier to navigate and search through photo galleries if the pictures are grouped into events [1]. This is supported by the research trend in both academia and industry to organize media collections following the concept of event [2][3]. In this context, being able to recognize an event related to a photo gallery becomes a crucial issue.

Several approaches have been proposed to analyze the relationship between events and related media, to enable event discovery, event-based tagging, or event-based media retrieval[4]. Effective approaches to manage large multimedia data collections need to both allow event detection and retrieve all corresponding media. This is very challenging due to the heterogeneity, multimodality and generally unstructured form of such content [5]. In this demo we address a major problem that is still unsolved: how to understand the inherent link between visual concepts and events, and in particular, which are the key elements that allow a human being to perceive from a photo what is the underlying event. In particular, we propose to face such a problem via a game-based approach. During last years the new trend of *gamification* [6], the use of game mechanics and game design techniques in non-game contexts, has proved to be a successful strategy for building user bases as well as to engage them in active participation [7].

A. Rosani, D.T. Dang-Nguyen, G. Boato, and F.G.B. De Natale are with the Department of Information Engineering Computer Science, University of Trento, Trento, I-38123 Italy e-mail: (andrea.rosani)@unitn.it.

Fig. 1. Example of an image where event-related salient areas are covered. This results in a not intuitive association of the image to a specific event.

We design a new game developed in order to gain insight of what we call *event-related saliency*. Saliency [8] usually represents parts of single images where users focus attention. Here we assume that there are key elements in an image connected to an event that allow human beings recognizing that particular event and do not necessarily correspond to the most important part from a perceptual point of view (i.e., image saliency). We define these information as event-related saliency.

The idea is illustrated in Figure 1. Which event represents this picture? Since there is a snow landscape, it is probably an event connected to the winter or mountains, like a vacation, but we don't have particular clue about the specific event. Indeed, we superimposed to the image a mask corresponding to the event-related details of the photo.

Our aim is to discover such an information for each event engaging users in a competition through a web interface which allow people playing with different photo galleries: on one hand to mask parts of images which they think will allow other players to understand the event type represented by that gallery, on the other hand to recognize events into galleries which are masked by other players. The analysis is carried out on a set of images representing different types of social events with a set of rules and an adequate score system forcing players to focus on real important details, thus allowing the detection of most salient parts of the media.

II. Scientific and technical description

The objective of this work is the identification of image details which support event-based classification of galleries of pictures. The result is an *event-related saliency map*, that indicates which are the regions of the images that allow recognizing a specific event (see for instance the white area in Figure 1).

Event saliency is different from visual saliency. Although the two concepts may partially coincide, the idea is very different: *Visual saliency* detects the parts of the image that attract attention (depend on color, contrast, foreground, position). *Event saliency* should detect all the visual contents that can lead to a recognition of the event, even if it is in the background or represented in a marginal detail. Event saliency is very difficult to extract automatically without a-priori knowledge.
The relationship between visual and event saliency is not always evident. Given the attitude of photographers to attract the attention on most important things sometimes they may partially overlap but often high informative parts are in details or in the background. (See example in Figure 2).
EventMask is in the form of an inversion problem: we do not ask people to indicate what is important for them to recognize an event, but to hide it to other people. It is competitive: to win they have to avoid that the other person recognize the event. Fundamental is to define a scoring system that prevents cheating. To the best of our knowledge this is the first proposal to apply gaming to this issue.

(a)

(b)

Fig. 2. Samples of event- and visual-saliency: (a) image with event-saliency mask, (b) same image with visual saliency mask. As can be seen, the visual saliency of the image corresponds to the most important part of the photo, but details are ignored. Viceversa event-saliency is strongly related to them.

Games used for the evaluation and curation of the underlying data are effectively incentives-driven tools, as discussed in [9], employing ease of use, fun and competition as incentives for users to perform what would otherwise be unrewarding and demanding manual tasks. Social gaming [10] has been one of the emerging trends in the last years, and both research and industrial efforts, as well as mainstream interests have grown quickly. The evolution of social networks such as Facebook or Twitter has helped introducing the content generation idea as a common mainstream concept, encouraging users to both produce and consume content generated by other users.

III. Implementation and use

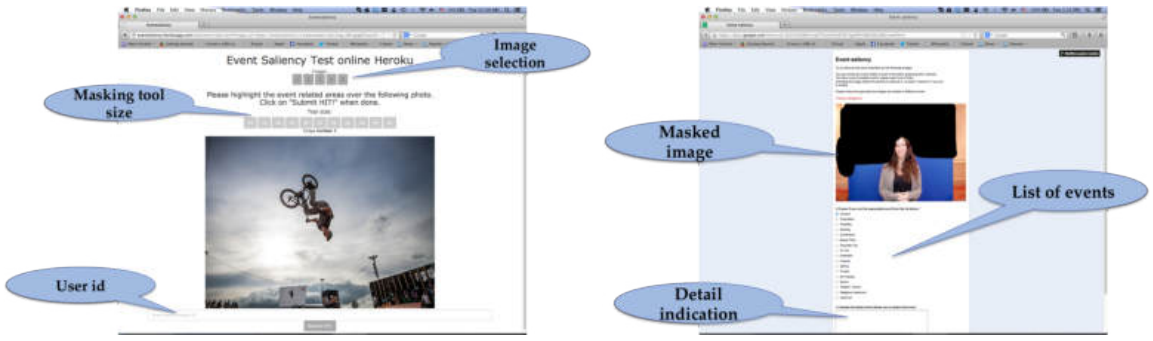
EventMask is a game with a hidden purpose (GWAP), which is a web based serious game that allows to involve several users, experts and non-experts, from various countries and cultures, in order to gather high quality data for image content analysis. To this aim, it is very important that the real goal is concealed by the game, to avoid cultural/personal bias.

A. Basic game rules

The game has two roles: masking users and discovering users. Two players participate in the game, taking different roles - the first player is the Masker, the second is the Discoverer. During the match the roles are switched following the game rules. Players are not requested to be connected at the same time.
Maskers are presented images related to events, and have to hide parts of them so as to make the event unrecognizable. Discoverers are presented the images masked by other players and have to classify them into a list of related events.
The goal of the game is for the Masker to hide all the details in the images that can connect the photo to a specific event, so that the Discoverer is not able to guess the associated event. The Discoverer would like to recognize the event looking to a masked image, using the information not hided by the mask that is still available in the photo.
Within the masking role, the player's goal is to make it difficult to recognize the event associated to the images. Players are presented a few images related to different events (Figure 3 (a)): they have to hide parts of them so that the event is no more recognizable. Users playing as Discoverers have to discover the event associated to a gallery of masked pictures, acting as evaluators of the job carried out by the Maskers. Players are requested to evaluate a certain number of images (different from the ones they masked). They should assign every image to a known set of potential events (Figure 3 (b)). They can choose not to classify an image, if they are not sure. This role is enabled only after the player acted as a Masker for one time.

B. Game points

We implemented a point mechanism for EventMask in order to keep players engaged as well as to avoid cheating. The score gained by the Masker in a game session is inverse proportional to the masked area (to discourage cheating or hiding the whole image). We fix an amount of 100 points per masked image, diminished proportionally respect to the percentage of the image area that has been covered using the tool provided to mask it. As an example, if an image has 40% of its area hided, the Masker gains for that image 60 points. Every masking procedure is subject to the validation of the Discoverers, who can guess the event related to a masked image using the details still visible. Discoverers can gain the same amount of points



(a) (b)

Fig. 3. Example of masking and unmasking procedure. Images are prompted to players and they should first hide with a tool the most relevant parts related to the event corresponding to that photo. Then they are requested to try to guess the related event, selecting it from a list, also indicating eventually the detail that allowed the recognition.

gained by the Masker who did the job of hiding image parts, if the correctly guess the event. To avoid cheating, if they do not discover the correct event, they are penalized by the same amount of points. All players scores are saved and prompted to the users of the game, so that they can discover their position in the list and see what is the best player of the game.

C. Map generation

The event saliency map, that corresponds to the region of the image that are *salient* in an event based perspective, is generated from a suitable combination of the above results. Event saliency maps are obtained taking into account both the masking and unmasking results. After masking, the j image is associated to a set of masks $M_{i,j}$ for each masking player i as described in Eq. 1, where T_D is the total number of users involved in the discovering procedure of a certain image and P_D represents the number of users that were still able to discover the event represented it.

$$M_{i,j}(x,y) = \begin{cases} 1 & \text{if pixel(x,y) is marked} \\ 0 & \text{if pixel(x,y) is unmarked} \end{cases} \quad (1)$$

After the discovery procedure, each map is converted into $M^*_{i,j}$ such that:

$$M^*_{i,j}(x,y) = \begin{cases} P_D/T_D & \text{if } M_{i,j}(x,y) = 0 \text{ AND event discovered by } P_D \text{ players out of } T_D \\ 1 & \text{if } M_{i,j}(x,y) = 1 \end{cases} \quad (2)$$

Finally, the i maps are multiplied pixel-wise and normalized to $[0,1]$.

$$M_j(x,y) = \prod_i M^*_{i,j}(x,y) \quad (3)$$

IV. Conclusions and Future Developments

In this work we presented a novel approach to identify the key elements that permit the connection of a collection of pictures to a particular event, i.e., what we have defined *event-based saliency*, using human knowledge. The problem of the identification of the visual concepts strongly related to events has been recast into an original game, with a hidden purpose, where different users have been engaged in a competition. The results of their work are the saliency maps of the various images, in an event based perspective, where the most important event-related details are highlighted. The results of this analysis could be exploited to gain knowledge about the inherent link between visual concepts and events, thus supporting automatic learning systems for concept recognition and consequently event detection.

References

[1] U. Westermann and R. Jain, "Toward a common event model for multimedia applications," *IEEE MultiMedia*, vol. 14, no. 1, pp. 19–29, Jan. 2007.
[2] R. Mattivi, G. Boato, and F. De Natale, "Event-based media organization and indexing," *Infocommunication Journal*, vol. 2011, no. 3, 2011.
[3] R. Troncy, B. Malocha, and A. Fialho, "Linking events with media," in *I-Semantics*, 2010.
[4] T. Reuter, S. Papadopoulos, G. Petkos, V. Mezaris, Y. Kompatsiaris, P. Cimiano, C. de Vries, and S. Geva, "Social event detection at mediaeval 2013: Challenges, datasets, and evaluation," in *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop Barcelona, Spain, October 18-19, 2013*, 2013.
[5] G. Petkos, S. Papadopoulos, and Y. Kompatisiaris, "Social event detection using multimodal clustering and integrating supervisory signals," in *Proceedings of ACM ICMR 2012*, 2012.
[6] L. V. Ahn and L. Dabbish, "Designing games with a purpose," *Communications of the ACM*, vol. 51, no. 8, pp. 58–67, August 2008.
[7] L. V. Ahn, R. Liu, and M. Blum, "Peekaboom: A game for locating objects in images," in *SIGCHI conference on Human Factors in computing systems*. ACM, 2006.
[8] O. Muratov, G. Boato, and F. G. De Natale, "Diversification of visual media retrieval results using saliency detection," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 86 670I–86 670I.
[9] K. Siorpaes and E. Simperl, "Human intelligence in the process of semantic content creation," *World Wide Web Journal*, vol. 13, March 2010.
[10] G. Abramson and M. Kuperman, "Social games in a social network," *Physical Review E*, vol. 63, no. 3, 2001.

EventMask: a game-based analysis of event-related saliency in photo galleries

UNIVERSITY OF TRENTO - Italy
Department of Information Engineering and Computer Science

Andrea Rosani, Duc-Tien Dang-Nguyen, Giulia Boato, and Francesco G.B. De Natale
Multimedia Signal Processing and Understanding Lab
DISI - University of Trento
Via Sommarive 9 – 38123 Povo - Trento (Italy)
http://mmlab.disi.unitn.it/

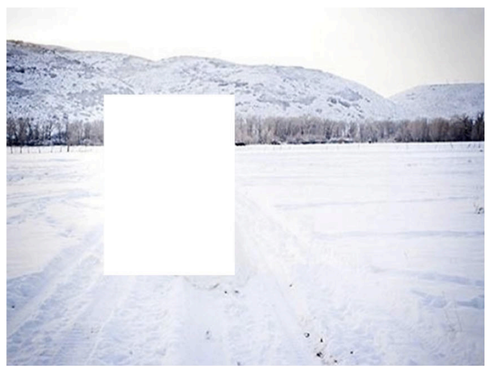
1. Introduction

Events emerged in the last years as a **driving concept to efficiently index and retrieve media**. Several approaches have been proposed to analyze the relationship between events and related media, to enable event discovery, event-based tagging, or event-based media retrieval.

Notwithstanding the outstanding work done by several researchers in this area, a major still unsolved problem is how to understand the inherent link between visual concepts and events, and in particular, **which are the key elements that allow a human being to perceive from a media collection what is the underlying event**.

In this work, **we address this problem** in an original way by exploiting human knowledge, **designing a game with a hidden purpose**.

2. Problem Statement



Which event represents this picture? Key elements that allow a human being to perceive the underlying event are masked resulting in a not intuitive association of the image to a specific event. We define these information as **event-related saliency** which do not correspond to the classical visual saliency.

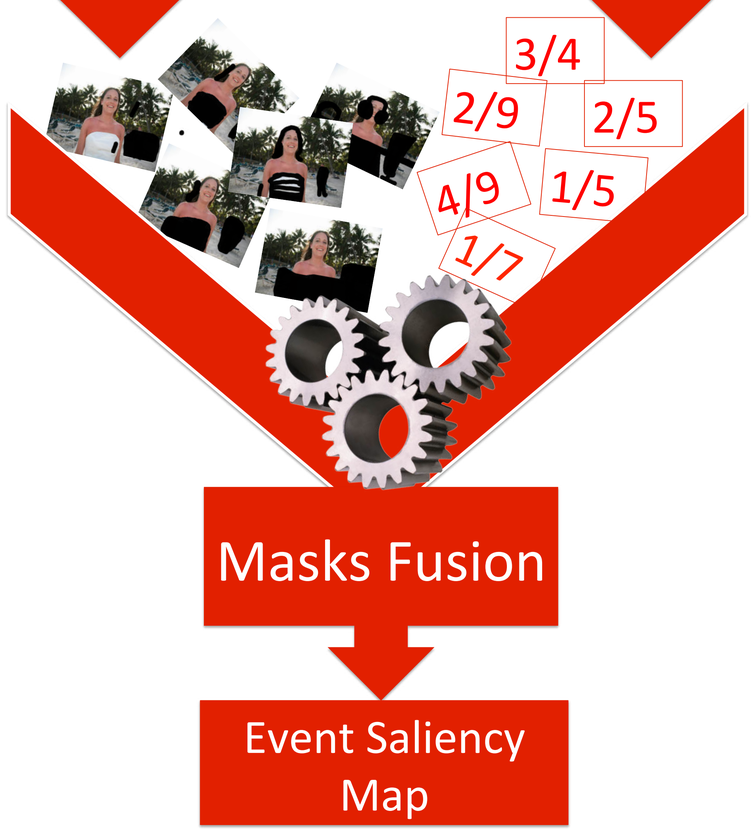
3. Solution based on a GWAP

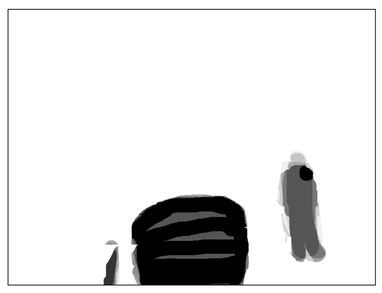
EventMask is a Game With A (Hidden) Purpose: Users are engaged in a competition using a web-application (**http://eventmask.azurewebsites.net**) where they can alternatively mask a photo collection to prevent other users to recognize the related event (**Masker role**) or observe a photo partially masked and try to guess the event connected (**Discoverer role**).
A set of **rules** and an adequate **score system** force players to focus on real important details, thus allowing the detection of most event-related salient parts of the media.
The final result is a saliency map that, differently from the typical concept of saliency, **highlights the visual concepts that are considered fundamental to perceive the nature of an event**.

4. GWAP and Event-Saliency Map generator

MASKER'S TASKS
- To hide event related details
- To cover the smallest area
- To gain points

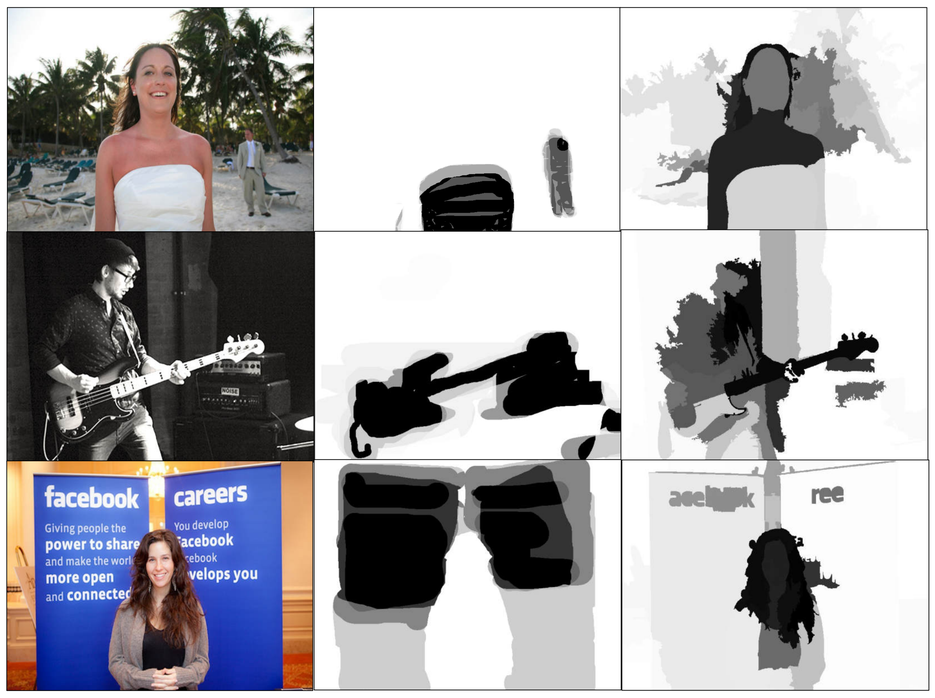
DISCOVERER'S TASKS
- To discover the event masked
- To evaluate Masker's work
- To gain points





5. Results

(Original Image) (Event Saliency) (Visual Saliency)



(Original Image) (Event Saliency) (Visual Saliency)

FLORENCE, ITALY, MAY 4-9, 2014
FIRENZE FIERA / FORTEZZA DA BASSO

Show and Tell

Show & Tell demonstrations were aimed at strengthening the interactions between researchers and practitioners. They offered an exciting opportunity for participants to demonstrate their state-of-the-art results to a wide audience of professionals in the area of signal processing. The Show & Tell demos of this year were 19 in total and were viewed and appreciated by most of the participants to ICASSP.

Show & Tell Schedule

All demos were evaluated for the Best Demo award of 1000€ offered by B&C Speakers.

Two demos tied for the best demo award and will split the prize:

- Sub-Nyquist Cognitive Radio System
- A Sub-Nyquist Radar Prototype: Hardware and Software

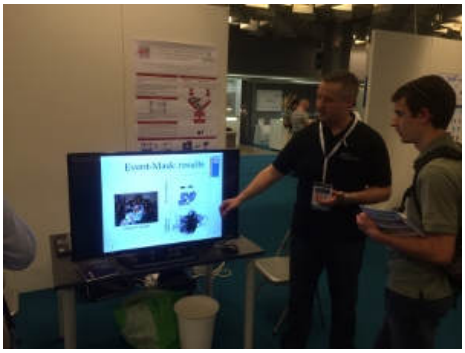
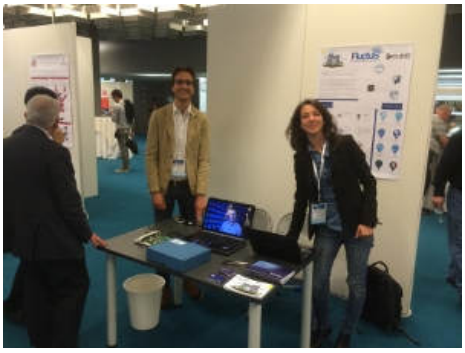
the winners are closely followed by

- Teaching Digital Signal Processing on Smartphones: A Mobile DSP Laboratory
- Object recognition in visual sensor networks based on compression and transmission of binary local features
- Cryptographically Secure Radios Based on Directional Modulation

Special appreciation goes to the demo:
- Interactive Dynamic Soundfield Rendering with Visual Feedback
which was kept off the competition.

Tuesday, May 6

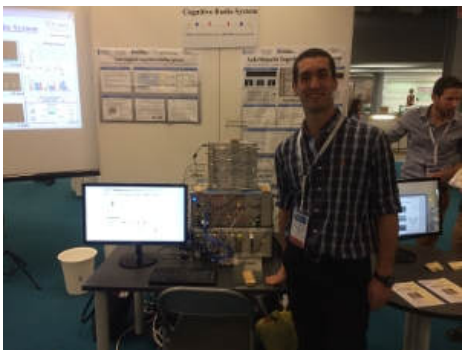
Afternoon, 16:15 - 18:15

Title	Authors	Picture
The Flexible Audio Source Separation Toolbox Version 2.0 [download demo description]	Yann Salaün (Inria, France), Emmanuel Vincent (Inria Nancy-Grand Est, France), Nancy Bertin (IRISA - CNRS UMR6074, France), Nathan Souviraà-Labastie (Université Rennes 1, France), Xabier Jaureguiberry (Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, France), Dung Tran (INRIA/LORIA, France), Frédéric Bimbot (IRISA (CNRS & INRIA), France)	
ALADIN: The self-taught vocal interface [download demo description]	Jort Gemmeke (KU Leuven, Belgium)	
EventMask: a game-based analysis of event-related saliency in photo galleries [download demo description]	Andrea Rosani, Duc Tien Dang Nguyen, Giulia Boato, Francesco De Natale (University of Trento, Italy)	
FluctuS - Intelligent Sensor System, and the IoTPrise Community [download demo description]	Marco Magnarosa (CUBIT Scarl, Italy)	

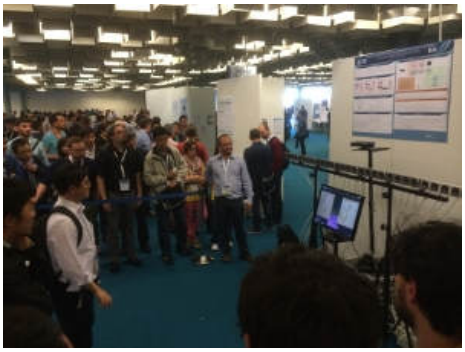
Wednesday, May 7

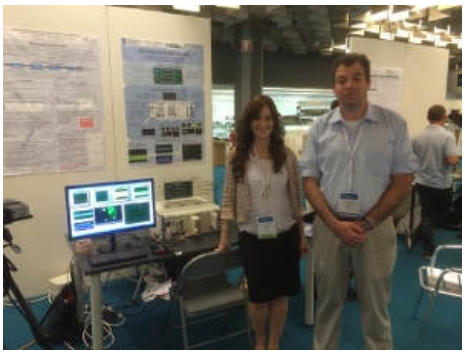
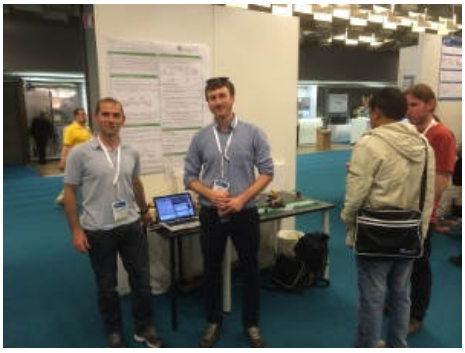
Morning, 11:00 - 13:00

Title	Authors	Picture
Cryptographically Secure Radios Based on	Vincenzo Pellegrini, Fabio Principe, Giacomo	

Directional Modulation [download demo description]	de Mauro, Rodolfo Guidi, Valerio Martorelli, Riccardo Cioni (IDS-Ingegneria dei Sistemi SpA, Italy)	
Sub-Nyquist Cognitive Radio System [download demo description]	Shahar Tsiper, Etgar Israeli, Deborah Cohen, Eli Shoshan, Alex Raisenson, Yonina C. Eldar (Technion-Israel Institute of Technology, Israel)	
A Demonstration of a Single Channel Blind Noise Reduction Algorithm with Live Recordings [download demo description]	Narimene Lezzoum, Ghyslain Gagnon, Jeremie Voix (ETS, University of Quebec, Canada)	
Multi-Sensor Ambient Assisted Living System for Fall Detection [download demo description] [download poster]	Ahmet Yazar (Bilkent University, Turkey), Fatih Erden (Hacettepe University, Turkey), A. Enis Cetin (Bilkent University, Turkey)	

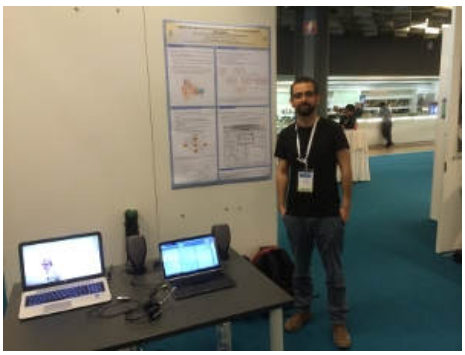
Afternoon, 16:15 - 18:15

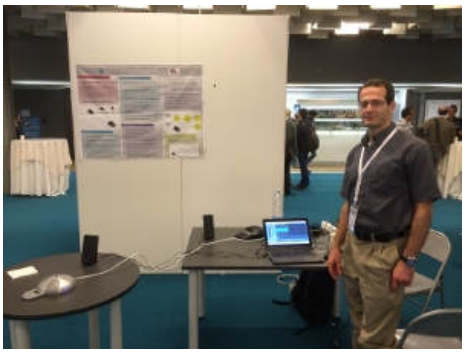
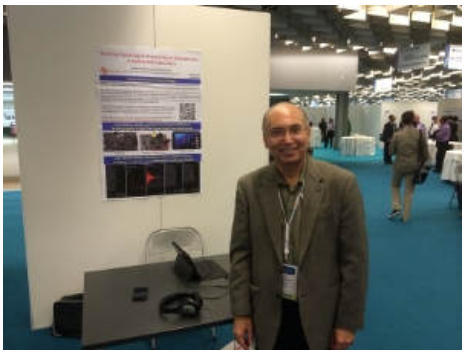
Title	Authors	Picture
Interactive Dynamic Soundfield Rendering with Visual Feedback [download demo description]	Fabio Antonacci, Lucio Bianchi, Dejan Marković, Antonio Canclini, Augusto Sarti, Stefano Tubaro (Politecnico di Milano, Italy), Roberto Magalotti (B&C Speakers, Italy)	

A Sub-Nyquist Radar Prototype: Hardware and Software [download demo description]	Tal Nagar, Idan Shmuel, Yonina C. Eldar, Eli Shoshan, Omer Bar-Ilan (Technion - Israel Institute of Technology, Israel)	
Object recognition in visual sensor networks based on compression and transmission of binary local features [download demo description]	Antonio Canclini, Luca Baroffio, Matteo Cesana, Alessandro E. C. Redondi, Marco Tagliasacchi (Politecnico di Milano, Italy)	
Real-Time Near-End Listening Enhancement for Mobile Phones [download demo description]	Bastian Sauert, Florian Heese, Peter Vary (RWTH Aachen University, Germany)	

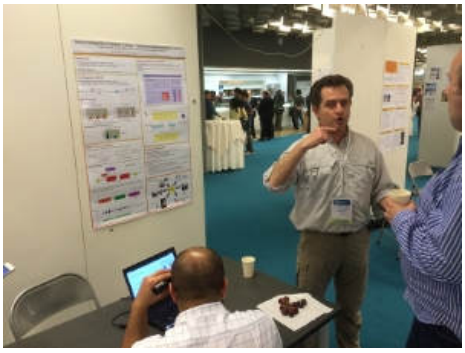
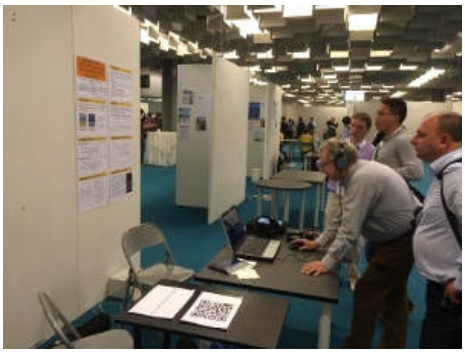
Thursday, May 8

Morning, 11:00 - 13:00

Title	Authors	Picture
Speech Reception Threshold Measurement Using Automatic Speech Recognition [download demo description]	Emre Yilmaz, Joris Pelemans (KU Leuven, Belgium), Stefan Lievens (Cochlear Technology Center, Belgium), Hugo Van hamme (KU Leuven, Belgium)	
The AXES Research video search system [download demo description]	Kevin McGuinness (CLARITY: Centre for Sensor Web Technologies, Ireland), Robin Aly (University of Twente, The Netherlands), Ken Chatfield, Omkar Parkhi, Relja Arandjelovic (University of Oxford, United Kingdom), Matthijs Douze (INRIA Rhône Alpes, France), Max Kemman, Martijn Kleppe (Erasmus University Rotterdam, The Netherlands), Peggy van der Kreeft, Kay Macquarrie (Deutsche Welle, Germany), Alexey Ozerov (Technicolor Research & Innovation, France), Noel E O'Connor (Dublin City University, Ireland), Franciska de Jong (University of Twente, The Netherlands), Andrew Zisserman (University of Oxford, United Kingdom), Cordelia Schmid	

	(INRIA, France), Patrick Perez (Technicolor, France)	
Real-time microphone selection in noisy reverberant environments for teleconferencing systems [download demo description]	Israel Cohen (Technion, Israel), Baruch Berdugo (MRD Technologies Ltd., Israel), Joseph Marash (Phoenix Audio Technologies, USA)	
Teaching Digital Signal Processing on Smartphones: A Mobile DSP Laboratory [download demo description]	Nasser Kehtarnavaz, Shane Parris(University of Texas at Dallas, USA)	

Afternoon, 16:15 - 18:15

Title	Authors	Picture
Real-Time Speech-in-Noise Intelligibility Enhancement based on Spectral Shaping and Dynamic Range Compression [download demo description]	Vassilis Tsiaras, Tudor Cătălin Zorilă, Yannis Stylianou (University of Crete, Greece), Masami Akamine (Toshiba, Japan)	
Singing Voice Correction System with Smartphone Interface [download demo description]	Elias Azarov, Maxim Vashkevich, Denis Likhachov (Belarusian State University of Informatics and Radioelectronics, Belarus), Alexander Petrovsky (Bialystok Technical University, Poland)	
HEVC Hardware Decoder Implementation for UHD Video Applications [download demo description]	Seunghyun Cho, Hyunmi Kim, Kyungjin Byun, Nakwoong Eum (Electronics and Telecommunications Research Institute (ETRI), Korea)	

Contacts

For further questions, please contact the Show &Tell committee chair Augusto Sarti at the address:
showandtell@icassp2014.org

IEEE ICASSP - International Conference on Acoustics, Speech and Signal Processing

We wish to thank the following sponsors for their contribution to the success of this conference:

Main Sponsor:

Selex ES
A Finmeccanica Company

Sponsors of the ICASSP 2014 Student Travel Grants:

REGIONE TOSCANA
NSF
QUALCOMM
UNIVERSITÀ DI PISA

Tricolor sponsors

Springer
amazon.com
IBM WATSON

Red sponsors

NUANCE
MathWorks
Google
Microsoft
THALES

White sponsors

MITSUBISHI ELECTRIC
IDS INGEGNERIA DEI SISTEMI
TEXAS INSTRUMENTS

Green sponsors

Consilium RADARS When Safety Matters
speechocean 海天瑞声
MBDA MISSILE SYSTEMS
elital ELETTRONICA ITALIANA

Baidu 百度
Telespazio A Finmeccanica/Thales Company
ThalesAlenia Space

Patrons


REGIONE TOSCANA
PROVINCIA DI FIRENZE
COMUNE DI FIRENZE
Ministero dello Sviluppo Economico
UNIVERSITÀ DI PISA
UNIVERSITÀ DEGLI STUDI FIRENZE
UNIVERSITÀ DI SIENA