Getting to know...

Sergio Escalera, IAPR Fellow

Sergio Escalera is a Full Professor at the Department of Mathematics and Informatics of the Universitat de Barcelona, where he is the head of the Informatics degree. He leads the Human Behavior Analysis Group. He is a Distinguished Professor at Aalborg University. He has been a visiting professor at TU Delft and Aalborg Universities. He is also a member of the Computer Vision Center. He is vice-president of the ChaLearn Challenges in Machine Learning, leading the ChaLearn Looking at People events. He is co-creator of the Codalab open source platform for challenges organization and co-founder of the NeurIPS competition and Datasets & Benchmarks tracks.

He is also a Fellow of ELLIS, the European Laboratory for Learning and Intelligent Systems working within the Human-centric Machine Learning program. He has been chairing and is the current vice-Chair of IAPR TC12: Multimedia and Visual Information Systems. He participated in several international funded projects and received an Amazon Research Award.

He has published more than 300 research papers and participated in the organization of several scientific events. He received a CVPR best paper award nomination and a CVPR outstanding reviewer award. His research interests include inclusive and transparent analysis of humans from visual and multi-modal data.

Sergio Escalera, IAPR Fellow
ICPR 2022, Montreal

For contributions in pattern recognition and human behavior analysis

Since I started my PhD in 2003 at the Computer Vision Center in Barcelona, I have been always interested in the creation of Pattern Recognition models that can discriminate a large set of visual categories. My PhD was focused on visual multi-class classification. At that time, we were seeking a proper trade-off between discriminative visual descriptors and accurate classification strategies. While I started with the research of some point of interest detectors and visual descriptors, I moved quickly to the design of ensemble of experts based on error correcting output codes (ECOC). The idea was seminal by TG Dietterich and G Bakiri in 1994 in JAIR, and we worked on its generalization to any dense and sparse design, to the combination of an arbitrary number of binary or multi-class classifiers, including neural networks, to the automatic generation of ensembles based on data (S Escalera et at. TPAMI 2008, S Escalera et at. TPAMI 2010).

More recently we showed how deep learning architectures could also benefit from our research based on error correction output coding and designed general-purpose deep architectures with a low dimensional target embedding to discriminate many categories, "Beyond One–hot Encoding: lower dimensional target embedding, Image and Vision Computing", IMAVIS in 2018.

At the time of defending my PhD in 2008 I got very interested in automatically analyzing a much more complex visual "object", the human being. My interest was 2-fold, methodological and practical. From a methodological point of view, the articulated nature of the human body, its changes in appearance, presence of self-occlusions, the background vs body relationships (scene information, human-object, human-human interactions, etc.), and the huge variability and multi-modality nature of human behaviors in video were exciting challenges to research. From a practical point of view, if intelligent systems can understand our behavior, they will be able to assist us in an accurate and personalized way. With the previous aims in mind, in 2009 I created the Human Behavior Understanding group, HuPBA, at the University of Barcelona and the Computer Vision Center.

During the last 15 years, I have focused on analyzing human faces, hands, bodies, behaviors (e.g. actions and gestures) from visual and multi-modal data, and thanks to the support of a great team of students and national and international collaborators, we have been pushing research on human behavior understanding (HBU) and affective computing. In 2020, I had the privilege to chair the IEEE Faces and Gestures as the general co-chair. The developed research has been transferred to society in projects to support emotional evolution of children with stroke in neuro-rehabilitation game scenarios, emotional analysis in mental health treatment with schizophrenia diseases, diagnosis support in children with attention deficit disorder, risk event recognition for people with reduced autonomy, automatic recognition of sign language for the deaf, and virtual coaching for the elder for healthy living, among others.

From 2013, and after winning one of the challenges organized by ChaLearn related to gesture recognition at ICPR 2012, I became co-president of ChaLearn, a non-profit organization in Berkeley for the organization of scientific challenges. Since 2013, I have been leading the ChaLearn Looking at People (LAP) events (*https://chalearnlap.cvc.uab.cat/*) organizing competitions and associating them to international events to push research on HBU while approaching real application needs.

Until now, together with ChaLearn President Prof. Isabelle Guyon and collaborators, we have created and made public for research purposes more than 20 large annotated human-centered databases to support applications for good related to human behavior understanding. We have organized associated scientific competitions, workshops, and special issues, making all material public under proper licenses. Thousands of participants joined our competitions, which were also supported by several sponsors, including Microsoft Research, Google, NVIDIA Corporation, Disney Research, Facebook, and Amazon. We also co-developed the Codalab open source platform for challenges organization (*https://codalab.org/*), and I participated as co-founder of the NeurIPS competition track and advisory board for the creation of the NeurIPS Datasets & Benchmarks tracks. These initiatives aim to increase sensitivity to the importance of data. There is no bias-free data, so it is very important to properly design benchmarks and understand their research versus practical impact, in my case always considering a human-centered perspective. I also push this human-centered perspective as a Fellow of the European Laboratory for Learning and Intelligent Systems, ELLIS, working within the Human-centric Machine Learning program.

Clearly, it is important to comment on HBU research progress that was allowed thanks to the deep learning revolution. Since 2012, deep learning architectures have been defining the state-of-the-art in several pattern recognition and computer vision problems, including those related to HBU. Now, the robustness of our HBU technology allows us to transfer it to several real scenarios. However, this new incursion of intelligent systems into daily living scenarios is generating a global alarm associated with ethics and data protection. From my point of view part of the technology developed by the research community has been transferred rapidly to society without proper democratization and regulation. We sometimes, unfortunately, delivered works without proper ethics analyses about their implications when applied to real cases, possible negative uses or possible associated bias and lack of generalization they may have under specific conditions. However, I am pleased to see that recently our community is seriously considering accountability in research, and that trustworthy AI is becoming a mandatory standard.

In relation to the current state of the technology, while the automatic recognition of fine-grain events in video is increasing its robustness, still the field of human behavior understanding is in its infancy. We should keep pushing interdisciplinary research to better understand human behavior that can be transferred into intelligent systems. In this regard, our group's current research interests on HBU are around self-supervision (to benefit from large non-annotated data), domain knowledge injection (to use some partial knowledge we have about the world), bias detection and mitigation (to allow for fair and inclusive systems), uncertainty estimation and explainability (to allow human-in-the-loop systems and promote acceptability of AI), continuous learning (to update model knowledge in a changing world), human behavior forecasting and behavior anticipation (for the early update of AI behavior), multimodal video understanding (to allow for fine-grain HBU), and domain adaptation and synthetic data (to allow models to generalize to different environmental conditions).

I hope progress on these research lines will further help in the design of inclusive and transparent applications of HBU and affective computing.