

# BMVA News

The Newsletter of the British Machine Vision Association and Society for Pattern Recognition

Volume 28 Number 2  
December 2017



**Editor:** Professor Roy Davies  
Department of Physics  
Royal Holloway, University of London  
Egham, Surrey, TW20 0EX  
tel: +44(0)1784 435064  
email: e.r.davies@rhul.ac.uk

<http://www.bmva.org/>

**BMVA** News<sup>1</sup> is published every three months. Contributions on any activity related to machine vision or pattern recognition are eagerly sought. These could include reports on technical activities such as conferences, workshops or other meetings. Items of timely or topical interest are also particularly welcome; these might include details of funding initiatives, programmatic reports from ongoing projects and standards activities. Items for the next edition should reach the Editor by 10 March 2018.

## Contents

Editorial: <i>What Sort of Brain do we need to cope with Vision and Visualisation at the Same Time?</i> .....	1
The Sullivan Thesis Prize and the BMVA Thesis Archive .....	2
IAPR Internship Brokerage .....	2
Report on CVPR 2017 .....	3
Report on CVPR 2017: Workshop on Deep Learning for Robotic Vision... ..	4
Report on MVA 2017 .....	6
Meeting Report: Human Activity Monitoring and Recognition.....	8
Meeting Report: Computer Vision and Modelling in Cancer.....	9
Cognitively Inspired Explainable Perception-Based AI.....	10
Upcoming Meetings.....	11
Travel Bursaries for International Conference Attendance .....	11
5 <sup>th</sup> Edition of Book on Computer Vision .....	11
BMVC 2018 to be held in Newcastle upon Tyne.....	12
MIUA 2018 – Call for Papers .....	13
Report on ICCV 2017 .....	14

## Editorial: *What Sort of Brain do we need to cope with Vision and Visualisation at the Same Time?*

When I was young my father (head of maths in Cardiff High School) seemed to know everything that mattered – how electromagnets work, how cars work, how planes stay up,

how to wire landing light switches so that any of them can switch the lights, what stars are, and all about science and maths – not to mention why the angles of a triangle add up to 180°. As I got older, his knowledge seemed to expand to cover a huge amount more – atoms and molecules, biology, and even how electronic valves work. His capabilities as a teacher and mentor lasted at least until I had my doctorate, though at that stage they deviated, as he wanted me to be a physicist, whereas somehow I wanted to be an inventor (this came about because at the age of six the Inventor in *Tales from Toytown* had made a big impression on me). In the end I metamorphosed into an electronic engineer, and, following the flow of technology, a microprocessor aficionado, then a pattern recogniser, and finally a vision guru. In the circumstances it would have been surprising if I could have followed my father's idealised path past my thirties and his sixties. Indeed, it would be surprising if many people could follow their own youthful aspirations right through their lives: coal mining has long since gone out; bus driving is about to; and what about the fire brigade people in *Trumpton*? Robots will undoubtedly make huge changes in their lives too.

But where did my father get his phenomenal knowledge? He seemed not to have had any notable teachers or mentors, and his parents had no knowledge of science and maths. His secret was his immense love of reading. This also covered detective stories and many 'good yarns', including quite a bit of SF, even though in the 1930s that subject was in its infancy. Going from conversations we had long, long ago, *Alf's Button* (based on a portion of Aladdin's lamp having been melted down to make buttons for the military) was a book that had made an impression on him, as did *The Death of Metal*, in which all metals suddenly became soft and useless; and of course *The War of the Worlds*; and the well-worn story *Vice Versa*, in which a father and son's minds are suddenly swapped over.

Perhaps the one tale he told that made the greatest impression on me was the one about a man who couldn't see what was in front of him – but instead saw scenes from elsewhere. This was obviously a nightmare and far worse than being blind. Yet as time went on, I realised that it might in one sense be not far from the truth. For we can all recall, imagine or dream scenes from 'elsewhere': our brains have an extra circuit that permits this to happen. Not only that,

<sup>1</sup> The British Machine Vision Association and Society for Pattern Recognition is a Company limited by guarantee, No. 2543446, registered in England and Wales. Registered Office: Granta Lodge, 71 Graham Road, Malvern, WR14 2JS. The Association is a non-profit-making body and is registered as charity No. 1002307.

but it interleaves neatly with our normal vision. Clearly, what is happening is that much of the wiring that allows us to see can be usurped by a visualisation sense, bypassing the visual input and flowing around the brain in the normal stream of consciousness. In fact, it would be a great pity if the chance to ‘overload’ the brain’s wiring in this way did not exist. However, to work properly, there has to be a switch allowing visual inputs to be replaced by other inputs from within the brain, and also for these other inputs to have the same topology as the visual channels they are replacing. But all this demands that we can validly switch off the visual channels without getting confused. That is a very demanding condition. For example, if we were driving a car, how would it be if we started dreaming or even imagining maps of the road ahead? At the minimum, this would have to be concerted very carefully, or we would be crashing our cars and running over pedestrians. Nevertheless, there are many instances of people going to sleep at the wheel – and it is possible that some of these correspond to dreaming or living temporarily in parallel worlds of our own making. And what is easier than conversing on a mobile phone and picturing the person talking to us? So we need to ask what internal safeguards we have against this. Unfortunately, the human mind has too few thinking channels. Basically, we have one, and this is normally used for vision, though clearly it can temporarily be swapped for important (visual) thoughts. Here, the cerebellum is vitally important to us as it allows instructions to be downloaded from the brain to our motor functions (e.g., “walk along the pavement”, “move the steering wheel to follow the road”, “hit the tennis ball back again”): with training, these instructions can be very accurately implemented without further conscious effort. Indeed, further conscious effort would slow down our reactions so much as to virtually incapacitate us.

To add substance to these musings, note that when people are talking to us, they move their eyes in very telling ways – often not looking at us but letting their eyes dart around autonomously and apparently randomly. In fact, they sometimes fix their gaze on points in the sky, which can be construed as allowing their thoughts free rein while thinking more deeply about what they are saying. This is accepted behaviour and is not regarded as rude, even though it may appear contrary to the eye-contact behaviour we expect when we are trying to assess the other person’s state of mind. [Actually, eye-contact itself is seldom maintained for more than a second or two at a time, as it tends to be interpreted as unacceptable dominance-related behaviour.]

I doubt whether we could cope with more thought channels than just the one current one. However, one view is that this consciousness channel is a serial processing channel which is accompanied by huge numbers of parallel processing channels working away in the background. How else could we get sufficient thought processing bandwidth out of the rather slow neurons in our brains? And don’t forget that we have upwards of  $10^{10}$  neurons in our brains, a great proportion of which are connected to  $\sim 10^4$  other neurons: it is a puzzle indeed that there isn’t the possibility of a goodly number of conscious thought channels available to us. But that is not the way that evolution has built us. It remains to be seen whether the future *really* deep neural networks will be able to overcome this limitation, or even whether it is a real limitation. Perhaps the truth is that we will need more powerful cerebella to cope with the highly

complex world that we will find ourselves in in the future: or should I say that our robot overseers will need this capability?

*Addendum 1:* In fact, it has been said that the human brain contains nearly  $10^{11}$  neurons, of which 80% are in the cerebellum controlling motor functions – which means that we are mechanically clever rather than intellectually clever.

*Addendum 2:* I imagine that very few readers will realise that the David Marr who is famous for his theories of early vision is the same as the David Marr who first unravelled the workings of the cerebellum in 1969.

Professor Roy Davies  
Editor, BMVA News  
email: e.r.davies@rhul.ac.uk

## The Sullivan Thesis Prize and the BMVA Thesis Archive

Every year, the BMVA awards a prize for the best thesis out of those brought to its attention as having been examined in the previous calendar year. The prize is awarded in the name of Geoff Sullivan, who played a significant role in the early days of the BMVA.

If you are in the final stages of writing up your thesis, please consider submitting it to the BMVA’s thesis archive: all the information is on the BMVA website. If you are supervising a PhD student who you think has done particularly well, please consider entering him or her for the Sullivan prize; again, the procedure is on the website.

Dr Adrian Clark  
BMVA Chair  
email: alien@essex.ac.uk

## IAPR Internship Brokerage

The BMVA is affiliated to the International Association for Pattern Recognition (IAPR – <http://www.iapr.org/>). The IAPR’s Industrial Liaison Committee wishes to promote opportunities for students to undertake internships at companies working in the areas of Pattern Recognition, AI, Computer Vision, Data Mining, Machine Learning, etc.

The IAPR is supporting a web-based internship listing service, where companies can present their internship opportunities. Students can browse the listings and contact companies directly. A first version service appears at: <http://homepages.inf.ed.ac.uk/rbf/IAPR/INDUSTRIAL/>

Companies are welcome to advertise new internship positions: you can see examples at the URL given above.

Please forward this information to anyone who you think would find it useful.

Bob Fisher  
Chair, IAPR Industrial Liaison Committee  
email: rbf@inf.ed.ac.uk

## Report on CVPR 2017

The 30<sup>th</sup> IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) was held this year at the Hawaii Convention Center from 21 to 26 July in Honolulu, Hawaii. The main conference took place during 22–25 July, while its co-located tutorials and workshops were organized on 21 and 26 July. CVPR is the premier annual computer vision event in the field of computer vision and pattern recognition. Based on the conference h5-index provided by Google Scholar Metrics, CVPR ranked first with 158 values of h-index in both Image Processing and Computer Vision and Computer Science. With the most advanced, cutting-edge research in the world, CVPR provides exceptional value and drives innovations for both academics and industry researchers. It was my pleasure to take part in CVPR this year with over 5000 people for such a feast. As the technology heavyweights said, “the best time is yet to come”.

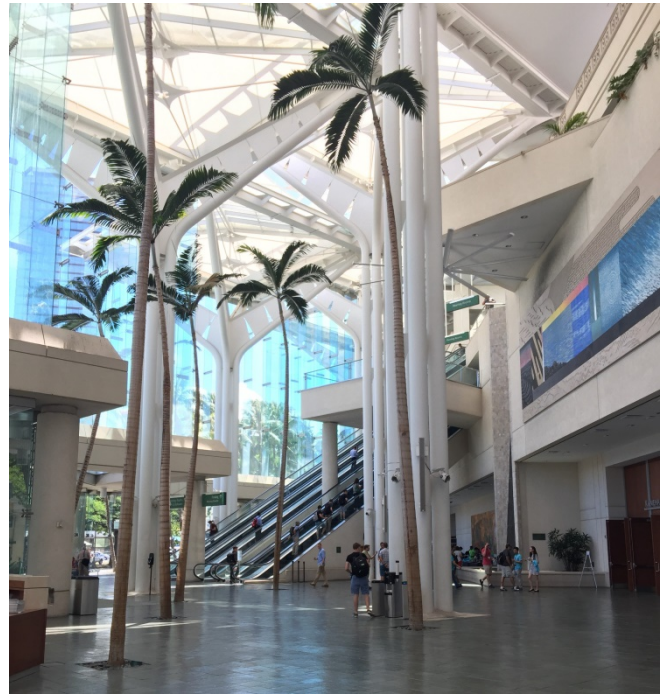


Outside the Hawaii Convention Center

For the detailed statistics of CVPR this year, out of 2680 valid submissions to the main conference, 2620 were fully reviewed, as the others were administratively rejected for technical or ethical reasons or were withdrawn before review. The reviewing committee took 85 researchers as Area Chairs and recruited a record number of experienced reviewers. From the 2620 fully reviewed papers, 783 were accepted with a 29% acceptance ratio of valid submissions. It is worth noting that the number of accepted papers at CVPR 2017 was 22% more than at CVPR 2016. My work “Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding” was placed in the biomedical image/video analysis field (1.53% of accepted papers) presented on 25 July.

In particular, 71 of all accepted papers were allocated as oral presentations (2.65%), each of which had 12 minutes, and 144 were selected as spotlight presentations (5.35%) each of which had 4 minutes. Therefore, a total of 8% of valid submissions had 2 types of oral presentation. Beside live presentations, all papers were presented in the iterative poster sessions to ensure more lively discussions.

The first day of the main conference started with an opening ceremony including a brief welcome and paper award presentation in the Kamehameha III Theatre after the morning registration and breakfast. Award announcement was very high profile, with continuous applause in the venue. This year, there were 2 best paper awards, 2 best paper



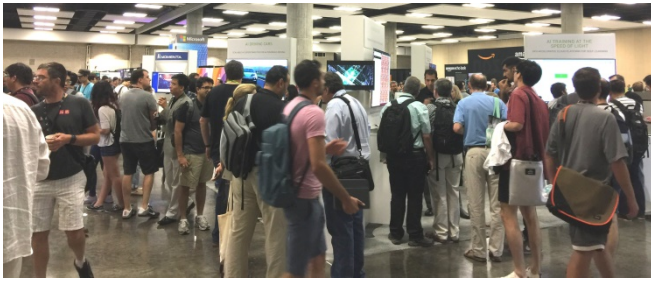
Conference venue

honourable mention awards and 1 best student paper award. The first best paper was awarded to Gao Huang, Zhuang Liu, Laurens van der Maaten and Kilian Q Weinberger for “Densely connected convolutional networks”. The second was awarded to Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang and Russell Webb for “Learning from simulated and unsupervised images through adversarial training”. Then, three parallel spotlight and oral sessions about machine learning, 3D vision and low and mid-level vision were distributed in Kamehameha III, Kalakaua Ballroom A-B and Kalakaua Ballroom C respectively, followed by a two-hour poster session. Exhibitions and demos ran parallel to the poster sessions. CVPR 2017 Expo provided a first-hand look at advanced computer vision and pattern recognition technologies and career opportunities for each attendee.



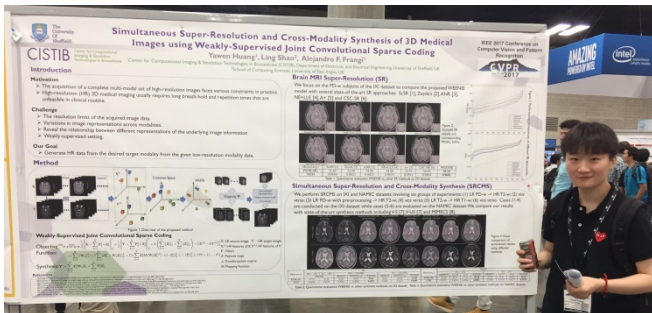
Opening ceremony (welcome and paper awards) in Kamehameha III theatre

After the coffee break and delightful lunchtime, there were three parallel spotlight and oral sessions about object recognition and scene understanding—computer vision and language, analysing human and image motion and tracking; video analysis followed by a corresponding poster session in the afternoon. The first day ended with a plenary talk which presented “Extracting social meaning from language” and was given by Professor Dan Jurafsky, professor and chair of linguistics and professor of computer science at Stanford University.



Exhibitions of advanced hardware systems and software

The following three days of the main conference permitted three parallel spotlight and oral sessions and a poster session in the morning and afternoon respectively. The plenary talk during the second day had a focus on computer vision and was given by Dr Harry Shum, who is the executive vice president of Microsoft's Artificial Intelligence and Research group, in a great talk named "Commercializing computer vision: success stories and lessons learned". My poster was presented on 25 July (the last day of the main conference) in the morning. During over 2 hours of presentation, I demonstrated the results of my research and discussed relevant topics with other researchers. I was especially encouraged that my poster presentation was very well attended. The inspiring discussions around my topic (which was about image synthesis and image super-resolution of medical images) provided some constructive feedback and interesting feature trends.



Yawen Huang presenting a poster at CVPR 2017



Plenary talk by Professor James J DiCarlo

The main conference ended with an interesting and valuable keynote talk on "The science of natural intelligence (NI): reverse engineering primate visual perception" given by Professor James J DiCarlo, Professor of Neuroscience and head of the department of brain and cognitive sciences at MIT. In his talk, he outlined how the efforts of brain and cognitive scientists can help to solve the challenges of reverse engineering the mind and forward engineering, aiming to emulate the mind. With more than 5000 attendees from many countries, CVPR 2017 was very well attended and provided an opportunity for each attendee to exchange

ideas, share resources, meet external brilliance and create connections with researchers of similar interests. Personally, I benefited a lot from both oral presentations and interactive poster sessions.

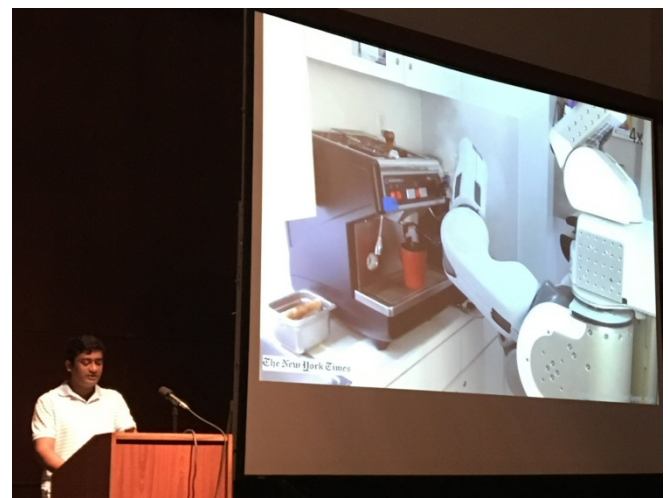
Finally, I would like to sincerely thank the BMVA for a travel bursary giving me the financial support to attend CVPR 2017.

Yawen Huang  
University of Sheffield  
email: yhuang36@sheffield.ac.uk

## Report on CVPR 2017: Workshop on Deep Learning for Robotic Vision

Last July I attended the workshop on Deep Learning for Robotic Vision, which was held in Honolulu, Hawaii (USA) in conjunction with the IEEE Conference on Pattern Recognition (CVPR 2017). The aim of the workshop, organised by researchers from Google and the Australian Centre for Robotic Vision, was to present recent advances in deep learning techniques for robotics applications, which present unique challenges such as real-time processing, accurate 3D understanding of the world, and scalability. The workshop consisted of several interesting keynote talks by renowned experts on their recent work, a poster session, and a discussion panel by the keynote speakers.

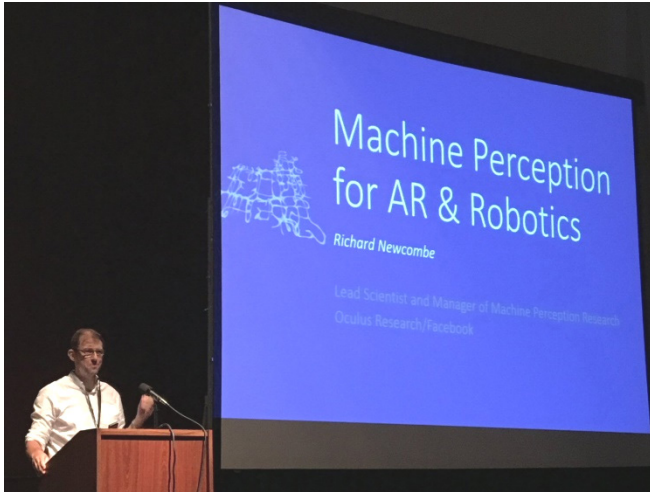
The first talk was given by Ashutosh Saxena from Caspar.AI (USA), who presented a futuristic view on how homes should intelligently adapt to the people living in them. Saxena highlighted the importance of three key factors to achieve smart homes: perception, control and communication between components. As an example, he presented the 'Robobarista', a robot which can learn how to make coffee simply by observing humans.



Ashutosh Saxena presenting the 'Robobarista'

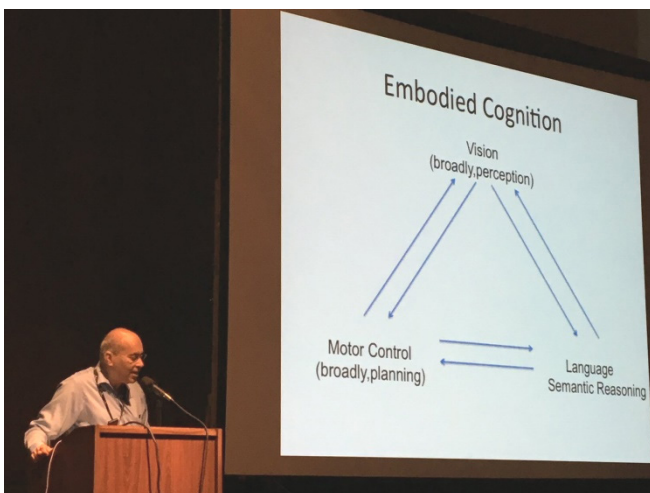
Richard Newcombe from Oculus Facebook (USA) followed with a talk on augmented reality (AR) and how, according to him, its final purpose is to make super-human powers available to everyone. This includes deeper memory (the computer will remember where you forgot your keys),

more accurate senses (the computer will tell you the exact speed of the car in front of you) and social assistance (the computer will tell you more about who is around you). Newcombe presented three key problems to solve: information understanding, synchronization with reality, and ‘always on’ perception. He also emphasised the challenge of modelling the state of the world, particularly in robotics where the environment changes through human interaction.



Richard Newcombe on the future of AR and robotics

Professor Jitendra Malik from University of Berkeley (USA) started with a historical overview on human development arguing that – paraphrasing Anaxagoras – humans are the most intelligent animals, thanks to their hands. Because of this, he believes that robots need to learn as humans do: with external supervision (parents, teachers) and with self-supervision (experimenting by themselves, playing). As an example of self-supervision, he presented a recent work consisting of a robot that learns how to move objects by randomly poking them and using deep neural networks to learn the underlying physics of the world from the resulting images.



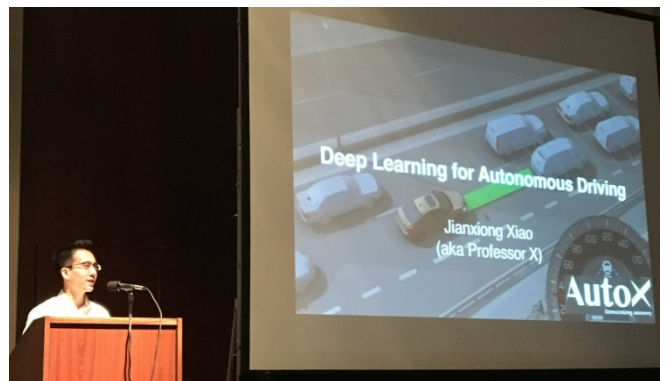
Professor Jitendra Malik: robots should learn like humans (or babies)!

The poster session took place before the lunch break. Among the most visited posters was the winner of the best workshop paper award: “Learning robot activities from first-person human videos using convolutional future regression” by Jangwon Lee and Michael S Ryoo (Indiana University, USA). Their work presented a neural network model that

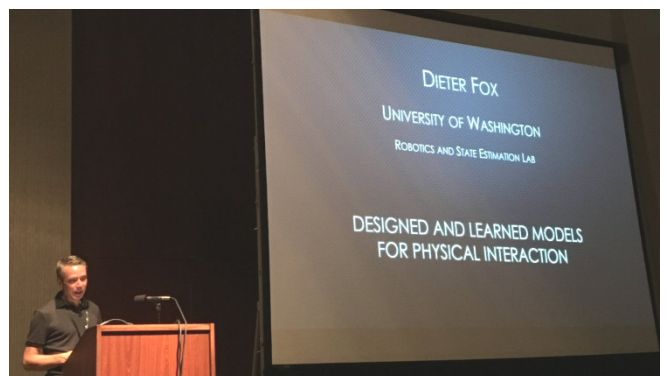
allows robots to learn how to perform activities by watching first-person videos from humans executing the same activities.

After the poster session and the lunch break, Professor Honglak Lee, from Google and U. of Michigan (USA), talked about learning disentangled representations for geometry and robotic grasping, where the key challenge is to tease apart the factors that generate images – such as pose, shape and illumination for 3D objects. According to Professor Lee, robot vision needs to mimic the human capability of being able to ‘imagine’ the geometry of a 3D object given a single 2D image. To illustrate this, he presented how a perspective transformer neural network is able to generate the 3D volume of an object given only a single-view image.

Jianyong Xiao (AutoX, USA) and Professor Raquel Urtasun (University of Toronto and Uber, Canada) presented, in two separate talks, recent advances in autonomous driving from their respective companies. Xiao predicted that autonomous driving will be one of the most invested applications of robotic vision and proposed a low-cost system using only a standard camera and neural networks. On the other hand, Professor Urtasun argued that using more sophisticated sensors can lead to better performance, and price should not be a problem when producing them on a large scale.



Jianxiang Xiao: democratising autonomous driving with deep learning



Professor Dieter Fox: Designed and learned models for physical interaction

Professor Dieter Fox, from the University of Washington (USA), emphasised the importance of good perception models to advance the field of robotics and talked about both manually designed and learned models for robot interaction. Professor Fox described deep learning as a wrapper around models or a glue that connects models to

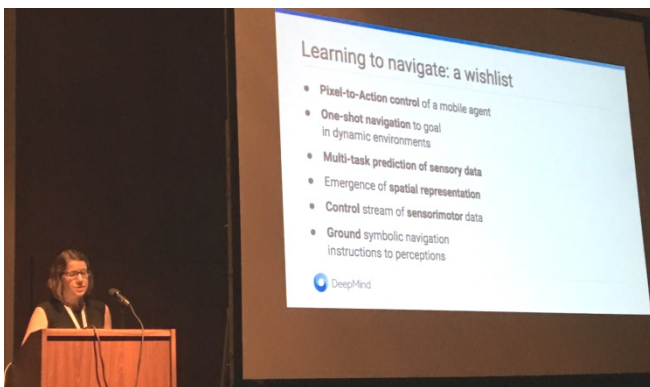
data but, in his opinion, model constraints should be integrated in deep learning to make the learning more general.

Professor Abhinav Gupta from Carnegie Mellon University (USA) discussed one of the central problems of robotic vision: labelling data manually is difficult, so how can we scale the process to take full advantage of deep learning? Professor Gupta presented a novel idea to train robots how to make robust grasps – by using two adversarial robots at the same time. The first robot (the protagonist) tries to grasp an object. The second robot (the adversary) tries to destabilise the grasp. This competition between the robots leads the protagonist to learn how to grasp objects in a solid way without requiring human intervention in the process, and thus is potentially scalable.



Professor Abhinav Gupta presented an adversarial approach for scalable robotic manipulation.

The last talk was given by Raia Hadsell from Google DeepMind (UK) where she presented recent advances in robot maze navigation using deep reinforcement learning. One interesting result was that learning can be accelerated through auxiliary processes such as depth prediction, reinforcing the assumption that good 3D perception is important for successful robotic navigation.



Raia Hadsell: Teaching agents how to navigate

The workshop concluded with a panel discussion between seven of the keynote speakers. One important question addressed was about the major differences between robot vision and computer vision. Most of the panellists agreed that the main differences are the amount of available data for learning, the absence of common benchmarks (every robot is different) and the active component of the agents (the world changes through interaction).



Panel discussion (left to right): Kevin Murphy (moderator), Honglak Lee, Raquel Urtasun, Raia Hadsell, Jitendra Malik, Dieter Fox, and Richard Newcombe

All in all, the workshop was a great success, the quality of the talks was excellent and both the speakers and the audience were passionate. The topics covered were interesting, up-to-date, and diverse – robot manipulation, autonomous driving, augmented reality, and so on. Work presented by the speakers was impressive, but what I enjoyed most was the ‘futuristic vibe’ of the talks. Promises and expectations on the future of robotics were set very high!

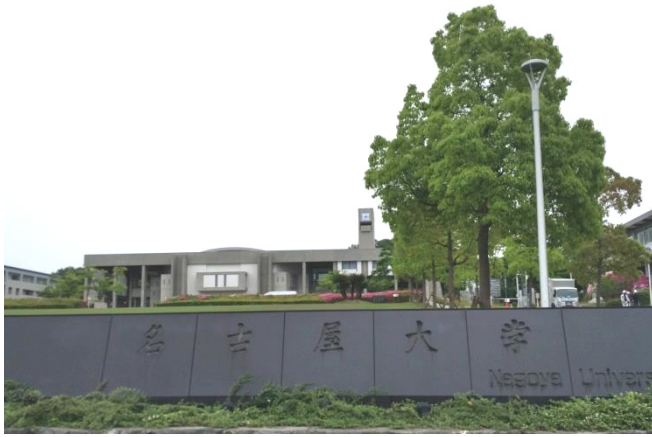
Let me conclude this report by thanking BMVA for sponsoring my conference trip to CVPR 2017 where I presented my work, attended many interesting talks, and socialised with fellow researchers. Should you have any questions about the workshop or the conference, please don’t hesitate to contact me.

Guillermo Garcia-Hernando  
Imperial College London  
email: [g.garcia-hernando@imperial.ac.uk](mailto:g.garcia-hernando@imperial.ac.uk)

## Report on MVA 2017

The 15<sup>th</sup> IAPR International Conference on Machine Vision Applications (MVA) was held on 8–12 May 2017, in Nagoya, Japan. The MVA conference aims to highlight research in machine vision algorithms and their applications, with topics covering a wide range of research areas such as 3D vision, image recognition and visual tracking. The MVA conference was first introduced in 1988 as an IAPR workshop on computer vision. In 2005, a modification was put forward to transform the workshop into an academic conference following an increase in both the number of papers and the overall quality. The MVA conference is scheduled to be held every other year and is based mainly on Japanese academia and industries. This year, the conference was held at the Higashiyama Campus of Nagoya University, one of the most prestigious universities in Japan.

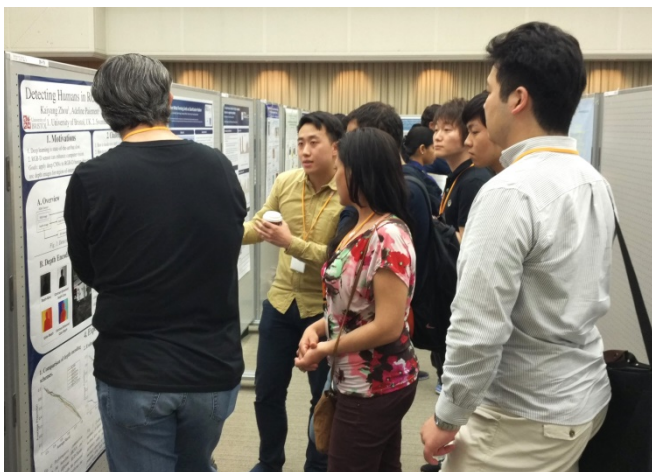
The main conference was held on 8 May for three consecutive days. Two tutorials and technical exhibitions around the computer vision labs at Nagoya University were organized on 11 and 12 May. Similar to other computer vision conferences, the main conference comprises oral sessions, poster presentations, lectures, demos, exhibitions and social events. In the demo sessions, nine internationally renowned companies were invited to introduce and present each of their cutting-edge machine vision products and technologies with the aim of motivating and securing potential collaborations. For example, a HUAWEI team showed the latest 3D face modeling feature that has been deployed in the HUAWEI mobile phones.



Entrance of the Nagoya University

The first day of the main conference on 8 May started with an opening speech given by Professor Hiroshi Ishikawa (Waseda University), the General Chair of MVA 2017. After the opening speech, three oral sessions took place, followed by corresponding poster presentations. In the last oral session, Yuzuko Utsumi from the Osaka Prefecture University presented her paper titled “Fast search based on generalized similarity measure”. The paper proposed a method to accelerate a generalized similarity measure for general recognition tasks, and was eventually given the Best Paper Award.

The subsequent two days of the main conference on 9 and 10 May followed an oral-poster-oral pattern where each poster session was assigned approximately 2 hours of time for researchers to communicate with each other and exchange valuable ideas. I presented my poster “Detecting humans in RGB-D data with CNNs” in the afternoon poster session. During the presentation I introduced the main ideas of my paper to other researchers and also discussed recent advances in related research areas. The full presentation session went smoothly. I had the opportunity to meet a number of experts, from whom I received constructive feedback which should be useful for my future research.



Kaiyang Zhou presenting his paper to other researchers

In addition to the oral and poster sessions, three IAPR distinguished lectures were delivered by Professor Daniel Cremers (Technical University of Munich), Professor Jianbo Shi (University of Pennsylvania) and Professor Hiroshi Murase (Nagoya University) during the main conference days. Professor Cremers gave a brief review on the

background of 3D reconstruction and visual SLAM. Related projects conducted in his lab, evolving from RGB image-based methods to recent RGB-D image-based ones, were also introduced. Professor Shi discussed his research on first person camera perception, with a main focus on attention-based modeling. Lastly, Professor Murase talked about his research on image recognition technologies for driving assistance.

Apart from the academic activities, there were also opportunities for socializing and interactions during lunch and evening meals. On 9 May, a Young Researchers’ Meeting was organized during the lunch break. This special session served as a platform for academic researchers and company recruiters to build connections and discuss potential collaborations and career opportunities. On the evening of the same day, the MVA 2017 banquet was held in the Westin Nagoya Castle Hotel. The beautiful scenery of the Castle could be seen through the giant transparent floor window in the banquet room. All guests including academic researchers and industrial engineers were served with traditional Japanese food. Near the end of the banquet, two awards were presented by the General Chair for papers with a significant influence on machine vision technologies over the decade. The first award went to the paper “One fish, two fish, butterflyfish, trumpeter: recognizing fish in underwater video”, authored by Andrew Rova, Greg Mori, and Lawrence M Dill (Simon Fraser University); the second award went to the paper “A video motion capture system for interactive games”, authored by Ryuzo Okada, Nobuhiro Kondoh, and Bjorn Stenger (Toshiba Corporation).



An early view in the MVA 2017 banquet held in the Westin Nagoya Castle Hotel – in front of the Nagoya Castle



Night view of the Nagoya Castle seen from the banquet room

Overall, this first experience of attending an international conference was a remarkable journey which has enabled me to expand my understanding and broaden my horizons on the computer vision field. I believe that the observations and interactions that I have gained and learned will add significant value to my academic path. Finally, I would like to thank my MSc supervisors, Professor Majid Mirmehdi

(University of Bristol) and Dr Adeline Paiement (Swansea University), for their constant support during my MSc project. I would also like to thank the BMVA for providing funds to support this extraordinary conference experience.

Kaiyang Zhou  
University of Bristol  
email: kz15291@my.bristol.ac.uk

## Meeting Report: Human Activity Monitoring and Recognition

BMVA Symposium at the BCS London: Wednesday 8 November 2017



The conference was chaired by Dr Ardhendu Behera (Edge Hill University), Dr Nicola Bellotto (University of Lincoln) and Dr Charith Abhayaratne (University of Sheffield) and took place in four sessions.

### Session I

Professor Yiannis Demiris from ICL kicked off the first session with his keynote speech making the case for human modelling for personalized robot assistance. He introduced modelling human activity at different levels of abstraction, collaboration to maximize learning, e.g., for robotic wheel chair assistance, assistive dressing or interacting with robot on cognitive tasks. Learning was divided into three parts, perceiving action, machine learning algorithms to build hierarchical user models and predictive assistance.

Dr James Charles from Cambridge talked about Personalizing human video pose estimation. He talked about using personal items such as hat or watch for discrimination. He talked about using X-factor, a background subtraction method, to obtain the human silhouette as a feature for real-time factored content for pose estimation. Next was a series of five poster presentations. Gertjan Burghouts of TNO intelligent imaging, The Netherlands, presented a method for aggression detection in footage from surveillance cameras by dividing pictures into multiple boxes of several sizes and using spatio-temporal optical flow information within those boxes. Rob Wright from the University of Kingston presented a method for exploring the depth information of static points of view to infer visual focus of attention. Lining Zhang from the University of Portsmouth presented a Multiview human activity recognition method for assisted living by combining CNN features from

different cameras. Bappaditya Debnath presented Dr Hui Fang's work on human activity recognition through CNNs where the sequential probability ratio test method was used for adaptive sampling of video frames to reduce computation. Dr Joana Silva from Fraunhofer, Portugal presented a fall detection method using sensor information to extract features. The features were used to build a supervised random forest method.

### Session II

The second session was kicked off by a keynote speech given by Professor Ian Craddock of Bristol University, in which he talked about SPHERE (Sensor platform for healthcare in a residential environment). SPHERE involves interdisciplinary research for behavioural analysis in the context of a data fusion platform. The aim is to address healthcare issues for patients suffering from chronic diseases like dementia, obesity, etc. where patients need to be monitored around the clock to analyse their behaviour. He talked about long-term data collection and monitoring by installing camera sensors at the patient's home. Next, Girmaw Abebe of Queen Mary University presented his doctoral work on Inertial vision: Cross-domain knowledge transfer for human activity recognition using wearable sensors. He talked about fusing data from inertial units and first-person videos to recognize proprioceptive activities. The method used a CNN followed by a long short-term memory (LSTM) network, transfer learning and a merit-based fusion of IMU and/or FPV streams.

### Session III

After lunch, the third session started with Professor David Hogg's keynote speech. It was about learning about human behaviour for interactive robotics and relied on a part-based approach. His method comprised a mapping environment with a routing RGBD, detection of face and body pose, learning facial identities, crowd sourcing descriptors for video clips and associating words. Yangdi Xu from the University of Bristol was the next to present his PhD research on unsupervised long-term routine modelling using dynamic Bayesian networks. His core idea was: assuming a person always performs purposeful activities at corresponding locations, spatial, pose and time-of-day information are used as sources of input for routine modelling. The work assessed variations of the independence assumptions within the DBN model among selected features. It automatically selects the number of hidden states for fully unsupervised discovery of a single person's indoor routine.

### Session IV

The fourth and final session started with Glasgow University Professor Alessandro Vinciarelli's keynote speech entitled "Body language without body: Social signals in technology mediated communication". The talk was about how nonverbal communication – one of the main channels through which people convey socially and psychologically relevant information – plays a role in settings where natural nonverbal cues (facial expressions, vocalizations, gestures, etc.) are no longer available. He discussed questions such as: can a non-verbal behavioural queue automatically synthesize an artificial robot's



instructions? Can a non-verbal behavioural queue be useful in e-commerce? He also discussed the SSPNet mobile corpus and the chat corpus used for research in social signal processing and more generally in human communication; and he discussed automatic recognition of behavioural signals using techniques such as conditional random fields.

The next presentation was by Claudio Coppola and related to his doctoral research on automatic detection of human interactions from RGB-D data for social activity classification. His main argument was that in realistic scenarios a model should be able to recognise social or individual interactions from unclipped videos. In addition to classification of the social interaction, the research carried out temporal segmentation of social activities. It used a combination of SVM, HMM and GMM to achieve its objectives.

The final presentation was made by Georgios Mastorakis on fall detection using myo-skeletal simulation. He discussed using Opensim for simulating falls. Feature selection included measuring the bounding box, height, etc. The method consisted of a simulation engine and a person detector whose combined output was fed into a fall detector. The detection method relied on metrics to compare velocity profiles created using the selected features. He talked about customizing methods for each person to see if it improves fall detection accuracy as a future direction of research.

Dr Ardhendu Behera  
Edge Hill University  
email: beheraa@edgehill.ac.uk

## Meeting Report: Computer Vision and Modelling in Cancer

BMVA 1-day meeting, 11 October 2017

Chairs: Constantino Carlos Reyes-Aldasoro and Greg Slabaugh (City, University of London)

The one-day meeting was dedicated to presenting technological advances in Cancer Research from two areas – Computer Vision and Mathematical Modelling – to bring together researchers to discuss theoretical and practical viewpoints to spur further advances in the field.

The event kicked off with a brief introduction from the chairs, Drs Constantino Carlos Reyes-Aldasoro and Greg Slabaugh – Senior Lecturers at City, University of London. Carlos shared some insights about the rise of cancer research compared to other topics and the impact of Computational and Mathematical approaches to Cancer, which motivated this meeting.

The first speaker, Jola Mirecka from the University of Oxford, talked about tumour subregional analysis, and the impact of local variations and local similarities. Following on from this presentation, the first Keynote was given by Dr Yinyin Yuan – Institute of Cancer Research – where histology deep learning is being used to decipher the Tumour Ecosystem. Through the research she presented, tumours have been found to be heterogeneous with different types of cancer cells, vessels, etc. Approaching tumours in

this way should have an impact on how we understand the lifespan of the tumour.

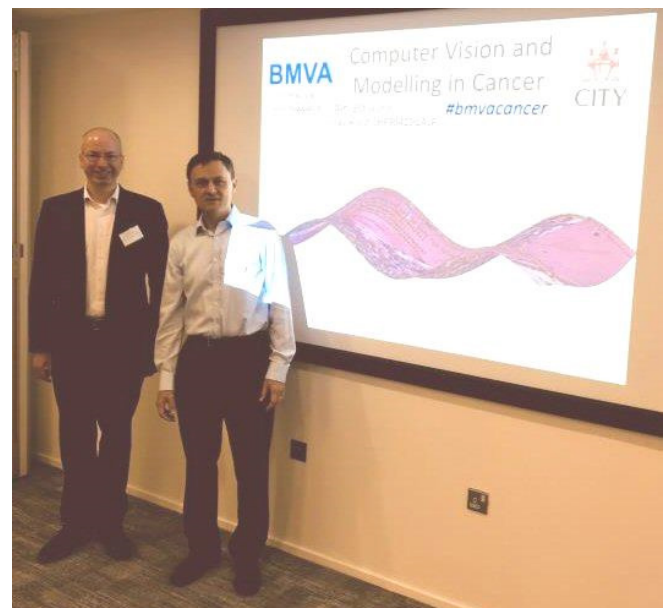
Antonia Creswell presented her doctoral work at Imperial College London in classifying skin lesions through a combination of labelled (Conditional Denoising Adversarial Autoencoder) and unlabelled data.

The second keynote speaker was Dr Ben Glocker from Imperial College London, who presented an interesting talk on brain tumour segmentation with deep neural networks. Ben discussed the agony of choice the field has encountered thanks to the widespread use of deep learning technologies and tried to propose a resolution to it through combinations of approaches. Finally, some challenges for the future of the field presented include: (i) learning the right features, (ii) detecting when the model goes wrong, and (iii) exceeding human-level performance.

A collection of 19 posters (listed below) were presented in conjunction with lunch. Posters spanned a variety of data analysis methodologies and clinical applications in cancer.

The second half of the meeting kicked off with a keynote from Professor Nasir Rajpoot, reviewing the work being done in the Computational Pathology Research Group at Warwick University. The talk reviewed computer vision algorithms used in pathology image analysis, spanning the areas of pre-processing, analysis, synthesis, biomarker detection and AI analytics.

Regarding the mathematical modelling aspect of the workshop, Tim Ingham-Dempster from the University of Sheffield introduced an agent-based model of the human colon at a cellular level, and the latest work he has been doing to extend it to include the crypts in the colon and hence to model the spread of colon cancer. Continuing with the topic of colorectal cancer, Zhaoyang Xu from Queen Mary University of London provided a talk on multi-scale semantic segmentation of colorectal cancer liver metastasis (CLRM). The main objective is to aid diagnosis and help with the treatment plan.



The two meeting chairs, (left) Greg and (right) Carlos

The final keynote was given by Professor Helen Byrne from the University of Oxford. Her talk focused on study cases where a sound mathematical model was useful in interpreting cancer-related applications. Professor Byrne

explained how the algorithms they developed could benefit from validation and interpretation of real-use cases. Two cases she presented involved aids in the design of an experimental protocol for estimating interaction terms of equations, as well as a usual case in which synthetic models helped to reduce complexity, producing the same qualitative behaviour.

As a follow-on to the meeting, the organisers have arranged for a Special Issue of IET Computer Vision dedicated to Computer Vision in Cancer Data Analysis. Submission deadline: 31 Dec 2017.

Nathan Olliverre and José Alonso Solís-Lemus  
City, University of London



The oral presentations and poster session covered a wide variety of topics in cancer image analysis and modelling and provided excellent opportunities for scientific exchange and detailed discussions.

## Cognitively Inspired Explainable Perception-Based AI

One-day BMVA Symposium in London: Wednesday 7 February 2018

Chairs: Serge Thill, University of Plymouth, Maria Riveiro, University of Skövde



### Introduction

AI systems are increasingly present in everyday society, from simple computer systems to agents such as autonomous vehicles or social robots. In this context, several researchers have noted that it is critical to understand how human users perceive such systems: in particular, the degree to which they understand how the system works, and what mental models they build of the underlying algorithms. ‘Explainable AI’ thus refers to AI systems that behave or provide the necessary information so that their working becomes comprehensible to the human user.

### Programme

- 09.00 Registration
- 09.30 Organisers’ introduction
- 09.40 Keynote 1: Alessandra Sciutti
- 10.20 Elin A Topp, University of Lund  
Generating meaningful clarification requests
- 10.40 Swen Gaudl, Metamakers Institute, Falmouth  
A genetic programming approach to derive new reasoning agents that can be interpreted by human users
- 11.00 Coffee Break
- 11.30 Szonya Durant, Royal Holloway, Univ. of London  
Integrated Cognitive User Assistance for Multiple-Display Systems
- 11.50 Tarek Besold, City, University of London  
Four notions of explainable AI
- 12.10 Henrik Svensson, University of Skövde  
Situated and distributed cognition for explainable AI
- 12.30 Lunch break
- 13.30 Keynote 2: Brad Hayes
- 14.10 Tove Helldin, University of Skövde  
Metrics for explainable AI: A brief introduction
- 14.30 Sylvester Kacmarek, Imperial College, AI  
Transparency
- 14.50 Chris Russell, University of Surrey/Alan Turing Institute  
Automated Decisions and the GDPR
- 15.10 Coffee break
- 15.50 Keynote 3: Yiannis Demiris
- 16.30 Discussion and conclusions

## Registration

Book online at [www.bmva.org/meetings](http://www.bmva.org/meetings)  
 £16 for BMVA Members, £36 for Non Members (prices include lunch).

Andrew Gilbert  
 University of Surrey  
 email: [a.gilbert@surrey.ac.uk](mailto:a.gilbert@surrey.ac.uk)

## Upcoming Meetings

We have four exciting upcoming meetings in 2018, as highlighted below: they include the BMVA showcase which will have presentations from key conference works; the bursary students; the BMVA Distinguished Fellow; and the Sullivan thesis prize winner.

Please go to [bmva.org/meetings](http://bmva.org/meetings) to register or submit work to these meetings.

7 Feb	Autonomous Perception, Serge Thill and Maria Riveiro
6 Jun	Computer Vision for smart environments and assisted living, Francesca Odone
7 Jul	Robots SLAM, Oscar Mendez-Maldonado and Mihai Bujanca
Sept/Oct	Meeting needed – could you chair a meeting?
28 Nov	BMVA Showcase 2018, Andrew Gilbert.

We are now looking for volunteers for meetings in late 2018: please contact me if you are interested in finding out more.

Andrew Gilbert  
 University of Surrey  
 email: [a.gilbert@surrey.ac.uk](mailto:a.gilbert@surrey.ac.uk)

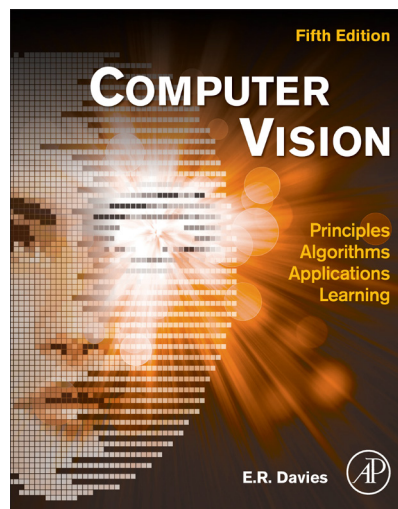
## Travel Bursaries for International Conference Attendance

The new arrangements for BMVA Travel Bursaries will carry forward into 2018. Note that there will be a fixed number of deadlines, as indicated on the bursary link to the BMVA website:

<http://www.bmva.org/bursaries>

Professor Lourdes Agapito  
 University College London  
 email: [l.agapito@cs.ucl.ac.uk](mailto:l.agapito@cs.ucl.ac.uk)

## 5<sup>th</sup> Edition of Book on Computer Vision



### Description

The 5<sup>th</sup> edition of this well-known book on Computer Vision was published in November 2017 under the revised title *Computer Vision: Principles, Algorithms, Applications, Learning*. At first sight, the 5<sup>th</sup> edition covers similar ground to previous editions: however, it also brings the text sharply up to date by including not only significantly more material on machine learning but also the very latest developments in deep learning: these are then exemplified in a new chapter on face detection and recognition.

### Key Features

- Three new chapters on machine learning emphasise the way the subject has been developing; two chapters cover basic classification concepts and probabilistic models, and the third covers the principles of deep learning networks and shows their impact on computer vision: a new chapter on face detection and recognition reflects and exemplifies the new emphasis.
- A new chapter on object segmentation and shape models reflects the methodology of machine learning and gives further practical demonstrations of its application.
- Amid much updating and reorganisation of the text, in-depth discussions have been included on geometric transformations, the EM algorithm, boosting, semantic segmentation, face frontalisation, RNNs and other key topics.
- Examples and applications – including the location of faces, eyes, road lanes, surveillance, vehicles, pedestrians, food products and foreign bodies – give the ‘ins and outs’ of developing real-world vision systems, showing the realities of practical implementation.
- Necessary mathematics and essential theory are made approachable by careful explanations and well-illustrated examples.
- Many new colour illustrations enhance the text and make it easier to understand the basic concepts.

For the full contents and further details, see the publisher’s website:

<https://www.elsevier.com/books/computer-vision/davies/978-0-12-809284-2>



BMVC 2018 website: <http://www.bmvc2018.org>

BMVC 2018 will be held in Newcastle upon Tyne, UK at Northumbria University. BMVC is the BMVA annual conference on machine vision, image processing, and pattern recognition. It is one of the major international conferences on computer vision and related areas held in the UK. Because of its increasing popularity and quality, it has established itself as a prestigious event on the vision calendar.

There have been 28 successful BMVCs since 1985. They were organized in many major universities in the UK, such as Cambridge (1987), Oxford (1990), Birmingham (1995), UEA (2003), Edinburgh (2006), Leeds (2008) and Imperial College London (2017). The scale of the conference has increased significantly in the past decade. In particular, the 2017 conference attracted 450 attendants and 635 full paper submissions, which was a record high. 188 high-quality research papers were accepted, in which around one-third of the workers came from Europe, one-third from Asia and one-third from North America. Apart from international researchers, the conference also attracted a wide spectrum of international companies, which sponsored the conference and used it to demonstrate their products.

#### Dates

- Conference date: 3–6 Sept 2018
  - Tutorial: 3 Sept 2018
  - Main conference: 4–6 Sept 2018
  - Workshop: 6 Sept 2018 (afternoon)
- Submission deadline: 30 April 2018
- Acceptance notification: 2 July 2018

#### Call for Papers

Authors are invited to submit full-length high-quality papers on image processing and machine vision. Papers covering theory and/or application areas of computer vision are invited for submission. Submitted papers will be refereed on their originality, presentation, empirical results, and quality of evaluation.

All papers will be reviewed doubly blind, normally by three members of our international programme committee. Please note that BMVC is a single track meeting with oral and poster presentations and will include two keynote presentations and one tutorial.

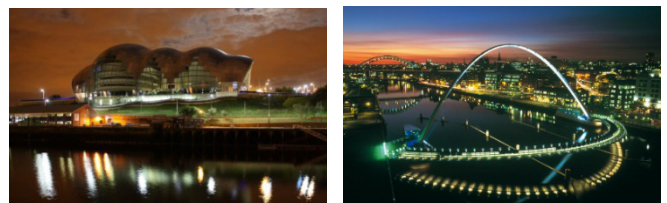
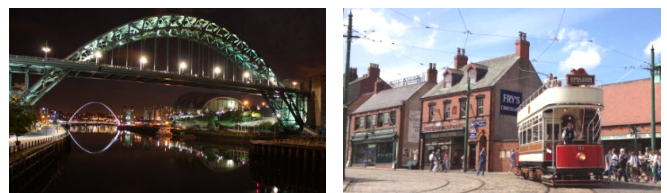
Topics include, but are not limited to:

- Statistics and machine learning for vision
- Stereo, calibration, geometric modelling and processing
- Face and gesture recognition
- Early and biologically inspired vision
- Motion, flow and tracking
- Segmentation and grouping
- Model-based vision
- Image processing techniques and methods
- Texture, shape and colour
- Video analysis
- Document processing and recognition
- Vision for quality assurance, medical diagnosis, etc.
- Vision for visualization, interaction, and graphics
- Object detection and recognition
- Shape-from-X
- Video analysis for action and event recognition
- Illumination and reflectance
- Deep learning for vision
- 3D computer vision
- RGBD analysis.

#### About Newcastle upon Tyne

Newcastle upon Tyne – or ‘Newcastle’ as it is most commonly referred to – is one of the most iconic cities situated in the North East of England, famous for its industrial heritage, eponymous brown ale, popular nightlife as well as its distinct regional ‘Geordie’ dialect. The city history dates back almost 2,000 years (since 120 AD) when the Romans built the first bridge across the River Tyne. The city was originally known by its Roman name, Pons Aelius, during the period it was under the control of Romans, the Saxons and the Danes (amongst others); however, the name Newcastle has been used since the Norman conquest of England.

Local attractions include Tyne Bridge, Beamish Museum, SAGE Gateshead, Millennium Bridge, and Newcastle United Football Club (NUFC) – as shown in the photographs below:



Dr Hubert PH Shum  
Northumbria University  
email: [hubert.shum@northumbria.ac.uk](mailto:hubert.shum@northumbria.ac.uk)

## MIUA 2018 – Call for Papers

9–11 July 2018  
Southampton, UK

MIUA is a forum for communicating research progress within the community interested in biomedical image analysis. Its goals are the dissemination and discussion of research in medical image processing and analysis, and it aims to encourage the growth and raise the profile of this multi-disciplinary field.

### Scope

- Big data processing
- Clinical and Scientific Evaluation of Imaging Studies
- Computer-Aided Pathology
- Computer-Aided Radiology
- Computer-Assisted Surgery
- Data Compression and Anonymization
- Data Fusion
- Decision Support
- Deep Learning in Medical Imaging
- Discovery of imaging biomarkers
- Human Computer Interaction
- Image Enhancement
- Image Interpretation
- Image-Guided Intervention
- Image Formation and Reconstruction
- Image Perception
- Image Registration
- Image Segmentation
- Intelligent Imaging Systems
- Machine Learning in Imaging
- Modelling and Simulation
- Motion Analysis
- Multi-Modality Image Analysis
- Pattern and feature recognition
- Protocol Development and Standardization
- Texture Analysis
- Time Series Analyses
- Virtual Reality
- Visualisation.

### Technical Paper Submission

Papers should describe original and unpublished work on the topics of the conference. Prospective authors should prepare a full paper, written in English, between 8 and 12 pages and must submit it electronically.

Each paper will be blind-reviewed by at least two reviewers: acceptance will be based on originality, significance and clarity. For accepted papers to be published, at least one author has to register and present the work at the conference.

### Clinical Abstract Submission

Submitted abstracts should fit on one A4 page. Calibri font with minimum font size of 11 is preferred. The page margins should be no less than 1 cm on left and right hand sides and no less than 2 cm on top and bottom. Abstracts

could be uploaded via the Conference website. However, if the abstract contains figure(s) and/or table(s), it needs to be sent via e-mail to [miua2018@soton.ac.uk](mailto:miua2018@soton.ac.uk) in MS Word 2013 format.

The abstract page should contain: Title, Authors and Affiliations, Introduction (i.e., Background and Purpose), Methods, Results, Discussion (or Conclusion), References and Acknowledgements if any. Tables and Figures are optional, and can be included as long as they fit within the page.

### Publication

Accepted full papers will appear in the conference proceedings and, in addition, will be published by Springer in the Communications in Computer and Information Sciences series, Springer CCIS – indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago and Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

Accepted clinical abstracts will only be included in the conference proceedings and published online. In the conference, they will be presented as traditional posters in portrait A0 format. A selection of abstracts will also be invited for a 2-minute oral presentation on a dedicated session at the event.

### Special Session

Proposals for Special Sessions are welcome in order to complement the regular program with new or emerging topics of particular interest to the Medical Imaging community. More information can be found on the conference website. Also, please find the call for papers and further information on the conference website:

<https://miua2018.soton.ac.uk/index.html>

### Key dates

Paper submission opens:	1 December 2017
Paper submission deadline:	12 February 2018, 17:00 GMT
Notification of Acceptance:	19 March 2018
Camera-ready papers due:	2 April 2018
Early-bird registration due:	~ May 2018

email: [miua2018@soton.ac.uk](mailto:miua2018@soton.ac.uk)

## Report on ICCV 2017

The International Conference on Computer Vision (ICCV) was held on 22–29 October on beautiful Lido island of Venice, Italy. The biennial premier computer vision event was hosted at the historic Palazzo del Cinema (Venice Convention Center), recognized worldwide as the setting for the oldest film festival in the world – the Venice International Film Festival.



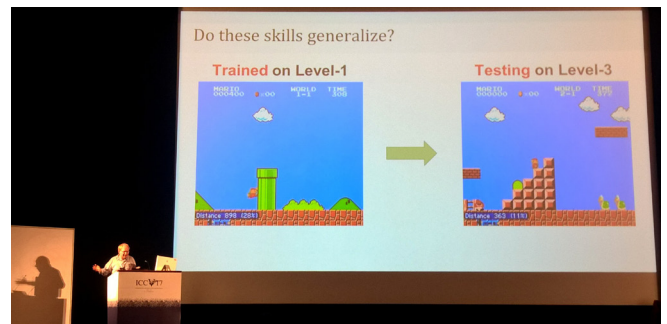
Outside the conference venue

Similar to CVPR and ECCV, ICCV is considered one of the top-quality conferences in computer vision. Co-sponsored by the IEEE and the Computer Vision Foundation (CVF), ICCV this year comprised the main conference, 44 co-located workshops, 60 exhibitors and 9 tutorials. It had 3107 attendees, 2143 submissions and 621 accepted papers, all highest in the long-lasting 30-year history of ICCV. This was a week-long conference organised and completed with a high degree of excellence: it offered oral and spotlight presentations as well as poster sessions. Here, I would like to provide an overview of some of the important studies.



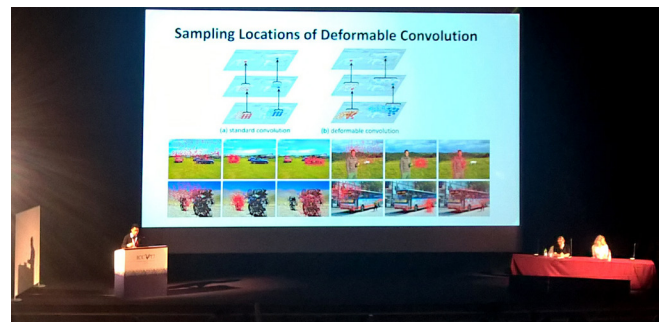
Michael Black on the history of human body modelling

Michael J Black from the Max Planck Institute for Intelligent Systems, where he leads the Perceiving Systems department, presented the latest developments in the field of human body modelling as an invited speaker at *PoseTrack Challenge: Human Pose Estimation and Tracking in the Wild*. A well-known figure in the research community as well as a co-founder of Body Labs Inc., which was recently acquired by Amazon, he outlined his vision for the open problems in full-body avatars, accurate 3D human pose and shape from single images and body shape estimation from clothed scans.



Alexei Efros on “The Revolution will not be supervised”

Alexei A Efros from the University of California, Berkeley gave one of his inspiring talks on reinforcement learning, aptly entitled “The Revolution will not be supervised”, at the *Beyond Supervised Learning* Workshop. He also remarked on the recent study of his research group at the Berkeley Artificial Intelligence Research Lab (BAIR) that they can turn horses into zebras with the power of generative adversarial networks.



Jifeng Dai on “Deformable Convolutional Networks”

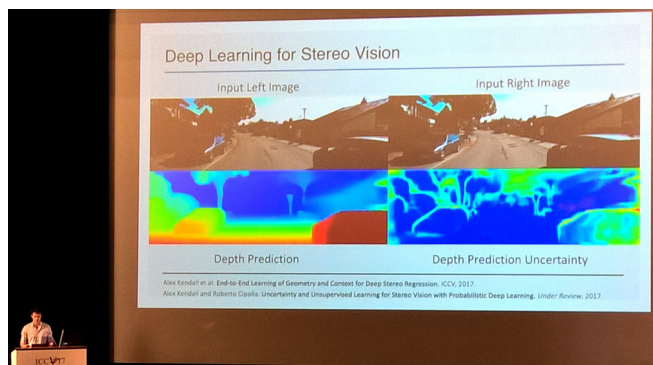
Jifeng Dai from Microsoft Research Asia presented an approach to model geometric transformations within Convolutional Neural Networks (CNNs). The study proposes new modules to enhance the limited capability of traditional neural networks. These new modules, called deformable convolution and deformable RoI pooling, learn 2D offsets from the target tasks without additional supervision.



Kaiming He on “Mask R-CNN”

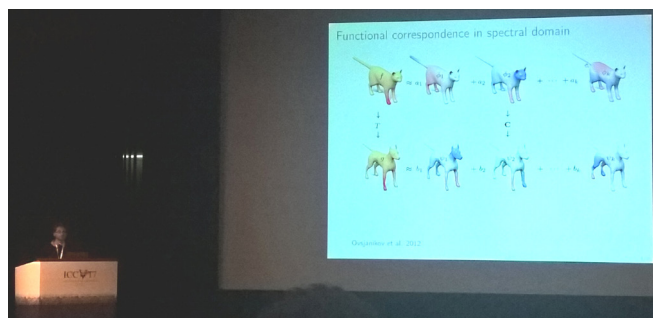
This year’s Best Paper Award (Marr prize) went to “Mask R-CNN” by Kaiming He from Facebook AI Research (FAIR), along with Georgia Gkioxari, Piotr Dollár and Ross Girshick. The study offers a general framework for instance segmentation by predicting object masks in instance-level as an extension to the famous paper, “Faster R-CNN”. Tsung-Yi Lin et al. from Facebook AI Research (FAIR), won the Best Student Paper Award with their study,

“Focal Loss for Dense Object Detection”. Tomaso Poggio – one of the founders of computational neuroscience – received the Azriel Rosenfeld Lifetime Achievement Award while the Distinguished Researcher Award was given to Luc van Gool and Richard Szeliski whose research has significantly contributed to the field of computer vision.



Alex Kendall on geometry meets deep learning.

Alex Kendall from the University of Cambridge focused on the understanding of classical computer vision geometry to design better deep convolutional neural network architectures at the workshop on *Geometry Meets Deep Learning*. In the invited talk, entitled “Has end-to-end deep learning killed geometry or can we do better?”, he presented an overview of his work in semantic segmentation, stereo vision and multi-task learning in the domain of scene understanding.



Alex Bronstein on the correspondence problem

Alex M Bronstein from Technion (Israel Institute of Technology) discussed several geometric notions to solve the correspondence problem efficiently. He showed very appealing examples on how to compute the correspondence between 3D objects combined with deep learning techniques.



A packed poster session

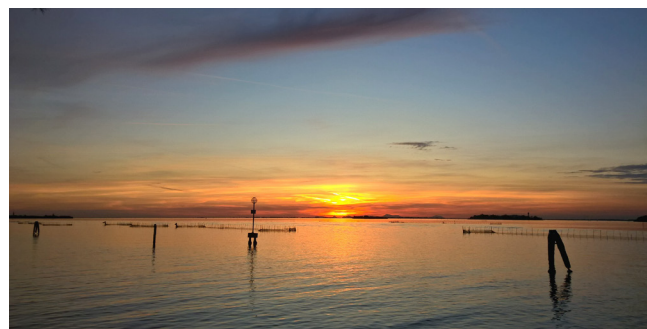
As an upward trend, deep learning studies were dominant throughout the conference, similar to other machine learning and computer vision conferences. In fact,

*learning, network, image* and *deep* are the most used words in the title of all submitted papers at ICCV, respectively. Moreover, according to the statistics from the program committee, *video and language* and *vision for autonomous driving* have the highest acceptance rate and were seen as hot topics at ICCV 2017. It would not be wrong to say that the hype around deep learning will not be settling anytime soon.



Anil Bas explains his work to delegates.

Overall, ICCV was the perfect hub for bringing researchers from academia and industry together to share and collaborate on their ideas. More information about the conference can be found at <http://iccv2017.thecvf.com>. The next ICCV will be held in Seoul, South Korea, in 2019 and after that, it will take place in Montreal, Canada, in 2021.



The sun sets on Venice.

Finally, I would like to thank the co-authors of our collaborative study between the University of York and the University of Surrey – Patrik Huber, William AP Smith, Muhammad Awais and Josef Kittler. I would also like to thank the BMVA for their generous travel bursary which made it possible for me to attend this great conference.

Anil Bas  
 University of York  
 email: [ab1792@york.ac.uk](mailto:ab1792@york.ac.uk)