University of BRISTOL

13th Vision Researchers Colloquium

Monday 4 July 2022

Abacws Building, University of Cardiff

Keynote speaker: Darren Cosker

Principal Scientist at Microsoft's Mixed Reality and AI laboratory

In partnership with:

GW4

UNIVERSITY OF EXETER

BRISTOL VISION INSTITUTE

Hosted by:

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

UNIVERSITY OF BATH

Programme

09:30	**Registration, tea and coffee**
10:00	**Welcome by Dr Oktay Karakus, Cardiff University**
	Chaired by Professor Natalie Hempel de Ibarra, University of Exeter
10:05	**How, M; Vincent Venables, S.; Drerup, C; Powell, S.; Marshall, J.; Herbert-Read, J**– University of Bristol: **A new function for polarization vision: mitigation of visual noise from dynamic illumination**
10:25	**Gonzales, D; Hempel de Ibarra, N; Anderson, K.** – University of Exeter: **Remote Sensing of Floral Resources for Pollinators – New Horizons From Satellites to Drones**
10:45	**Santon, M; Chivers, S; Hancock, G; How, M, J** – University of Bristol: **Jamming visual systems with stripes: the 'mesmerising' hunting display of the broadclub cuttlefish**
11:05	**Kjernsmo,, K; Lim, A; Middleton, R; Hall, J; Costello, L; Whitney, H; Scott-Samuel, N; Cuthill, I** – University of Bristol: **Iridescence as deceptive warning coloration**
11:25	**Break – tea and coffee**
	Chaired by Professor Innes Cuthill, University of Bristol
11:40	**Skoczek, K P; Acton, J H; Greenwood, J A; Redmond, T** - Cardiff University: **Target-flanker similarity modulates the strength, not spatial extent, of visual crowding**
12:00	**Martin, N. D; Rushton, S. K; Brossard, M; Redmond, T** – Cardiff University: **Integration of visual signals across space and time**
12:20	**Hammond, H; Armstrong, M; Thomas, G; Gilchrist, I** – University of Bristol: **Continuous measures of audience immersion: validating behavioural and physiological measures against self-report**
12:40	**Levordashka, A; Stanton Fraser, D; Gilchrist, I** – University of Bristol: **Watching the Audience: Motion Tracking Analysis of Audience Behaviour**
13:00	**Lunch and poster presentations**
	Chaired by Professor David Bull, Director, BVI, University of Bristol
14:10	**Keynote:** Darren Cosker, Principal Scientist at Microsoft's Mixed Reality and AI laboratory - **Creating Presence in Mixed Reality and the Metaverse**
	Chaired by Professor Simon Rushton, Cardiff University
15:10	**Deane, J; Kearney, S: Kim, K. I; Cosker, D** – University of Bath: **Synthetic Labels for Animals in Computer Vision**
15:30	**Fletcher, G; Cosker, D** – University of Bath: **Going beyond homeomorphism in neural motion retargeting**
15:50	**Finn. A.; Punit, S; de la Rosa, S; Teufel, C; von dem Hagen, E** - Cardiff University: **Individuals with ASD show an increased influence of body posture on facial expression perception**
16:10	**Arthur, T; Buckingham, G; Brosnan, M; Harris, D; Vine, S** – University of Exeter: **Visuomotor Differences in Autism: An evaluation of potential mechanisms**
16:30	**Closing the colloquium, Professor David Bull**
16:40	**Informal drinks reception**

Keynote

Creating Presence in Mixed Reality and the Metaverse

Darren Cosker
Principal Scientist Microsoft Mixed Reality and AI laboratory, Cambridge
Professor of Computer Science, University of Bath

Abstract
Imagine being able to have a conversation with someone who is hundreds of miles away, but it feels like they are actually there with you in the same room. Technologies which could achieve this - delivering true Presence experiences - would change the world. They would bringing distant family and friends closer together, transform the way we work and reduce carbon footprints. However, creating compelling interactive experiences involving other people in mixed reality and the metaverse is a challenging task combining expertise in computer vision, graphics, AI and engineering. In this talk I will examine some of the technologies required to make this a reality, the progress we have made, and the challenges ahead.

Biography
Darren Cosker is a Principal Scientist at Microsoft's Mixed Reality and AI laboratory, Cambridge. At Microsoft he is helping design and create Microsoft's vision for the Metaverse through products such as Microsoft Mesh and HoloLens. Darren is also a Prof. at the University of Bath, where he was previously the Director and founder of CAMERA (www.camera.ac.uk). The aim of Darren's research is to create presence in mixed reality. This is the feeling that you can be with someone in the same room - working or socialising together - even though they may be hundreds of miles away.

https://www.microsoft.com/en-us/research/people/coskerdarren/

Please connect on social media by following @VisionInstitute.

Presentations

Session 1 – Professor Natalie Hempel de Ibarra

Presentation 1

How, Martin - University of Bristol

A new function for polarization vision: mitigation of visual noise from dynamic illumination

In shallow water, downwelling light is refracted from surface waves onto the substrate creating bands of light that fluctuate in both time and space, known as caustics. This dynamic illumination can be a visual hindrance for animals in shallow underwater environments. Polarization vision may have evolved underwater to enable discrimination of objects, while ignoring the variations in illumination caused by caustics. To explore this possibility, crabs (Carcinus maenas) and cuttlefish (Sepia officinalis), both of which possess polarization vision, were presented with moving stimuli overlaid with caustics. Dynamic caustics inhibited the detection of an intensity-based stimulus, but not when these stimuli were polarized. This study is the first to demonstrate that polarization vision reduces the negative impacts that dynamic illumination can have on visual perception.

Vincent Venables, S.; Drerup, C; Powell, S.; Marshall, J.; Herbert-Read, J.; How, M.

Presentation 2

Gonzales, Dunia- University of Exeter

Remote Sensing of Floral Resources for Pollinators – New Horizons From Satellites to Drones

Insect pollinators are affected by the spatio-temporal distribution of floral resources, which are dynamic across time and space, and also influenced heavily by anthropogenic activities. There is a need for imagery describing the time-varying spatial distribution of flowers, which can be used within behavioural and ecological studies. Traditional field techniques for mapping flowers are often laborious and limited to relatively small areas, making it difficult to assess how floral resources are perceived by pollinators to guide their behaviours.

Recently, lightweight drone technology has been adopted by the ecological community, offering a capability for flexible deployment in the field, and delivery of centimetric resolution imagery, providing a clear opportunity for capturing fine-grained information on floral resources at key times of the flowering season. In this review, we answer three key questions of relevance to pollination science – can remote sensing imagery deliver information on (a) how isolated are floral resources? (b) What resources are available within a flower patch? And (c) how do floral patches change over time? We explain how such information has potential to deepen ecological understanding of the distribution of floral resources that feed pollinators and the parameters that determine their navigational and foraging choices based on the sensory information they extract at different spatial scales. We provide examples of how such data can be used to generate new insights into pollinator behaviours in distinct landscape types and their resilience to environmental change.

Gonzales, D., Hempel de Ibarra, N., Anderson, K.

Presentation 3

Santon, Matteo- University of Bristol

Jamming visual systems with stripes: the 'mesmerising' hunting display of the broadclub cuttlefish

High contrast stripes are renowned for generating motion illusions in humans by altering the estimation of trajectory and speed of targets (Stevens and Merilaita, 2009, Merilaita et al., 2017, How and Zanker, 2014). Striped patterns are also widespread in nature, from insects to mammals. Yet, there is still little experimental evidence about whether, and how, animals evolved such patterns to dazzle, jam, or disrupt target visual systems. We research the use of motion dazzle in nature by investigating the effects of high contrast stripes in a real biological predator-prey interaction, the broadclub cuttlefish hunting crabs. This cuttlefish has a unique hunting technique. When approaching prey, it displays highly contrasting stripes moving in a downward direction across the head and arms

(How et al., 2017). By tethering shore crabs over a Styrofoam treadmill in front of an LCD monitor (Smithers et al., 2019), we tested if an expanding stimulus with moving stripes - representing the approaching cuttlefish - elicits a weaker escape response in prey crabs than a control stimulus with stationary stripes. Our preliminary results not only show that a greater contrast is required to elicit an escape response in crabs when an approaching stimulus features moving stripes compared to stationary stripes, but also that this effect depends on stripe speed. Implementation of motion detection algorithms further suggests that crab responses may be altered by an overlaid downward motion cue very different from that expected by an approaching predator, thus masking the looming cuttlefish.

Santon M., Chivers S., Hancock G. & How M.J

Presentation 4
Kjernsmo, Karin- University of Bristol
Iridescence as deceptive warning coloration

Protective coloration including various forms of camouflage, aposematism, and mimicry provides several ways for prey to escape predation. However, the functions and adaptive value of one striking form of structural colouration, iridescence, is not yet fully understood. It was recently discovered that iridescence can act as a highly effective form of camouflage, particularly at longer viewing distances and against a glossy green background. However, can iridescence also confer a survival benefit to prey post detection and, if so, which optical properties of iridescent prey are important for this putative anti-predator function?

Here, using real and artificial jewel beetle (Sternocera aequisignata) wing cases, we tested the effects of both iridescence and gloss on the attack behaviour of prey-naïve avian predators. Intriguingly, we found that iridescence significantly reduced the birds' willingness to attack the prey, even when presented up close against a mismatching background. Importantly, we also found that the main factor explaining this avoidance behaviour displayed by the birds was iridescence, not having multiple colours per se, with surface gloss also having an independent effect. Our results suggest that reduced avian predation could facilitate the initial evolution of iridescence in many species of prey, and that it is the very hallmark of iridescence, its colour changeability, that is important for this effect.

Kjernsmo, K, Lim A, Middleton R, Hall J, Costello, L, Whitney, H, Scott-Samuel, N & Cuthill, I.

Session 2 – Professor Innes Cuthill

Presentation 5
Skoczek, Kristian- Cardiff University
Target-flanker similarity modulates the strength, not spatial extent, of visual crowding

Visual crowding occurs when nearby distractors (flankers) influence target identification in normal human vision. Errors in identification follow the appearance of flankers that occur within a 'crowding zone' around the object of interest. Crowding has been demonstrated with stimuli differing in orientation, colour, and motion direction. Perceived similarity between flankers and the target have been reported to influence visual crowding, but the mechanism of this influence is unclear. We systematically explored the combined effects of eccentricity and target-flanker similarity (via relative orientation difference) in a psychophysical task with a continuous feature report method. Participants (n=3) viewed a crowded Landolt-C target at each of three eccentricities (4.2°, 12.7°, 21.2° scaled for cortical magnification) and reported its perceived orientation on a continuous scale. Crowding was induced with a pair of identical Landolt-C flankers presented with 3 orientation differences relative to the target and at 5 levels of centre-to-centre target-flanker separation. Relative proportions of crowded and uncrowded responses were calculated by modelling errors as a mixed pair of vonMises distributions. Modelling the shift in relative proportions of responses across target-flanker spatial separation produced logistic functions with a shared midpoint at each eccentricity. These midpoints were unaffected by stimulus orientation difference, but the slopes of the functions were steeper with larger target-flanker orientation differences. We propose these findings together indicate target-flanker similarity influences the strength of crowding effects within an invariant crowding zone.

Variable crowding strength is a potential confounding influence on zone extent estimates when inconsistent definitions of extent are used between studies.

Skoczek, K P; Acton, J H; Greenwood, J A; Redmond, T

Presentation 6

Martin, Nick- Cardiff University

Integration of visual signals across space and time

Integration of signals across space and time by the visual system is well studied behaviourally, and, within specific limits, is described by Ricco's area (Ricco, 1877; integration over space) and Bloch's law (Bloch, 1885; integration over time). However, spatial and temporal integration cannot be assumed to be independent: e.g., Ricco's area is smaller when probes presented for 8ms than for 930ms (Barlow, 1958; Owen, 1972). We investigate integration across a range of spatial and temporal separation values by varying probe luminance in a 4AFC detection task. We used 2-probe stimuli, with which the probe size and duration can be held constant while varying the spatial or temporal separation between probes. This allowed us to explore the time-course of integration. Contrary to predictions that integration would always be maximal when probes are presented simultaneously, across all participants we find that as spatial as separation increases, the probes are integrated more efficiently (e.g., detected at lower luminance thresholds) if they are also separated temporally compared to when they are presented simultaneously. This finding raises intriguing questions about the underlying neural processes ect.

Martin, N. D., Rushton, S. K., Brossard, M., Redmond, T.

Riccò, A. (1877). Relazione fra il minimo angolo visuale e l'intensità luminosa. Memorie Della Societa Degli Spettroscopisti Italiani, 6, B29-- B58.

Bloch, A. M. (1885). Experiences sur la vision. Comptes Rendus de La Societé de Biologie, 37, 493.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. The Journal of Physiology, 141(2), 337–350.

Owen, G. W. (1972). Spatio-temporal integration in the human perhipheral retina. Vision Research, 12, 1011–1026

Presentation 7

Hammond, Hugo -University of Bristol

Continuous measures of audience immersion: validating behavioural and physiological measures against self-report

When an audience member becomes immersed, attention is shifted towards the media and story, and this is accompanied by the allocation of mental resources to represent events and characters. Here we investigate whether it is possible to measure immersion using continuous behavioural or physiological measures. Using television and film clips, we validated dual-task reaction times, heart rate, and skin conductance against self-reported narrative engagement. We find that reaction times in the dual-task were strongly positively correlated with immersion: slower reaction times were indicative of greater immersion, particularly emotional engagement. Synchrony in heart rate between participants was associated with attentional and emotional engagement with the story, although we found no such relationship with skin conductance. These results establish both dual-task reaction times and heart rate as candidate measures for the real-time, continuous, assessment of audience immersion.

Hammond, H; Armstrong, M; Thomas, G; Gilchrist, I

Presentation 8

Levordashka, Ana -University of Bath

Watching the Audience: Motion Tracking Analysis of Audience Behaviour

Immersion is a state of deep mental involvement, accompanied by reduced awareness of the physical world (Agrawal et al., 2020); it is a powerful experience, linked to artistic appreciation, memory, task performance, and behaviour change. We examine immersion in home audiences, with the aim of identifying behavioural markers,

specifically testing ones suggested by prior research: stillness (Seli et al., 2014; Theodorou et al., 2019) and movement synchrony (Farmer et al., 2020). We used a mixed-method approach, including face tracking and experience sampling, to measure momentary and retrospective immersion in a diverse sample of individuals (N = 113) who watched a 30-minute segment of a theatre play in their homes. Our results provide evidence that psychological immersion is associated with head stillness. These findings shed light on the processes underlying immersion and ways to measure it effectively.

Levordashka, A; Stanton Fraser, D; Gilchrist, I

Agrawal, S. et al. (2020) 'Defining immersion: Literature review and implications for research on audiovisual experiences', AES: Journal of the Audio Engineering Society, pp. 404–417. doi: 10.17743/jaes.2020.0039.

Farmer, H. et al. (2020) 'Did you see what I saw?: Comparing attentional synchrony during 360° video viewing in head mounted display and tablets.', Journal of Experimental Psychology: Applied. doi: 10.1037/xap0000332.

Seli, P. et al. (2014) 'Restless mind, restless body', Journal of Experimental Psychology: Learning Memory and Cognition, 40(3), pp. 660–668. doi: 10.1037/a0035260.

Theodorou, L., Healey, P. G. T. and Smeraldi, F. (2019) 'Engaging with contemporary dance: What can body movements tell us about audience responses?', Frontiers in Psychology, 10(FEB), pp. 1–13. doi: 10.3389/fpsyg.2019.00071.

Session 3 – Professor Simon Rushton

Presentation 9
Deane, Jake -University of Bath
Synthetic Labels for Animals in Computer Vision

Humans are one of the most common computer vision subjects with large scale datasets available for complex problems such as pose estimation or segmentation (Lin et al. 2014; Johnson and Everingham 2010). But what about animals? There is a significant lack of equivalent animal data: Such data is time consuming and difficult to produce even for a single species of animals such as dogs, requiring time consuming manual annotation. There do exist some datasets (Biggs, Roddick, et al. 2018; Biggs, Boyne, et al. 2020; Yu et al. 2021; Del Pero et al. 2015) but these are limited in terms of the number of data points and variety of data labels. Synthetic data has been used to make up for the lack of real labelled ground truth data (Cao et al. 2019; Mu et al. 2020), but these synthetic datasets suffer from the domain gap and are still limited in terms of the number and variety of data labels. Even then the data ignores important keypoints which would be highly useful in applications such as animation. We introduce a method to create multi-task dense data labels for canine images, by fitting a novel parametric model. With this method we can generate dense 2/3D keypoints, part-segmentation maps, bounding boxes and depth maps. We present a series of experiments using this data evaluating our data for training models for common body analysis tasks such as pose estimation and part-segmentation.

Deane, J; Kearney, S: Kim, K. I; Cosker, D

Presentation 10
Fletcher, George -University of Bath
Going beyond homeomorphism in neural motion retargeting

Motion retargeting, the transferring of motion from one articulated character to another, is an important technique in the application of motion capture data to many applications in computer vision and graphics.

Traditionally retargeting is performed semi-manually, often tediously, by animators which has led to interest in automating the task. Recently, deep learning has succeeded in automatic motion retargeting, by framing it as an unpaired domain translation problem. However, such methods are limited to the same skeleton up to homeomorphism (Aberman et al., 2020).

In this work we categorise motion retargeting into three types: proportional, homeomorphic and non-homeomorphic. These are skeletons with the same number of joints but different lengths, skeletons that are homeomorphic as graphs, and those that are not respectively.

We first describe a study on the perception of retargeted motion of quadrupeds (Fletcher et al., 2021) in the proportional retargeting regime where we discovered that the average observer cannot distinguish retargeted motions from real ones in terms of “naturalness”.

Given that this is the case, we then consider the open problem of non-homeomorphic neural motion retargeting and discuss possible approaches. on.

Fletcher, G; Cosker, D.

Aberman et al. (2020), Skeleton-aware networks for deep motion retargeting. ACM Trans. Graph. 39, 4, Article 62 (August 2020), 14 pages.

Fletcher et al. (2021) Exploring the Perception of Quadruped Motion Retargeting. 14th ACM SIGGRAPH Conference on Motion, Interaction and Games (November 2021), 2 pages.

Presentation 11

Finn, Abi -Cardiff University

Individuals with ASD show an increased influence of body posture on facial expression perception

Facial expression perception is influenced by context. For instance, affective body postures bias facial expression perception toward the emotion conveyed by the body. Amongst typical adults, the magnitude of this context effect is determined by the precision with which observers represent facial expressions, with lower precision leading to a greater influence of body posture. Individuals with Autism Spectrum Disorder (ASD) show reduced facial expression discrimination abilities, which should lead to larger influences of body posture on facial expression perception. However, individuals with ASD also exhibit a local processing bias, which should limit their ability to incorporate contextual cues into global percept’s, and would thus lead to an opposite effect, i.e., a smaller influence of body posture. Here, we address these competing hypotheses. We used online psychophysical methods to characterise the precision of isolated facial expressions and the influence of body posture in autistic and neurotypical individuals. In line with previous research, we found that autistic individuals had poorer facial expression discrimination ability than neurotypical individuals. Both groups showed the expected influence of body context on facial expression perception, but this effect was larger in the ASD group. Crucially, we found that, in both groups, the magnitude of the context effect was related to the observers’ facial expression discrimination ability. These results suggest that, in autistic individuals, similar principles may govern the integration of facial expression and body posture information as in neurotypicals. n.

Finn. A; Punit, S; de la Rosa, S; Teufel, C; von dem Hagen, E

Presentation 12

Arthur, Tom - University of Exeter

Visuomotor Differences in Autism: An evaluation of potential mechanisms

Many autistic people struggle with controlling their bodily movements and processing sensory cues. These difficulties can lead to clumsiness and poor balance, while limiting personal independence and quality of life. Despite these important implications, autism-related visuomotor difficulties are relatively understudied in the field.

To pave the way for new evidence-based programmes, we investigated the aetiology of visuomotor differences in autism via a series of ‘vision in action’ methodologies. Specifically, eye-tracking analyses were incorporated within a novel virtual-reality paradigm to assess aspects of hand-eye coordination during naturalistic interception actions. Here, autistic and non-autistic participants were required to hit virtual tennis balls that bounced with varying levels of elasticity while we measured key variables relating to gaze behaviour and swing kinematics. Contrary to a number of high-profile theories of autism, we found that autistic and non-autistic individuals comparably used prior expectations to guide their visuomotor responses. Indeed, all participants employed ‘predictive’ eye movements that shifted visual attention ahead of the virtual balls (towards their anticipated future locations). However, the

manner in which participants dynamically adjusted these predictive eye movements differed between groups. Specifically, autistic individuals demonstrated atypical gaze responses to unexpected and volatile visual cues. These atypicalities were related to visuomotor interception abilities, with autistic participants exhibiting poorer performances under unstable ball bounciness conditions. Taken together, findings suggest that autistic people may have difficulties adjusting predictive action behaviours during dynamic hand-eye coordination tasks. Results support the development of evidence-based programmes that make visuomotor environments feel 'more predictable' for autistic people.

Arthur, T; Buckingham, G; Brosnan, M; Harris, D; Vine, S

Posters

1	**Alshewaier, H; Sun, X; Qin, Y** – Cardiff University: **Weakly Supervised Methods for Medical Image Segmentation**
2	**Chapman, K; Cozma, N; Hodgkinson, A; English, R; Gaston, K; Hempel de Ibarra, N** – University of Exeter: **Bumblebees exploit known sources but return with partial pollen loads when foraging under low evening light**
3	**Feng, C; Zhang, A; Bull, D**– University of Bristol: **Deep VQA based on a Novel Hybrid Training Methodology**
4	**Cipolina-Kun, Lucia; Caenazzo, S; Mazzei, Gaston** – University of Bristol: **Multi-Modal diffusion models for the preservation of cultural heritage.**
5	**Danier, D; Zhang, F; Bull, D** – University of Bristol: **A Subjective Quality Study for Video Frame Interpolation**
6	**Davies, J; Leonards, U; Stevanov, J** – University of Bristol: **What makes an environment restorative? New insights from multi-arrangement and representational similarity methods.**
7	**Dickson, G; Anisimova, E; Burn, J.F; Leonards, U** – University of Bristol: **The impact of visual discomfort of floor patterns on human gait kinematics**
8	**Dong,Z; Wu, X; Zhao, X; Zhang, F; Liu, H** – Cardiff University: **Identifying Pitfalls in the Evaluation of Saliency Models for Videos**
9	**Jones, K; Leonards, U; Metatla, O** – University of Bristol: **Joint attention experiences for children with visual impairments**
10	**Kerr, W; Li, W; Haines, T**– University of Bath: **Autonomous Filming Systems: Towards Empathetic Imitatio**n
11	**Liang, Y; Wu, J; Lai, Y-K; Qin, Y** – Cardiff University: **Exploring and Exploiting Hubness Priors for High-Quality GAN latent sampling**
12	**Lou, J; Lin, H; Marshall, D; White, R; Yang, Y; Shelmerdine, S; Liu, H** – Cardiff University: **Predicting radiologist attention during mammogram reading with deep and shallow high-resolution encoding**
13	**Ma, Y; Tanguy, J-V; Corcoran, P; Liu, H** – Cardiff University: **Evaluation of Image Quality Assessment Metrics for MR Images**

14	**Ma W; Karakus O; Rosin P** – Cardiff University: **Attention based multi modal deep learning approach for land cover mapping**
15	**Ming, Y; Yang, X; Zhang, G; Calway, A** – University of Bristol: **CGiS-Net: Aggregating Colour, Geometry and Implicit Semantic Features for Indoor Place Recognition**
16	**Samano, N; Zhou, M; Calway, A** – University of Bristol: **Geolocation by Embedding Images and Maps**
17	**Tian, X, Anantrasirichai, N, Nicholson, L, Achim, A** – University of Bristol: **Optimal Transport-based Graph Matching for 3D retinal OCT image registration**
18	**Wang, H; Luice, L; Liu, H** – Cardiff University: **CUID: A New Study of Perceived Image Quality and it's Subjective Assessment**
19	**Wu, X; Dong, Z; Zhang, F; Rosin,P, L; Liu, H** – Cardiff University: **Analysis of video quality induced spatio-temporal saliency shifts**
20	**Zhao, Z; Hicks, Y; Sun, X** – Cardiff University: **Deep learning for fruit ripeness determination**

Poster 1

Alshewaier, Hateef – Cardiff University

Weakly Supervised Methods for Medical Image Segmentation

In recent years, numerous researchers have investigated the topic of brain tumour segmentation and developed new approaches and methodologies to segment the tumour from the healthy tissues in the human brain images. This inspired us to explore this subject and develop a method to perform a segmentation task for weakly annotated medical images. Since the fully annotated data costs a long time and it needs experts in the labelling process, it is expensive to annotate images. Nonetheless, weakly supervised learning can solve these issues, which attracts many researchers' interests because it requires only a small amount of the data to be annotated. We propose a new technique to segment tumours from human brains using weakly annotated data, and develop a new method to calculate the loss function. For the proposal, we implement the suggested method using both bounding-box and scribble annotations on brain MR images. For the proposed loss function, we consider and produce weight according to two aspects: 1), the scribble pixels for the foreground will have higher weight than the background, 2), the pixel position inside the bounding-box determines the amount of weight given while calculating the loss.

Alshewaier, H; Sun, X; Qin, Y

Poster 2

Chapman, Katy – University of Exeter

Bumblebees exploit known sources but return with partial pollen loads when foraging under low evening light

Despite the changing visual conditions at sunset, some diurnal foragers such as the bumblebee Bombus terrestris continue to forage. Key decisions regarding the daily cessation of visually-guided foraging behaviour are made during this period of dimming light levels, and foraging during this time could be costly for individuals. Resource acquisition may be reduced due to daytime resource depletion or visual constraints on flight control, navigational performance, or location and exploitation of resources, which could vary amongst individuals with differently sized apposition eyes. Evening foragers may also benefit from avoiding competition or from accessing different flower species than daytime foragers. We conducted a field experiment in a mixed-agricultural landscape to investigate pollen foraging

at sunset, compared to the daytime, by recording activity at the colony entrance and collecting pollen from returning foragers. We found that foragers returned with smaller pollen loads in the evenings than in the day, but that the same flower taxa were being exploited. This suggests a cost to evening foraging as a result of resource depletion or visual limitations, however despite this, bees continue to forage, indicating an influx-maximisation goal at the colony level. We found no effect of individual size on the likelihood of evening foraging or on load size, although bees returning in lower light carried smaller loads, confirming that light is a key limiting factor for pollen collection. This study adds to our understanding of motivation and decision-making in relation to a daily change in the sensory environment in an important pollinator.g the loss.

Chapman, K; Cozma, N; Hodgkinson, A; English, R; Gaston, K; Hempel de Ibarra, N.

Poster 3

Feng, Chen – University of Bristol

Deep VQA based on a Novel Hybrid Training Methodology

In recent years, deep learning techniques have been widely applied to video quality assessment (VQA), showing significant potential to achieve higher correlation performance with subjective opinions compared to conventional approaches. However, these methods are often developed based on limited training materials and evaluated through cross validation, due to the lack of large scale subjective databases. In this context, this paper proposes a new hybrid training methodology, which generates large volumes of training data by using quality indices from an existing perceptual quality metric, VMAF, as training targets, instead of actual subjective opinion scores. An additional shallow CNN is also employed for temporal pooling, which was trained based on a small subjective video database. The resulting Deep Video Quality Metric (based on Hybrid Training), DVQM-HT, has been fully tested on eight HD subjective video databases, and consistently exhibits higher correlation with perceptual quality compared to other deep quality assessment methods, with an average SROCC value of 0.8263.

Feng, C; Zhang, A; Bull, D

Poster 4

Cipolina-Kun, Lucia – University of Bristol

Multi-Modal diffusion models for the preservation of cultural heritage

Within the field of Cultural Heritage, image in-painting is a conservation process that fills in missing or damaged parts of an artwork to present a complete image. Multi-modal diffusion models have brought photo-realistic results on image in- painting where content can be generated by using descriptive text prompts. However, these models fail to produce content consistent with a particular painter's artistic style and period, being unsuitable for the reconstruction of fine arts and requiring laborious expert judgement. Moreover, genera- tive models produce many plausible outputs for a given prompt. This work presents a methodology to improve the inpainting of fine art by automating the selection process of inpainted candidates. We propose a discriminator model that processes the output of inpainting models and assigns a probability that indicates the likelihood that the restored image belongs to a certain painter.

Cipolina-Kun, L; Caenazzo, S; Mazzei, G

Poster 5

Danier, Duolikun – University of Bristol

A Subjective Quality Study for Video Frame Interpolation

Video frame interpolation (VFI) is one of the fundamental research areas in video processing. While the development of novel interpolation algorithms has been widely researched, there is an important topic which has not been sufficiently investigated. This is the quality assessment of interpolated content. In this paper, we have conducted a subjective study for VFI based on a newly developed video database, BVI-VFI, which contains 36 reference at three different frame rates and 180 distorted videos generated using five conventional and learning based VFI algorithms. Subjective opinion scores have been collected from 60 human participants, and then

employed to evaluate eight popular quality metrics, including PSNR, SSIM and LPIPS which are commonly used for assessing VFI methods. The results indicate that none of these metrics provide acceptable correlation with the perceived quality on interpolated content, with the best-performing metric, LPIPS, offering a SROCC value below 0.6. Our findings show that there is an urgent need to develop a bespoke perceptual quality metric for VFI.

Danier, D; Zhang, F; Bull, D

Poster 6
Davies, Jay – University of Bristol
What makes an environment restorative? New insights from multi-arrangement and representational similarity methods

Nature environments are generally thought to be psychologically and physiologically more restorative than urban environments, with most literature arguing that nature possesses fundamentally different qualities to other environmental categories. However, little is known about what these qualities might be, as experiments are biased in stimulus selection and suffer from design flaws that limit the number of comparisons that can be sensibly made between different environments. Here, we implemented spatial multi-arrangement methodology to obtain comparisons between 92 images containing a wide range of nature and urban scenes from across the world (46 nature, 46 urban). Across three counterbalanced sessions, all participants (N=18) compared the 92 images for visual discomfort, liking and fascination as these are qualities suggested in the literature to be linked to restoration. Data (N=54 across all three criteria) are presented in the form of dissimilarity matrices and were analysed using multi-dimensional scaling and representational similarity analysis. Dissimilarity matrices exhibit a nature and urban similarity distinction for all three criteria, with analyses revealing that all three criteria are related. Furthermore, multi-dimensional scaling confirmed that liking and visual discomfort are the most tightly linked, while fascination diverges. Implications for the use of multi-arrangement methodology in aesthetics and vision research are discussed, as well as the significance of these findings to existing theories and experimental results in the field.

Davies, J; Leonards, U; Stevanov, J

Poster 8
Dong, Zhengyan – University of Bristol
Identifying Pitfalls in the Evaluation of Saliency Models for Videos

Saliency prediction has been extensively studied for natural images. In the area of video coding and video quality assessment, researchers attempt to integrate a saliency model to individual frames of a video sequence. In selecting best-performing saliency models for these applications, the evaluation only considers the average model performance over all frames of a video. This may mask the defects of a saliency model and consequently hinder further improvement of the model. In this paper, we present the identification of pitfalls in the evaluation of saliency models for videos. We demonstrate the importance of considering the video content classification and temporal effect. Building on the findings, we make recommendations for saliency model evaluation and selection for videos.

Dong, Z; Wu, X; Zhao, X; Zhang, F; Liu, H

Poster 9
Jones, Katherine – University of Bristol
Joint attention experiences for children with visual impairments

We aim to establish a model that outlines the processes underlying social interactions through joint attention in children with visual impairments.

Shared experiences permeate almost every part of our lives, but little is known about what children who are blind or visually impaired experience. These children do not have the same opportunities to engage in social interaction with their sighted parents or peers. They cannot easily identify who is in their immediate vicinity, let alone what others' focus of attention is, what is being gestured at or which facial expressions others around them might have.

Furthermore, sighted parents and peers can't easily recognise when a blind child is paying attention or wanting to share attention.

These problems contribute to children who are blind or visually impaired being more likely to be diagnosed with autism, having fewer friends, being more likely to be bullied and being 60% less likely to work for more than six months in adulthood.

Using classical grounded theory to analyse videos of children who are blind or visually impaired engaging in joint attention with others, we are developing a model of joint attention that provides ways to more easily recognise signals to initiate and maintain social interactions. By doing this, we hope to provide more opportunities for children who are blind or visually impaired. For example, through increased autonomy in initiating and maintaining social situations, improved social connections, independence when entering rooms, and learning more about the nuances of social interaction.

Jones, K; Leonards, U; Metatla, O

Poster 10

Kerr, Will – University of Bath

Autonomous Filming Systems: Towards Empathetic Imitation

Film making is an artistic and resource intensive process. The visual aesthetic of a finished film is the product of many departments, with significant expert and stylistic contributions from directors and cinematographers in the planning (pre-production / pre-visualisation) and filming (production) stages.

More commonly, pre-visualisation is performed in virtual environments, to allow ideas to be tested and visualised before expensive real-world filming takes place. In the pursuit of efficiency saving, there is scope to autonomise some aspects of these virtual planning stages, such that professional-level camera poses can be recommended to film makers limited by finance or expertise. This project takes the above problem forward by asking 2 research questions:

1)How can the intent of a professional cinematographer be learnt from existing finished film content?

2)How can these learnings be imitated back to new virtual filming tasks, empathetic to the story?

Existing work in autonomous virtual film-making has focussed on actor and camera positioning, but there remains a gap in how the composition of the frame is designed, particularly how the background elements (shape, colour, focus etc) play a part in the aesthetics of the footage in an empathetic style to the story. Work so far has developed a suite of visual analysis tools and film datasets, providing some evidence of cinematographic styles that were applied to films. Future work will develop a virtual filming environment, apply style to virtual shot composition, and offer comparisons to existing film footage.

Kerr, W; Li, W; Haines, T

Poster 11

Liang, Yuanbang – Cardiff University

Exploring and Exploiting Hubness Priors for High-Quality GAN latent sampling

Despite the extensive studies on Generative Adversarial Networks (GANs), how to reliably sample high-quality images from their latent spaces remains an under-explored topic. In this paper, we propose a novel GAN latent sampling method by exploring and exploiting the hubness priors of GAN latent distributions. Our key insight is that the high dimensionality of the GAN latent space will inevitably lead to the emergence of hub latents that usually have much larger sampling densities than other latents in the latent space.

As a result, these hub latents are better trained and thus contribute more to the synthesis of high-quality images. Unlike the a posteriori "cherry-picking", our method is highly efficient as it is an a priori method that identifies high-quality latents before the synthesis of images. Furthermore, we show that the well-known but purely empirical

truncation trick is a naive approximation of the central clustering effect of hub latents, which not only uncovers the rationale of the truncation trick, but also indicates the superiority and fundamentality of our method. Extensive experimental results demonstrate the effectiveness of the proposed method.

Liang, Y; Wu, J; Lai, Y-K; Qin, Y

Poster 12
Lou, Jianxun – Cardiff University
Predicting radiologist attention during mammogram reading with deep and shallow high-resolution encoding

Radiologists' eye-movement during diagnostic image reading reflects their personal training and experience, which means that their diagnostic decisions are related to their perceptual processes. For training, monitoring, and performance evaluation of radiologists, it would be beneficial to be able to automatically predict the spatial distribution of the radiologist's visual attention on the diagnostic images. The measurement of visual saliency is a well-studied area that allows for prediction of a person's gaze attention. However, compared with the extensively studied natural image visual saliency (in free viewing tasks), the saliency for diagnostic images is less studied; there could be fundamental differences in eye-movement behaviours between these two domains. Most current saliency prediction models have been optimally developed for natural images, which could lead them to be less adept at predicting the visual attention of radiologists during the diagnosis. We propose a method specifically for automatically capturing the visual attention of radiologists during mammogram reading. By adopting high-resolution image representations from both deep and shallow encoders, the proposed method avoids potential detail losses and achieves superior results on multiple evaluation metrics in a large mammogram eye-movement dataset. We also demonstrated the superiority of pre-training on a benchmark eye-movement dataset of natural images to estimate the saliency of diagnostic images.

Lou, J. Lin, H. Marshall, D. White, R. Yang, Y. Shelmerdine, S. Liu, H.

Poster 13
Ma, Yueran – Cardiff University
Evaluation of Image Quality Assessment Metrics for MR Images

Perceptual quality assessment of medical images has become increasingly necessary. Little research has been conducted at the algorithmic level in objective assessment methods of medical image quality, while the study of objective assessment methods for natural images has been better developed. It is imperative to extend the objective assessment methods for natural images to the medical image domain. In this study, a new subjective experiment was conducted in a radiology reading room environment at Angers University Hospital Center, France. The performance of 32 state-of-the-art traditional (i.e., as opposed to learning-based) objective IQA methods including 29 Full-Reference methods and 3 No-Reference methods was evaluated based on our MRI database using the Pearson Linear Correlation Coefficient (PLCC) as a measure. The results showed that nine FR methods have an accuracy of over 80 percent on MR images, with the best methods approaching 90 percent accuracy, while the NR method does not seem to perform well. In conclusion, according to the evaluation of the state-of-the-art traditional image quality assessment methods on medical images in our MRI database, it is feasible to apply the methodological ideas of natural image quality assessment to medical images, and there is certainly room for improvement. Our results shed some light on the differences in the details of feature extraction in medical and natural images. This will help us to propose new objective quality evaluation methods for medical images in the future.

Ma, Y; Tanguy, J-V; Corcoran, P; Liu, H

Poster 14
Ma, Wanli – Cardiff University
Attention based multi modal deep learning approach for land cover mapping

During the past a few decades, human activities have posed a serious threat to the environment, such as over-logging, over mining, illegal hunting, plastic pollution. This fact makes it necessary to monitor the earth to take

action against these and to prevent damages to the environment. Remote sensing is a widely used technology for the purpose of monitoring the earth but there are still many challenges for using remote sensing images in various environment applications. To name but a few: 1. It is expensive to obtain high-resolution images for all possible problematic areas. 2. passive remote sensors (e.g optical) are subject to weather conditions, such as cloud cover. 3. the content of remote sensing images are generally very complicated and difficult to analyse. In order to overcome the aforementioned challenges and beyond of using remote sensing images for environmental applications, taking advantages of fusing different remote sensing modalities, i.e. making use of strong capabilities of various sensor data, can be a potential solution, since fusing the information of different types of remote sensors can benefit from their complementary features. Thus, exploring multi-modality remote sensing data becomes very important to create highly powerful methods for the environmental analysis. Compared to traditional image fusion approaches, machine learning based methods generally show better performance for remote sensing image fusion in environment applications. This works proposes an attention based multi modal remote sensing image fusion approach for land cover mapping application in comparison to the benchmark deep learning approaches.

Ma, W; Karakus, O; Rosin, P

Poster 15
Ming, Yuhang – University of Bristol
CGiS-Net: Aggregating Colour, Geometry and Implicit Semantic Features for Indoor Place Recognition

We describe a novel approach to indoor place recognition from RGB point clouds based on aggregating lowlevel colour and geometry features with high-level implicit semantic features. It uses a 2-stage deep learning framework, in which the first stage is trained for the auxiliary task of semantic segmentation and the second stage uses features from layers in the first stage to generate discriminate descriptors for place recognition. The auxiliary task encourages the features to be semantically meaningful, hence aggregating the geometry and colour in the RGB point cloud data with implicit semantic information. We use an indoor place recognition dataset derived from the ScanNet dataset for training and evaluation, with a test set comprising 3,608 point clouds generated from 100 different rooms. Comparison with a traditional feature-based method and four state-of-the-art deep learning methods demonstrate that our approach significantly outperforms all five methods, achieving, for example, a top-3 average recall rate of 75% compared with 41% for the closest rival method.

Ming, Y; Yang, X; Zhang, G; Calway, A

Poster 16
Samano, Noe – University of Bristol
Geolocation by Embedding Images and Maps

Self-localization is an essential component for genuinely autonomous robots. Over the last years, diverse cross-view localization techniques that match visual observations with aerial imagery have been proposed. However, some disadvantages of aerial imagery are its dependence on capturing conditions and maintenance challenges. This research proposes a visual geolocation method that uses cartographic maps as the reference domain. Unlike aerial imagery, maps are a semantic representation of the world and are invariant to environmental conditions. The key idea of the method is to embed images and maps into a shared space where domains can be compared using Euclidean distance. The method allows a neural network to learn the most relevant features in a self-supervised manner and therefore generalizes previous approaches based on hand-crafted descriptors. Diverse experiments in extensive urban and suburban environments confirm the method's potential to geolocate street-level and aerial vehicles. Furthermore, a direct comparison between aerial imagery and cartographic maps when used as a reference for geolocation in the same geographic areas is presented. The results suggest that aerial imagery yields better results when environmental conditions during inference align well with those in training, but maps are more robust against variations in such conditions.

Samano, N; Zhou, M; Calway, A

Poster 17

Tian, Xin – University of Bristol

Optimal Transport-based Graph Matching for 3D retinal OCT image registration

Registration of longitudinal optical coherence tomography (OCT) images assists disease monitoring and is essential in image fusion applications. Mouse retinal OCT images are often collected for longitudinal study of eye disease models such as uveitis, but their quality is often poor compared with human imaging. This paper presents a novel but efficient framework involving an optimal transport based graph matching (OT-GM) method for 3D mouse OCT image registration. We first perform registration of fundus-like images obtained by projecting all b-scans of a volume on a plane orthogonal to them, hereafter referred to as the x-y plane. We introduce Adaptive Weighted Vessel Graph Descriptors (AWVGD) and 3D Cube Descriptors (CD) to identify the correspondence between nodes of graphs extracted from segmented vessels within the OCT projection images. The AWVGD comprises scaling, translation and rotation, which are computationally efficient, whereas CD exploits 3D spatial and frequency domain information. The OT-GM method subsequently performs the correct alignment in the x-y plane. Finally, registration along the direction orthogonal to the x-y plane (the z-direction) is guided by the segmentation of two important anatomical features peculiar to mouse b-scans, the Internal Limiting Membrane (ILM) and the hyaloid remnant (HR). Both subjective and objective evaluation results demonstrate that our framework outperforms other well-established methods on mouse OCT images within a reasonable execution time.

Tian, X; Anantrasirichai, N; Nicholson, L; Achim, A

Poster 18

Wang, Huasheng – Cardiff University

CUID: A New Study of Percieved Image Quality and it's Subjective Assessment

Research on image quality assessment (IQA) remains limited mainly due to our incomplete knowledge about human visual perception. Existing IQA algorithms have been designed or trained with insufficient subjective data with a small degree of stimulus variability. This has led to challenges for those algorithms to handle complexity and diversity of real-world digital content. Perceptual evidence from human subjects serves as a grounding for the development of advanced IQA algorithms. It is thus critical to acquire reliable subjective data with controlled perception experiments that faithfully reflect human behavioural responses to distortions in visual signals. In this paper, we present a new study of image quality perception where subjective ratings were reliably collected in a controlled laboratory environment and for a large degree of stimulus content variability. We investigate how quality perception is affected by a combination of different categories of images and different types and levels of distortions. The database will be made publicly available to facilitate calibration and validation of IQA algorithms.

Luice, L; Lui, H; Wang, H

Poster 19

Wu, Xinbo – Cardiff University

Analysis of video quality induced spatio-temporal saliency shifts

Human viewers' eye movements reflect their perceptual responses to visual signals. Previous research has shown that distortions in videos cause spatio-temporal gaze shifts, which means gaze behaviour is related to video quality perception. It would be highly beneficial to understand gaze behaviour of viewing videos of varying perceived quality. However, little is known about the interactions between gaze, video content and distortions. The study revealed that there is a significant difference in saliency between natural scene (i.e., original and pristine video content) and distorted scene (i.e., video content with visible distortions); and that the distortion/quality induced saliency shifts (QSS) significantly contribute towards the video quality assessment behaviour. In this paper, we conduct systematic analyses on the QSS in terms of the impact of video content (VC) and time order (TO). Building on the characteristics of gaze behaviour and quantitative methods, we make recommendations for future development of saliency methods in VQA metrics and video compression algorithms.

Wu, X; Dong, Z; Zhang, F; Rosin,P, L; Liu, H

Poster 20
Ziang, Ziao – Cardiff University
Deep learning for fruit ripeness determination

Deep learning models can be adopted as a non-destructive method to determine fruits ripeness stage, which would be deployed on some automatic harvest device to pick up the fruit at the right time. The aim of this research is to develop deep learning models to determine fruit ripeness stage based on images of fruits. Different deep learning techniques especially those segmentation networks such as U-net will be investigated to detect fruits and classify their ripeness stage. This research will propose deep learning models to achieve high classification accuracy for the tasks. In addition, the proposed model will be simplified without great accuracy trade-off, so that it would be deployed on mobile devices to facilitate fruit ripeness classification in field. Currently, some public fruit datasets are used to investigate the performance of different deep learning models, and the research is working on collecting images of several cultivars of peaches from China.

Zhao, Z; Hicks, Y; Sun, X