York University
Centre for Vision Research
International Conference on Perceptual Organization
June 23 – 26, 2015
Toronto, Canada

A principal challenge for both human and machine vision systems is to integrate and organize the diversity of cues received from the environment into the coherent global representations required to make good decisions and take effective actions. This conference brings together an interdisciplinary roster of leading researchers in both biological and computer vision to report and discuss the latest research on this process of perceptual organization.

Conference Chair: James Elder

Sponsors:

- Magstim
- York University
  - Office of the Vice President Research and Innovation
  - Lassonde School of Engineering
  - Faculty of Health
  - Faculty of Science
  - Department of Psychology
  - Department of Electrical Engineering and Computer Science
  - School of Kinesiology and Health Science

❖CONFERENCE SCHEDULE❖

Tuesday June 23rd — **Accolade West Main Floor Lobby**

5:00 Welcome Reception & Registration

Wednesday June 24th — **Accolade West 109 & Main Floor Lobby**

8:00 Coffee & Pastries

8:30 Conference Introduction

Multiple Sensory Streams
Session Chair: Doug Crawford

8:40 David Melcher, University of Trento
Constructing Objects across Multiple Time Scales and Frames of Reference

9:30 Laurence Harris, York University
The Role of the Body in Perceptual Organization

10:20 Marc Ernst, Bielefeld University
Getting in Touch with Multisensory Integration

11:10 Posters and coffee

12:00 Lunch (on your own)

Object Boundaries
Session Chair: Maz Fallah

1:20 Johan Wagemans, University of Leuven
Perceptual Organization at Object Boundaries: More than Meets the Edge

2:10 Laurie Wilcox, York University
Is depth from disparity a casualty of object perception?

3:00 Coffee Break

3:30 Pieter Roelfsema, Netherlands Institute for Neuroscience
The Virtues of Feedback Connections in the Visual Cortex for Perceptual Organization

4:20 Steve Silverstein, Rutgers University
Impaired Contour Integration in Schizophrenia: Contributions of Cortical Regions, and the Retina

5:10 Posters (canapés, cash bar)

Thursday June 25th **Accolade West 109 & Main Floor Lobby**

8:30 Coffee & Pastries

Motion
Session Chair: Rob Allison

9:00 Richard Born, Harvard Medical School
A (Relatively) New Look at an Old Problem: Motion through Apertures

9:50 Peter Tse, Dartmouth College
Interactions of Form and Motion in the Perception of Moving Objects

10:40 Coffee Break

11:10 Bill Freeman, Massachusetts Institute of Technology
A Big World of Tiny Motions

12:00 Lunch (on your own)

Visual Scenes
Session Chair: Rick Wildes

1:20 Axel Pinz, Graz University of Technology
Space-Time Representation and Recognition - from Dynamic Scenes to Actions in Context

2:10 Dirk Bernhardt-Walther, University of Toronto
2D Cues to 3D Structure Underlie Categorization of Real-world Scenes by Humans

3:00 Coffee Break

3:30 Derek Hoiem, University of Illinois at Urbana-Champaign
3D Scene Geometry from One View: Lessons in Computer Vision

4:20 Posters & Demos

6:15 Banquet (Black Creek Pioneer Village Patio)

Friday June 26th — **Accolade West 206 & Main Floor Lobby**

8:30 Coffee & Pastries

Object Shape
Session Chair: Hugh Wilson

9:00 Irving Biederman, University of Southern California
The Neural Basis of Shape Recognition

9:50 James Elder, York University
Perceptual Organization of Shape

10:40 Coffee Break

11:10 Sven Dickinson, University of Toronto
Detecting and Grouping Symmetric Parts in Cluttered Scenes

12:00 Lunch (on your own)

Light & Shading
Session Chair: Frances Wilkinson

1:20 Mike Langer, McGill University
Perception of Shape from Shading, Highlights, and Mirror Reflections

2:10 Richard Murray, York University
Lightness Constancy via Bayesian Anchoring

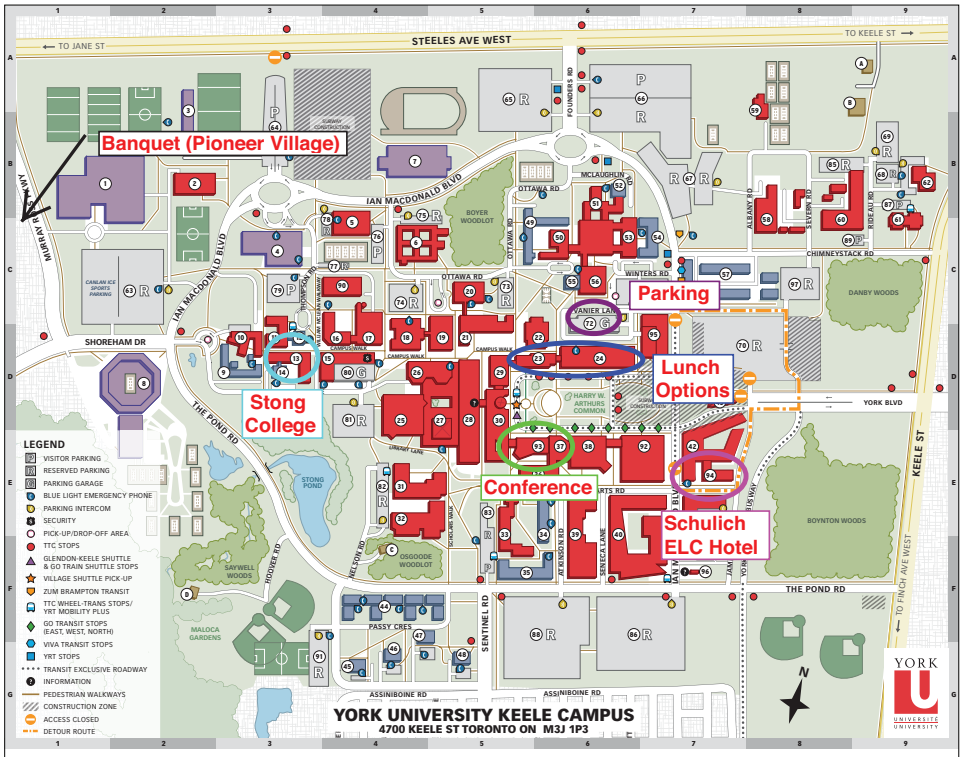
3:00 Coffee Break

3:30 Wendy Adams, University of Southampton
Shiny Things: Human Gloss Perception

5:00 Bus Pickup (in front of Vari Hall, to your left from Accolade West)

5:30 Reception at the McMichael Canadian Art Collection Gallery, Kleinburg

❖CAMPUS MAP❖



CAMPUS DIRECTORY

ACADEMIC, ADMINISTRATIVE & COMMERCIAL BUILDINGS

59	190 Albany Road	ALB	A8
92	Accolade East	ACE	E7
93	Accolade West	ACW	E6
95	Archives of Ontario*	AO	D7
33	Atkinson	ATK	E5
29	Behavioural Science	BSB	D5
41	**Bennett Centre for Student Services, Admissions**	BCS	E7
24	Bookstore, York Lanes	YL	D6
10	Calumet College	CC	D3
27	Central Square	CSQ	D5
58	Central Utilities Building	CUB	B8
38	Centre for Film & Theatre	CFT	E6
16	Chemistry	CB	D4
62	Computer Methods Building**	CMB	B9
26	Curtis Lecture Halls	CLH	D5
6	East Office Building	EOB	C4
94	Executive Learning Centre	ELC	E7
21	Farquharson Life Sciences	FRQ	D5
50	Founders College	FC	C6
31	Health, Nursing & Environmental Studies	HNE	E4
32	Ignat Kaneff Building, Osgoode Hall Law School	OSG	E4
36	Joan & Martin Goldfarb Centre for Fine Arts	CFA	E5
61	Kinsmen	K	C9
19	Lassonde Building	LAS	D5
90	Life Sciences Building	LSB	C4
43	Lorna R. Marsden Honour Court & Welcome Centre*	HC	D7
20	Lumbers	LUM	C5
51	McLaughlin College	MC	B6
11	Norman Bethune College	BC	D3
17	Observatory, Petrie	PSE	D4
17	Petrie Science & Engineering	PSE	D4
60	Physical Resources Building	PRB	B9
28	Ross Building	R	D5
25	Scott Library	SCL	D4
27	Scott Religious Centre, CSQ	SRC	D5
40	Seneca @ York, Stephen E. Quinlan Building**	SAY	E7
42	Seymour Schulich Building	SSB	E7
2	Sherman Health Science Research Centre	SHR	B2
18	Steacie Science & Engineering Library	STL	D4
22	Stedman Lecture Halls	SLH	D6
13	Stong College	SC	D3
23	Student Centre	STC	D6
4	Tait McKenzie Centre	TM	C3
39	Technology Enhanced Learning	TEL	E6
96	Temporary Welcome Centre		F7
56	Vanier College	VC	C6
30	Vari Hall	VH	D5
5	West Office Building	WOB	C4
15	William Small Centre	WSC	D4
53	Winters College	WC	C6
24	York Lanes	YL	D6
95	York Research Tower	YRT	D7

RESIDENCES & APARTMENTS

48	320 Assiniboine Road	AS2	G5
47	340 Assiniboine Road	AS4	F4
46	360 Assiniboine Road	AS6	G4
45	380 Assiniboine Road	AS8	G4
34	Atkinson Residence	AR	E6
12	Bethune Residence	BR	D3
9	Calumet Residence	CR	D3
49	Founders Residence	FR	C5
57	Harry Sherman Crowe Housing Co-op**	HCC	C7
44	Passy Gardens, 2–18 Passy Cres.	PASSY	F4
35	The Pond Road Residence	PON	F5
14	Stong Residence	SR	D3
52	Tatham Hall	TH	B6
55	Vanier Residence	VR	C6
54	Winters Residence	WR	C7

PARKING GARAGES – VISITORS

80	Arboretum Lane Parking Garage	ARB	D4
84	Student Services Parking Garage	SSP	E7
72	York Lanes Parking Garage	YLP	D6

PARKING LOTS – VISITORS

83	Atkinson Lot	E5
66	Founders Road East Lot	A6
64	Northwest Gate Temporary Lot	B3
79	Thompson Road Lot	C4
71	Vanier Lot	C7
76	West Office Building East Lot	C4
89	Physical Resources Building South Visitor Lot	C9
87	Kinsmen Lot	B9

PARKING LOTS – RESERVED

67	Albany Road Lot	B7
83	Atkinson Lot	E5
97	Chimney Stack Lot	C8
75	East Office Building Lot	B4
66	Founders Road East Lot	A6
65	Founders Road West Lot	A5
81	Library Lot	D4
73	Lumbers Lot	C5
82	Nelson Road Lot	E4
91	Passy Crescent Lot	G3
85	Physical Resources Building North Lot	D7
68	Physical Resources Lot	B9
69	Rideau Road Lot	B9
88	Sentinel Road Lot	F6
63	Shoreham Drive Lot	C2
74	Steacie Lot	C4
77	Tait McKenzie Lot	C3
86	The Pond Road East Lot	F7
71	Vanier Lot	C7
78	West Office Building West Lot	C4
70	York Boulevard Lot	B8

VISUAL PERFORMANCE ARTS FACILITIES

92	Art Gallery of York University	ACE	E7
37	Burton Auditorium*	BU	E6
92	Gales Gallery	ACW	E6
38	Joseph G. Green Studio Theatre	CFT	E6
92	McLean Performance Studio	ACE	E7
92	Price Family Cinema	ACE	E7
92	Sandra Faire & Ivan Fecan Theatre	ACE	E7
92	Tribute Communities Recital Hall	ACE	E7

SPORT & RECREATION FACILITIES

1	Canlan Ice Sports*	ICE	B1
4	Tait McKenzie Centre	TM	C3
8	Tennis Canada – Rexall Centre*	TC	D2
7	Track & Field Centre*	TFC	B4
3	York Stadium	STA	A2

HISTORICAL HOUSE

C	Hart House	HH	F4
D	Hoover House	HOH	F2
B	Stong Barn	SB	A9
A	Stong House	SH	A9

* Shared use
** Non-York facility
• Temporarily Closed

WiFi Access: **York is an Eduroam Institution. Please use AirYork or AirYorkPlus and log in with the credentials from your home institution.**

❖INVITED SPEAKERS❖

Constructing objects across multiple time scales and frames of reference

David Melcher
University of Trento

A basic challenge for perceptual organization is to parse the continuous flow of sensory information into discrete and stable objects. This process involves segregating objects (as separate, individual entities) but also integrating information about the same object over time and across the senses. Here, I will discuss recent work from my lab that investigates the temporal unfolding of this balance between integration and segregation. We have also examined how these processes account for shifts in coordinate frames caused by eye movements and by multisensory integration. Overall, our findings are consistent with theories that emphasize an active, multisensory and sensorimotor perception of coherent objects that interacts with ongoing (pre-stimulus) brain states.

David Melcher *received his PhD in Psychology from Rutgers University in 2001 and is now Associate Professor in the Faculty of Cognitive Sciences and Head of the Doctoral School in Cognitive and Brain Sciences at the University of Trento. He studies the interaction of perception, attention, memory and action within a cognitive neuroscience framework, examining the way that perception interacts with context, action (particularly eye movements), memory, emotion and tasks constraints.*

The role of the body in perceptual organization

Laurence Harris
York University

Incorporating the fact that the senses are embodied is necessary for an organism to interpret sensory information. Before a unified perception of the world can be formed, sensory signals must be processed with reference to a representation of the body. In this talk I shall examine some of the relationships between sensory and motor information, body representation, and perception of the world (including the body). I consider several examples of how the body affects perception including: how body orientation effects distance perception and object orientation; how visual-auditory crossmodal-correspondences depend on the orientation of the body; and how the perceived location of touch is affected both by the orientation of the head and eyes and the task that an observer is attempting to perform. The perceived attributes of the body, such as body size, affect tactile perception even at the level of detection thresholds and two-point discrimination, and long-range tactile masking may provide clues to the arrangement of our canonical body schema. Finally, our sense of ownership of a seen body part depends on its orientation and the perspective in which it is viewed. Together, these findings demonstrate how sensory and motor information, body representation, and our perception of our body and the world are highly interdependent.

Laurence Harris *obtained his PhD in Physiology and Psychology from Cambridge University in 1978. After a post doc in Dalhousie University, Nova Scotia, and an Addison Wheeler Fellowship at Durham (England), he took up a professorship in physiology at Cardiff University. He came to York University in 1990 where he is presently a professor of psychology, kinesiology and biology. He is the director of the Centre for Vision Research. His research is in multisensory interactions with an emphasis on the role of the body in all aspects of perception.*

Getting in touch with multisensory integration

Marc O. Ernst
Bielefeld University

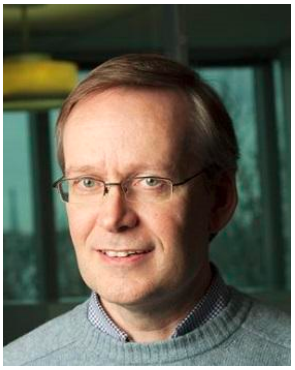
Vision is by far the best studied sensory modality. Compared to this, our knowledge about the sense of touch is very limited. On the one hand, this is due to the complexity involved in generating stimuli for the sense of touch in a controlled and largely automatic fashion. That is, for rendering better haptic stimuli, we would need a significant advance in haptic display technology. On the other hand, this is due to the complexity integral to the sense of touch. E.g., the sense of touch is inherently multisensory, and for a coherent representation of the external world the brain has to constantly combine tactile information with proprioceptive and kinesthetic information. Further, the sense of touch not only receives information about the world passively, but gathers information actively. And finally, compared to the eye, the primary organ of the human sense of touch–the hand–has many more degrees of freedom (28 for each hand compared to 3 for the eye). Despite these challenges, the human brain has to continuously integrate and perceptually organize the incoming information in order to form a robust and stable representation of the external world with witch we interact. In this talk I will briefly review some of our recent studies on the integration of information in the sense of touch. Furthermore, I will draw analogies in the processing of information between vision and touch, and demonstrate that many of the illusions well known in vision caused by eye movements have an analog in touch when we actively explore and scan objects with our hands. These results demonstrate the existence of common mechanisms in visual and haptic motion perception and for achieving spatial constancy. I will end my talk by outlining some of our recent findings of exploiting the redundancy of the motor system during manual interactions.

Marc Ernst *studied physics in Heidelberg and Frankfurt / Main. In 2000 he received his Ph.D. degree from the Eberhard-Karls-University Tübingen for investigations into human visuomotor behaviour, which he conducted at the Max Planck Institute for Biological Cybernetics. He spent almost 2 years as a research associate at the University of California, Berkeley working with Prof. Martin Banks on psychophysical experiments and computational models investigating the integration of visual-haptic information. In 2001, he returned to the MPI in Tübingen and became principle investigator of the Sensorimotor Lab in the Department of Prof. Heinrich Bülthoff. In 2007 he became leader of the Max Planck Research Group on Human Multisensory Perception and Action. Since 2011 he is Chair of the Department of Cognitive Neuroscience at the University of Bielefeld. Marc Ernst's scientific interest is in human multisensory perception, perceptual learning, sensorimotor integration and human-machine interaction.*

Perceptual organization at object boundaries: More than meets the edge

Johan Wagemans
University of Leuven

Objects are usually bounded in space. When we look at them from a single view, we usually have to figure out its shape and identity based on the occluding boundary. In most natural images, object boundaries are not given but must be derived from noisy, cluttered edge fragments. In a frequently used paradigm to study contour integration, the participants' task is to find a path or snake in a collection of Gabor patches (supposed to mimic edge fragments detected by V1 cells). The main focus of my talk is to indicate that perceptual organization at object boundaries is much richer than what is captured in this snake detection paradigm. I will discuss three lines of work from the literature and from my own lab to illustrate and support this idea: (1) the interplay between texture grouping and contour grouping, (2) surface completion versus edge linking, and (3) segmentation of object outlines into parts, based on local curvature properties as well as more global shape properties.



Johan Wagemans *holds a BA in psychology and philosophy and an M. Sc. And Ph. D. in experimental psychology, all from the University of Leuven (KU Leuven), where he is now a Full Professor and Director of the Laboratory of Experimental Psychology (founded in 1892) and Head of the Department of Brain & Cognition (founded in 2015). He is chief-editor of Perception, i-Perception, and Art & Perception and also editor of the Oxford Handbook of Perceptual Organization (2015). His main research interests are in so-called mid-level vision (perceptual grouping, figure-ground organization, depth and shape perception), but stretching out to low-level vision (contrast detection and discrimination) and high-level vision (object recognition and categorization). In addition to fundamental research in these areas, using a variety of methodological approaches such as psychophysics, modelling and neuroimaging, he is also investigating application in autism, arts and sports.*

Is depth from disparity a casualty of object perception?

Laurie Wilcox
York University

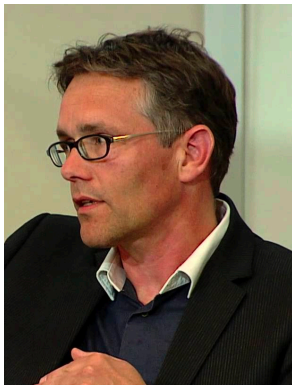
The perception of depth from binocular disparity is susceptible to configural changes. For example, stereoscopic thresholds for a pair of vertical lines increase dramatically when these lines are connected to form a closed object (McKee 1983). In a series of experiments we have shown that this phenomenon is also evident in *suprathreshold* depth magnitude estimates; perceived depth is reduced when elements are perceived as belonging to the same object. In this talk I will describe experiments in which we systematically evaluated the contribution of well-known 2D Gestalt grouping cues (connectedness, closure, similarity) to this phenomenon, and showed the close relationship between figural grouping and perceived depth. Further, I will discuss results which demonstrate that there is a 3D variant of good continuation that can modulate the strength of this grouping phenomenon. These data cannot be explained by local interactions between stimulus components. Instead they necessitate input from object or surface -based representations, which enhance object cohesion, and detectability, at the expense of within-object depth percepts.

***Laurie Wilcox** is a Professor of Psychology, and member of the Centre for Vision Research at York University, Toronto. She is also cross-appointed to the graduate program in Biology. Laurie completed her graduate degrees at the University of Western Ontario, and joined York University in 1996. Her research uses psychophysical techniques to reveal properties of the neural mechanisms which underpin stereoscopic (S3D) depth perception. In addition to basic research on S3D she has been actively involved in understanding the factors that influence the viewer experience of S3D film with collaborators including as IMAX and Christie Digital.*

The virtues of feedback connections in the visual cortex for perceptual organization

Pieter R. Roelfsema
Netherlands Institute for Neurosciences

Our visual system has to group the image elements that belong to one object and to segregate them from other objects and the background. I will discuss how perceptual grouping is implemented in the visual cortex. It is useful to distinguish between two mechanisms for grouping: base-grouping and incremental grouping. Base-groupings are coded by single neurons tuned to multiple features and are computed rapidly. However, not all conceivable feature combinations are coded by dedicated neurons. Therefore, a second, flexible form of grouping is required that is called incremental grouping. Incremental grouping is an iterative process that relies on horizontal and feedback connections that spread an enhanced response to all the neurons that code image elements of the same perceptual object. At a psychological level of description, this labeling process corresponds to object-based attention. Feedback connections can propagate information about the overall shape and the predominant features of relevant objects to lower level neurons. Similarly, horizontal connections can propagate the enhanced response to neurons coding image elements related to each other by Gestalt grouping rules, such as colinearity and connectedness. Thus, we start to understand the interactions among the neurons in the visual cortex that enable perceptual grouping. Importantly, we can also measure the time-course of the incremental grouping process in the reaction times of human observers.



Pieter Roelfsema *received his MD degree at the University of Groningen (1991). For his Ph.D. work he joined the Wolf Singer lab at the Max-Planck-Institute for Brain Research in Frankfurt (Germany) to work on the visual system of the cat and the role of neuronal synchronization. He received his Ph.D. at the University of Amsterdam in 1995. He then worked as postdoc in the department of Visual Systems Analysis at the University of Amsterdam. In 2002 he became Programme Leader at the Netherlands Ophthalmic Research Institute, which became part of the Netherlands Institute for Neuroscience in 2005. In 2007 he was appointed as General Director of the Netherlands Institute for Neuroscience, an institute of the Royal Academy of Arts and Sciences of the Netherlands. In 2005 he became Strategic Professor at the VU and in 2012 also Professor at the AMC in Amsterdam. His main interest lies in how different areas of the cortex co-ordinate their activity during perception and cognition. He uses highly precise measuring techniques to study the role of single neurons, cortical layers and receptor subtypes in various perceptual tasks.*

Impaired contour integration in schizophrenia: contributions of cortical regions, and the retina

Steven M. Silverstein
Rutgers University

Over 50 studies have now demonstrated impaired perceptual organization in schizophrenia, and the stimulus and task conditions under which this is most pronounced. Data are also consistent on which patient characteristics [state (symptoms) and trait (developmental course)] are related to the variability in task performance. There has been little work thus far, however, on the neurobiology of perceptual organization impairment in schizophrenia. This presentation will review data from 4 studies of contour integration that shed light on this issue. The first, an fMRI study with chronic and symptomatic patients, indicated hypoactivity in occipital (V2-V4), prefrontal and parietal regions. The second study, from a similar patient population, examined event-related potential (ERP) correlates of task performance, identified reduced P1 and N_{CL} amplitudes in patients, and localized contour integration impairment to reduced activity in V2 and the fusiform gyrus. The third study was a large, multi-site fMRI study of relatively asymptomatic patients. This study found no evidence for occipital lobe involvement, and found *increased* activity in patients in a number of parietal and frontal regions involved in higher level visual perception and spatial attention. Methodological differences between this study and the two earlier ones suggest that increased exposure to well-formed contours normalized occipital function, and that reduced task accuracy reflected failures in adaptation, stimulus enhancement and perceptual learning. The fourth study is examining correlations between retinal nerve fiber layer thinning, as revealed by spectral domain optical coherence tomography, and contour integration task performance, to evaluate low-level contributions to reduced perceptual organization. Taken together, these results suggest that contour integration impairment is a reliable feature of schizophrenia, and that it can involve both occipital and higher-level brain regions, depending on the task and patient characteristics. Potential retinal contributions will also be discussed.

Steven Silverstein *is Director of Research, and Director of the Division of Schizophrenia Research at Rutgers-University Behavioral Health Care (UBHC). Dr. Silverstein received his Ph.D. in Psychology from the State University of New York at Buffalo in 1989, and completed his internship and fellowship at Cornell Medical College from 1988-1990. He was on the faculty of the University of the Rochester Medical Center, the Weill Medical College of Cornell University, and the University of Illinois at Chicago (UIC) prior to moving to Rutgers in 2006. Dr. Silverstein's research interests are in cognitive and perceptual impairments in schizophrenia, rehabilitation of these problems, and psychometric and methodological issues related to their assessment.*

A (relatively) new look at an old problem: Motion through apertures

Richard Born
Harvard Medical School

How does the primate visual system solve the aperture problem for visual motion? The textbook explanation involves a purely one-dimensional (1D) representation of the motion of extended contours by V1 neurons followed by the nonlinear integration of 1D measurements made from two or more contours with different orientations by neurons in MT. I will present evidence from monkey neurophysiology and human psychophysics that argues for consideration of a different class of models, collectively referred to as "selectionist," in which mechanisms in V1, such as end-stopping, suppress 1D motion signals in favor of those arising from the motion of 2D features, or "terminators." The temporal dynamics of V1 end-stopping and its sensitivity to contrast are consistent with human motion perception, which reveals a vector average of 1D motion signals for short viewing times and low-contrast stimuli. For high-contrast stimuli, terminator motion exerts a powerful influence on human motion perception as evidenced by, for example, the barber-pole illusion. Moreover, the visual system's classification of terminators as "intrinsic" (i.e. belonging to the thing that is moving) vs. "extrinsic" (i.e. due to accidents of occlusion) also determines how strongly they will influence motion perception. I will show data demonstrating that the responses of direction-selective neurons in both MT and V1 are sensitive to the motion of terminators and to their classification, and will speculate (supplemented by data) on how such sensitivity might arise through an interaction between feedforward (V1 → V2 → MT) and feedback (V2 → V1 and/or MT → V1) pathways.

Richard T. Born, M.D., is a Professor of Neurobiology at Harvard Medical School and was Director of the Harvard PhD Program in Neuroscience from 2009 to 2014. He was raised in the Upper Peninsula of Michigan and attended DePauw University where he received a B.A. in Chemistry in 1983. He attended Harvard Medical School (HMS), where he discovered the joys of visual neurophysiology working with Professors David Hubel and Margaret Livingstone. After receiving the M.D. degree in 1988, he continued on as a postdoctoral fellow in the Hubel/Livingstone lab, undertook a second postdoc with William Newsome at Stanford and then returned to HMS in 1995 as an Assistant Professor in the Department of Neurobiology. He is currently a member of the Faculty of 1000 and serves on the "Sensory Processing and Cognition" Study Section at the NIH. His laboratory studies cortical visual processing in nonhuman primates, with a particular interest in the nature of cortico-cortical feedback.

Interactions of form and motion in the perception of moving objects

Peter Tse
Dartmouth University

I will summarize the past 20 years of our research demonstrating that there is 'motion from form' processing. It has long been known that the visual system can construct 'form from motion.' Less appreciated has been the degree to which perceived motion follows from processes that rely upon rapid analyses of form cues. Percepts that depend on such form-motion interactions reveal that form information can be processed and integrated with motion information to determine both the perceived velocity and shape of a moving object. These integration processes must be rapid enough to occur in the brief period, probably less than a quarter of a second, between retinal activation and visual experience. Data suggest that global form analyses subserve motion processing in at least five ways (Porter et al., 2011). Here, I describe three examples in which the analysis of form significantly influences our experience of moving objects. First, I describe Transformational Apparent Motion, a phenomenon that reveals how form analyses permit the figural segmentation dedicated to solving the problem of figure-to-figure matching over time (Hsieh and Tse, 2006; Tse, 2006; Tse & Caplovitz, 2006; Tse & Logothetis, 2002). Secondly, I describe how the size and shape of an object can influence how fast it is perceived to rotate. These interactions reveal the way in which form analyses permit the definition of trackable features whose unambiguous motion signals can be generalized to ambiguously moving portions of an object to solve the aperture problem (Caplovitz et al., 2006; Caplovitz & Tse, 2007a,b). Finally, we describe a number of peculiar ways in which the motions of individual elements can interact with the perceived shape and motion of a global object constructed by the grouping of these elements. These phenomena reveal that the form analyses that underlie various types of perceptual grouping can lead to the generation of emergent motion signals belonging to the perceptually grouped object that appear to underlie the conscious experience of motion (Caplovitz & Tse, 2006, 2007b; Hsieh & Tse, 2007; Kohler et al., 2010; Kohler et al., 2009).

Peter Ulric Tse *is a professor of cognitive neuroscience at Dartmouth College. He studied physics and math at Dartmouth from 1980 to 1984, and then worked as a schoolteacher in Nepal, and later as a businessman in Japan. He returned to America in 1992 to study the brain and cognitive psychology, getting his Ph.D. from Harvard in 1998. From 1999 - 2001 he was a postdoctoral fellow in brain imaging at the Max Planck Institute for Biological Cybernetics where he worked with macaques. His lab's research focuses on using brain and behavioral data to constrain models of the neural bases of attention and consciousness, unconscious processing that precedes and constructs consciousness, mental causation, and human capacities for imagination and creativity.*

A big world of tiny motions

Bill Freeman
Massachusetts Institute of Technology

We have developed a "motion microscope" to visualize small motions by synthesizing a video with the desired motions amplified. The project began as an algorithm to amplify small color changes in videos, allowing color changes from blood flow to be visualized. Modifications to this algorithm allow small motions to be amplified in a video. I'll describe the algorithms, and show color-magnified videos of adults and babies, and motion-magnified videos of throats, pipes, cars, smoke, and pregnant bellies. These algorithms are being used in biological, civil, and mechanical engineering applications. Having this tool led us to explore other vision problems involving tiny motions. I'll describe recent work in analyzing fluid flow and depth by exploiting small motions in video or stereo video sequences caused by refraction of turbulent air flow (joint work with the authors below and Tianfan Xue, Anat Levin, and Hossein Mobahi). We have also developed a "visual microphone" to record sounds by watching objects, like a bag of chips, vibrate (joint with the authors below and Abe Davis and Gautham Mysore).

Collaborators: Michael Rubinstein, Neal Wadhwa, and co-PI Fredo Durand.

William T. Freeman *is Professor of Electrical Engineering and Computer Science at MIT, and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL) there. He is currently on leave, starting a computer vision group at Google in Cambridge, MA. He received his PhD in Media Arts and Sciences from MIT in 1992. His current research interests include machine learning applied to computer vision, Bayesian models of visual perception, and computational photography. Previous research topics include steerable filters and pyramids, orientation histograms, the generic viewpoint assumption, color constancy, computer vision for computer games, and belief propagation in networks with loops.*

Space-Time representation and recognition - from dynamic scenes to actions in context

Axel Pinz
Graz University of Technology

It is well-known that contextual priming can substantially improve and ease image and object categorization tasks. How should contextual priming be tackled in the more complex case of video analysis? In this talk, I will introduce a particular representational framework for video and show its application to dynamic scene and to action recognition. Complementary spatial, spatio-temporal and color features are extracted from short video slices, coded into a mid-level representation and pooled in a novel dynamic space-time pyramid. Experiments provide evidence for the complementarity of the various features and indicate that this unified approach is well-suited to video analysis in general. In a second, more general part, I will discuss various important characteristics of video recognition, related terminology, and implications thereof. As a result, I will analyze current benchmark video datasets, their limitations and the benefit of putting actions in context of camera motion and general scene dynamics.

Axel Pinz *received the M.Sc. degree in electrical engineering in 1983 and the Ph.D. degree in computer science in 1988 from Vienna University of Technology. In 1995 he received the habilitation in computer science from Graz University of Technology.*
He worked in high-level image analysis in Remote Sensing at the University of Natural Resources in Vienna (1983-1990, Institute of Surveying and Remote Sensing IVFL). From 1990 to 1994 he was with the Institute for Automation, Department of Pattern Recognition and Image Processing (PRIP), Technical University of Vienna, working as an assistant professor in Computer Vision. From 1994-1999 he was a visiting scientist at the Institute for Computer Graphics and Vision (ICG), Graz University of Technology, where he was building up the Computer Vision Group of the Institute.
In 1996/97, he served as the academic head of the ICG, and from Oct 1997 - July 1999 he was a visiting professor in computer vision and computer graphics at Graz University of Technology, Austria. Since Oct 1999 he has been with the Institute of Electrical Measurement and Measurement Signal Processing (EMT), Graz University of Technology, Austria, where he has built up and is now heading the vision-based measurement group (vmg).

2D Cues to 3D structure underlie categorization of real-world scenes by humans

Dirk Bernhardt-Walther
University of Toronto

People can categorize their natural environment as, say, a forest, a highway, or an office, very quickly and accurately. What image properties underlie this remarkable ability? We extracted a number of candidate features from line drawings of natural scenes and measured their relevance for this task. Computational analysis showed that the statistical distribution of contour orientation is the most informative feature. Comparison of the structure of categorization errors between computational and behavioral scene categorization, however, revealed that contour junctions and curvature but not contour length or orientation are related to human performance in categorizing scenes. When we manipulated the line drawings such that contour junctions got perturbed, however, participants resorted to using contour length and orientation. Comparisons of error patterns between the computational analysis and decoding of scene categories from fMRI activity showed that the parahippocampal place area (PPA) primarily relies on junction properties and curvature in its representation of scene categories, whereas primary visual cortex is mostly related to contour length and orientation. We conclude that people are able to opportunistically use a range of features for scene categorization, but given the choice, they will use contour junctions and curvature. These features also underlie the representation of natural scene categories in the PPA. Importantly, these features are well suited to constructing 3d structural descriptions of shape rather than global statistics of image texture.

Dirk Bernhardt-Walther *obtained his Ph.D. in Computation and Neural Systems at the California Institute of Technology, working with Christof Koch on modeling visual attention and object recognition. After a few months with John Tsotsos at York University in Toronto Dirk became a Beckman Postdoctoral Fellow at the Beckman Institute at the University of Illinois at Urbana-Champaign. There he worked with Diane Beck and Fei-Fei Li on natural scene perception and on decoding natural scene categories from fMRI data. He started his first own lab at The Ohio State University, before moving to the University of Toronto in 2014. Dirk and his lab employ visual psychophysics, eye tracking and functional magnetic resonance imaging as well as computational modeling in order to decode how our brain manages to make sense of the complex visual information around us.*

3D scene geometry from one view: lessons in computer vision

Derek Hoiem
University of Illinois at Urbana-Champaign

Within computer vision, the problem of inferring 3D scenes has mainly been addressed as a mathematical, quantitative problem: if we can find corresponding points in images from two or more views, we can solve for the 3D positions of those points. This approach of geometry as algebraic constraint satisfaction has led to successful and useful systems, but it leaves open that nagging question, how can humans close one eye or examine a photograph and still experience a strong impression of depth and geometric structure? In the last ten years, we've seen remarkable progress in 3D interpretation from one image. This progress has required a completely different approach to how we represent and infer geometry: from triangulation to recognition, from metric reconstruction to abstractions, and from reporting what we see to inferring what is there. In this talk, I'll describe the key lessons learned as the computer vision approach to single-view 3D interpretation has evolved and show applications to recognition and visualization.

Derek Hoiem *is an Assistant Professor of Computer Science at the University of Illinois at Urbana-Champaign, since January 2009. Derek received his PhD in Robotics from Carnegie Mellon University in 2007 and completed a postdoctoral fellowship at the Beckman Institute in 2008. His main research interests are 3D scene understanding and object recognition. Derek's work has been recognized with an ACM Doctoral Dissertation Award honorable mention, CVPR best paper award, Intel Early Career Faculty award, Sloan Fellowship, and PAMI Significant Young Researcher award.*

The neural basis of shape recognition

Irving Biederman
University of Southern California

In a fraction of a second--from a single visual fixation--humans can comprehend novel images of objects and scenes, often under varied or degraded viewing conditions. The representation that allows this extraordinary access to scene understanding can be achieved from a single image by coding view-invariant properties of contours that mark orientation and depth discontinuities, that is, through *shape*. Discounted are surface properties, such as color, texture, or illumination gradients. Recent behavioral, neural imaging and single-unit recording experiments now provide strong confirmation that this achievement can be mediated by representing objects as an arrangement of simple, viewpoint-invariant, shape primitives. Although the incorporation of robust image statistics has been a guiding principle in understanding the development of early (i.e., V1) image coding, we present the first evidence (to our knowledge) that higher order shape coding is similarly constrained by image statistics.

Irving Biederman *received his Ph.D. degree from the University of Michigan in 1966. In addition to being professor of psychology and computer science at the USC College of Letters, Arts and Sciences, he is holder of the Harold Dornsife Chair in Cognitive Neuroscience, and is also a member of the USC Program in Neural, Informational and Behavioral Sciences. Dr. Biederman specializes in the study of brain processes underlying humans' ability to quickly recognize and interpret what they see. While best known for his Recognition by Components Theory that focuses on object recognition, his more recent work has tended to examine the recognition of human faces. Biederman argues that face recognition is separate and distinct from the recognition of objects. Biederman Biederman recently appeared on an episode of Penn & Teller: Bullshit!, explaining the thought process of UFO hunters.*

Perceptual organization of shape

James Elder
York University

Humans are very good at rapidly detecting salient natural shapes such as animals in complex scenes. This suggests that our visual system is finely tuned to the statistics of natural shape, and this is borne out by recent evidence for efficient coding of local shape information. In computational models the prevailing view has been that these local regularities drive the perceptual organization of global shape in feed-forward fashion. However, feed-forward computer vision algorithms fall well below human levels of performance, and recent results suggest that perceptual organization is also drive by global cues. This suggests a recurrent architecture in which higher areas of the object pathway generate shape hypotheses that condition grouping processes in early visual areas. As a candidate generative model for shape coding, I will explore a framework in which shape is progressively represented as a sequence of deformations called formlets. I will compare this formlet representation with alternative theories in terms of accuracy in completing fragmented shape hypotheses that may be generated by the early visual system, and in terms of capturing the shape information required for recognition.

James Elder *is a Professor in the Department of Electrical Engineering & Computer Science and the Department of Psychology at York University, and a member of York's Centre for Vision Research. He obtained his PhD in Electrical Engineering from McGill University in 1996, working with Steve Zucker. His research seeks to improve machine vision systems through a better understanding of visual processing in biological systems. Current research is focused on natural scene statistics, perceptual organization, contour processing, shape perception, single-view 3D reconstruction, attentive vision systems and machine vision systems for dynamic 3D urban awareness.*

Detecting and grouping symmetric parts in cluttered scenes

Sven Dickinson
University of Toronto

Perceptual grouping played a prominent role in support of early object recognition systems, which typically took an input image and a database of shape models and identified which of the models was visible in the image. Using intermediate-level shape priors, causally related shape features were grouped into discriminative shape indices that were used to prune the database down to a few promising candidates that might account for the query. In recent years, however, the recognition (categorization) community has focused on the object detection problem, in which the input image is searched for a specific target object. Since indexing is not required to select the target model, perceptual grouping is not required to construct a discriminative shape index. As a result, perceptual grouping activity at our major computer vision conferences has diminished. However, there are clear signs that the recognition community is moving from appearance back to shape, and from detection back to multi-class object categorization. Shape-based perceptual grouping will play a critical role in facilitating this transition. In this talk, I will describe our recent progress on detecting and grouping symmetric parts in cluttered scenes.

Sven Dickinson *received the B.A.Sc. degree in Systems Design Engineering from the University of Waterloo, in 1983, and the M.S. and Ph.D. degrees in Computer Science from the University of Maryland, in 1988 and 1991, respectively. He is currently Professor and Chair of the Department of Computer Science at the University of Toronto. Prior to that, he was a faculty member at Rutgers University where he held a joint appointment between the Department of Computer Science and the Rutgers Center for Cognitive Science (RuCCS). His research focuses on object recognition, perceptual grouping, shape representation, and algorithms for inexact graph indexing and matching.*

Perception of shape from shading, highlights, and mirror reflections

Michael Langer
McGill University

The human visual system has a remarkable ability to perceive three-dimensional surface shape from visual cues such as texture, shading and specular reflections. Here I will present several experiments that examined qualitative shape perception of surfaces that were rendered using a variety of material models, ranging from diffuse to shiny to mirror-like. The experiments re-examined the light-from-above and viewpoint-from-above priors, which are known to be used for shape from shading under diffuse reflectance, and asked whether these priors also are used for surfaces that have specular or mirror reflectance. I will also discuss theories of what these priors imply about how 3D shape can be inferred from image cues such as shading and texture. Finally I will address the more general question of whether lighting and material models are needed to explain shape from shading perception, or whether purely image-based cues can suffice.

Michael Langer *is an Associate Professor in the School of Computer Science and a member of the Centre for Intelligent Machines at McGill University. He received a B.Sc. in Mathematics from McGill in 1986, an M.Sc. in Computer Science from the University of Toronto in 1988, and a Ph.D. from McGill in 1994. He was a postdoctoral fellow at the NEC Research Institute in Princeton NJ, and a Humboldt Research Fellow at the Max-Planck-Institute for Biological Cybernetics in Tuebingen, Germany.*

Lightness constancy via Bayesian anchoring

Richard F. Murray
York University

It has been surprisingly difficult to develop successful theories of how people perceive simple surface properties such as reflectance, shape, and illuminance. Image ambiguity is a fundamental obstacle to perceiving these properties, which suggests that Bayesian approaches may be useful. I will show that a simple Bayesian theory that makes natural assumptions about surface reflectance and lighting conditions can account for a surprisingly wide range of phenomena in lightness perception. I will also describe classification image experiments that investigate what parts of an image contribute to lightness percepts. The results of these experiments are inconsistent with most current theories of lightness perception, such as the low-level ODOG model, and the idea that certain lightness illusions are largely driven by X junctions. The results are most consistent, but not completely consistent, with Alan Gilchrist's anchoring theory, which says that image patches are first grouped into lighting frameworks, and then assigned lightnesses based on the ratio of their luminance to the highest luminance in the relevant frameworks. Finally, I will show that an interesting special case of lightness perception, namely glow perception, is closely bound up with 3D shape perception: we can turn glow percepts on and off, simply by using disparity cues to change the perceived 3D shape of a fixed 2D image.

Richard Murray *is an Associate Professor of Psychology at York University. He studied philosophy at the University of Toronto for a B.A. and at the University of Cambridge for an M.Phil., and then studied experimental psychology at the University of Toronto for his Ph.D. under Patrick Bennett and Allison Sekuler. After a postdoc with Wilson Geisler at the University of Texas at Austin and a faculty position at the University of Pennsylvania, he joined York University in 2005. His research interests centre on low-level spatial vision and mid-level visual phenomena such as lightness perception and perceptual organization.*

Shiny things: human gloss perception

Wendy Adams
University of Southampton

From a single glance, we are pretty adept at judging what material something is made of. This is useful for recognizing objects and planning our interactions with them. For example, how glossy an object looks might predict how it will feel.
However, judging gloss from an image is challenging – the pattern of specular highlights in the image depends not only on the object's reflectance, but also on its shape and how it has been illuminated. I will discuss:
i. How our perception of gloss is affected by changes in illumination
ii. How vision and touch interact when we estimate gloss, and
iii. Measurements of our natural environment that can inform models of gloss perception.

Wendy Adams *is Associate Professor of Psychology at the University of Southampton, UK. She received a PhD in Psychology from the University of Sheffield, and then held postsdoctoral positions in the laboratory of Martin Banks at the University of California, Berkeley and Pascal Mamassian at the University of Glasgow. Her research interests include shape and depth perception, material perception, visual-haptic interactions and natural scene statistics.*

❖POSTERS❖

1. No otoacoustic evidence for a peripheral basis underlying absolute pitch

Christopher Bergevin[1], Larissa McKetton[1], Victoria S[2], Jessica Grahn[2], & David Purcell[2]
[1]York University, [2]University of Western Ontario, Canada

Absolute pitch (AP) is the ability to identify or produce the perceived pitch of a sound (e.g., fundamental frequency of a piano note) without an external reference. This ability is relatively rare (~1/10000 individuals possess it) and the mechanisms underlying AP are not well understood. This study examined whether there was evidence for a peripheral (i.e., cochlear) basis for AP based upon otoacoustic emissions (OAEs). The chief motivations were that both AP and spontaneous emissions (SOAEs) appear to have genetic components and anecdotal observations of prevalence in certain populations (e.g., relatively higher incidence of both in Asians). We examined SOAEs and stimulus-frequency emissions (SFOAEs) in both control (N=21) and AP (N=13) normal-hearing populations. We found no substantial differences in SOAE activity between groups (e.g., no evidence for one or more strong SOAEs that could act as a cue). SFOAE phase-gradient delays, measured using several probe levels (20-50 dB SPL), also showed no significant differences. This latter observation argues against sharper frequency selectivity in AP subjects. Taken together, these data support the prevailing view that AP mechanisms arise at a processing level in the central nervous system at the brainstem or higher (e.g., optimized neural coding).

2. Enhancing access to online financial charts: Considerations in translating visual line-graphs into sound for low-vision or visually impaired populations

Michael James Carnevale, Ambrose Li, Damon Pfaff, & Peter Coppin, OCAD University, Canada

Low-vision populations have difficulty accessing digital media visually. To solve this problem, authors of digital media manually translate visual graphics into text-to-speech descriptions that are presented through auditory screen reader technology. Text-to-speech descriptions, however, cannot fully convey the shape contours afforded by directly interpreting visual graphics such as charts and diagrams through visual perception. Here we present a pair of prototypes for translating financial charts (i.e., line graphs of stock price over time) into an accessible auditory format by employing (a) sonification—the presentation of information using non-speech audio, and (b) a text-to-speech generator—an algorithm for automatically translating novel data sets into text-to-speech. Using lower-level perceptual features for sonification such as pitch, spatial audio, and sound intensity, the perceptually specific shape information afforded by line graphs can be conveyed. In contrast, the text-to-speech generator can provide a general overview of the graph as well as conceptual specifics. Differences between presenting information via sonification and text-to-speech revealed through the process of design will be discussed, and psychophysical experiments intended to test and optimize these prototypes are outlined.

3. Cognitive inference and the resolution of audio-visual stream/bounce displays

Philip M. Grove, & Mick Zeljko, School of Psychology, The University of Queensland, Australia

The "stream/bounce" illusion is a tractable stimulus used to probe questions about audio/visual interactions in perception. A typical display depicts two identical discs that approach each other on a collision course which can be seen as two discs streaming through each other or bouncing off each other. Which perception dominates, is influenced by a brief transient, usually a sound, presented around the time of simulated contact. Several theories have been proposed to account for the switching in dominance based on sensory processing, attention and cognitive inference, but a universally applicable, parsimonious explanation has not emerged. Data from our laboratory are most consistent with the proposal that the perceptual system employs a form of cognitive inference to formulate representations of events by taking into account information accumulated over time – over the course of one's life or at least over the course of an experiment. We present three lines of evidence in line with the cognitive inference hypothesis. In study 1, we discount the role of low-level sensory processes via a reexamination of studies by Bertenthal et al. (1993) and Sekuler and Sekuler (1999), showing that manipulations of the motion profile of the moving targets do not influence perception. In study 2 we show that information acquired during the course of an experiment influences responses to these stimuli such that dominance of streaming or bouncing is relative to the context of the experimental stimuli. Lastly, in study 3, we show that the subjective binding of auditory and visual events is associated with criterion shifts rather than changes in sensitivity in a signal detection paradigm. Together, these observations discount bottom up processes and favor cognitive inference as the major driver determining whether a bounce or streaming is perceived and may form the basis for a universal explanation of instances of the stream/bounce illusion.

4. Decoding the neural representations of real-world soundscapes

Yaelan Jung[1], Bart Larsen[2], & Dirk Bernhardt-Walther[1]

[1] Department of Psychology, University of Toronto, Canada, [2] Department of Psychology, University of Pittsburgh, USA

The categories of visually presented natural scenes can be classified from the pattern of brain activity in visual cortex (Walther et al., 2009). In this study, we examine whether the categories of acoustically presented scenes can be decoded from fMRI activity in auditory cortex. Participants were asked to listen to the sounds of four different scene categories (beach, forest, city, and office) while they were in an MRI scanner. Their neural activity was used to train a classifier to predict the scene categories of the sounds. As visual scene categories can be decoded in various visual areas, we found that auditory scene categories can be decoded in primary auditory cortex (A1) as well as the superior temporal sulcus and the right inferior and middle frontal gyrus. In order to better characterize the nature of scene information represented in these brain areas, we examined how closely error patterns from the decoding analysis were related to the errors from behavioral judgments on sound categories or the errors of a classifier based on the frequency spectra as they are represented in the cochlea. A multilevel regression analysis showed that the error patterns in A1 are predicted by the errors of behavioral categorization, but not by the errors reflecting sound structure. This pattern of results was only found in the posterior parts of A1 but not in the more anterior fourth and fifth sub-divisions of A1, suggesting that each sub-division of A1 may be involved in different

modules in auditory processing (Norman-Haignere, Kanwisher, & McDermott, 2013). Based on these findings, we suggest that a brain network from primary auditory cortex to pre-frontal cortex contributes to human categorization on auditory information from natural environments.

5. The role of graphical fidelity on learning within a virtual learning environment

Zain Khan, & Bill Kapralos, University of Ontario Institute of Technology, Canada

Problem: Despite the great computing hardware currently available, real-time high fidelity rendering of complex virtual environments across all modalities is still not feasible even with today's computational resources. Striving for high fidelity environments (as typically done), can burden our computational resources (particularly when considering portable computing devices), increase the probability of lag and subsequent simulator sickness. What effect does the fidelity of both the audio and graphics components of a virtual simulation have on learning?

Purpose: We have developed a serious game for medical-based cultural competence training intended for medical students, practitioners, and professionals. The goal is for the trainee to successfully complete a specific scenario (dialogue script), focusing directly on the mood and the cultural background of the virtual patient (VP). The serious game includes a dialogue authoring system and scenario editor that can be used to not only facilitate medical-based scenarios, but a variety of non-medical scenarios including hostage negotiation, interview preparation, and language learning systems.

Method: In addition to its use as a training tool, the serious game provides us the ability to alter various parameters of a simulation including graphical fidelity thus allowing it to be used a test-bed to examine the effect of fidelity and audio-visual interactions on learning. Within the scope of this work, are examining the use of low fidelity visuals (graphics) and their effect on learning.

Conclusion: We have developed a serious game that facilitates the dialogue between one or more players (taking on the role of medical professionals) and one or more VPs (representing the patient and possibly loved ones). The serious game will allow us to examine how different interactions can help identify and resolve communication barriers. The serious game allows various simulation parameters to be easily adjusted thus allowing us to methodically examine the effect of fidelity and multi-modal interactions on learning.

6. Representation of statistical charts and graphs without the use of visual perception

Damon Pfaff, Ambrose Li, Michael Carnevale, & Peter Coppin
Perceptual Artifacts Lab - Accessible Graphics Initiative, OCAD University, Canada

Blind or low vision individuals who require access to statistical learning materials, such as textbooks, do not have access to the graphical visualizations of statistical distributions within. This poster will present an exploratory prototype in the sonified representation of a statistical normal (Gaussian) distribution. The prototype investigates the use of sonic cues to convey relevant aspects of this graphical representation to a user without the use of visual perception. Sonic qualities such as volume, multichannel pan, reverb, equalization, and pitch are modulated in order to contribute to the sonic localization of a simulated sound source in four dimensions. This localization effect is employed to communicate—through the everyday capabilities that humans use to locate themselves and the objects around them in space and time—the pictorial qualities of statistical graphs. This work

differs from other efforts in that it does not make use of text-to-speech in order describe visual topological relationships, nor does it use the analogy of rising and falling tones to communicate the height of a graphical line. As a result, these analogical cues are left free to convey other types of information such as colour or multiple data sets.

7. Asymmetrical medial geniculate nucleus body in people with one eye

Stefania S. Moro[1,2,4], Krista R. Kelly[5], Larissa McKetton[2,3] , & Jennifer K.E. Steeves[1,2,4]

[1] Department of Psychology, [2] Centre for Vision Research, [3] Department of Biology, York University, Canada, [4]The Hospital for Sick Children, Canada, [5]Retina Foundation of the Southwest, USA

Structurally, people with one eye have decreased lateral geniculate nuclei volume (LGN; thalamic visual relay station) that is less severe in the LGN contralateral to the remaining eye, indicating altered structural development. The medial geniculate body (MGB; thalamic auditory relay station) plays a central role in auditory processing with both efferent and afferent tracts to primary auditory cortex. We investigated whether structural MGB changes are also present. MGB volumes were measured in adults who had undergone early unilateral eye enucleation and were compared to binocularly intact controls using the current gold standard methodology for anatomical localization of the MGB. Unlike controls, people with one eye had a significant asymmetry with a larger MGB volume in the left compared to the right hemisphere, independent of eye of enucleation. The volume asymmetry in the MGB in people with one eye may represent increased interaction between the left MGB and primary auditory cortex as compensation for the loss of one half of the visual inputs early in life.

8. Does gaze position attract or shift perceived touch location

Lisa M Pritchett, & Laurence R Harris, Centre for Vision Research, York University, Canada

We have previously shown that perceived location of touch on the torso is affected by gaze position. We put forward a model that could explain why touch localization is shifted in opposite directions by gaze depending on whether touch is coded relative to the orientation of gaze or of the body. Such models predict that all touches are coded in one reference frame shift by an amount proportional to gaze eccentricity. Here, an alternative model is considered where gaze position acts as an attractor for the perceived position of a touch. Nine participants reported the perceived locations of eight vibrotactile stimuli arranged across the front of the torso. Vibrations were delivered while gaze was directed at one of seven locations between ±45°. Before reporting perceived location, participants returned their gaze to center. Three response methods were used: a visual method (reporting the location on a line), a numerical method (reporting a number for the part of skin stimulated), and a motor method (pointing to the perceived location). The effect of gaze on perceived touch location depended on the response method used, the distance between gaze and touch location, and the location of the touch on the body. These results test predictions of our previous model and will have implications for our understanding of the reference frames and mechanisms used for coding touch location.

9. Space-fixed, retina-fixed, and frame-independent mechanisms of trans-saccadic feature integration: An fMRIa paradigm

B.-R. Baltaretu[1, 2, 3], B. T. Dunkley[6], S. Monaco[7], Y. Chen[1, 2, 4], & J. D. Crawford[1, 2,3,4,5]
[1] Centre for Vision Research, [2] Canadian Action and Perception Network (CAPnet), [3] Department of Biology, and Neuroscience Graduate Diploma Program, [4] Department of Kinesiology, [5] Departments of Psychology, Biology, and Kinesiology and Health Sciences, and Neuroscience Graduate Diploma Program, York University, Canada, [6] Department of Diagnostic Imaging, Hospital for Sick Children, Canada, [7] Center for Mind/Brain Sciences, University of Trento, Italy

To date, the neural mechanisms of feature information integration across saccades, also known as trans-saccadic integration (TSI), of low-level object features are relatively unknown. Using fMRI adaptation (fMRIa), we found that the right inferior parietal lobule (IPL; specifically, SMG) and extrastriate cortex (putative V4) are sensitive to stimulus orientation in a space-fixed reference frame (Dunkley & Crawford, Society for Neuroscience Abstracts, 2012). To identify the neural mechanisms of underlying TSI in multiple reference frames, we employed fMRIa to probe three spatial conditions: 1) Space-fixed, 2) Retina-fixed and 3) Frame-independent (neither Space-fixed, nor Retina-fixed). Functional data were collected across 12 participants while they observed an obliquely oriented grating (45° or 135°), followed by a grating at the same ('Repeat' condition) or different angle ('Novel' condition). Participants were instructed to decide via 2AFC if the subsequent grating was repeated or novel. Repeat vs. Novel contrasts showed repetition suppression (RS) and enhancement (RE). RS showed condition-specific patterns within a parieto-frontal network. Distinct areas of activation were identified for the three conditions (i.e., SMG for Condition 1; middle and inferior frontal gyri (MFG, IFG) for Condition 2; and FEF and area 7 for Condition 3) as well as common clusters (i.e., posterior middle intraparietal sulcus, M1and pre-supplementary area). RE was observed in occipitotemporal areas. Specifically, RE in Condition 1 was observed in lateral occipitotemporal gyrus (LOtG) in the left hemisphere. RE in Condition 2 was not observed. In Condition 3, RE was found in LOtG in the right hemisphere. Overall, TSI of orientation activated different cortical patterns (with some parietal overlap) in the three frames. Specifically, suppression occurred in a 'cognitive-sensorimotor,' parieto-frontal network, whereas enhancement occurred in occipitotemporal regions.

10. Action-induced updating of visual working memory

Anna Heuer[1], J. Douglas Crawford[2,3,4] & Anna Schubö[1]
[1] Experimental and Biological Psychology, Philipps-Universität Marburg, Germany, [2] Centre for Vision Research, [3] Department of Psychology, [4] Canadian Action and Perception Network (CAPnet), York University, Canada

Although it seems highly intuitive that what we are doing modulates which visual information from our environment is relevant and should be maintained in visual working memory (VWM), action-induced effects on VWM maintenance have not yet been systematically investigated. It has been shown that attention is automatically drawn towards the goal of a movement, and that the deployment of attention towards representations held in VWM improves memory performance for the respective items. We combined these two insights and tested whether memory items previously presented at an action-relevant location benefit from the action-related stronger attentional engagement at that location in a similar manner as when attention is explicitly cued to be deployed to certain representations. During the retention interval of a VWM task, participants

performed a pointing movement to one of the locations at which the memory items had previously been presented. Indeed, memory performance for items presented at the location of the pointing goal was better than for items presented at action-irrelevant locations. Importantly, this was only observed when participants actually pointed towards the location at which the memory item had been presented, but not when the movement was performed towards fixation. These findings indicate that our actions contribute to the flexible updating of VWM: Information which is potentially action-relevant, simply due to the spatial correspondence between the respective memory representation and our action goal, is preferentially maintained.

11. A computational model to study spatial updating of remembered visual targets across eye movements.

Yalda Mohsenzadeh[1], & Douglas Crawford[1,2]

[1] Centre for Vision Research, Canadian Action and Perception Network and NSERC CAN-ACT CREATE Program, [2] Neuroscience Graduate Diploma Program and Departments of Biology, Psychology, and Kinesiology & Health Sciences, York University, Canada

In the oculomotor system, spatial updating is the ability to aim a saccade toward a remembered visual target position despite intervening motion of the eyes. Although efforts have been made to discover the mechanism underlying spatial updating in humans and animals, there are still many unanswered questions about the neuronal mechanism of this phenomenon. We propose a state space model (SSM) for updating target-related spatial information in gaze-centered coordinates. We considered three types of input in our proposed model: 1) an efference copy signal, inspired by motor burst signal in SC, 2) an eye position signal, found in many LIP, VIP, MT and MST neurons and 3) 2D visual topographic maps of visual stimuli, located in SC. To model the internal neuronal behaviour of the system, we developed a radial basis function (RBF) neural network which can be trained with sequences of input-outputs using a version of the Kalman filter known as Extended Kalman filter. This neural network represents the state space from which we can obtain a topographic map of the remembered target in the hidden layer. Finally, the output of our proposed model is the decoded location of the remembered target. In order to find the hidden states in our model, we employed Kalman filtering approach.

To explore the internal mechanism underlying updating process, we trained this model on a double step saccade-saccade or pursuit-saccade task. After training, the receptive fields of state-space units replicated both predictive remapping during saccades (Duhamel et al. *Science* 1992) and continuous eye-centered updating during smooth pursuit (Dash et al. *Current Biology*, in press). In addition, during trans-saccadic remapping, receptive fields also expanded (to our knowledge, this predicted expansion has not yet been reported in the published literature.)

12. Cortical substrates for allocentric vs. egocentric representation of remembered saccade targets in humans

Y. Chen[1], & J. D. Crawford[1, 2]
[1]School of Kinesiology and Health Science, Centre for Vision Research, Canadian Action and Perception Network, [2]Deparments of Psychology and Biology, York University, Canada

The location of a remembered target can be defined in egocentric or allocentric reference frames, but the neural mechanisms for allocentric saccade coding in humans are essentially unknown. Here we employed an event-related fMRI design same as our recent reach study (Chen et al. Journal of Neuroscience 2014) to investigate the brain areas supporting these two types of representation in twelve participants. The target and the landmark were always presented briefly, but at the beginning of each trial, participants were instructed to ignore the landmark and remember target location (Ego) or remember target location relative to the landmark (Allo). During the delay phase participants had to remember the target location in the appropriate reference frame. In a non-spatial Control participants remembered and reported the target color. We found that during the delay phase Ego and Allo elicited higher activation as compared to the Control in left precuneus, bilateral midposterior intraparietal sulcus and frontal eye field. Inferior parietal lobes showed higher activation for Ego vs. Allo, whereas temporal and occipital cortex showed higher activation for Allo vs. Ego. Egocentric directional selectivity was observed in superior and inferior occipital cortex (IOG). Allocentric directional selectivity was observed in calcarine, IOG and precuneus. These results confirm different cortical mechanisms for egocentric vs. allocentric target memory, but comparing this to our previous study, the detailed mechanisms also depend on the motor effector (eye vs. hand).

13. Electrophysiological investigation of human hippocampal-neocortical networks associated with goal-directed scene search

Omid Talakoub[1,2,3], Taufik Valiante[4, 5], & Kari L Hoffman[1,2,3]
[1]Centre for Vision Research, [2]Department of Psychology, [3]Department of Biology, York University, Canada, [3]Division of Fundamental Neurobiology, Toronto Western Hospital Research Institute, Canada, [4]Division of Neurosurgery, Department of Surgery, University of Toronto, Canada

Vision is the predominant modality of exploration in primates. The way we visually scan the environment during exploration changes as a function of learning and experience. Whereas the hippocampus plays a critical role in learning about and remembering visual scenes, the neural mechanisms that support this role are still unclear. Furthermore, guidance of search based on memory is a type of top-down influence on search, leading to the prediction that brain regions involved in goal-directed visual search may also show neural correlates of this top-down influence. We recorded the field potentials at hippocampal and neocortical sites in individuals with medially refractory epilepsy who underwent surgical implantation of subdural surface electrodes and depth macroelectrodes to localize epileptogenic regions. Participants searched for a changing object -"target"- embedded within a natural scene while their gaze was tracked. Target detection was faster for repeated presentations in which targets were remembered than for the first presentation or for repeated presentations in which targets were forgotten, indicating a correspondence between detection times and explicit recall. Next, we studied and qualitatively compared memory-related neural correlates. Hippocampal and several neocortical sites showed changes in power in two frequency bands during search:

alpha/beta (8-20 Hz) and gamma (30-150 Hz). These task-related changes were both site- and memory-dependent. For example, alpha/beta power decreased in the hippocampus and temporal lobe during the visual spatial exploration, whereas alpha/beta power was enhanced during search in the medial-frontal regions. In hippocampal and specific cortical contacts, we observed higher gamma power after finding a novel target but also in the time preceding detection of remembered targets, i.e., the periods attributed to memory encoding and retrieval, respectively. Our results suggest that gamma-band power may index involvement and possible interactions of hippocampus and neocortical regions in a memory-guided visual search task. In addition, local 30-150 Hz gamma power may be inversely related to 8-20 Hz alpha/beta during top-down guidance of vision.

14. Inter-network interactions: Impact of connections between oscillatory neuronal networks on oscillation frequency and pattern

Oscar J. Avella Gonzalez[1,2], Karlijn I. van Aerde[1,3], Huibert D. Mansvelder[1], Jaap van Pelt[1], & Arjen van Ooyen[1]
[1]Department of Integrative Neurophysiology, CNCR VU University Amsterdam, The Netherlands
[2]Centre for Vision Research, York University, Canada
[3]Department of Translational Neuroscience, UMC Utrecht, The Netherlands

Oscillations in electrical activity are a characteristic feature of many brain networks including the visual cortex, and display a large variety of temporal patterns. In a certain area, oscillations may vary in frequency and power or even multiple oscillations may be alternatively present (Fig. 1). However, the origin of this complex repertoire of activity remains unclear. Interactions between oscillatory networks may contribute, but the effects of these interactions are poorly known. Here we present a model of the interaction between cortical subnetworks, that sheds some light on the dynamical structure of the resulting oscillations.

15. Differential synchronization of alpha peak and beta power during motor imagery of a newly learned dance with eyes open versus closed: Novel evidence for genre specificity

Paula Di Nota[1,2,3], Rodrigo Montefusco-Siegmund[1,2], & Joseph F.X. DeSouza[1,2,3,4]
[1]Centre for Vision Research, [2]Department of Psychology, [3]Neuroscience Graduate Diploma Program, [4]Canadian Action and Perception Network (CAPNet), York University, Canada

Increased alpha (8-13Hz) power, known as event-related synchronization (ERS), in is found when eyes are closed and during inhibition of task-irrelevant brain areas, while ERS of beta (13-20Hz) is observed in task-relevant brain areas representing cognitive processing. The aim of the present study was to examine individual alpha peak and beta power during learning and kinesthetic motor imagery (KMI) of a novel ballet dance sequence among high (n=25 ballet dancers), intermediate (n=21 dancers from other genres), and low (n=15 non-dancers) familiarity groups. Additionally, subjects performed KMI with eyes open (EO) and eyes closed (EC) to reveal possible changes in power by eye state. Ballet dancers showed significantly higher alpha peak frequencies during observational learning ($P<.05$), indicating faster oscillatory processing of complex, naturalistic dance stimuli among familiar experts. Log alpha peak power was highest in the intermediate group relative to non-dancers ($P<.01$), and log beta power was highest in ballet dancers during learning ($P<.000$), reflecting greater recruitment of visuomotor

networks with increased familiarity to the observed motor repertoire. Relative alpha event-related desynchronization (ERD) was observed during KMI-EC in parietal and occipital sites (P<.05) while KMI-EO elicited relative ERS (P<.000). Similar effects were observed for beta, with EC conditions eliciting higher power than EO during both tasks (Baseline: P<.000, KMI: P<.05). However, a significant four-way interaction reveals higher beta power during KMI-EO relative to KMI-EC in ballet dancers only in bilateral temporal sites (P<.01), putatively reflecting greater recruitment of regions involved in action simulation during effortful processing of KMI-EO, including superior temporal cortex. Together, these results demonstrate a novel pattern of ERS/D in the alpha and beta frequency bands during observational learning and KMI of a complex dance sequence among highly familiar experts, and as manipulated by eye state.

16. Entrainment of movement kinematics as the default mode of human joint actions: A cooperative grasping study

Lin Guo[1], & Matthias Niemeier[1, 2]
[1]University of Toronto Scarborough, [2]Centre for Vision Research, York University, Canada

Humans can be highly effective in coordinating their actions with others. To perform such "joint actions," the co-actors need to align temporal and spatial aspects of their motor control with great precision. But to study interpersonal coordination most research has investigated rhythmic limb movements during socially coordinated behaviour with many degrees of freedom in the way the tasks could be performed, and largely the tasks could be performed by one person alone. To probe joint actions affording a maximum amount of cooperation, here we used a task that was irreducible to a single person's actions with low degrees of freedom. We asked two participants at a time to cooperatively grab an object with a precision grasp such that each person used their index and middle fingers of one hand to push from opposite sides of the object and to lift it. As a control condition, we asked the participants to compete in speed as they reached for the opposite sides of the object. We hypothesized that cooperative grasping should yield higher degrees of coordination than competitive reaching, in particular over time. Using infrared motion capture, we recorded the hand movements of 11 pairs of participants. As expected we found that movement coordination for cooperation vs. competition became different over the course of the experiment, i.e., towards the end, hand velocities were more correlated during cooperative grasping. However, the change came from the competitive condition where correlations started high and declined over time, whereas the collaborative condition showed high correlations throughout. Our data suggest that joint actions entrain people's movement kinematics into temporal synchrony, arguably based on forward models that anticipate the movements of the co-actor. Moreover, during competition actors might need to learn to de-synchronize to overcome the default mode of cooperation.

17. Separating predicted and perceived sensory consequences of motor learning
Holly A.Clayton, Bernard Marius't Hart & Denise P. Henriques
Centre for Vision Research, York University, Canada
During motor adaptation the discrepancy between expected and actually perceived sensory feedback is thought to be minimized, but it can be difficult to measure predictions of the sensory consequences of one's actions. Studies attempting to measure changes in predicted sensory consequences (Synofzik et al (2008); Izawa et al (2012) have found that self-directed, unseen hand movements are mis-localized in the direction of the altered visual feedback. However, our lab has shown that such learning also leads to changes in perceptual estimates of static hand position when the target hand is guided or passively displaced. We attribute these changes to a recalibration of hand proprioception; in the absence of a volitional hand movement, efferent (or predictive) signals are less likely to be involved. The goal here is to test the extent by which changes in hand localization reflect a change in the predicted sensory consequences, or a change in the perceived (proprioceptive) consequences. We did this by comparing changes in localization produced when the hand movement was self-generated ('active localization') versus robot-generated ('passive localization') to the same locations following visuomotor adaptation to a rotated cursor. In this passive version, there should be no predicted consequences of these robot-generated hand movements. We found that, although changes in localization were somewhat larger in active localization, the passive localization task also elicited substantial changes. Our results suggest that the change in localization following visuomotor adaptation may not be based entirely on updating predicted sensory consequences, but also largely reflects changes in our proprioceptive state estimate.

18. Concurrent reach and tracking adaptations of moving and static targets
Maria N. Ayala, Priyanka Sharma, & Denise Henriques
Centre for Vision Research, York University, Canada
Given that the neural networks and behavioural parameters subserving saccadic and smooth pursuit eye movements are independent of one another, we wanted to explore whether a similar analogy exists for ballistic reaching and tracking arm movements. Does adaptation to perturbed tracking movements generalize to that of ballistic reaching movements? In the following experiments, we explored whether training by tracking a moving target with a perturbed hand-cursor produces motor aftereffects (AE) and if these AE differ from those produced in a typical perturbed ballistic reaching task with a static target. Adaptation to perturbed tracking movements produce significant reach AE although to a smaller extent; tracking AE were about half the size (on average 9°) of those produced after ballistic reach training (on average 19°). Additionally, we looked at whether neural processing of adaptation to tracking and reaching paradigms are independent of one another and would thus allow for concurrent adaptation to opposing perturbations (i.e. dual adaptation). Tracking trials were associated with a 30° CCW rotation while reach trials were associated with a 30° CW rotation. The 'single' perturbation groups adapted to either a CW or CCW perturbation while the 'dual' group experienced both perturbations concurrently. We found significant AE following dual training of about 7°, which was substantially smaller than that produced when reach training was not concurrent with tracking training. The size of reduction of AE is consistent with the extent of the interference from tracking training as measured by the reach AE produced when only that condition was performed. Additionally, both tracking and reaching performance in response to visuomotor rotation significantly improved across training for both single and

dual tracking groups, with tracking errors saturating at the same level but with different learning curves, with only the single group fully returning to baseline levels.

19. Time course of reach adaptation and proprioceptive recalibration during visuomotor training

Jennifer E. Ruttle[1,2], Erin K. Cressman[4] , & Denise Y. P. Henriques[1-3]
[1]Centre for Vision Research, [2]Department of Psychology, [3]School of Kinesiology and Health Science, York University, [4]School of Human Kinetics, University of Ottawa, Canada

Training to reach with rotated hand-cursor feedback, results in changes or adaptation of these hand movements during training, which continues to persist when the perturbation is removed (reach aftereffects), as well as changes in felt hand position, which we refer to as proprioceptive recalibration. The rate by which motor and proprioceptive changes arise throughout training is unknown. Here, we aim to determine the timescale of these changes and their relationship in order to gain insight into the processes that may be involved in learning. We measured reach aftereffects (no-cursor reaches) and perceived hand position after every 6 reach-training trials with a 30° rotated-cursor to 3 radially located targets. To assess proprioceptive recalibration, the right adapted hand was passively moved to one of the three target sites by a robot, and its perceived location was indicated by the left untrained hand. Participants trained with both a clockwise and a counter-clockwise cursor rotation in sessions a week apart to determine if the original training led to any retention or interference of these motor and sensory changes. Results suggest that both motor and proprioceptive recalibration occurred simultaneously and immediately after only 6 or 12 rotated-cursor training trials (13.08° & 4.50° respectively) and did not increase much or at all with further training. Moreover, there was no retention or interference present one week after training. This suggests that the implicit changes in both motor and sensory systems do not simply reflect changes in performance made with a cursor during training.

20. Perceptual organization of web pages: A study for assistive technology

Michael Cormier[1], Richard Mann[1], Robin Cohen[1], & Karyn Moffatt[2]
[1]Computer Science, University of Waterloo, [2]Information Sciences, McGill University, Canada

As the Internet has become a vital part of everyday life, ensuring that web page content is accessible has become correspondingly important, for users with a variety of assistive needs. Systems designed to achieve this goal generally use the structure of the source code to infer the semantic structure of the webpage's content, in order to coherently present this information to the user. We argue, however, that the appropriate representation from which to infer this semantic structure is the rendered page, analyzed using vision-based techniques. The visual appearance of the page is designed to convey the semantic structure of its content to the user, whereas the source code is a designer's effort to produce the desired appearance effectively through a browser's rendering engine.
We present progress toward a “back end” system for interpreting the semantic structure of the page contents using an image of the rendered page alone and exploiting assumptions about web page design. This system will provide rich information about page structure to the “front end” system, which will present the page to the user with appropriate assistive adaptations. First, our system will produce a hierarchical segmentation of an image of the page using an edge-based segmentation algorithm developed specifically for this context.

It will then classify regions according to their semantic role using a hidden Markov tree with hidden states representing region classes, observations representing features of each region, and a large-scale structure based on the segmentation tree.
A rendered web page is a designed image produced as an abstract representation of information rather than a natural scene produced by perspective projection, but one designed to be interpreted by a human rather than by an algorithm. This is therefore a restricted but nontrivial domain for determining object boundaries and for interpreting perceptual organization, as vision research.

21. Persistent contour integration by the association field

Zhiheng Zhou, & Lars Strother, Department of Psychology, University of Nevada Reno, USA

Contour integration relies on multiple neural mechanisms. Here we introduce a method of studying contour integration that involves the perceptual fading of a previously visible contour despite an unchanging stimulus. Observers viewed contours that were visible upon onset but became camouflaged by the background within a few seconds, even though the elements comprising the contours remained superimposed on the background. Previous studies found that the duration of this fading depends on global form properties, including familiarity (Strother et al., 2011). Here we were primarily interested in the prospective influence of inter-element orientation, which determines contour smoothness. As in a previous study (Strother & Alferov, 2014), we found that smooth contours were slower to fade than contours comprised of increasingly randomly oriented elements. We propose that this implicates the 'association field' of contour integration (Field et al. 1993). Further, we conducted an addition contour detection experiment to test whether or not individuals' sensitivities to contour smoothness were correlated (within observers) between the two paradigms (fading and detection)—we found that sensitivities, but not bias, were strongly correlated. The results of our experiments suggest that (a) neural mechanisms related to stimulus familiarity interact with the association field in visual cortex; and (b) the contour fading paradigm used here is a promising means of studying to the neural basis of contour integration, and the interplay between top-down and bottom-up neural processes.

22. On the interaction between 2-D and 3-D grouping cues in determining depth from disparity

Lesley M. Deas, & Laurie M. Wilcox, Center for Vision Research, York University, Canada

The amount of depth perceived between a vertical line pair is markedly and consistently reduced when horizontal lines connect the pair to form a closed object (Deas & Wilcox, 2014). We have shown that this phenomenon reflects the operation of Gestalt grouping principles, specifically perceptual closure. Here we systematically assess the impact of specific 2-D cues and properties - including connectedness, collinearity and proximity – to understand their role in perceptual closure in depth.
Quantitative depth magnitude was measured for a pair of vertical lines in which the object interpretation was manipulated. The baseline comparison consisted of a set of four isolated vertical lines contrasted with a 'closed object' version whereby the central pair was connected by horizontal lines. In subsequent conditions, the closed object was manipulated by i) introducing gaps (0.5 to 1.5deg) at different locations along the horizontal contours, ii) altering the position of the horizontal lines between the vertical test

lines. Note that in all conditions the properties of the central vertical test lines were not changed. We also assessed the role of a new stereoscopic grouping principle, which we call 'good stereoscopic continuation'.
Our results show that perceptual closure in depth does not require a physical connection between components, but critically depends on good continuation (collinearity) and corners (L-junctions). Moreover, good stereoscopic continuation is necessary to finalize/maintain grouping. When the inducing lines satisfy both 2-D and 3-D constraints, their relative separation in depth is dramatically degraded; however, if either set of constraints are not satisfied, then grouping does not occur and estimates are restored to the level of isolated lines. Taken together, our results highlight the interaction between 2-D and 3-D grouping cues in stereoscopic object formation.

23. Stereoscopic surface interpolation from illusory contours
Brittney A. Hartle, Richard F. Murray, & Laurie M. Wilcox
Centre for Vision Research, York University, Canada
Disparity interpolation is a method of stereoscopic surface interpretation that involves estimating the shape and position of regions without unique disparity signals by assigning depth values based on the explicit disparity of neighbouring elements. Stereoscopic Kanizsa figures are a unique case of three-dimensional (3D) illusory surface interpolation, where the corresponding disparity signals at the inducing elements appear to define the shape of the illusory surface across a region of uniform luminance. To investigate how the disparity signal at the inducing elements defines the shape of interpolated 3D contours, we measured the magnitude of depth perceived from a set of Kaniza figures with a range of disparities defined by illusory contours. We compared depth estimates for these stimuli with depth percepts generated by figures with luminance-defined contours using both crossed and uncrossed disparities. Observers estimated the perceived distance between the fronto-parallel plane containing the inducing elements and the peak of the stereoscopic surface using a purpose-built haptic sensor strip. We found no difference in the magnitude of perceived depth between illusory and luminance-defined surfaces regardless of the direction of the depth offset. However, in both conditions, depth from uncrossed disparities was consistently underestimated compared to theoretical predictions. We propose that the observed reduction in perceived depth is the result of cue conflict between 2D occlusion information and binocular disparity for interpolated surfaces, irrespective of whether their bounding contours are luminance-defined or illusory.

24. Common onset masking: Does the mechanism depend upon type of mask?
Marwan Daar, & Hugh Wilson, Centre for Vision Research, York University, Canada
Object substitution masking (OSM) is typically studied in peripheral visual field using a common onset four dot mask. Here, we first establish common onset four dot masking in central visual field. Next, we compare common onset masking functions (performance as a function of mask duration) using a four dot mask and a contour mask, in order to assess whether these two mask types operate through similar mechanisms. Our data suggest that a dissociation of masking mechanism for these two mask types.

25. Attention deployment in visual contexts of varying homogeneity

Anna Schubö, & Tobias Feldmann-Wüstefeld, Philipps University Marburg, Germany

The degree of efficiency with which observers find a target object in a visual scene depends both on the features of the target and on the structure of the background elements. Dissimilarity between target object and surrounding context elements usually increases search efficiency, and so does similarity between the individual context elements. In several experiments, we investigated the role of context element homogeneity in the deployment of visual attention. We hypothesized that context homogeneity modulates attention guidance in addition to other attentional control mechanisms such as the observer's current intention. We used texture stimuli composed of simple line elements that were arranged in such a way as to form homogeneous or heterogeneous contexts. Observers were to search for and report an orientation target. Behavioral search efficiency (accuracy and response times) as well as the neural processes (such as the N2pc component, an attention-related marker in the event-related brain potential) involved in target and background processing were measured. Results showed that (pre-attentive) context grouping preceded target detection and had a strong impact on the way visual attention was deployed in the scene. Both the observer's intention and the homogeneity of context elements guided visual attention. The relative contribution of top-down and bottom-up processing on attention deployment depended on the homogeneity of the context elements.

26. Fast detection yet slow recognition of emerging images

Barbara Nordhjem[1], Constanza I. Kurman Petrozzelli[1], Nicolás Gravel[1, 3], Remco Renken[2], & Frans W. Cornelissen[1], [1]Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, The Netherlands, [2]BCN Neuroimaging Center, University Medical Center Groningen, University of Groningen, The Netherlands, [3]Laboratorio de Circuitos Neuronales, Centro Interdisciplinario de Neurociencia, Pontificia Universidad Católica de Chile, Chile

Visual object recognition typically happens very fast and it has therefore been difficult to disentangle its constituent processes. Extended recognition times have been observed for images with emergent properties, suggesting these may help examine perceptual and cognitive processes in object recognition. Until now, their use has been constrained by limited availability. Here, we used a new set of 15 stimuli with emergent properties – akin to the famous Gestalt image of a Dalmatian – in combination with eye tracking to examine the processes underlying human object recognition. Recognition times were relatively long (median ~ 5 s) –confirming the objects' emergent properties. Surprisingly, already within the first 500 ms, the majority of fixations were aimed at the hidden object. Contemporary saliency models – that emulate the early feature processing stages of human vision – fail to predict these eye movements. The fast detection suggests that observers selected potentially relevant image sections based on image statistics presently not captured by these saliency models. The quick detection yet slow recognition of emergent images points towards discrete contributions of perceptual and cognitive processes to human object recognition. This may eventually aid the development of better biologically plausible computer vision algorithms.

27. Visual distortions induced by simple and complex shapes

Galina Goren, & James H. Elder, Centre for Vision Research, York University, Canada

Some theories of shape posit that shape can be neurally represented by deformation processes. Motivated by these theories, we recently discovered that complex natural contours induce perceptual distortions in the surrounding space (Goren & Elder VSS2013). In a follow up study, we found that simpler shapes (lines and circles) also induce distortions (Goren & Elder CSBBCS2014). However, the exact nature and magnitude of these distortions remains unclear. Here, we quantitatively assess the size of these distortions as well as their direction relative to the shape.

Methods. Observers were presented with a triplet of collinear points in proximity and aligned orthogonally to a contour shape. Observers used a mouse to move the central dot so that it appears to bisect the distance between the two outer dots. Systematic displacement of the central dot from the point of bisection indicates a local deformation of perceptual space. In a control condition, observers made the same judgment in the absence of the contour shape.

Results. We found that perceptual space was expanded near natural, complex shapes and compressed near simple shapes. We discuss possible explanations for this interesting pattern of results.

28. Psychophysical evaluation of planar shape representations for object recognition

Ingo Fründ, & James H. Elder, Center for Vision Research, York University, Canada

Intermediate areas of the object pathway appear to represent shape in terms of features of moderate complexity, however the precise nature of this distributed code remains unclear. Here we use a novel method to evaluate the efficiency with which three candidate representations (Fourier Descriptors, Shapelets and Formlets) capture the planar shape information required for humans to reliably recognize objects.

The Fourier Descriptor representation is the Fourier transform of the points defining the object boundary, represented as complex numbers; a good approximation to a shape is attained by truncating this Fourier sequence. Shapelets are a wavelet version of Fourier Descriptors; each component is localized in both frequency and position along the curve, and these are computed by matching pursuit. Formlets represent shape as a series of smooth localized deformations applied to an embryonic shape (an ellipse in our case), also computed using matching pursuit.

We employed a database of 77 animal shapes from 11 categories. In objective terms (Euclidean error), these shapes are more efficiently coded by Fourier Descriptors and Shapelets. To evaluate subjective efficiency, shapes were rendered using each of these three representations; the observer's task was to identify the category of each shape from four alternatives. For each representation, the number of shape components was increased over trials from 1 to 10.

For all 6 observers, Shapelets required fewest components to support threshold performance (mean of 1.8+/-0.4 components), followed by Fourier Descriptors (5.0+/-0.4 components), and finally Formlets (6.5+/-0.7 components). Interestingly, however, recognition rates for the Formlet approximation of shape reached threshold at a lower objective fidelity than for the other codes, suggesting that Formlets, while having lower objective fidelity, selectively capture shape information of high perceptual value.

29. Girl or Boy? Cues to identifying the gender of children from their faces

F. Wilkinson, A. Harrington, L. Lillakas, E. Kelly, J-J Kim, N. Slavat, & Y. Haque
Center for Vision Research, York University, Canada

Adult faces carry information about gender that is easily decoded in the absence of other cues such as posture, clothing or hairstyle. Is this also true of children's faces? If so, what facial cues are most critical in gender determination? We report the findings of two studies using colour photographs of children's faces 7-11 years of age (hair covered; Dartmouth Database of Children's Faces). In Exp 1, a single face was presented on each trial and the participants (N=40) had to classify the face as "girl" or "boy". In Exp 2, one male and one female face of the same age were presented simultaneously, and the participant's task was to indicate either the male face (N=40) or the female face (N=40). Exp 1 revealed a very strong bias in favour of male faces, and very weak evidence of discrimination. Exp 2 revealed clear evidence of gender discrimination. Overall, mean correct performance was 75%; however, discrimination of gender for 11 year old faces was significantly worse than for other ages ($p<0.001$). Surprisingly, performance was significantly better ($p <0.001$) when the task was to choose the female face. Neither the gender of the participant ($p=0.15$) nor the amount of experience with children of this age group ($p=0.45$) was related to performance. Self-report of attended cues suggested that eyes (82%), eyebrows (79%) and mouth (74%) were most influential; however, there was considerable variability in the described details of these features.

30. Conjunctive versus single-feature visual processing relate to autistic traits

Naomi Hazlett, Ryan A. Stevenson, Lok-Kin Yeung, Susanne Ferber, & Morgan D. Barense
Department of Psychology, University of Toronto, Canada

Visual face perception is typically achieved using a conjunctive-processing strategy; individuals do not perceive each individual feature but rather a single, unified face. Individuals also become more likely to use this strategy with non-face visual presentations of higher ambiguity. Individuals with Autism Spectrum Disorders are thought to perceive faces in an atypical manner, and are likewise thought to default towards the details of visual presentations as opposed to the global whole. It is undetermined whether the underlying issue in ASD is with face perception, or with conjunctive processing more generally.

Here, we used an eye-tracking paradigm to test whether individuals with high levels of ASD symptoms differentially used visual conjunctive-processing strategies across two variables; (1) face versus non-face processing, and (2) level of visual ambiguity. If such individuals processed faces differently regardless of ambiguity, face perception is therefore likely distinctly altered in ASD. Conversely, if individuals with high ASD symptomatology showed differences in visual processing regardless of stimuli type, this would suggest that the issues in face perception commonly seen in ASD may be the result of a more generalized difference.

A sample of adults without a clinical ASD diagnosis (n=15) compared novel objects and faces of high and low ambiguity while eye movements were recorded. Afterward, they completed a series of questionnaires measuring ASD-symptomatology, including the Autism Quotient.

Accuracy was matched for each level of ambiguity across stimulus types. A 2x2, repeated-measures ANOVA revealed a main effect of ambiguity, with more high ambiguity stimuli being processed more configurally. Additionally, a stimulus-by-ambiguity interaction was

observed, where configural processing of faces was less impacted by the level of ambiguity. Lastly, a significant correlation between differential configural processing and ASD symptomatology was found. These results indicate that individuals with high ASD symptomatology demonstrate an atypical processing strategy for face stimuli.

31. Stable sensory phenotypes in autism influence ASD symptomology

Karen R. Black[1], Ryan A. Stevenson[1], Busisiwe L. Ncube[2], Magali Segers[2], Sol Z. Sun[1], James Bebko[2], Morgan D. Barense[1], & Susanne Ferber[1], [1] Department of Psychology, University of Toronto, [2] Department of Psychology, York University, Canada

Autism spectrum disorder (ASD) is a neuropsychiatric disorder characterized by difficulties with social communication and restricted and repetitive behaviours. Recent evidence suggests that differences in sensory perception may underlie these symptoms. One such hypothesis suggests that sensory sensitivity (SS) and the tendency to avoid sensory stimulation (AV) impacts the ASD symptom of insistence-on-sameness behaviours (I/S) and related anxiety. Furthermore, 40% of children with ASD suffer from co-morbid anxiety, which has been linked to increased aggression and self-injury. While I/S have been associated with increased anxiety, repetitive motor movements (RMM) do not show this same relationship. To tease apart the relationship between sensory perception, repetitive behaviours, and anxiety, we asked parents of 73 children (ASD, n=39, M_{age}=12.0, SD=3.1; TD, n=34, M_{age}=11.6, SD=3.1) to complete the Child Sensory Profile-2 (CSP-2), the Repetitive Behavior Questionnaire-2 (RBQ-2) and the Spence Children's Anxiety Scale (SCAS). We predicted:

1) I/S would be positively correlated with SS, AV and anxiety symptoms overall and for individual subscales of the SCAS.
2) I/S-sensory-anxiety correlations would be significantly different between ASD and TD groups.

Findings indicated that both SS and AV were positively correlated with I/S but not RMM. Additionally, the relationship between SS/AV and I/S was present only for children with ASD and not their typically developing peers. Furthermore, both SS and AV were positively correlated with the panic subscale of the SCAS, again only for ASD and not TD children. Finally, a mediation analysis revealed that scores on the panic subscale mediated the relationship between SS and I/S. These results suggest that sensory sensitivities may lead to increases in particular types of anxiety, which in turn result in I/S behaviours in children with ASD, thus opening the possibility for future development of targeted interventions which treat underlying sensory sources of Autism symptoms.

32. Charles Bonnet syndrome: A description of a man whose visual hallucinations correspond to his neuropsychological impairment

Josée Rivest[1, 2], & Teodora Dundjerovic[2], [1]Centre for Vision Research, [2]Glendon College, Neuropsychology and Cognitive Health Program, Baycrest, York University, Canada

About 11% of individuals with vision lost (and without psychological disorders) experience visual hallucinations while being aware that their perceptions are unreal: They have Charles Bonnet Syndrome (CBS). This study describes a man (RL: retired lawyer born in 1923) who has CBS following a history of glaucoma (he has low vision in both eyes). The nature of his visual hallucinations is unusual as it corresponds to his cognitive difficulty: RL hallucinates illegible text and has alexia. In 2013, RL started to see columns and/or rows

of illegible text standing on top of any surfaces at which he looks; his hallucinations are particularly prominent when falling asleep. Frustrated by his inability to read normal texts, RL started writing large black print notes so that he could see his written text. To his surprise and dismay, despite writing and seeing his written notes, he was unable to read them. Neuropsychological evaluation (2014) shows that RL's verbal intellectual skills are superior. He has good verbal memory, and his daily functioning is normal (other than having to deal with his low vision). Despite his strong problem solving verbal skills, normal spelling and writing, RL is a letter-by-letter reader (he has alexia without agraphia). We investigated whether the nature of his hallucinations would change after successfully reading printed words (while looking and/or tracing the letters), and when looking at frames tilted at different orientation. The prominence of his hallucinations did not vary with an increased ability to read printed words. The orientation of his hallucinated illegible texts systematically varied with the orientation the frame. It is suggested that his visual hallucinations are related to activation of visual areas coding orientation, and not of areas related to reading. RL's cognitive profile has not changed since 2014; his CBS does not appear to be a symptom of dementia.

33. Differential processing of words and pseudowords: A repetition priming and TMS study

Mikaeel Valli, Lily M. Solomon-Harris, Jennifer K.E. Steeves, & W. Dale Stevens
Centre for Vision Research & Department of Psychology, York University, Canada

The neural mechanisms underpinning reading are contested among researchers and remain elusive despite extensive research. In the present behavioural study, we asked whether words and pseudowords are processed fundamentally differently. Pseudowords are pronounceable, legal letter strings that follow the lexical rules of a language, yet they have no meaning in the lexicon. Thus, while pseudowords are meaningless, they are perceptually indistinguishable from real words, and differences in their processing challenge bottom-up models of reading. We used a repetition priming paradigm, measuring participants' reaction time to judge whether letter strings were words or pseudowords in response to repeated presentations of these stimuli. We observed expected priming effects, as reaction time for both letter string types decreased with repeated presentations. Interestingly, when "orthographic neighbors" (i.e., one letter was changed) of studied words were presented, reaction time was increased, relative to repeated words, as if these stimuli were entirely novel. Conversely, reaction time for orthographic neighbours of the studied pseudowords decreased as if these stimuli were repeated. These findings suggest that whereas pseudowords are processed perceptually, real-words are processed at the conceptual level. Preliminary pilot data were also acquired from two participants to investigate the effects of disrupting processing in the superior temporal gyrus (Wernicke's area) with transcranial magnetic stimulation (TMS). These behavioural findings serve as a pilot study for a larger neuroimaging project currently under development.

34. Pattern separation deficits in a patient with bilateral dentate gyrus lesions

Stevenson Baker[1], & R. Shayna Rosenbaum[1, 2], [1]York University, [2]Baycrest, York University, Canada

Neuroimaging and behavioural studies support various theories on the nature of hippocampal contributions to episodic memory, the ability to remember past personal events in one's life that are specific to time and place as well as items that have been previously encountered in a laboratory setting. In order to represent episodic memories as distinct, a mechanism of pattern separation is needed to reduce interference among similar neural inputs by using non-overlapping representations. In recent years, computational models and animal studies have prompted speculation that pattern separation can be localized to the dentate gyrus (DG) and CA3 subfields at the core of the hippocampus. A challenge with studying pattern separation in humans, however, is the fact that individuals with hippocampal damage typically have lesions that extend into and beyond the hippocampus, making it difficult to localize pattern separation to the DG. In the current study, we investigated pattern separation in a rare person, B.L., who, in relation to electrical injury and cardiac arrest, suffered bilateral hippocampal damage localized to the dentate gyrus. B.L.'s performance was compared to that of age-matched controls on a widely used pattern separation task. Results of impaired pattern separation provide the first evidence in humans of a selective role of the DG in creating and maintaining distinct visual memories.

35. TMS to object preferential LO and scene preferential TOS reveals modulation effects to remote regions in the network

Sara A. Rafique, Lily M. Solomon-Harris, & Jennifer K.E. Steeves
Centre for Vision Research & Department of Psychology, York University, Canada

Viewing the world involves many computations across a great number of regions of the brain. We sought to determine the connectivity of object and scene processing regions of cortex through the influence of transient interruption of discrete nodes within these networks. We consecutively paired repetitive transcranial magnetic stimulation (rTMS) with functional magnetic resonance-adaptation (fMR-A) to measure the effect of rTMS on functional response properties at the stimulation site and in remote regions. In separate sessions, rTMS was applied to the object preferential lateral occipital region (LO) and scene preferential transverse occipital sulcus (TOS). Pre- and post-stimulation responses were compared using fMR-A. In addition to modulating activity at the stimulation site, TMS disrupted remote regions revealing inter and intrahemispheric connections between LO, TOS, and the posterior parahippocampal place area (pPPA). Moreover, we show connections from object preferential LO to outside the ventral perception network in parietal and frontal areas.

36. Distracting pixel noise amplifies spatial frequency-specific cross-over of the perceptual bias: A psychophysical and Monte Carlo simulation study

Jiaqing Chen, & Matthias Niemeier, University of Toronto at Scarborough, Canada

Rarely noticed in daily life, attention may prefer the left side of space. Such perceptual biases offer key insights into functions of spatial attention and visual awareness because they complement pathological biases in patients with spatial neglect who become largely unaware of the left side after right-brain damage. Yet there is little comprehensive understanding of these normal and pathological biases and how they relate to other perceptual/ attentional functions. Here we used a grating-scales task (GST) to test

whether leftward biases and their spatial frequency-dependent cross-over interact with attentional mechanisms of distractor removal. We asked healthy participants to make perceptual judgements to capture attentional biases in a high and a low spatial-frequency condition (GST-HI and GST-LO), and we degraded stimuli with distracting pixel noise. We found that with distracting pixel noise, cross-over grew while biases remained positively correlated. Using Monte Carlo simulations we probed the feasibility of three models and conclude that our data can only be explained by two, or more, biasing mechanisms, arguably interacting with each other through interhemispheric competition. Our study sets the stage for a new systematic approach to investigating the visuospatial mechanisms of the right hemisphere.

37. Detection of chromatic and luminance distortions in natural scenes

Ben J. Jennings[1], Karen Wang[1, 2], Samantha Menzies[1, 2], & Frederick A.A. Kingdom[1]
[1]McGill Vision Research, Department of Ophthalmology, McGill University, Canada.
[2]Optometry & Vision Science, University of Waterloo, Canada.

A number of previous studies have measured visual thresholds for detecting spatial distortions applied to images of natural scenes. One of these measured sensitivity to sinusoidal spatial modulations of image scale (Bex, 2010). Here we consider the relative contribution of the chromatic and luminance layers of natural-scene images to sinusoidal scale distortion. We first report that when the distortion was applied to both luminance and chromatic layers, sensitivity did not depend on whether the undistorted comparison image was of the same or of a different scene. Next we report the effects of individually distorting one or other layer. When only the luminance layer was distorted, performance was the same irrespective of whether the chromatic layer was present, absent or phase scrambled, revealing that the chromatic layer in whatever form did not affect sensitivity. However when the chromatic layer was distorted, sensitivity was higher when the luminance layer was intact compared to when absent or phase-scrambled. Moreover, even when the chromatic layer was visibly distorted, it appeared to contribute little to the sense of distortion. We conclude (a) that observers have an in-built sense of how a normal image of a natural scene should appear, and (b) that both the appearance and detection of distortion is mediated predominantly by the luminance as opposed to chromatic layer.

38. Neural decoding of architectural styles from scene-specific brain regions

Heeyoung Choo[1], Bardia Nikrahei[2], Jack Nasar[2], & Dirk Walther[1], [1]Department of Psychology, University of Toronto, Canada, [2]Department of City and Regional Planning, The Ohio State University, USA

The human visual cortex can elicit neural activity patterns that are distinctive for basic-level scene categories (e.g., highways) as well as for superordinate level scene categories (e.g., man-made). Can the human brain also elicit category-specific neural activity patterns for scenes at a subordinate level? Using functional magnetic resonance imaging (fMRI) we recorded brain activity of participants viewing scene categories at a basic level - mountains, pastures, highways, and playgrounds, and scene categories at a subordinate level - buildings in byzantine, renaissance, modern, and deconstructive architectural styles, and buildings designed by four well-known architects of which categorization is likely to be guided more by participants' knowledge than coherent perceptual structure in buildings. Using multi-voxel pattern analysis, we could decode viewed buildings of different architectural styles significantly better than chance from the parahippocampal place area

(PPA), retrosplenial cortex (RSC), occipital place area (OPA), and lateral occipital complex (LOC). Decoding of buildings by different architects was also successful in the PPA and OPA. Consistent with previous findings, we could successfully decode viewed scene categories. Comparison of error pattern between a simple V1 model and decoding from fMRI data showed that categorical information about architectural styles in the PPA cannot be explained solely by high low-level visual similarity. On the other hand, our simple V1 model was able to discriminate between buildings by different architects to some extent. Our results suggest that the PPA can maintain categorical representation at multiple levels, including subordinate categories, such as architectural styles of buildings, relying on complex structural characteristics of scenes rather than low-level visual similarity. Top-down influences such as domain knowledge or geographical familiarity may thus capitalize on these fine-grained categorical activation patterns in the PPA.

39. Gaze patterns are predictive of scene category across line drawings and photographs

Claudia Damiano[1], Jay M Schmidt[2], & Dirk B Walther[1], [1]Department of Psychology, University of Toronto, Canada, [2]Department of Psychology, The Ohio State University, USA

Recent research from our lab has shown that gaze patterns of subjects viewing colour photographs are predictive of scene category. Here we ask if this result extends to grayscale photographs and line drawings. 77 participants viewed grayscale photographs and line drawings of real-world scenes. In a leave-one-subject-out cross validation analysis, viewed scene category was predicted from gaze patterns by computing the best match between the gaze path of the left-out subject in a given trial and a set of fixation density maps (FDMs) that were derived from all other subjects. Gaze patterns are predictive of scene category for photographs (accuracy=33.6%, significantly above chance level of 16.7% with p=3.1·10^-52) and line drawings (accuracy=30.9%, p=4.5·10^-43). Predicting category of line drawings from FDMs of photographs was accurate for 28.5% (p=6.6·10^-39) of the trials, and predicting photographs from FDMs of line drawings for 29.9% (p=5.0·10^-50). However, prediction accuracy across image types was significantly lower than accuracy within image type (p=4.5·10^-11). This pattern of results suggests that gaze patterns between line drawings and photographs are compatible to a limited extent. We investigated the temporal aspect of gaze patterns by restricting the analysis to time bins with a width of 300 ms. In all four cases, prediction accuracy increased for the first few hundred milliseconds, peaked at 600ms after stimulus onset, and then slowly decreased. What are the image properties that drive gaze behavior that is so distinctive of scene category? We have recently shown that localized features, such as contour junctions, are important for scene categorization. Using the line drawings, we are currently investigating whether subjects look at those important features and use them to guide their gaze patterns.

40. Convergence to points on and off the horizon, foreshortening, and overestimation of angles

Wnuczko, M., Singh, K., & Kennedy, J. , University of Toronto, Canada

Observers viewed pictures of a flat ground plain with a z-line (a line extending into depth) intersected by oblique lines. Lines higher on the picture surface depicted obliques farther away on the ground. In one condition, the lines standing for obliques converged to the picture's horizon line. This depicted parallel obliques on the ground. In the other condition, they converged to points off the horizon, depicting obliques that were not parallel. In both conditions, the more foreshortening increased, the more observers overestimated acute angles on the ground between the z-line and the obliques. Observers made very little use of the information about parallel obliques provided by convergence to the horizon line.

41. A dynamic programming approach to line segment detection

Emilio J. Almazan, Ron Tal, & James H. Elder, Centre for Vision Research, York University, Canada

Line segment detection is important in computer vision for a wide range of applications such as image segmentation and 3D reconstruction. Typical methods for line segment detection use grouping heuristics to grow chains of roughly collinear edges. One disadvantage of such methods is the absence of a global constraint and objective function on this growth process. Here we present a novel global method for line segment detection that conditions the search on the extended lines detected by a probabilistic Hough method (Tal & Elder, 2012). The benefits of this approach include: (1) Limiting the search of line segments to the lines identified by peaks in the Hough map reduces the search space to $O\left(k\sqrt{n}\right)$, where k is the number of extended lines and n is the number of pixels in the image; (2) The identified extended lines have a natural order according to their significance in the Hough map, allowing the line segment search to be further limited to the most significant lines; (3) In urban scenes, co-linear line segments are extremely common, as they arise from architectural repetition seen in cladding, tiling, windows, etc. Thus multiple segments may potentially be recovered from the analysis of each line; (4) Each of the extended lines identified by a peak in the Hough map results from careful accumulation of the global evidence for the line, and thus will more accurately identify the (ρ,θ) parameters of the line segment than will a few local edges; (5) Limiting search to a line allows the problem of determining maximum probability segments to be solved exactly, using dynamic programming. We evaluate and compare this novel method with competing methods on the York Urban Database http://www.elderlab.yorku.ca/YorkUrbanDB/.

42. Multiple indoor layout generation using a single image

Ali Baligh Jahromi, & Gunho Sohn, Department of Earth and Space Science and Engineering, York University, Canada

Indoor scene layout generation from a single image is inherently an ambiguous problem. However, most of the man made scenes are comprised of orthogonal planes that provide valuable information for understanding the indoor scenes. Most of the current methods define the scene layout as a single cubic primitive. This domain-specific knowledge is often not valid in many cases of indoor corridors where multiple corridors are linked each other. Here, we aim to address this problem by hypothesizing-verifying multiple cubic primitives representing the indoor scene layout. This method utilizes middle-level perceptual organization, and relies on finding the ground-wall and ceiling-wall boundaries. A comprehensive interpretation of these edge relations is often hindered due to shadows and occlusions. To handle this problem, the proposed method makes use of actual line segments and virtual rays originated from orthogonal vanishing points to aid the creation of a physically valid multi-cubic structure. Many scene layout hypotheses are created which will be evaluated by an edge correspondence objective function to find the best fitting hypothesis to the image. The best layout hypothesis offered with the highest score is then converted to a 3D model. The proposed method is fully automatic and no human intervention is necessary to obtain an approximate 3D reconstruction.

43. Place recognition system for localization of mobile robots

Raghavender Sahdev, Asheer Bachoo, & John K. Tsotsos, Centre for Vision Research, York University, Canada

This poster focuses on development of a Place Recognition and Categorization system for localization of mobile Robots. The Robot learns the places from experience and then recognizes previously observed topological places in known environments and categorizes previously unseen places in new environments. This system has been practically tested with a novel dataset built by us to validate the theoretical results of the proposed system. A HOUP (Histogram of Oriented Uniform Patterns) descriptor has been developed which is used to represent an image and then appropriate classifiers have been used to perform the classification tasks. It is shown in the work that our developed system not only performs well on the existing datasets but also performs remarkably well on the dataset developed by us. The proposed system has been tested with the 2 robots Virtual Me and Pioneer under different illumination conditions.

44. Sequential modeling of building rooftops by integrating airborne LIDAR data and optical imagery

J. Jung, & G. Sohn, Department of Earth and Space Science and Engineering, York University, Canada

As a virtual replica of real world, photorealistic rooftop models have been considered as critical element of urban space modeling for supporting various applications. Recently, emerging geospatial technologies urgently demand advanced methods of rooftop modeling, producing more accurate, cost-effective and large scale virtual city models. However, developing a "universal" intelligent machine enabling the massive generation of highly accurate rooftop models in a fully-automated manner still remains as a challenging task. In particular, difficulties for reconstruction of highly accurate rooftop model are caused by scene complexity, incomplete cue extraction, and sensor dependency. One of promising approaches to address these problems is to combine the modeling cues

detected from multiple sensors, with expectations that the limitations inherited from one sensor can be compensated by the others. Therefore, this paper presents a sequential rooftop modeling method to refine initial rooftop models derived from airborne LiDAR data by integrating it with linear cues retrieved from single imagery. A cue integration between two datasets is facilitated by creating new topological features connecting between the initial model and image lines, with which new model hypotheses (variances to the initial model) are produced. We adopt Minimum Description Length (MDL) principle for competing the model candidates and selecting the optimal model by considering the balanced trade-off between the model closeness and the model complexity. The Monte Carlo Markov Chain (MCMC) coupling with a simulated annealing (SA) is used to solve optimization problem. Our preliminary results demonstrate the image-driven modeling cues can compensate the limitations posed by LiDAR data in rooftop modeling.

45. Kalman filter based railway tracking from mobile LIDAR data

Y. Jwa, & G. Sohn, Department of Geomatics Engineering, York University, Canada

A precise and effective maintenance of railway infrastructure should be guaranteed to find a solution for high operating safety and low maintenance costs. For this, the key operation is to perform the evaluation of as-built condition by analyzing the geometric and topological information of tracks. This study introduced a new method for 3D model reconstruction of railway tracks as a main object in the railway corridor scene using mobile LiDAR data. The proposed approach starts to approximate the orientation of railway track trajectory from the raw data and extract a strip which can be defined as a local search region in the orthogonal direction of the railway orientation. A main track region in which the LiDAR system attached to a train is expected to run on the track is then detected based on the Bayesian decision after extracting possible features such as HOG (Histogram of Oriented Gradients) from the strip. Once the main track region is localized, track components (head, web, and foot) are segmented based on the region growing approach from initial track points detected in the feature extraction step and initial 3D track models are reconstructed using a third-degree polynomial function. Based on the initial modelling result, a potential track region with a various length is predicted and track models are updated using points representing the track head in the Kalman Filter framework. The key aspect is that the proposed approach is able to enhance the efficiency of the railway tracking process by reducing the complexity for detecting track points and reconstructing track models based on the use of the track model previously reconstructed. An evaluation of the proposed method is performed over an urban railway corridor area containing a few railway track pairs.

46. Coregistration of video images with 3D models for UAS localization

Julien Li-Chee-Ming, & Costas Armenakis, Geomatics Engineering, GeoICT Lab, Department of Earth and Space Science and Engineering, York University, Canada

First-person view (FPV) unmanned aerial systems (UAS) are equipped with a forward-looking camera and a transmitter to downlink the video signal wirelessly in real-time to a ground station. FPV gives the pilot a perspective from the UAS's 'cockpit'. This allows the aircraft to be piloted more intuitively than by visual line-of-sight and beyond the pilot's visual range. FPV systems are used solely as a visual aid in remotely piloting the UAS. This work presents a method to further extend the application of this system by estimating the position and orientation, or pose, of the UAS from the FPV video in near real-time as it travels through a known 3D environment. The estimated pose is used to augment the

autopilot's position and orientation. The need for precise navigation is increased in urban missions, where the possibility of crashing is high, as UASs fly at low altitudes among buildings, avoid obstacles, and perform sharp maneuvers. The developed self-localization process is based on the matching of video features to a metric 3D map of the environment. The main steps are: a) feature extraction from the online video frame, b) geometric hashing-based matching of video features with those of the 3D model using synthetic images, and c) estimate the camera pose as a function of the matched features using photogrammetric bundle adjustment. To demonstrate the approach, images were captured from an Aeryon Scout quadcopter's camera. The Scout flew over York University, approximately 40 metres above the ground. The 3D virtual building model of York University's Keele Campus was used as the known environment. The model consists of photorealistic buildings, trees, and terrain. The obtained results show improvement in the accuracy of the estimated camera pose to that of the UAS's autopilot solution, derived from the onboard navigation sensors (single frequency GPS receiver and MEMS-IMU).The findings of this work are especially useful for video-based navigation in GPS-denied environments or in dense-signal multipath environments such as urban canyons.

47. Multi-range context based terrestrial laser scanning data classification

Chao Luo, & Gunho Sohn, Department of Earth and Space Science, York University, Canada

Terrestrial Laser Scanning (TLS) rapidly becomes a primary surveying tool due to its fast acquisition of highly dense three-dimensional point clouds. For fully utilizing its benefits, developing a robust method to classify many objects of interests from huge amounts of laser point clouds is urgently required.

Only relying on appearance features, local classifiers are likely to produce classification error due to ambiguities of appearance features among classes in varying vision conditions. Conditional Random Field (CRF) is a well-known discriminative classifier, which integrates local appearance of the observation (laser point) with spatial interactions among its neighbouring points in classification process. Typical CRFs employ generic label consistency using short-range dependency only, which often causes locality problem. In this research, we present a multi-range and asymmetric Conditional Random Field (CRF) (maCRF), which adopts a priori information of scene-layout compatibility addressing long-range dependency. Long range dependency is a regularity of spatial arrangement between adjacent objects. Scene layout indicates relative location of line segments both in vertical ('above-bellow' relation) and horizontal ('front-behind') directions. To capture directional spatial relation, asymmetric pairwise potential was used. We incorporate multi-range context, short range smoothness constraint, and long range vertical/horizontal scene layout regularity into a probabilistic graphical model, multi-range asymmetric conditional random field (maCRF). To make comparative research, three single range CRF model, srCRF (short range), lrvCRF(long range vertical) and lrhCRF(long range horizontal) were also constructed. We validated maCRF's performance with TLS point clouds acquired from RIEGL LMS-Z390i scanner using cross validation. Experiment results demonstrate that synergetic classification improvement can be achievable by incorporating two CRF models.

48. Dense depth scene reconstruction using RGB-D TOF Kinect V2 sensors
Logan Jeya, & John Zelek, University of Waterloo, Canada
The depth of a given scene contains useful and meaningful information for both complex and simple computer vision related tasks such as object recognition, segmentation, scene understanding and robotic navigation etc. There exists many approaches to constructing a 3d model of the scene. State of art algorithms and methods are often both computationally and monetarily expensive. The 2010 release of the Kinect allowed for relatively cheap, real time depth reconstruction of a scene. It projected IR patterns and made inferences about the depth to model the 3d scene. The flaws with using a structured light approach are the following: narrow field of view, lighting distortions and restricted range. In July of 2014, Microsoft released Kinect V2. This new version uses a Time of Flight (TOF) method to reconstruct the depth scene. Its IR sensor makes it lighting invariant and the wide angle lenses mitigates the narrow field of view. However, it is still impossible to capture and reconstruct the full scene using a single Kinect V2 camera due to occlusion and limited depth scope. This work presents a novel and simple method to reconstruct a 3d scene in its entirety with cheap Kinect V2 sensors. The algorithm will be sufficiently modular that it will be able to combine n number of the sensors to build a rich 3d model. The proposed research uses both 3d and 2d features (SURF) as descriptors and subsequently a Best Bin First KD tree to determine correspondences. A Markov Random Field is then formulated to enforce perceptual features such as edges and colours to avoid pixel collisions and improve accuracy. Early qualitative results demonstrate that the proposed method is both efficient in real time and can densely and richly reconstruct a given 3D scene from 3 sensors.

49. Plenoptic imaging: From inception to state of the art – a review
David Abou Chacra, University of Waterloo, Canada
Plenoptic, or Light Field, images differ from regular images captured with ordinary 2D cameras, in a sense that they record information about the spatial and directional rays of light. This allows the retention of much more information about the scene that would have been lost if using ordinary photographic techniques. The introduction of the Lytro light field camera has allowed researchers and consumers alike to capture plenoptic images by utilizing a simple to use handheld device. This camera employs a lenticular array to capture the surrounding light field. Other approaches to capturing the light field include multi-camera and multiplexing approaches among others. Since its introduction into the consumer market, there have been numerous applications to using plenoptic images. A lot of applications leverage the extra information to easily calculate depth and perform 3D reconstructions of scenes. Other applications include 3D displays, and refocusing a 2D image after its acquisition.

50. Is there a sun?
Andrew J Schofield[1], Paul Rock[1], Peng Sun[1,2] , & Giacomo Mazzilli[1], [1]School of Psychology, University of Birmingham, UK, [2]Department of Psychology, New York University, USA
In the absence of well articulated lighting cues humans are thought to adopt a default assumption that scenes are lit by a directional light source, from above (Ramachandran, Nature, 332, 163-165, 1988). However, some studies have suggested that the default assumption might be for diffuse illumination (Tyler, J. Image Science & Technology, 42, 319-325, 1998). Further, the default assumption is easily overridden by cues to actual

lighting direction (Morgenstern et al, J.Vision, 14(9):15, 1-17, 2011). We will present a summary of our work on this topic as follows. 1) The default illuminant is directed from above but has a strong diffuse component (Schofield et al, Vis Res, 51, 2317-2330, 2011). People see sinusoidal gratings as if they are sinusoidal surfaces, but the phase relationship between luminance and perceived surface peaks varies with grating orientation. This would not happen under purely punctuate or diffuse lighting but can be easily modeled with a mixed light source assumption. 2) Low level stimulus cues influence the lighting assumption even in highly reduced stimuli (Sun & Schofield, J.Vision, 12(1):12, 2012). The perceived shape of sine, square and ramp gratings depends upon the number of visible grating cycles and edges. Both directional and diffuse lighting interpretations are revealed depending on the features present. 3) Only local cues to actual light sources influence peoples' lighting assumption. We illuminated a complex scene with two, directional, localized light sources and had observers make shape estimates at various locations in the scene. When the probed location was surrounded by objects giving strong support to the illumination, shape interpretations were biased away from each observer's default illumination assumption towards the locally active light source. When the surrounding objects were removed observers reverted to their default illumination assumption even when more distal objects provided support for an alternative illumination profile.

51. Classification images reveal that local grouping within lighting frameworks drives the argyle illusion

Minjung Kim[1, 2], Jason M. Gold[3], & Richard F. Murray[2], [1]New York University, USA, [2]Centre for Vision Research, York University, Canada, [3]Indiania University, USA

The argyle illusion (Adelson, 1993) is a brightness illusion in which lines, triangles, and diamonds are arranged to resemble patterns on knitted garments. The standard explanation is that, due to shapes forming x-junctions, the visual system perceives strips of dark and light transparent filters over the argyle. Since diamonds under the two different filters have the same luminance, the visual system infers that the diamonds must be of different reflectances, thus perceiving a brightness difference.

Using classification images, a technique which reveals image components that influence perceptual decisions, we examined whether the x-junction explanation was correct, i.e., whether x-junctions would be the most influential parts of the classification image. We chose two test diamonds, one from each of the filtered regions, and set their luminances to observers' PSEs. Other argyle elements were corrupted with white, Gaussian noise (SD=.18). Observers (n=3, 10,200 trials each) chose which test diamond appeared brighter (2AFC).

Surprisingly, classification images did not show that x-junctions were the most influential stimulus elements; rather, local contrast better explained observer choices. We also found that diamonds immediately neighbouring the test diamonds contributed to the illusion, but only if the neighbours were under the same filters as the test diamonds. This suggests a localized grouping effect, where neighbours sharing the same lighting framework moderate perceived brightness of the test diamonds.

We conclude that x-junctions do not play a decisive role in the argyle illusion, and that instead, stimulus patches are compared to local grouped elements. Our experiments point towards the importance of lighting frameworks for brightness perception, and demonstrate classification images as a general technique for examining mid-level lighting effects, such as shading and transparency.

52. Shape from shading under inconsistent lighting

John Wilder, Wendy Adams, & Richard F Murray, Centre for Vision Research, York University, Canada

Shape from shading models traditionally assume that observers estimate a lighting direction and use this estimate to infer shape from shading. In real world scenes, local lighting direction varies in unpredictable ways. How locally consistent must lighting be to accurately perceive shape from shading? We manipulated local lighting directions and measured the effect on shape from shading.
In exp. 1a subjects were shown surfaces that varied in depth and judged the relative depth of the surface at two probe locations. The depth profiles of the surfaces were filtered Gaussian white noise with a kernel of one of three widths (sigmaS). The lighting direction varied smoothly from place to place. Lighting directions were generated using Gaussian noise filtered with one of six kernels (sigmaL). There was also one uniform-lighting-direction condition. Performance decreased smoothly as sigmaL decreased (high sigmaL = less lighting variation) but even with quite rapid changes in local lighting direction, performance was still well above chance. In Expt. 1b a window of uniform lighting was centered around the probes; outside this window lighting direction varied. Window size varied each trial. If local lighting is consistent over more than two cycles of surface change observers can recover shape from shading. In Expt. 2a subjects viewed a surface in which three quadrants were lit from one direction and the lighting direction of the fourth differed by a tilt of 90°. Between quadrants, lighting direction changed smoothly from one direction to the other. Subjects tried to identify the odd quadrant. All subjects performed at chance. However, in Expt. 2b subjects successfully performed a relative depth judgment task, demonstrating that their shape from shading mechanism accurately estimates local lighting information. These results suggest that shape from shading mechanisms tolerate rapid variations in local lighting direction, and furthermore observers cannot detect strong lighting inconsistencies.

53. Evaluating anticipatory licking as a reward predicting signal in macaque monkeys during
reversal learning of feature values

Sood K, Westendorff S, & Womelsdorf T., Centre for Vision Research, York University, Canada

Goal-directed behaviour is enabled through selective attention towards specific information that is considered most relevant to a goal. Attention is likely based on an internal prediction system that is a function of learning (i.e. environmental experience). This project serves to identify and validate overt behavioural signatures of this learning system to provide insight on the choices made by a subject prior to being made. We set out to determine whether licking was indicative of anticipation for liquid-rewarded choices as a progression with learning and further, whether this signature was restricted to rewarding choices during a selective attention foraging task in a macaque monkey.

54. Attentional blink with moving stimuli

Marta Kryven, & William Cowan
David R. Cheriton School of Computer Science, University of Waterloo, Canada

Attentional Blink occurs when the detection of the second of the two visual targets presented in close succession is impaired. The subjects respond to the first target, T1, but if the second target, T2, is presented within about 500ms of T1, it is often missed. In a typical attentional blink experiment targets are letters or digits (MacLean & Arnell, 2012). We investigate whether attentional blink can occur with simple animated stimuli, such as motion detection tasks. The result will help us understand how viewers distribute attention across elements of animated graphics. Stage magicians often use techniques resembling those used in attentional blink experiments to hide motion during a trick. For example, a magician may ask a spectator a difficult question, T1, and immediately perform a movement, T2, missed by the spectator. Alternatively, a magician may use a large trick-irrelevant gesture, T1, to hide a smaller trick-relevant motion, T2. Our study aims to recreate the same effect using a simple 2D animation.

We first hypothesized that since motion, unlike text, is easily perceived, a motion detection task will not be blinked. However, our results indicate the contrary. Second, we hypothesized that presenting a moving background during an attentional blink task attenuates attentional blink for moving stimuli, as was observed for literal stimuli (Arend, Johnston & Shapiro, 2006). In our experiment, attenuation of attentional blink did not occur: moreover, attentional blink became deeper. Possibly, the distractors moving in the background are harder to suppress while attending to a moving stimulus during T2. Finally, we hypothesized that attentional blink does not occur when both task require a response to simple moving graphics. However, our results show a strong attentional blink for a moving stimulus T2 after subjects attend to a motion discrimination task T1.

55. The effects of motion cues on figure-ground perception in healthy aging

Jordan W. Lass[1], Patrick J. Bennett[1], Mary A. Peterson[2], & Allison B. Sekuler[1]
[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada, [2]Department of Psychology and Cognitive Science Program, University of Arizona, USA

Figure-ground (FG) perception involves segmenting adjacent regions sharing a border into figure and background. Border convexity is one static cue that influences FG perception in a context-dependent manner: The probability of perceiving the figure on the convex side of a border increases with the number of alternating convex and homogeneously filled concave regions (Peterson & Salvagio, J Vision, 2008). This Convexity Context Effect (CCE) is reduced in older adults compared to younger adults (Lass, et. al., VSS, 2013). The reduced CCE in older observers may result from decreased competition resolution in FG patterns, making it more likely that the stimuli are interpreted as flat patterns. If so, then adding cues that indicate depth in the stimulus may enhance the CCE in seniors. We examined this hypothesis by testing younger (M=22.1 years) and older (M=65.9 years) observers in a FG task using 100 ms static displays consisting of 2 or 8 alternating lighter and darker regions of random dot textures, and dynamic displays in which the textures in adjacent regions moved in opposite directions. Froyen, Feldman, and Singh (J Vision, 2013) found that such motion evokes a strong percept of depth in younger observers. The FG task was to indicate the colour of the region that appeared to be in the foreground. Contrary to the hypothesis, the CCE exhibited by older observers was not larger for moving stimuli compared to static stimuli. This result is consistent with the idea that

healthy aging alters the way that configural cues influence FG organization, even when additional cues signal depth in the displays. Currently, we are examining the possibility that seniors require longer presentation times to accurately perceive the motion (Bennett, Sekuler & Sekuler, Vis Res, 1997). We are also investigating the relationship between individual differences in perceived depth and CCEs.

56. Estimating the response properties of the human lateral geniculate nucleus using a spatiotemporal population receptive field model

Kevin DeSimone[1,2] , & Keith A. Schneider[2,3], [1]Department of Psychology, [2]Department of Biology, [3]Centre for Vision Research, York University, Canada

The population receptive field (pRF) model has been widely adopted in the imaging community as a quantitative approach for estimating the response properties of populations of neurons. The initial formulation of the pRF model represented the spatial tuning of each voxel in terms of a 2D gaussian (Dumoulin & Wandell, 2008). This pRF model has characterized the retinotopic organization of cortex and more recently in multiple subcortical nuclei (DeSimone et al., in revision). We sought to extend the spatial pRF model to include a temporal component for characterizing the functionally and anatomically distinct layers of the human lateral geniculate nucleus (LGN). Using a flickering visual stimulus, we found evidence that the magnocellular and parvocellular layers of the LGN may be functionally differentiated using a spatiotemporal pRF model.

57. Age-related shift in E-I balance in dorsal visual pathway

Siu CR[1], & Murphy KM[1,2], [1]McMaster Integrative Neuroscience Discovery and Study (MiNDS) McMaster University, [2]Psychology, Neuroscience and Behaviour, McMaster University, Canada

Many aspects of visual perception decline in older adults, but surprisingly they perform better than young adults on motion discrimination tasks that require surround suppression. The main hypothesis explaining this is a global loss of GABA-mediated surround suppression in aging. However, there are not global changes in all visual tasks that rely on surround suppression. Local changes in other neurotransmitter systems might contribute to the loss of surround suppression. For example, visual cortex feedforward-driven activity is strongly mediated by AMPA receptors, while feedback modulation is mediated by NMDA receptors. Perhaps expression of these glutamatergic receptors is maintained in the aging dorsal/motion visual pathway but lost in other visual areas. We used Western blotting to quantify expression of GAD65, the enzyme that makes the on-demand pool of GABA, AMPA (GluA2) and NMDA (GluN1) receptor subunits in V1, V3 (dorsal), and V4 (ventral) of human postmortem cortical tissue. We studied 15 cases ranging in age from 5-80 years.

We found similar age-related losses in expression of GAD65 in V1, V3, and V4. We compared expression of GluA2 and GluN1 and found similar losses in V1 and V4, but surprisingly no loss of these in V3. Furthermore, only the dorsal pathway had a change in GAD65 vs GluA2 and GluN1 that could shift the excitatory-inhibitory balance in favor of excitation.

These findings provide new information about pathway specific neurobiological changes in the aging human visual system. The maintenance of high levels of glutamate receptors in the dorsal pathway suggests that this excitatory neurotransmitter system is likely to have a greater contribution to motion perception in aging. These findings point toward the balance

between excitation and inhibition, as opposed to absolute levels of excitation or inhibition alone, as being important for surround suppression in aging visual perception.

58. Effect of contrast on perceptual and decision making processes in the dorsal stream

Carolyn J. Perry[1,2], Henna Asrar[2,3], Prakash Amarasooriya[1,2], & Mazyar Fallah[1,2,3,4]
[1]School of Kinesiology and Health Science, [2]Centre for Vision Research, [3]Departments of Biology and Psychology, [4]Canadian Action and Perception Network, York University, Canada

Evidence supports the integration of both dorsal and ventral stream information into motion computation processes. This gives rise to intermediate object representations in the dorsal stream which in turn facilitate object selection and decision making mechanisms. Previous work, using superimposed moving surfaces, has found that colour (a ventral stream feature) is integrated into dorsal stream object representations only after direction computations in MT (middle temporal visual area) and allows for object selection which speeds decision-making in the dorsal stream (Perry and Fallah, 2012). However, speed (a dorsal stream feature) is integrated prior to direction computations in MT, improving direction discriminations, and also reducing processing time (Perry et al., 2014). Here we investigate at what stage of processing contrast is integrated into dorsal stream object representations and used to facilitate perceptual and decision making processes. Using contrast levels to which both the ventral and dorsal stream are sensitive, we determined the relative influence that varying the contrast of a second surface has on participants' abilities to correctly determine surface direction (perceptual task) and/or the time it takes to process and decide on the directions of both surfaces. In general, as the contrast of the second surface is reduced, speed of processing increases. However, contrast levels, to which the dorsal stream is sensitive, reduces processing time but does not improve direction discrimination. This suggests that in spite of the early contribution of contrast to visual processing, integration into dorsal stream object representations occurs after direction computation in MT.

59. The effects of high frame rate on perception of 2-D and 3-D global coherent motion

Yoshitaka Fujii[1], Robert S. Allison[1], Lili Shen[1,2], & Laurie M. Wilcox[1]
[1]Centre for Vision Research, York University, Canada, [2]School of Electronic Information Engineering, Tianjin University, China

Digital technologies allow movies to be exhibited at frame rates much higher than the traditional 24 fps. High frame rate (HFR) movies being released in theaters and it is assumed that HFR will reduce artifacts and enhance quality of motion in 2-D and 3-D media. The goal of this project is to assess this assumption empirically by basic measurement of motion perception. In a series of experiments we measured lateral (2-D) and in depth (3-D) global motion coherence thresholds using random-dot patterns in a mirror stereoscope and a 3D projection system. The refresh rate of the display was fixed at 96 Hz, and we manipulated the flash protocol to create 96 (single flash), 48 (double flash) and 24 (quadruple flash) frames per second. Simulated linear velocity of the elements through space was equated in the 2-D and 3-D conditions. Conditions were randomly interleaved using the method of constant stimuli and a two-interval forced-choice procedure to measure the proportion of coherent elements required to reliably detect global motion. Results showed no consistent effect of flash protocol on coherence

thresholds in either the 2-D or the 3-D conditions in both the stereoscope and 3D projection system. Our results show that while frame rate influences local 2-D motion processing, it has no apparent impact on global lateral, or in depth, motion coherence perception. This indicates that progression in quality of motion signal does not always enhance perception.

60. Effects of head orientation and scene rigidity on vection

Pearl S. Guterman, & Robert, S. Allison, Centre for Vision Research, York University, Canada

Changing head tilt relative to gravity changes the dynamic sensitivity of the otoliths to linear accelerations (gravitational and inertial). We explored whether visually induced self-motion (vection) is influenced by varying head tilt and optic flow direction with respect to gravity. We previously found that vection was enhanced when upright observers viewed vertical optic flow (i.e., simulating self-motion along the spinal axis) compared to horizontal flow. We hypothesized that if this benefit was due to aligning the visual motion signal with gravity, then inter-aural lamellar flow while laying on the side would provide a similar vection advantage. Observers stood and lay supine, prone, left and right side down, while viewing a translating random dot pattern simulating self-motion along the spinal or inter-aural axis. Vection magnitude estimates, onset, and duration were recorded. The results showed that aligning the direction of visual motion and gravity enhanced vection in side-laying observers, but when gravity was irrelevant—as in the supine and prone posture—spinal axis motion enhanced vection. However, perceived scene rigidity varied with head orientation (e.g., dots were seen as floating bubbles), so the issue of scene rigidity was examined by comparing vection in two environments: a rigid pipe structure which looked like a complex arrangement of plumbing pipes, and a field of dots. The results of varying head, motion direction, and perceived scene rigidity, will be discussed and may provide insight into whether self-motion perception is determined by a weighted summation of visual and vestibular signals.

61. Subjective sleepiness correlates with vection strength

Takeharu Seno, Masaki Ogawa, Kouhei Matsumori, & Shigekazu Higuchi
Faculty of Design, Kyushu University, Japan

We examined the effect of sleep deprivation on self-motion perception (vection). When exposed to a visual motion field that simulates the retinal optical flow generated by movement, we can perceive the subjective movement of our own body, “vection”. We measured the strength of vection in two conditions: sleep-deprivation and normal sleep. In the sleep deprivation condition, participants did not sleep for about 20 hours. The experiment used a within-subjects design with 16 men. Stimuli were generated and controlled by a computer (MacBookPro, MD101J/A; Apple) and presented on a plasma display (3D Viera, 50 inch, Panasonic, 1920 × 1080 resolution with 60 Hz refresh rate). The experiment was conducted in a dark chamber. Optic flow displays consisted of 1240 randomly positioned dots per frame with projected global dot motion that simulated forward self-motion (20 m/sec). Stimulus duration was 40 s. The stimuli subtended 72° (horizontal) × 57° (vertical) at a viewing distance of 57 cm. Two conditions were counter-balanced over the 16 participants. When viewing the computer-generated radially expanding pattern, participants were asked to press a designated button with their dominant hand as soon as they perceived forward self-motion. After each trial, the participants rated their subjective vection strength using a 101-point rating scale ranging

from 0 (no vection) to 100 (very strong vection). Results showed that vection tended to be stronger in the sleep deprivation condition. We also obtained subjective sleepiness. As expected, subjective sleepiness significantly increased following sleep deprivation. Magnitudes of vection and subjective sleepiness were significantly positively correlated. It has been reported that sleep deprivation can alter human vision (e.g., Schneider et al., 1996; Paavonen et al., 2010; Araujo et al., 2013; Chee & Chuah, 2007). Our current findings also support this claim.

62. Dense depth reconstruction using stereo-inertial fusion with multi-layer conditional random field

Francis Li, John Zelek, & Alexander Wong, University of Waterloo, Canada

3D models of scenes and objects are useful for tasks such as robotic navigation, object recognition, parts inspection, artefact preservation, graphical modelling, etc. Existing 3D scanning solutions can be extremely expensive, costing thousands of dollars, or are bulky and not portable. With technological trends favouring mobile solutions, there is a push for 3D scanners to be in a portable, compact, and small form factor at an affordable price. For short range depth sensing, compact stereo cameras are suitable for this task. This work presents a novel and simple method to densely reconstruct the surface geometry of an object from stereo motion using a low cost and compact stereo camera system. The end goal of the project is to enable devices such as smart phones to cheaply acquire usable 3D models of world objects. The proposed research first combines inertial and stereoscopic measurements (stereo-inertial fusion) to estimate the motion between captured camera frames for depth alignment. Dense 3D reconstruction and refinement is then performed through the use of a multi-layer conditional random field (MCRF). The MCRF approach formulates depth reconstruction as a maximum a posterior (MAP) problem and uses photometric measurements along with sparse depth measurements to reconstruct a dense point cloud representing the object surface. Qualitative results show that the proposed method can densely reconstruct a surface patch captured from different viewpoints. The method shows promise and can be extended to perform full 3D scans of objects.

63. Automatic inference of highway structure from video data

Eduardo R. Corral-Soto, & James H. Elder, Centre for Vision Research, York University, Canada

Many highway systems have extensive camera networks providing video data that can potentially be used for automatic traffic analytics. A limiting factor in the accuracy and reliability of video analytics algorithms is variability both within and across sites in terms of the structure of the highway as well as the height, placement, pose (pan/tilt) and focal length (zoom) of the camera. Here we study automatic methods for inferring key parameters of the camera and the highway that can lead to more reliable traffic analytics systems. First, we develop and evaluate a novel method that exploits the parallelism of highway contours to infer tilt angle and focal length of the camera and thus rectify the imagery. Second, we show how these contour algorithms can be combined with vehicle detection algorithms to infer lane structure.