

Marseille, October, 18th 2008

**SPHCV 2008**



First International Workshop on Shape Perception in Human and Computer Vision

**Part of the 10<sup>th</sup> European Conference on Computer Vision (ECCV 2008)**



**First International Workshop on Shape Perception in Human and Computer Vision**

SPHCV @ ECCV08

October 18, 2008

## Organization

The format of the one-day workshop will be 12 invited speakers (six human vision, six computer vision). Each talk will last **25 min plus 5 min** for discussion. The speakers have been chosen to represent a broad cross-section of shape perception research, representing the major paradigms in both the human and computer vision communities. Speakers will be encouraged to reflect on their experience, identify critical challenges, etc., rather than present snapshots of their latest research results.

## Speakers

### *Human Vision*

Irving Biederman  
Jacob Feldman  
Jan Koenderink  
Zygmunt Pizlo  
Johan Wagemans  
Qasim Zaidi

### *Computer Vision*

Ronen Basri  
Daniel Cremers  
Pedro Felzenszwalb  
Martial Hebert  
Jean Ponce  
Steve Zucker

### Organizers

**Sven Dickinson**  
University of Toronto  
Toronto, Canada  
[sven@cs.toronto.edu](mailto:sven@cs.toronto.edu)

**Zygmunt Pizlo**  
Purdue University  
West Lafayette, USA  
[pizlo@psych.purdue.edu](mailto:pizlo@psych.purdue.edu)

Supported by US Air Force Office of Scientific Research

## Abstracts - Human Vision

---

### **The Neural Basis of Shape Recognition. Irving Biederman. University of Southern California**

Harold W. Dornsife Professor of Neuroscience, Departments of Psychology and Computer Science and the Neuroscience Program, University of Southern California

About 20 years ago, a proposal was advanced that a considerable range of behavioral phenomena associated with human object recognition can be understood in terms of a representation positing an arrangement of simple part primitives distinguished by viewpoint invariant properties (= geons). Recent research on optical imaging as well as single unit activity of cells in macaque IT and behavioral and fMRI studies in human lateral occipital complex (the likely homolog to IT) provide a surprisingly strong confirmation of this proposal. Specifically, the evidence supports preferential coding for: a) simple parts (vs. whole object templates or irregular forms), b) nonaccidental (vs. metric) properties, c) relative (vs. absolute) position, and edges corresponding to orientation and depth discontinuities (vs. surface properties).

---

### **"Explaining" a shape by estimating its generating skeleton. Jacob Feldman. Rutgers University.**

Shape representations that accurately reflect intuitive part structure remain an elusive goal. Skeletal or medial-axis representations of shape have long been regarded as essentially unsuitable for part decomposition because of their tendency to include spurious axes that don't correspond to any perceptually natural part. I will discuss a new way of thinking about the shape skeleton that avoids many of the classic problems. We think of a shape as the result of a stochastic generative process by which a skeletal structure "grows" a shape around it. This conceptualization recasts shape representation as an inverse probability problem, in which the goal is to estimate the skeleton most likely to have generated the observed shape. We adopt a Bayesian approach to this problem, defining a prior over skeletons, and a likelihood function quantifying the probability that a given shape would be generated by a given hypothetical skeleton. The maximum a posteriori (MAP) skeleton, meaning the skeleton most likely to have generated the observed skeleton under the assumed prior and likelihood, can be regarded as the best "explanation" of the shape as the result of a generative growth process, and generally has an intuitive structure with each axis corresponding to a perceptually natural part of the shape. Reconceiving shape as the result of a probabilistic generative process opens up new ways of thinking about a number of key related problems. Shape similarity can be thought of as reflecting the probability that one shape could be generated by the model (skeleton) of the other shape, and vice versa; this gives a metric that correlates well with human similarity judgments. The approach also extends fairly naturally to 3D, requiring only an extension of the likelihood function, but leaving unchanged the logic of the estimation problem. Finally, the probabilistic formulation allows a natural connection to the statistics of natural shapes, which form the basis of ecological shape priors.

Joint work with Manish Singh.

---

---

## Shape in Relief Space.

Jan J. Koenderink. Delft University of Technology.

'Relief-space' is experienced when you look into a picture, it is also known as 'pictorial space'. In relief-space you are faced with surfaces that appear as the 'frontal surfaces' of solid objects, but differ from these because there is no 'backside' to them. Unlike Euclidean space, objects in relief-space cannot make a full turn. Relief-space is a 'hallucination' constrained through image structures interpreted as '(depth-)cues' by the spectator. An analysis of the 'inverse optics' theories of cues reveals that the cues constrain relief-space only up to a group of ambiguities. This group can be interpreted as the group of motions and similarities of the space. The perceiver experiences relief space only up to arbitrary motions and/or similarities, these define the 'beholder's share'. From the perspective of Klein's 'Erlangen Program' the group of motions (or congruences) defines the structure of the geometry of relief-space. It turns out to be the singly isotropic 3D Cayley-Klein space studied in detail by Strubecker and later by Sachs. 'Shape' is formally understood as the invariant under arbitrary motions and similarities, in the case of 'local shape' (curvature) these are the differential invariants. Relief shape is different from Euclidian shape, although there exist many analogies. Relief shape is important in monocular static viewing, for instance pictorial viewing. This is why relief shape is important in the visual arts (the first formal treatment is due to the 19th c. German sculptor Hildebrand) and why it must be expected to have important implications for computer graphics.

---

## Shape constancy and shape recovery: wherein human and computer vision meet.

Zygmunt Pizlo. Purdue University.

Recovery of 3D scenes, based on information provided by one or more 2D images, is an inverse problem. Its solution depends critically on knowledge about the family of possible solutions (this knowledge is represented by priors of 3D scenes). Priors have been identified for many perceptual properties. Examples include: isotropic surface texture, Lambertian reflectance, the light source above the observer, spatially continuous and piecewise rigid objects, piecewise smooth object's surfaces, small curvatures and orientations of surfaces, as well as black-body approximations of daylight. Priors can be defined for actual objects and scenes, or for abstract properties of objects and scenes. The former are "easier" in the sense that an algorithm can learn such priors from examples. The latter, however, are potentially more effective because abstract priors generalize easily to unfamiliar scenes and objects. We found 4 very effective, general-case *shape* priors (there are no equally-effective *surface* priors): 3D symmetry, maximal 3D compactness, planarity of contours, and minimum surface area. Performance of the model is very similar to the performance of human subjects. Both recover 3D shapes very well (they achieve shape constancy), and in the rare cases when they do not, their errors are the same. Our model's shape recovery *does not*: (i) depend on surface recovery, (ii) require depth cues such as binocular disparity or motion, or (iii) use familiarity with objects. Our model's shape recovery *does* depend on applying our 4 shape priors to 2D *shapes* in the image. Our model's success in recovering 3D shape shows that finding objects in the image and characterizing their shapes is of fundamental importance for the perception of 3D shape.

---

---

**On the role of curvature singularities in the perception of outline drawings of objects.  
Johan Wagemans. University of Leuven.**

It is remarkable how easy it is for humans to recognize objects in a few lines on a flat canvas. In a classic paper of more than half a century ago, Attneave (1954, *Psychological Review*) argued that most of the information about objects is concentrated along the object's contour and, more specifically, in points along the contour where curvature changes most strongly (i.e., curvature extrema). In this talk, I will present an overview of a whole series of studies aimed at testing this proposition in many different ways. We have developed an extensive stimulus set derived from line drawings of everyday objects (Snodgrass and Vanderwart, 1980, *Journal of Experimental Psychology: Human Memory and Learning*), consisting of outlines with known curvature values. In this way, we could create different stimulus conditions with specific manipulations at the curvature singularities (extrema and inflections), for example, straight line versions connecting either extrema or inflections and fragmented versions with contour fragments positioned on either extrema or inflections. In addition to these large-scale identification experiments, I will also describe a segmentation study, testing whether curvature minima are used as segmentation points (as proposed by Hoffman and Richards, 1984, *Cognition*). The overall conclusion of this research program is that the role of local contour properties like curvature singularities must be understood in relation to more global shape factors like complexity, homogeneity, symmetry, etc

---

**Early Neural Processes that facilitate 3-D shape perception.  
Qasim Zaidi. State University of New York.**

In retinal images of 3-D surfaces, the statistics of the texture pattern change with the curvature of the surface. Shape-from-texture models assume that the texture on the surface is statistically homogeneous, but under generic conditions the texture on a carved or stretched surface is not homogeneous, and the inhomogeneity may change as the surface deforms. Estimating the projective transform from texture inhomogeneity and reversing it, thus may not infer the correct 3-D shape of the surface. By parsing images into orientation and frequency patterns, we show that correct 3-D percepts of curvature/slant arise from perspective generated orientation flows, irrespective of texture homogeneity. Spatial frequency flows give percepts of correct relative depth in images where frequency gradients result from relative distance, but incorrect depths where frequency gradients result from surface slant. We then examined whether cortical neurons in V1 & V2 facilitate the extraction of 2-D orientation flows. Slanting a textured planar surface generally enhances the visibility of the component parallel to the slant, which facilitates 3-D slant perception. Using contrast thresholds, we show that this enhancement results from a decrease in cross-orientation suppression when 3-D slant creates a frequency mismatch between texture components. The frequency-specific component of suppression cannot be simulated by existing LGN-based models, thus implicating cortical interactions. In four anesthetized macaques, 29 neurons in V1 and V2 were isolated with tetrode recordings, and presented with fronto-parallel and slanted gratings and plaids. Compared to optimal single gratings, flat plaids induced significant suppression in 78% of the neurons. However, suppression was significantly reduced in 45% of the neurons for slanted plaids. Since cross-orientation suppression reduces responses to patterns in natural scenes, stimuli that undermine these sources of suppression allow V1/V2 to signal areas containing 3-D shape. In addition, 28% of V1 and 56% of V2 neurons showed enhanced responses to orientation flows per se, indicating that some early cortical processes facilitate the decoding of 3-D shape.

---

### **On Poisson Descriptors and Prior Shape Knowledge. Ronen Basri. The Weizmann Institute of Science.**

The role of shape in visual object recognition has long been acknowledged. Shape provides a signature, invariant to viewing conditions that can readily be used to identify objects in images. In this talk I will introduce methods for extracting shape information in images by means of hierarchical image segmentation and for representing shapes by means of Poisson-based descriptors. I will then discuss how these representations can be used for object recognition. If time permits, I will discuss also the use of prior shape knowledge in 3D shape perception as is exemplified by two-tone ("Mooney") images.

---

### **Combinatorial Algorithms for Image Segmentation with Elastic Shape Priors. Daniel Cremers. University of Bonn.**

Numerous efforts have been made to impose statistical shape priors into image segmentation processes. The resulting segmentation process favors segmentations which are consistent with previously observed shape instances, it is therefore robust to missing or misleading low level information due to noise, background clutter and partial occlusions. While statistical and energy minimization methods allow for a transparent integration of shape prior and image information, the subsequent optimization schemes typically only provide locally optimal solutions with very little insight as to how far the computed solutions are from the globally optimal one. In my presentation, I will present existing approaches to impose shape priors, I will discuss limitations of local optimization methods and introduce efficient algorithms to impose shape priors in a globally optimal manner. The proposed algorithms find optimal shape-consistent segmentations in the space of all conceivable closed curves.

This is joint work with Thomas Schoenemann.

---

### **Hierarchical Models for Shape Recognition. Pedro Felzenszwalb. University of Chicago.**

Shape recognition has proven to be a challenging task for computer vision systems. One of the main difficulties is in developing representations that can effectively capture important shape variations. A classical approach for addressing this problem is to use deformable models, where each shape in a class is viewed as a deformed version of an ideal object. By using a hierarchical representation we are able to develop simple elastic matching algorithms that can take global geometric information into account. These algorithms are based on a dynamic programming procedure similar to CKY parsing, and can be used both for comparing pairs of objects and for detecting objects in cluttered images. I will also discuss how the hierarchical representation can be extended into a shape grammar that can explicitly capture both geometric deformations and structure variation. In this case the coarse structure of a shape is defined by a context-free-grammar while the precise geometry is defined by an abstract deformation model.

---

---

**Martial Hebert. Contour fragments for recognition and scene analysis.****Carnegie Mellon University.**

We discuss the use of contour fragments extracted from images or image sequences in recognition and scene analysis. We show how noisy contour fragments can be used to extract groupings and object hypothesis from the input image, and suggest how they can be used for recognition. The approach relies on two main ingredients. First, we show how individual fragment suggests a possible grouping of the image pixels into regions corresponding to objects. While each fragment contains limited information and can contribute only a hint at a possible segmentation, the combination of these hints can be used to generate sensible hypotheses. Second, we show how fragments can be grouped by using higher-order relations. We show how the parameters used to represent the relations between fragments can be estimated from training data. Finally, we show how the grouping information can be used in a matching framework for recognition.

---

**Sparse Discriminative Models for Local Image Analysis.****Jean Ponce.**

Ecole Normale Supérieure, Willow project-team ENS/INRIA/CNRS UMR 8548.

Sparse signal models have been the focus of much recent research, leading to (or improving upon) state-of-the-art results in signal, image, and video restoration. I will show in this talk that they can also be used in a new framework for classification tasks in local image analysis. I will present two variants of this approach: In the first one, an energy function including both reconstruction and discrimination components with  $l_0$  or  $l_1$  sparsity terms is used to learn one dictionary per class. In the second one, a single learned dictionary is shared by all the classes, but the parameters of multiple decisions functions are learned at the same time as this dictionary. Both approaches have been implemented, and I will present applications to feature selection, edge and object detection, texture classification, and handwritten digit recognition.

Joint work with Julien Mairal, Francis Bach, Martial Hebert, Marius Leordeanu, Guillermo Sapiro, and Andrew Zisserman.

---

**Perceptual Organization In Support of Shape Representation.****Steven W. Zucker. Yale University.**

The history of early visual descriptions for shape representation is reviewed. Based on both computational and neurobiological evidence, it is argued that the early stages necessarily involve inferences about boundaries, qualitative surface features, and their interactions in key neighborhoods; and that the later stages incorporate more global information, such as provided by the distance map. The result is a richer descriptive system than is normally sought in computer vision and a more consistent one than postulated in neurobiology.

---