Observables and Invariance for Early Cognitive Vision

G.J. Burghouts, J.M. Geusebroek, A.W.M. Smeulders
Intelligent Sensory Information Systems, University of Amsterdam.
Email: burghout@science.uva.nl

Abstract

This paper presents the visual measurement of physical object properties that characterize the perceived object including: size, shape, surface properties, cover reflectance properties, distance, and motion. We provide an overview of complete set of local visual measurements. We derive photometric, geometrical, and temporal invariants to counteract unwanted transformations in the observation including: illumination spectrum and intensity, scene setting causing shadow, shading and highlight effects, and variation due to object position, pose and distance.

1 Introduction

The objective of any cognitive visual system is to obtain an accurate representation of its environmental context and from properties designated as objects therein. For biological systems more true but also in effect for artificial systems, the representation depends on the system's environment and purpose (Gibson, 1979).

A cognitive visual system is in equilibrium with the characteristics of the environment it processes. Consequently, vision is tuned to the physical environment it operates in, as can be derived from statistics (Barlow, 1997) and physics (Foster and Nascimento, 1994) of the visual stimulus. Sensing and representing the visual stimulus enables a visual system to describe operationally its surroundings.

2 Local Visual Measurements

Due to the physiological structure of human vision (Hubel, 1988), we constrain ourselves to a visual stimulus in four physical dimensions: spatial coordinates x and y, wavelength spectrum λ, and time t. Representing the visual stimulus by the front-end therefore boils down to measuring the energy density volume in these physical variables. Measuring the stimulus energy, that is, plenoptics of the scene, a visual system requires spatial, spectral, and temporal 'receptive fields' (Adelson and Bergen, 1991). We construct receptive fields from spatial, spectral, and temporal receptive fields in parallel. The local visual measurement $\hat{E} : \mathbb{R}^4 \mapsto \mathbb{R}$ of the (change of) energy E of the irradiance falling onto a Gaussian receptive field $G : \mathbb{R}^4 \mapsto \mathbb{R}$ probing physical variables x, y, λ, and t simultaneously, becomes:

$$\hat{E}_{x^i y^j \lambda^k t^l}(x, y, \lambda, t) \equiv E(x, y, \lambda, t) * G^{\sigma_x, \sigma_y, \sigma_\lambda, \sigma_t}_{x^i y^j \lambda^k t^l}(x, y, \lambda, t), \tag{1}$$

where $(*)$ is a linear correlation operator and σ_x, σ_y, σ_λ, σ_t denote the scales.

Together with Gabor receptive fields $\tilde{G}$ probing spatial frequency (u, v) and temporal frequency w, a complete set of receptive fields is constructed from the spectral, spatial, and temporal receptive fields, see Table 1.

Table 1: Physical variables and the complete set of receptive fields.

Physical variable	*Receptive field*
λ	$G_{\lambda^i}(x, y, \lambda, t)$
x, y	$G_{x^i y^j}(x, y, \lambda, t)$
t	$G_{t^i}(x, y, \lambda, t)$
λ, x, y, t	$G_{\lambda^i x^j y^k t^l}(x, y, \lambda, t)$
λ, u, v, t	$\tilde{G}^{u_0, v_0}_{\lambda^i t^j}(x, y, \lambda, t)$
λ, x, y, w	$\tilde{G}^{w_0}_{\lambda^i x^j y^k}(x, y, \lambda, t)$
λ, u, v, w	$\tilde{G}^{u_0, v_0, w_0}_{\lambda^i}(x, y, \lambda, t)$

3 Visual Observables and Invariance

Whereas receptive fields measure the 2-dimensional light field, a visual system aims at representing: object size and macro-, meso-, and micro-shape, its surface properties, cover reflectance properties, distance and motion, that is, 'visual observables' characterizing perceived objects (Smeulders et al., 2001).

The problem now is how to arrive at observables given a set of visual measurements. Answering this question requires knowledge of the physical laws involved in the visual stimulus formation. The set of laws concerning stimulus formation are constant for a particular ecology. We consider atmospheric vision. Inferring the physical sources of variation enable to a visual system to relate visual measurements and observables under various transformations. An invariant relates the visual measurements to an observable ignoring the irrelevant component in the perceived variation.

In short, visual measurements probe the 4-dimensional energy density in x, y, λ, and t. Hence, a limited set of object properties can be represented. Visual measurements have limited accuracy and consequently impose a metameric class of observables. Given the limited set of visual measurements and the limited physical sources of variation, only a fixed set of invariants can be derived. The purpose of a visual system is to obtain representations invariant to unwanted transformations, tailored to the retained specific object variant conditions. Table 2 provides an overview of the variants and invariants, following an incremental approach of application of invariants.

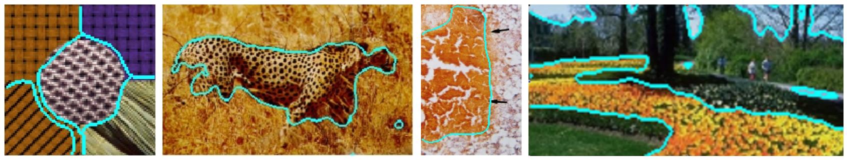
Samples from the COREL image collection are weakly segmented, confirming the discriminative power of the invariants, see Figure 1.

4 Conclusion

We have derived a complete set of local invariant visual measurements. The measurements are performed by receptive fields probing space, wavelength spectrum, and time. We identified physical object properties to metamerically characterize a perceived object. Various unwanted transformations in the visual stimulus formation disturb the measurement of observables. Invariants eliminate systematically deviations of illumination spectrum and intensity, scene setting causing shadow, shading and highlight effects, and variation due to object position, pose and distance.

Table 2: Receptive fields, observables, unwanted transformations and invariants.

Physical variable	*Receptive field(s)*	*Observable(s)*	*Directly measurable*	*Unwanted transformation(s)*	*Invariant(s)*
Wavelength spectrum	$G_{\lambda^i x^j y^k}$ (Geusebroek et al., 1999)	Object cover reflectance	Reflectance	Illumination intensity	$\mathcal{W}$
				Illumination intensity, shadow, shading, highlights	$\mathcal{H}$
				Illumination intensity and spectrum, shadow, shading	$\mathcal{N}$ (Geusebroek et al., 2001)
Local geometry	$G_{\lambda^i x^j y^k}$ (Koenderink and van Doorn, 1987)	Object macro-shape	Edges, curvature, junctions, corners	Object position, pose, distance	$\mathcal{I}$ (Florack, 1997)
		Object size, surface/cover granularity			$\max\arg(\sigma_{xy})\mathcal{I}^{norm}$ (Lindeberg, 1998)
		Object distance	Far distance		$\mathcal{Z}$ (Pentland, 1987)
		Object meso-shape	Principal curvatures		κ (Koenderink and van Doorn, 1987)
			Shading		w (Lee and Rosenfeld, 1985)
		Object cover reflectance type	Highlights		w
Spatial frequency	$\tilde{G}^{u_0,v_0}_{\lambda^i}$ (Bovik et al., 1990)	Surface/cover regularity	Spatial frequency	Object position, pose, distance	$\mathcal{S}$
		Object distance	Relative near distance		
Time	$G_{\lambda^i x^j y^k t^l}$ (Adelson and Bergen, 1985)	Object motion	Projected object motion	Object distance	$\mathcal{V}$ (Horn and Schunck, 1981)
Temporal Frequency	$\tilde{G}^{w_0}_{\lambda^i}$	Object motion periodicity	Temporal frequency		



(a) (b) (c) (d)

Figure 1: Segmentation of a benchmark image and images from the COREL collection. Data from the intensity measurement $\hat{E}$, and invariants $\mathcal{W}_\lambda$, $\mathcal{W}_{\lambda\lambda}$, and $\mathcal{S}$, $\mathcal{S}_\lambda$ and $\mathcal{S}_{\lambda\lambda}$ are k-means clustered based on an Euclidean distance measure.

References

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA-A*, 2(2):284–299.

Adelson, E. H. and Bergen, J. R. (1991). The plenoptic function and the elements of early vision. In Landy, M. and Movshon, J. A., editors, *Computational Models of Visual Processing*, pages 3–20. MIT Press, Cambridge.

Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Phil. Trans. R. Soc. London B*, 352:1141–1147.

Bovik, A. C., Clark, M., and Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE T PAMI*, 12:55–73.

Florack, L. M. J. (1997). *Image Structure*, volume 10 of *Computational Imaging and Vision Series*. Kluwer Academic Publishers, Dordrecht.

Foster, D. H. and Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proc. Roy. Soc. London*, Series B(257):115–121.

Geusebroek, J. M., Boomgaard, R., Smeulders, A. W. M., and Dev, A. (1999). Color and scale: The spatial structure of color images. In *IEEE ECCV*, volume 1, pages 331–341.

Geusebroek, J. M., Boomgaard, R., Smeulders, A. W. M., and Geerts, H. (2001). Color invariance. *IEEE T PAMI*, 23(12):1338–1350.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton-Mifflin, Boston.

Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *AI*, 16:185–203.

Hubel, D. (1988). *Eye, brain, and vision*. W. H. Freeman, New York.

Koenderink, J. J. and van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics*, 63:291–297.

Lee, C. H. and Rosenfeld, A. (1985). Improved methods of estimating shape from shading using the light source coordinate system. *AI*, 26:125–143.

Lindeberg, T. (1998). Feature detection with automatic scale selection. *IJCV*, 30(2):117–154.

Pentland, A. P. (1987). A new sense of depth of field. *IEEE T PAMI*, 9:523–531.

Smeulders, A. W. M., Geusebroek, J. M., and Gevers, T. (2001). Invariant representation in image processing. In *IEEE ICIP*, volume 3, pages 18–21.