

# Image Statistics: From Data Collection to Applications in Graphics

Erik Reinhard\*  
University of Bristol, UK

Tania Pouli  
University of Bristol, UK

Douglas Cunningham  
Brandenburg Technical University, Germany



Figure 1: The left image can easily be recognised as a natural scene. The image on the right however appears as random noise despite both images consisting of the same pixels.

## Abstract

Natural images exhibit statistical regularities that differentiate them from random collections of pixels. Moreover, the human visual system appears to have evolved to exploit such statistical regularities. As computer graphics is concerned with producing imagery for observation by humans, it would be prudent to understand which statistical regularities occur in nature, so they can be emulated by image synthesis methods. In this course we introduce all aspects of natural image statistics, ranging from data collection to analysis and finally their applications in computer graphics, computational photography, visualization and image processing.

**CR Categories:** I.3.3 [Computer Graphics]: Picture/Image Generation—Display Algorithms;

**Keywords:** Natural Image Statistics, Computational Photography, Rendering

## 1 Outline

The field of natural image statistics studies images and their statistical regularities. The human visual system has evolved in natural environments, i.e. those without man-made structures. Human vision, it is argued, has therefore evolved to be particularly good at observing natural scenes, as opposed to random images [van der Schaaf 1998]. As an example, the left image in Figure 1 shows a scene that is easy to interpret. Scrambling the pixels produces the image on the right, which does not have any structure that can be recognised. However, the number of random images is many orders of magnitude larger than the number of natural images.

Thus, natural images form a very small subset of all possible images, and contain structure to which the human visual system is optimised. By studying the input to the human visual system, it is thought that the human visual system can be understood better. This has led to a rich literature on various aspects of natural image statistics [Pouli et al. 2010].

\*e-mail: reinhard@cs.bris.ac.uk

The typical process by which natural image statistics are computed, is to pass each image of an ensemble through a particular transform. The transforms typically employed tend to correspond to a component of human vision. It is, for instance, possible to compute the power spectra of each image, and from these compute the average power spectral slope for the whole set. This course reviews this and many other transforms and statistics.

Thus, the reason for this course is that, aside from learning about human vision and the implications for perception, some of these statistical regularities have transferred to computer graphics: several applications now successfully employ statistics of natural images to help solve engineering problems. As an example, deblurring algorithms rely on optimisation involving priors that help push candidate images towards visually plausible solutions. Of course, visually plausible solutions are those that appear to humans in some sense natural. Image priors have therefore been designed with the aid of specific knowledge of natural image statistics, in particular regarding the average distribution of image gradients.

We think that the use of natural image statistics has so far shown great promise, with interesting applications making use of some of the available statistical regularities. We also see excellent scope for further future developments. We hope to educate the graphics community and make researchers and practitioners in this field aware of the exciting possibilities of natural image statistics.

To this end, we will review the current state of natural image statistics, including data collection and calibration techniques and the different types of transforms and statistics encountered. We will thoroughly discuss specific areas of application within computer graphics, computational photography, visualization and image processing with the aid of specific examples and finally, we will draw conclusions and hypothesize about how the use of natural image statistics can be extended in the future.

## 2 Target Audience

The target audience for this course includes researchers and practitioners in computer graphics, computer vision, as well as image processing who are interesting in understanding how to compute

and use image statistics in the context of image processing, or in any algorithm aiming at producing natural images. More specifically, the course should be useful to computational photography and image processing researchers interested in incorporating information about human vision and natural environments in their work.

### 3 Topics

The course will incorporate all aspects of natural image statistics, from the capture of image ensembles, their analysis, currently known statistical regularities, as well as applications that make use of such statistics.

**Introduction** We will introduce image statistics in the context of human vision research, explain why such findings are useful in computer graphics and related disciplines and will categorise different types of statistics.

**Data collection and calibration** Here, we will focus on the process of collecting the image ensembles and calibrating them in order to remove artefacts due to lens and camera inaccuracies [Willson 1994; Shih et al. 1995; Shah and Aggarwal 1996; Brady and Legge 2009]. The areas covered will include radiometric calibration, geometric and chromatic lens distortions and noise. Image capture will also be discussed covering areas such as high dynamic range capture [Healey and Kondepudy 1994; Mann and Picard 1995; Mitsunaga and Nayar 1999; Kim and Pollefeys 2004] and panorama stitching [Ostia et al.]. Finally, the different types of scenes that can be used for image statistics will be presented. Although image statistics work has traditionally focused on natural images [Ruderman 1997b; van der Schaaf 1998; Huang and Mumford 1999], urban and artificial scenes have also been explored [Ziegeus and Lang 1997; Ziegeus and Lang 1998] while Dror et al. [Dror et al. 2001] have been the first to look at the statistics of high dynamic range, spherical environment maps.

**First order statistics** First order statistics look at single pixels in the image and do not take into account relations between neighbours. They correspond to the distribution of values within the analysed data set and its properties, making them easy to both compute and interpret [van der Schaaf 1998; Huang and Mumford 1999]. Typical examples include histogram moments such as the mean, standard deviation, skew and kurtosis.

**Second order statistics** Second order statistics capture relations and regularities over pairs of pixels in the image. The power spectra of images have been extensively analysed with most work focusing on the slope resulting from the frequency-power relation over sets of natural images [Burton and Moorhead 1987; Tolhurst et al. 1992; Ruderman and Bialek 1994b; Ruderman and Bialek 1994a; van der Schaaf and Hateren 1996; Field and Brady 1997; Ruderman 1997a; Tolhurst and Tadmor 1997; Webster and Miyahara 1997; Rainville and Kingdom 1999; Reinhard et al. 2001b; Reinhard et al. 2004]. Power spectra are computed in Fourier space, but relate to the auto-correlation function in image space, thus effectively providing information on pairs of pixels. The power spectrum, when averaged over large numbers of natural images, tends to show a behaviour which can be well modelled with  $P = f^{-2}$ , where  $P$  is the power as function of frequency  $f$ . Similar results were found for the temporal aspects in video [Dong and Atick 1995a; Dong and Atick 1995b]. The  $1/f^2$  slope of power spectra constitutes perhaps the strongest result from the field of natural image statistics, and has several interesting consequences for both human vision, as well as applications

in visual engineering disciplines such as graphics and image processing.

Image gradients and their distributions have also been extensively analysed for natural image sets [Ruderman 1994; Huang and Mumford 1999], range images [Huang et al. 2000], volume visualization [Kindlmann and Durkin 1998] and high dynamic range environment maps [Dror et al. 2001]. Further, rich literature exists on gradients in the logarithmic domain as they represent contrast [Ruderman 1994; Reinagel and Zadow 1999]. The general finding is that gradient histograms show a very sharp peak at zero, falling off sharply away from zero, leading to long tails where higher gradients are located. This behaviour can be attributed, at least in part, to the existence of edges in scenes [Balboa and Grzywacz 2000; Balboa et al. 2001]. In volume visualization, in contrast, where the intensity values usually represent the density of tissue or other material present at that a given location, the gradient histograms tend to show peaks at the interface between materials or objects [Preim and Bartz 2007].

Although most studies assume that image statistics are stationary, i.e., they do not change significantly over the area of an images, in individual images this assumption does not tend to hold. This means that in individual images, the distribution of contrasts may lead to observers altering how they focus their attention. This effect has been studied by means of eye tracking [Reinagel and Zadow 1999].

**Higher order statistics** To capture higher level information about image structure, features and positions, statistics that measure relationships between more than two pixels can be used. Wavelets have been extensively studied [Field 1993; Hurri et al. 1997; Simoncelli and Portilla 1998; Simoncelli 1999b; Huang and Mumford 1999] as well as their relation to cortical cells [Field 1987; Field 1999]. Additionally, sparse coding of natural scenes has been investigated in the context of human vision [Olshausen and Field 1996b; Olshausen and Field 1996a; Field 1999; Willmore and Tolhurst 2001], aiming to derive a set of efficient basis functions that would better represent the response properties of neurons.

Principal [Baddeley and Hancock 1991; Hancock et al. 1992; Zetzsche and Rhrbein 2001] and Independent [Bell and Sejnowski 1996; Bell and Sejnowski 1997; van Hateren and van der Schaaf 1998; van Hateren and Ruderman 1998; Zetzsche and Rhrbein 2001] Component Analysis (PCA and ICA) have also been applied to image ensembles with interesting results, decomposing images to a set of independent components that closely resemble the responses of cortical cells. This has provided the insight that the human visual system appears to decorrelate signals as much as possible.

Although second order statistics can be derived from the amplitude spectrum of Fourier transformed images, it is generally thought that much of the image structure is encoded in the phase spectrum. This has led to the analysis phase statistics. To eliminate the effect of second order correlations, it is possible to whiten the images first, a process whereby the slope of the amplitude spectrum of each image is made zero. Higher order phase analysis has essentially led to the discovery of distributions with high kurtosis [Fleet and Jepson 1993; Thomson 1999b; Thomson 1999a; Thomson 2001].

**Colour statistics** Statistical regularities have also been found when studying distributions of colours [Burton and Moorhead 1987; Párraga et al. 1998; Chiao et al. 2000a; Chiao et al. 2000b; von der Twer and Macleod 2001]. A particularly interesting and useful insight was obtained with a set of spectral

images, which were subsequently converted to cone space, and subjected to Principal Components Analysis [Ruderman et al. 1998]. This rotates the data so that the three resulting colour axes maximally decorrelate the data. The surprising result is that the axes that were found correspond to colour opponency. As a result, the human visual system appears to decorrelate colour data when it computes colour opponent channels.

**Existing graphics applications** Statistical regularities of images have been exploited by a number of computer graphics and vision applications. Wavelets have been used extensively for a variety of purposes from compression and restoration [Simoncelli 1997], to denoising [Simoncelli 1999a; Portilla and Simoncelli 2000a; Wainwright et al. 2000] and texture synthesis [Portilla and Simoncelli 2000b]. Image processing algorithms such as deblurring [Fergus et al. 2006; Levin 2007; Shan et al. 2008] and separation of reflections [Levin and Weiss 2007] have used models of gradient distributions in natural images in order to achieve more plausible results. Finally, a variety of statistics have been utilised in scene classification and comparisons, from joint histograms [Pass and Zabih 1999], to eigenimages [Bischof and Leonardis 2000] and spectral signatures [Torralba and Oliva 2003].

Colour statistics in the form of colour opponent channels have been used extensively, for instance in the computation of colour differences. More recently, colour transfer between images has specifically used the fact that colour opponent channels are decorrelated for natural images [Reinhard et al. 2001a]. Further, colour statistics are used routinely in white balancing, mostly by making the grey-world assumption, i.e. the average reflectance of all objects in a scene is grey.

## 4 Syllabus

1. Introduction - Erik Reinhard (5 minutes)
2. Data Collection and Calibration - Tania Pouli (15 minutes)
3. First Order Statistics - Tania Pouli (15 minutes)
4. Second Order Statistics - Douglas Cunningham (25 minutes)
5. Higher Order Statistics - Douglas Cunningham (15 minutes)
6. Colour Statistics - Erik Reinhard (15 minutes)
7. Conclusions/Summary - Erik Reinhard (5 minutes)
8. Q&A - All (10 minutes)

## 5 Author Biographies

### Erik Reinhard

Department of Computer Science, University of Bristol, UK  
**Email:** reinhard@cs.bris.ac.uk

Erik Reinhard received his Ph.D. in Computer Science from the University of Bristol in 2000, having worked on his Ph.D. for three years at Delft University of Technology, prior to a further three years in Bristol. Following a post-doctoral position at the University of Utah (2000-2002) and assistant professor at the University of Central Florida (2002-2005), he returned to Bristol as a lecturer in January 2006. Erik founded the prestigious ACM Transactions on Applied Perception, and has been Editor-in-Chief since its inception in 2003, until early 2009. He is also guest editor for a special issue on HDRI for Elsevier's Journal of Visual Communication and Image Representation (2007). He

is lead author of 'High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting'. He was program chair for Eurographics PGV 2002 and ACM APGV 2006. He is also program co-chair for the short papers program of the Eurographics main conference in 2008. He is invited speaker for the Society for Information Display's main conference (2007) and keynote speaker for Eurographics 2010. His work on natural image statistics include an assessment of such statistics for use in computer graphics, as well as an image statistics based algorithm for the transfer of colour between two images.

### Tania Pouli

Department of Computer Science, University of Bristol, UK  
**Email:** pouli@cs.bris.ac.uk

Tania Pouli is a PhD student at the University of Bristol. She received a B.Sc. in Computer Software Theory from the University of Bath in 2003 and has been a member of the Interaction & Graphics group in Bristol since then. Her current research focuses on statistical analysis of high dynamic range images and novel image based editing applications using image statistics as priors.

### Douglas W. Cunningham

Brandenburg Technical University Cottbus, Cottbus, Germany  
**Email:** douglas.cunningham@tuebingen.mpg.de

Douglas Cunningham is guest professor for Graphical Systems at the Brandenburg Technical University Cottbus. His current research focuses on the conceptual and practical integration of computer graphics and perception research, with particular emphasis on facial animation and recognition as well as sensorimotor adaptation. Douglas studied Psychology, receiving a B.S. from the University of Washington, a M.S. from the University of Georgia, and in 1997 a Ph.D. from Temple University. In 2007 he received his habilitation in Cognition Science from the University of Tübingen (in Germany). He worked as a senior research scientist for Logicon Technical Services, the Max Planck Institute for Biological Cybernetics, and the University of Tübingen.

## References

- BADDELEY, R. J., AND HANCOCK, P. J. B. 1991. A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proc. Roy. Soc. Lond. B* 246, 219–223.
- BALBOA, R., AND GRZYWACZ, N. 2000. Occlusions and their relationship with the distribution of contrasts in natural images. *Vision Research*.
- BALBOA, R. M., TYLER, C. W., AND GRZYWACZ, N. M. 2001. Occlusions contribute to scaling in natural images. *Vision Research* 41, 7, 955–964.
- BELL, A. J., AND SEJNOWSKI, T. J. 1996. Edges are the 'independent components' of natural scenes. *Advances in Neural Information Processing Systems* 9.
- BELL, A. J., AND SEJNOWSKI, T. J. 1997. The independent components of natural scenes are edge filters. *Vision Research* 37, 3327–3338.
- BISCHOF, H., AND LEONARDIS, A. 2000. Recognizing objects by their appearance using eigenimages. *LECTURE NOTES IN COMPUTER SCIENCE*, 245–265.
- BRADY, M., AND LEGGE, G. 2009. Camera calibration for natural image studies and vision research. *Journal of the Optical Society of America A* 26, 1, 30–42.

- BURTON, G. J., AND MOORHEAD, I. R. 1987. Color and spatial structure in natural scenes. *Applied Optics* 26, 1 (January), 157–170.
- CHIAO, C.-C., CRONIN, T. W., AND OSORIO, D. 2000. Color signals in natural scenes: characteristics of reflectance spectra and the effects of natural illuminants. *J. Opt. Soc. Am. A* 17, 2 (February), 218–224.
- CHIAO, C.-C., OSORIO, D., VOROBYEV, M., AND CRONIN, T. W. 2000. Characterization of natural illuminants in forests and the use of digital video data to reconstruct illuminant spectra. *J. Opt. Soc. Am. A* 17, 10 (October), 1713–1721.
- DONG, D. W., AND ATICK, J. J. 1995. Statistics of natural time-varying images. *Network: Computation in Neural Systems* 6, 3, 345–358.
- DONG, D. W., AND ATICK, J. J. 1995. Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems* 6, 2, 159–178.
- DROR, R., LEUNG, T., ADELSON, E., AND WILLSKY, A. 2001. Statistics of real-world illumination. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2.
- FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S., AND FREEMAN, W. 2006. Removing camera shake from a single photograph. *SIGGRAPH '06: SIGGRAPH 2006 Papers*.
- FIELD, D. J., AND BRADY, N. 1997. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research* 37, 23, 3367–3383.
- FIELD, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 12, 2379–2394.
- FIELD, D. J. 1993. Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In *Wavelets, fractals and Fourier transforms*, M. Farge, J. C. R. Hunt, and J. C. Vassilicos, Eds. Clarendon Press, Oxford, 151–193.
- FIELD, D. 1999. Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions: Mathematical*.
- FLEET, D. J., AND JEPSON, A. D. 1993. Stability of phase information. *IEEE Trans. on PAMI* 15, 12, 1253–1268.
- HANCOCK, P. J. B., BADDELEY, R. J., AND SMITH, L. S. 1992. The principle components of natural images. *Network* 3, 61–70.
- HEALEY, G., AND KONDEPUDY, R. 1994. Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 3, 267–276.
- HUANG, J., AND MUMFORD, D. 1999. Statistics of natural images and models. In *Proc. CVPR 99*, 541–547.
- HUANG, J., LEE, A., AND MUMFORD, D. 2000. Statistics of range images. *Computer Vision and Pattern Recognition*.
- HURRI, J., HYVÄRINEN, A., AND OJA, E. 1997. Wavelets and natural image statistics. In *Proc. of 10th Scandinavian Conference on Image Analysis*, 13–18.
- KIM, S., AND POLLEFEYS, M. 2004. Radiometric alignment of image sequences. *Computer Vision and Pattern Recognition*.
- KINDLMANN, G., AND DURKIN, J. W. 1998. Semi-automatic generation of transfer functions for direct volume rendering. In *VVS '98: Proceedings of the 1998 IEEE symposium on Volume visualization*, ACM, New York, NY, USA, 79–86.
- LEVIN, A., AND WEISS, Y. 2007. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- LEVIN, A. 2007. Blind motion deblurring using image statistics. *Advances in Neural Information Processing Systems*.
- MANN, S., AND PICARD, R. 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. *Proceedings of IS&T 48th annual conference*, 422–428.
- MITSUNAGA, T., AND NAYAR, S. 1999. Radiometric self calibration. *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. 1.
- VAN HATEREN, H., AND VAN DER SCHAAF, A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. R. Soc. Lond. B* 265, 359–366.
- VAN HATEREN, H., AND RUDERMAN, D. L. 1998. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. Lond. B* 265, 2315–2320.
- OLSHAUSEN, B., AND FIELD, D. 1996. Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images. *Nature*.
- OLSHAUSEN, B., AND FIELD, D. 1996. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7: 333–339.
- OSTIAK, P. Implementation of hdr panorama stitching algorithm.
- PÁRRAGA, C. A., BRELSTAFF, G., TROSCIANKO, T., AND MOOREHEAD, I. R. 1998. Color and luminance information in natural scenes. *J. Opt. Soc. Am. A* 15, 3, 563–569.
- PASS, G., AND ZABIH, R. 1999. Comparing images using joint histograms. *Multimedia Systems* 7, 3, 234–240.
- PORTILLA, J., AND SIMONCELLI, E. P. 2000. Image denoising via adjustment of wavelet coefficient magnitude correlation. In *Proc. 7th IEEE Int'l Conf. on Image Processing*.
- PORTILLA, J., AND SIMONCELLI, E. P. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int'l Journal of Computer Vision* 40, 1 (December), 49–71.
- POULI, T., CUNNINGHAM, D., AND REINHARD, E. 2010. Image statistics and their applications in computer graphics. In *Eurographics State of the Art Report (STAR)*.
- PREIM, B., AND BARTZ, D. 2007. *Visualization in Medicine: Theory, Algorithms, and Applications (The Morgan Kaufmann Series in Computer Graphics)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- RAINVILLE, S. J. M., AND KINGDOM, F. A. A. 1999. Spatial-scale contribution to the detection of mirror symmetry in fractal noise. *J. Opt. Soc. Am. A* 16, 9 (September), 2112–2123.
- REINAGEL, P., AND ZADOW, A. M. 1999. Natural scene statistics at the centre of gaze. *Network: Comput. Neural Syst.* 10, 1–10.
- REINHARD, E., ASHIKHMIN, M., GOOCH, B., AND SHIRLEY, P. 2001. Color transfer between images. *IEEE Computer Graphics and Applications* 21 (September/October), 34–41.



- REINHARD, E., SHIRLEY, P., AND TROSCIANKO, T. 2001. Natural image statistics for computer graphics. Tech. Rep. UUCS-01-002.
- REINHARD, E., SHIRLEY, P., ASHIKHMIN, M., AND TROSCIANKO, T. 2004. Second order image statistics in computer graphics. *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*, 99–106.
- RUDERMAN, D., AND BIALEK, W. 1994. Statistics of natural images: Scaling in the woods. *Physical Review Letters*.
- RUDERMAN, D. L., AND BIALEK, W. 1994. Statistics of natural images: Scaling in the woods. *Physical Review Letters* 73, 6, 814–817.
- RUDERMAN, D. L., CRONIN, T. W., AND CHIAO, C. 1998. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A* 15, 8, 2036–2045.
- RUDERMAN, D. L. 1994. The statistics of natural images. *Network: Computation in Neural Systems*, 5, 517–548.
- RUDERMAN, D. L. 1997. Origins of scaling in natural images. *Vision Research* 37, 3385–3398.
- RUDERMAN, D. L. 1997. The statistics of natural images. *Network: Computation in Neural Systems* 5, 4, 517–548.
- VAN DER SCHAAF, A. 1998. *Natural image statistics and visual processing*. PhD thesis, Rijksuniversiteit Groningen, The Netherlands.
- SHAH, S., AND AGGARWAL, J. 1996. Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation\*. *Pattern Recognition* 29, 11, 1775–1788.
- SHAN, Q., JIA, J., AND AGARWALA, A. 2008. High-quality motion deblurring from a single image. *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*.
- SHIH, S., HUNG, Y., AND LIN, W. 1995. When should we consider lens distortion in camera calibration. *Pattern Recognition* 28, 3, 447–461.
- SIMONCELLI, E. P., AND PORTILLA, J. 1998. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proc. 5th Int'l Conf. on Image Processing*.
- SIMONCELLI, E. P. 1997. Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conference on Signals, Systems and Computers*.
- SIMONCELLI, E. P. 1999. Bayesian denoising of visual images in the wavelet domain. In *Bayesian Inference in Wavelet Based Models*, Springer-Verlag, New York, P. Müller and B. Vidakovic, Eds., vol. 141 of *Lecture Notes in Statistics*, 291–308.
- SIMONCELLI, E. P. 1999. Modelling the joint statistics of images in the wavelet domain. In *Proc. SPIE 44th Annual Meeting*, vol. 3813, 188–195.
- THOMSON, M. G. A. 1999. Higher-order structure in natural scenes. *Journal of the Optical Society of America A* 16, 7, 1549–1553.
- THOMSON, M. G. A. 1999. Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems* 10, 2, 123–132.
- THOMSON, M. G. A. 2001. Beats, kurtosis and visual coding. *Network: Computation in Neural Systems*.
- TOLHURST, D. J., TADMOR, Y., AND CHIAO, T. 1992. Amplitude spectra of natural images. *Ophthalmic and Physiological Optics* 12, 229–232.
- TOLHURST, D. J., AND TADMOR, Y. 1997. Discrimination of changes in the slopes of the amplitude spectra of natural images: Band-limited contrast and psychometric functions. *Perception* 26, 8, 1011–1025.
- TORRALBA, A., AND OLIVA, A. 2003. Statistics of natural image categories. *Network: Computation in Neural Systems* 14, 391–412.
- VAN DER SCHAAF, A., AND HATEREN, J. V. 1996. Modelling the power spectra of natural images: statistics and information. *Vision Research* 36, 17, 2759–2770.
- VON DER TWER, T., AND MACLEOD, D. 2001. Optimal nonlinear codes for the perception of natural colours. *Network: Computation in Neural Systems*.
- WAINWRIGHT, M. J., SIMONCELLI, E. P., AND WILLSKY, A. S. 2000. Random cascades of gaussian scale mixtures and their use in modelling natural images with application to denoising. In *Proc. 7th IEEE Int'l Conf. on Image Processing*.
- WEBSTER, M. A., AND MIYAHARA, E. 1997. Contrast adaptation and the spatial structure of natural images. *Journal of the Optical Society of America A* 14, 9 (September), 2355–2366.
- WILLMORE, B., AND TOLHURST, D. 2001. Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*.
- WILLSON, R. 1994. Modeling and calibration of automated zoom lenses.
- ZETZSCHE, C., AND RHRBEIN, F. 2001. Nonlinear and extra-classical receptive field properties and the statistics of natural scenes. *Network: Computation in Neural Systems*.
- ZIEGAUS, C., AND LANG, E. W. 1997. Statistics of natural and urban images. In *Proc. 7<sup>th</sup> International Conference on Artificial Neural Networks*, Springer-Verlag, Berlin, vol. 1327 of *Lecture Notes in Computer Science*, 219–224.
- ZIEGAUS, C., AND LANG, E. W. 1998. Statistical invariances in artificial, natural and urban images. *Z. Naturforsch* 53a, 1009–1021.

# Image Statistics: from Data Collection to Applications in Graphics

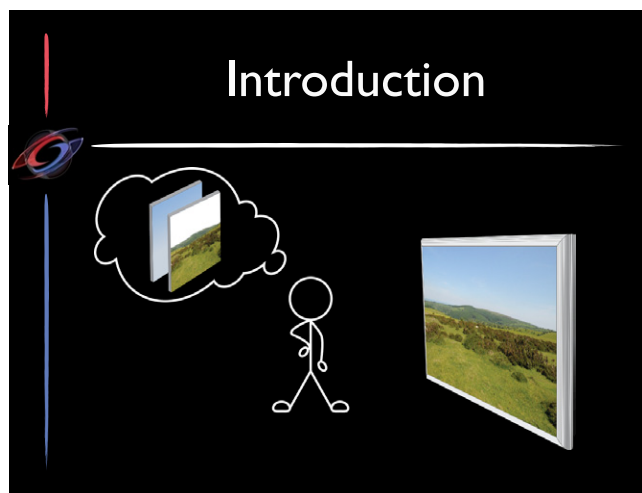
Erik Reinhard  
Tania Pouli  
Douglas Cunningham

**SIGGRAPH2010 Course**

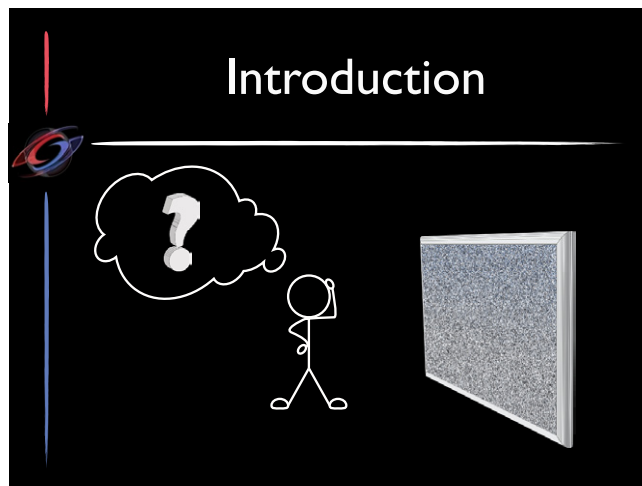
The statistics of natural images have attracted the attention of researchers in a variety of fields and have been used as a means to better understand the human visual system and its processes. A number of algorithms in computer graphics, vision and image processing take advantage of such statistical findings to create visually more plausible results. With this report we aim to review the state of the art in image statistics and discuss existing and potential applications within computer graphics and related areas.

## Syllabus

- Introduction - Erik (5 mins)
- Data collection & calibration - Tania (15 mins)
- First order statistics - Tania (15 mins)
- Second order statistics - Douglas (25 mins)
- Higher order statistics - Erik (15 mins)
- Color statistics - Erik (15 mins)
- Conclusions/Summary - Erik (5 mins)
- Q&A - All (10 mins)

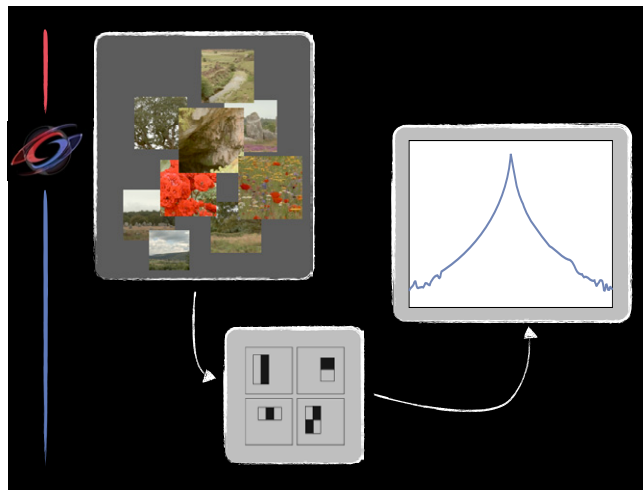


The field of natural image statistics studies images and their statistical regularities. The human visual system has evolved in natural environments, i.e. those without man-made structures. Human vision, it is argued, has therefore evolved to be particularly good at observing natural scenes...

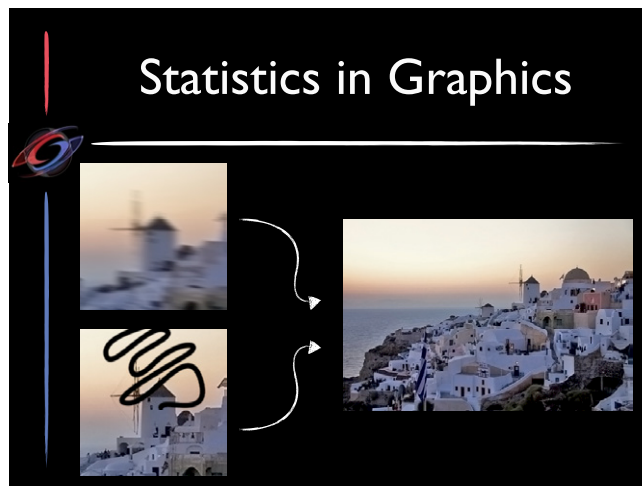


...as opposed to random images.

The number of random images is many orders of magnitude larger than the number of natural images. Thus, natural images form a very small subset of all possible images, and contain structure to which the human visual system is optimized. By studying the input to the human visual system, it is thought that the human visual system can be understood better.



The typical process by which natural image statistics are computed, is to pass each image of an ensemble through a particular transform. The transforms typically employed tend to correspond to a component of human vision. It is, for instance, possible to compute the power spectra of each image, and from these compute the average power spectral slope for the whole set. This course reviews this and many other transforms and statistics.



Thus, the reason for this course is that, aside from learning about human vision and the implications for perception, some of these statistical regularities have transferred to computer graphics: several applications now successfully employ statistics of natural images to help solve engineering problems. As an example, deblurring algorithms rely on optimization involving priors that help push candidate images towards visually plausible solutions. Of course, visually plausible solutions are those that appear to humans in some sense natural. Image priors have therefore been designed with the aid of specific knowledge of natural image statistics, in particular regarding the average distribution of image gradients.

We think that the use of natural image statistics has so far shown great promise, with interesting applications making use of some of the available statistical regularities. We also see excellent scope for further future developments.

# Types of Statistics

- **First order**

- Each pixel viewed independently

- **Second Order**

- Relations between pairs of pixels

- **Higher Order**

- How does a pixel relate to more than one other pixel in the image?

For the rest of this course, we will classify statistics depending on the number of pixels considered. First order statistics look at individual pixels, thus ignoring spatial relations. Second order statistics consider pairs of pixels while higher order statistics refer to any transform involving more than two pixels.

## Data Collection & Calibration

Tania Pouli



# What type of data?

- Depends on the purpose of the study
- LDR vs HDR:
  - What are the typical lighting conditions encountered?
- Field of view:
  - Normal/Wide/Spherical (360°)

The statistical analysis of natural images typically involves the collection of large ensembles. Various image data sets are already available but in order to create new ensembles, several points need to be considered. One might need to create new ensembles if the intended application involves a specific class of images, or if the study is to understand human vision under specific viewing conditions. Advances in camera technology have made the capture of higher resolution and higher dynamic range imagery possible. Each such type of image presents its own advantages in terms of information and its own specific issues regarding capturing and calibration. The type of image appropriate for each study largely depends on the type of information studied.

For most purposes, low dynamic range (LDR) imagery is sufficient, although HDR images can capture a wider range of luminance values, making them more appropriate for scenes of more extreme lighting (such as night scenes where artificial light sources are much brighter than the rest of the scene). Panoramas or wider angle images can capture a larger part of the field of view, or even the full scene in the case of spherical panoramas.

# What type of scene?



The type of scene to be captured is another important consideration, and again the answer depends on the purpose of the study. Typically, studies coming from the area of human vision tend to focus on natural images (i.e. those without manmade structures). If on the other hand the purpose of the analysis is to derive a scene categorization algorithm then different scene types may be appropriate.

# How many images?

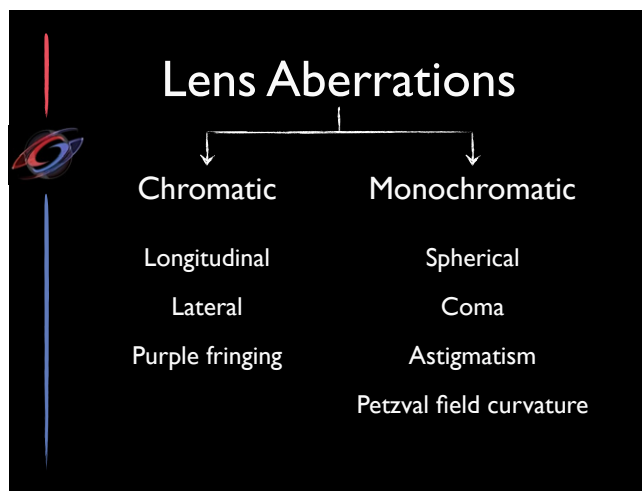
Study	Image Type	# of Images	Resolution
Ziegeus et al. (1998)	8 bit CCD Video Frames	100	256 x 256
Ruderman et al. (1998)	Spectral Images	12 scenes x 43 images	192 x 165
van Hateren & van der Schaaf (1998)	8 bit CCD Video Frames	276	768 x 512
Huang et al. (2000)	Range Panoramic Images	205	444 x 1440
Huang & Mumford (1999)	Calibrated Grayscale Images (from [vv98])	4000	1024 x 1536
Dror et al. (2001)	HDR Spherical Panoramas	95 (Teller:2003) + 9 (Debevec)	various

As for the number of images necessary, although generally larger data sets afford more accurate statistical findings, in the case of image statistics, the trade-off between number of images and data collection time often needs to be resolved. The number of images appropriate for a particular study largely depends on its aims. Here, some example numbers from existing studies are given.

## Calibration

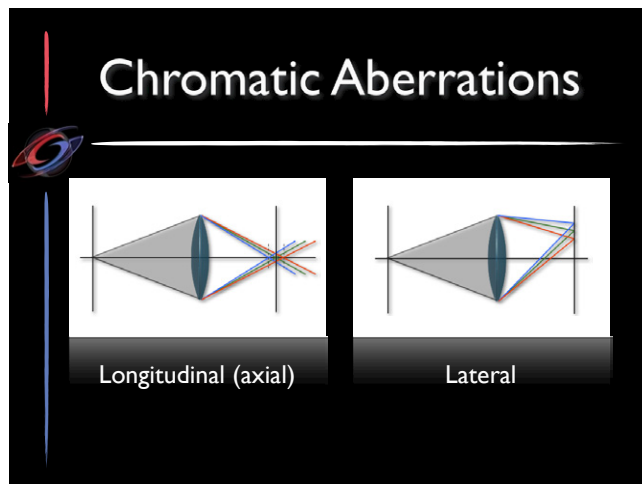
- Aberrations
  - Chromatic/Achromatic
- Noise
- Vignetting
- Camera response
- Merging exposures

When measuring the statistical regularities of an image ensemble, some care is necessary to ensure that irregularities and artifacts due to the camera, lens or shooting conditions do not affect the results. To avoid such an effect, the properties of the equipment used need to be accounted for. Moreover, to ensure the images are calibrated and therefore pixel values between them are directly comparable, properties of the scene may need to be measured.



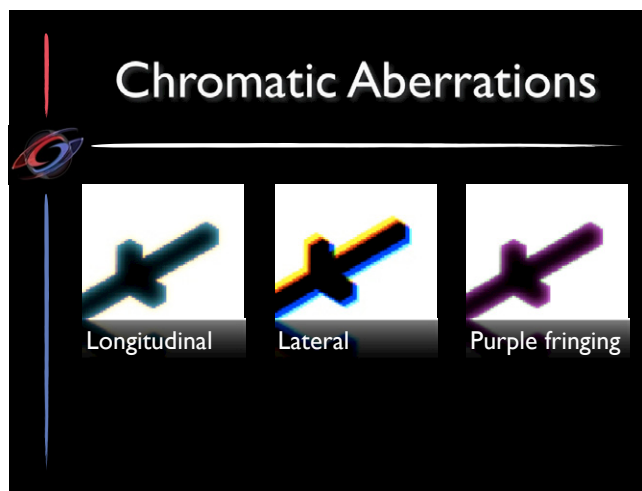
The lens is responsible for focusing the light coming from the scene to an image plane (which may be a sensor in digital cameras or the film). For many applications, it is sufficient to model the camera as a simple pinhole whereby no lens is present and the aperture through which the light enters is a point. To model the various effects of an optical system to the image however, more complex lenses need to be considered.

Imperfections in the design or construction of such lenses can cause them to violate the assumption of Gaussian optics. These deviations are known as aberrations and can be broadly categorized as chromatic, where the optical system behaves differently for different wavelengths, and monochromatic, where the distortions in the image are apparent in all channels.



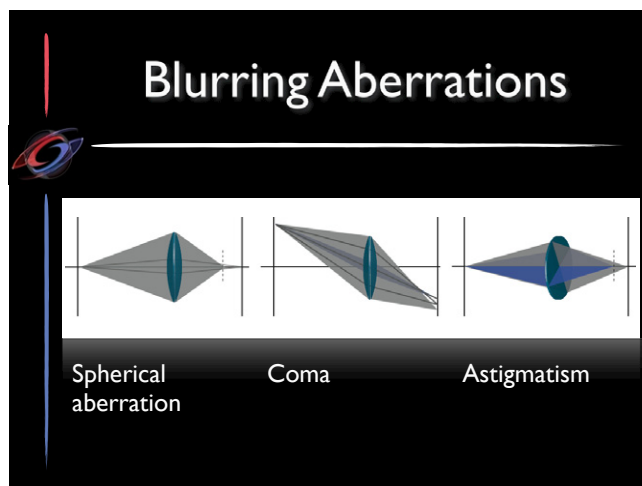
Lens elements can have wavelength-dependent indices of refraction, resulting in different colors focusing in slightly different places. This causes longitudinal or axial chromatic aberration, which appears as chromatic image blur. Typically this form of aberration is managed optically, using a pair of thin lenses with different refractive indices, known as achromatic doublets. Such systems can reduce longitudinal aberration effects but not always completely remove them.

Another type of chromatic aberration is lateral chromatic aberration, which is the result of different wavelengths hitting the image plane at slightly different positions. This is more commonly known as color fringing and can be reduced digitally. This can typically be achieved by realigning the color channels of the image which can compensate for the effect of the color shifting to some extent, or for a more accurate reconstruction, a more complete characterization of the optical system can be utilized.



One particular case of chromatic aberration specific to digital cameras, especially ones using a charge coupled device sensor (CCD), is known as purple fringing. This effect is caused by a combination of factors that operate in addition to the lens aberrations described earlier. Highlights in the scene that are overly bright can cause some of the CCD quantum-wells to flood, creating the visual effect of blooming. Additionally, a false color effect can be created by the demosaicing process.

A number of digital solutions are available for correcting the effects of chromatic aberrations in images with various levels of accuracy. DxO Optics handles chromatic aberrations by taking into account the characteristics of the lens where possible. A similar approach is also found in PTLens. Adobe Photoshop (version CS4 tested) provides simple sliders to account for red/cyan and blue/yellow shifting.



A wider range of artifacts in images appears as the result of monochromatic aberrations. These include blurring aberrations and geometric distortions. Blurring aberrations can be further divided into spherical aberration, coma, astigmatism and Petzval field curvature.

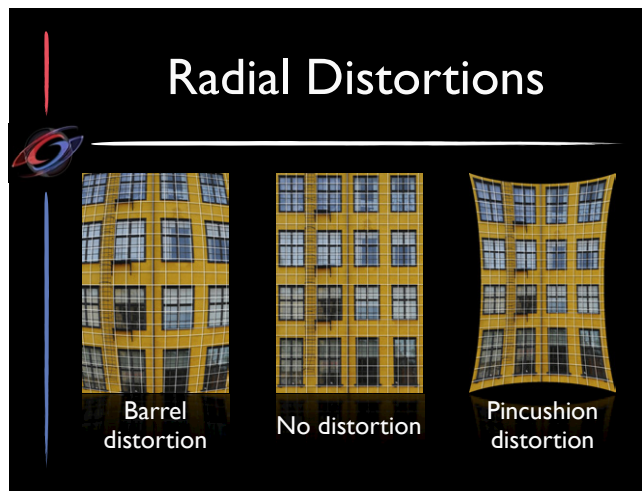
Spherical lenses are commonly used in optical systems as they are relatively easy to manufacture, but their shape is not ideal for the formation of a sharp image. Spherical aberrations are the consequence of light rays further from the lens axis (marginal) focusing at a different position than rays passing through the middle of the lens. When marginal rays focus nearer the lens than the principal rays, the effect is known as positive or undercorrected spherical aberration, while a marginal focus further away from the lens than the paraxial focus causes negative or overcorrected spherical aberration.

Comatic aberrations cause objects further from the optical axis to appear distorted. Positive coma occurs when marginal rays focus further away from the optical axis than principal rays, while negative coma is the opposite effect, namely the marginal rays focusing closer to the optical axis.

A third type of blurring aberration is known as astigmatism and is the result of off-axis rays of different orientations focusing at slightly different positions. The meridional plane for an off-axis object point is defined by that point and the optical axis. The plane orthogonal to the meridional plane is known as the saggital plane. Rays along these two planes focus at different positions on the optical axis. As such, if the lens is focused such that the meridional image is sharp, blur will occur in the saggital direction and vice versa.

A final aberration considered here is a consequence of the field curvature. The image and object planes are generally considered to be planar. However, in the case of spherical lenses they are in fact spherical, which is known as Petzval field curvature. Regions of the image further from the optical axis increasingly deviate from the planar assumption. This is more evident when a planar image plane is used (such as a sensor) in which case the image is less sharp towards the periphery.

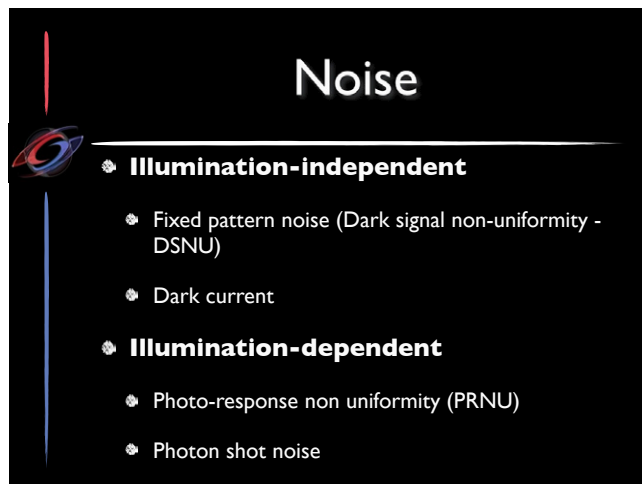
The effect of blurring aberrations can be reduced in post-processing by sharpening the image or during the image capture by using smaller aperture settings and thus only allowing rays closer to the optical axis to enter. Optical solutions do exist and generally involve the use of lens doublets or carefully designed lenses. These solutions however, complicate the design of the optical system and are as a result found in more expensive lenses.



A different class of monochromatic aberrations consists of radial distortions. These cause straight lines in the scene to appear curved in the resulting image and can be corrected digitally either manually or using information about the characteristics of the lens that was used. Radial distortions are the result of deviations from a linear projection model and are more pronounced for wider angle lenses. Although interactions between the various lens elements of a complex optical system can produce some less well defined distortions, the two main types considered here are barrel and pincushion distortions.


Barrel distortion is the result of decreasing lateral magnification and it causes straight lines to curve away from the image center. Pincushion distortion occurs in lenses with increasing lateral magnification and causes straight lines to curve towards the image center. Radial distortions can be adequately estimated as a fourth degree polynomial. Several methods have been proposed in the literature for estimating the distortion parameters of that polynomial. In the simplest case, by photographing a test scene containing straight lines, the distortion parameters can be adjusted until the lines in the resulting image are also straight.

Various software packages offer solutions that reduce or remove the effect of radial distortions. Options such as DxO Optics or PTLens use lens-specific information while other solutions exist that require more manual input, such as for instance Hugin or the Lens Correction module in Adobe Photoshop.



Noise can be a major cause of image degradation and can be defined as any random variation in brightness or color in the image caused by the sensor, circuitry or other parts of the camera. Noise can have various sources, each requiring different treatment. Generally, noise can be illumination-independent, in which case the amount of light hitting the film or sensor does not have an effect on the amount of noise, or illumination-dependent if there is a correlation between that type of noise and the illumination in the scene.





# Noise

---

## Fixed pattern noise:


“any noise component that survives frame averaging”

- DSNU - capture and subtract image with no illumination (lens cap on)
- PRNU - capture flat, constantly illuminated scenes

Fabrication errors can cause different pixels of a sensor array to have different responses to constant illumination. This is known as fixed pattern noise and appears in two forms.

Dark signal non-uniformity (DSNU) is the fixed offset from the average across each pixel of the image sensor, occurring even when no illumination is present.

Photo response non-uniformity (PRNU) is the variation in the responsivity of pixels under constant illumination.



# Noise

---

## Dark current:

Depends on temperature, time, exposure...

## Photon Shot noise:

- Due to random fluctuations in number of photons hitting the sensor

Both are more apparent in low light conditions.

Even if no light reaches the sensor, electrons can still reach the quantum-wells, generating dark current shot noise. The severity of this type of noise is affected by temperature and consequently can be reduced by cooling the sensor.

The number of photons collected by each individual quantum-well of an image sensor is not constant for a given intensity. Due to this randomness, a given pixel can receive more or fewer photons than the average. The distribution of photons arriving at the sensor can be modeled by Poisson statistics (mean  $\mu$  equals variance  $\sigma^2$ ). As such, if the intensity of the source increases, so will the variance in the signal.

Most of the noise sources discussed do not depend on illumination levels in the scenes. As such, in darker scenes, the signal-to-noise ratio becomes smaller, effectively amplifying the effect of noise in the image. Although dark current and fixed pattern noise can be minimized through the use of appropriate calibration, other noise sources (including photon shot noise) cause less predictable patterns. Refer to the work by Healey and Kondepudy for a detailed account on noise estimation and camera calibration.

- HEALEY G., KONDEPUDY R.: Radiometric ccd camera calibration and noise estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 3 (1994)

# Radiometric Calibration

- Cameras map scene radiance to pixel values in a **non-linear** manner
- **Camera characterization** - recovers relation between radiance and pixel values ( $XYZ \rightarrow RGB$ )
  - monochromator + radiance meter
  - color target + spectrophotometer



To map pixel values to real-world radiometric quantities, the response function of the camera, which is the mapping between sensor illuminances to pixel values, needs to be known. Although this information is not generally available, the non-linearities for a particular camera can be recovered through the use of calibrated input. The relationship between scene radiance and pixel values can be recovered in two ways and the process is known as **camera characterization**.

Several approaches exist for finding the transformation between the measured XYZ values and the captured RGB values. **Look-up tables store the data of such measured/captured pairs.** If a large enough number of samples is collected so that the camera gamut is densely populated, then XYZ values corresponding to RGB samples not in the look-up table can be estimated using three-dimensional interpolation. Alternatively, if the number of samples available is not sufficient for interpolation, **a function can be fitted to the available data.** This is however much more computationally expensive.

## Camera Response

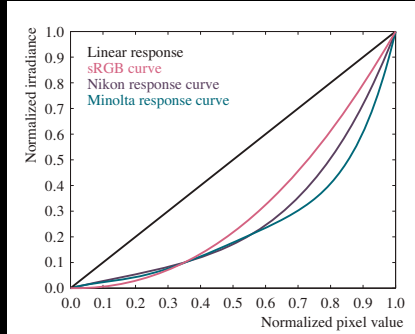


Linear

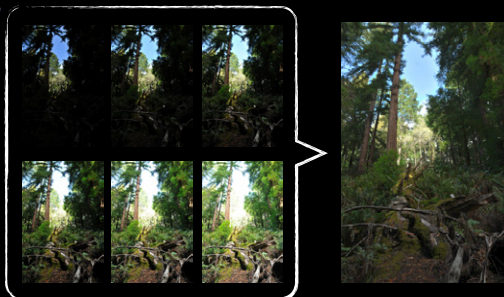
sRGB

As discussed, cameras are not linear light meters. Although these non-linearities do make more detail in the image visible (much like gamma correction or the sRGB curve), there are cases where a linear relation between pixel values and luminance values is required. One such example is merging exposures into an HDR image.

# Camera Response Curve



# HDR Merging



If the accurate representation of the scene illumination is necessary, a single exposure may not be sufficient to cover the full dynamic range in a scene. A series of exposures can be captured, which can then be merged into a single high dynamic range image. Each individual exposure follows the non-linearities of the camera. As such, before they can be merged into an HDR image, the different exposures need to be linearized using the response curve of the camera. The techniques described previously can be used to recover this response curve.

Alternatively, the sequence of differently exposed images can be used directly to recover the response curve. A commonly used method is proposed by Debevec and Malik (1997). By exploiting a property of both physical film and digital sensors known as reciprocity (meaning that halving the irradiance hitting the sensor  $E$  and simultaneously doubling the exposure time  $\Delta t$  will result in the same pixel values  $p$ ), a linear system is derived. The solution to that is the inverse function  $g^{-1}$  to the camera response curve.

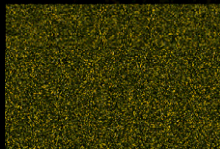
Once the individual exposures are linearized, a set of weights can be used to define their contribution for the final HDR image. Several weighting functions have been proposed in the literature. Mann and Picard merge the exposures using certainty functions computed from the derivatives of the response curves for each differently exposed image. A simpler approach is used in where a hat function suppressing values that are likely to be under- or over-exposed is applied. Finally, inspired by signal theory, Mitsunaga and Nayar propose a weighting function that takes into account the signal-to-noise ratio. Despite their differences, any of these approaches will yield satisfactory results.

In terms of software, many packages are capable of merging exposures into an HDR image. Photoshop offers a 'Merge to HDR' function since its CS2 version that allows some manual control over the image alignment. Other packages include HDRSoft's Photomatix and HDRShop, both of which offer batch processing functionality. Finally, a free option that is only available for Apple computers is Greg Ward's Photosphere, which is also an excellent viewer for HDR image collections.

# First Order Statistics

Tania Pouli

## First Order

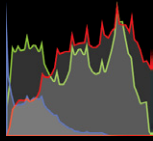


- Each pixel considered independently
- Location invariant
- Easy to compute & interpret
- First order statistics:
  - histogram moments
  - contrast\*

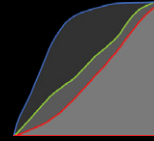
The simplest regularities in images that can be explored through statistical analysis are properties of single pixels. The distribution of pixel values in images is typically represented with histograms. First order statistics do not capture relationships between pixels and as such are not suitable for studying spatial regularities within images. These two images show an example of two very different images that would result in identical first order statistics. The right image is constructed by randomly permuting the pixels of the left image, yet it appears unnatural. Statistical analysis of the distribution of pixel intensities can however lead to interesting insights.

# Intensity Histograms

- How often does each intensity value occur?
- Histograms can be very important in data analysis, image analysis, and visualization



Individual frequency  
of occurrence



Cumulative frequency  
of occurrence

# Histogram Moments

- They give information about the shape of the distribution
- A measure of Gaussianity

$$m_k = \sum_{p=1}^N \frac{(I(p) - c)^k}{N}$$

Statistical moments are commonly employed to quantitatively describe the shape of a distribution. The  $k$ th moment of a distribution can be computed as shown where  $c$  can be any constant. Generally, if  $c = 0$  then the above equation computes the raw moments of the distribution, while setting  $c = \mu$  gives us the central moments (i.e. centered at the mean). The first raw moment is the mean  $\mu$  of the distribution and the second is the variance  $\sigma^2$ .



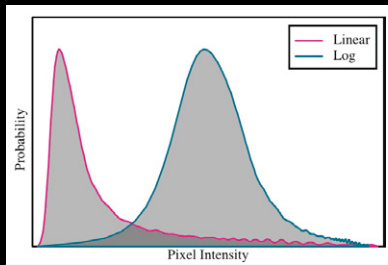
# Histogram Moments

- ❁ **1st moment:** mean
- ❁ **2nd moment:** variance
- ❁ **3rd moment:** relates to skewness  $S = \frac{m_3}{\sigma^3}$
- ❁ **4th moment:** relates to kurtosis  $\kappa = \frac{m_4}{\sigma^4}$

The meaning of further moments is less straight forward but the skewness  $S$  and kurtosis  $\kappa$  of a distribution relate to the third and fourth moments respectively. More specifically, the skewness and the kurtosis are defined as shown, where  $m_3$  and  $m_4$  are the third and fourth central moments respectively.

Huang and Mumford analyzed more than 4000 greyscale images of natural scenes (taken from the database created by J. H. van Hateren) by computing central moments of the log histograms. The value of the skewness shows that the distribution is not symmetrical, which can be attributed, at least partly, to the presence of sky in many of the images, resulting in a bias towards higher intensities. In addition to that, the values of both the skewness and the kurtosis show that the distribution is **non-Gaussian**.

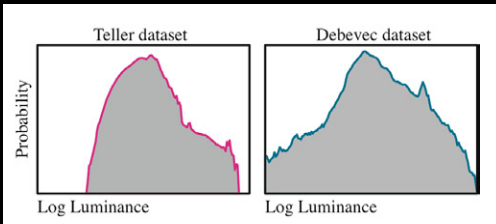
# Luminance Histograms



Local contrast in natural images: normalization and coding efficiency - Brady & Field (2000)

A less skewed distribution was found by Brady and Field (2000) in their analysis of 46 logarithmically transformed natural images, while the linear images resulted in a distribution skewed towards darker values. Although no exact values were provided for the moments of the distributions, the resulting histograms for both cases are shown here. As can be seen from the above examples, although in both cases natural images were used and analyzed, the results vary. Generally, the distribution of log intensities for natural images does not deviate far from symmetric. Results do depend however on the choice of images.

# Luminance Histograms



Surface reflectance estimation and natural illumination statistics - Dror et al. (2001)

One particular case of interest was explored by Dror et al. (2001) where high dynamic range spherical panoramas were analyzed. Two separate datasets were used consisting of 95 images taken from the Teller database (<http://city.lcs.mit.edu/data>) and 9 from Debevec's Light Probe Image Gallery (<http://www.debevec.org/Probes/>). The resulting log intensity histograms for the two datasets are shown here.

## Histograms relate to surface properties:

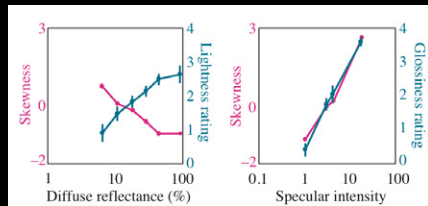
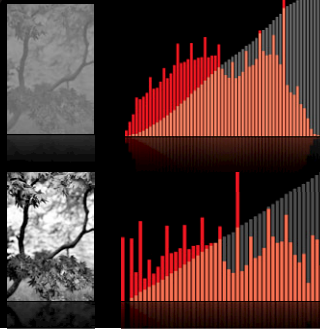


Image statistics and the perception of surface qualities - Motoyoshi et al., 2007

The shape of a histogram can also be associated with particular qualities of the depicted surface. Motoyoshi et al. studied the perceived surface qualities of various materials and analyzed their relation to the associated histograms. They found that materials that are perceived as glossier tend to have a positively skewed distribution while matte surfaces result in a negatively skewed distribution. The correlation found between the surface properties and the skewness of the histograms is shown here. As the diffuse reflectance of the surface increases, the lightness rating as perceived by the observers also increases while the corresponding histogram skewness decreases (left). An increase in specular intensity also results in an increased rating of glossiness as well as higher skewness value (right).

- MOTOYOSHI I., NISHIDA S., SHARAN L., ADELSON E. H.: Image statistics and the perception of surface qualities. Nature 447, 7141 (2007)

# Histogram Equalization

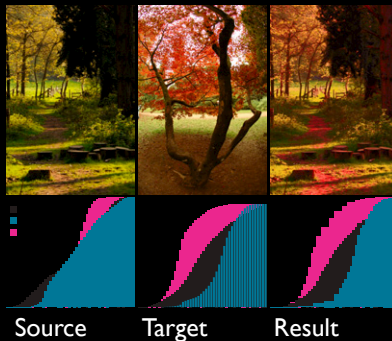


- Images do not use intensity values evenly
- Equalizing the values in the **cumulative** histogram can improve contrast

Despite their simplicity, first order statistics have now found several applications in image processing. Studies have shown correlations between first order statistical regularities in images and properties of the illuminant which has proved useful in areas such as white balancing. Moreover, transferring statistical moments between images in appropriate color spaces has been demonstrated in what is now known as color transfer. These color-related application will be discussed in more detail later in the course.

First order statistics can also be computed on a single image basis. By manipulating the distribution of values of a single image, a variety of effects can be achieved. For instance, the contrast of an image that only covers a small portion of the available range of intensities can be increased by adjusting the pixel values such that the full range of intensities is more equally represented. This is known as **histogram equalization**.

# Histogram Matching



A more general version of histogram equalization is known as histogram matching and it allows the histogram of a source image to be matched to that of a given target. First, the cumulative histograms  $C$  of the source and target are computed. The inverse function  $C^{-1}$  is then computed, which acts as a reverse lookup on the histogram, returning the bin index corresponding to a given count. An example of this technique applied on a source-target pair of color images is shown here together with the corresponding histograms for the 3 channels.

# Is 1st Order Enough?

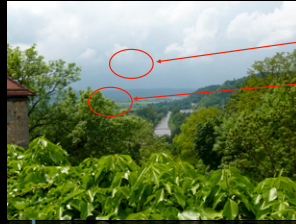
- Simple to compute and interpret BUT...
- No spatial information
- No information on relations between pixels
- We need 2nd/higher order statistics for that!

Due to their limitations to single pixel properties, first order statistics are the simplest to compute and interpret as already discussed. To further explore the relations between pixels, their spatial configurations and their associations with aspects of human vision more complex transforms are necessary. Second and higher order statistics can be used to provide the additional information and they will be discussed in the following parts of this course. It is important though to note that the statistical tools presented in this part of the course can be used directly on the intensities of an image, leading to first order statistical results, or on distributions resulting from other transforms, leading to 2nd or higher order statistical findings.

## Second Order Statistics

Douglas Cunningham

# Second Order



- Two main features in natural images:
  - Large homogeneous sections:  
**Surfaces**
  - Large gradients between surfaces:  
**Edges**
- **Relations between Neighboring pixels**
- Second order Statistics:
  - Gradients
  - Power Spectrum

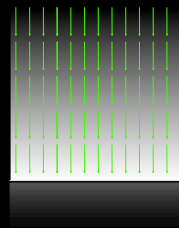
While the first order statistics of an image provide considerable information, their focus on individual pixels restricts them to the “unstructured” aspects of the image equivalent to an investigation of individual rays of light. As James J. Gibson (The ecological approach to visual perception. 1979) famously pointed out, since the interaction of light rays with an object systematic alters light according to the surface properties of the object, the pattern of change across a set of light rays – which he referred to as the structured optic array – provides direct, reliable information about the object. this higher-order structure is of central interest to the human visual system. It should not be surprising, then, that considerable attention has been directed towards investigating second and even higher order statistics.

Natural images consist of two main features: large homogeneous surfaces where there is little variation in intensities and large gradients between such surfaces that correspond to edges. Second order statistics capture the relations between pairs of pixels, allowing us to study these features in images.

# Gradients

**Greatest rate of change**

$$\nabla I(x, y) = \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix}$$



Perhaps the simplest way to examine information contained in more than a single pixel, is to look at the relationship between pairs of pixels. This is a discrete approximation of the first derivative or gradient of the intensity function of an image and represents the rate of change. There are three different discrete approximations: forward/backward difference, central difference, and the sobel operator.



# Gradient Computation



## Forward Difference

**Taylor Series:**

$$I(i+h, j) = I(i, j) + hI'(i, j) + \frac{h^2}{2}I''(i, j) + \dots$$

$$I(i+h, j) - I(i, j) = hI'(i, j) + \frac{h^2}{2}I''(i, j) + \dots$$

$$\frac{I(i+h, j) - I(i, j)}{h} \approx I'(i, j)$$

$$\nabla I_{i,j} = \begin{bmatrix} \frac{\partial I}{\partial i} \\ \frac{\partial I}{\partial j} \end{bmatrix} \approx \begin{bmatrix} (I_{i+1,j}) - (I_{i,j}) \\ (I_{i,j+1}) - (I_{i,j}) \end{bmatrix}$$

		$I_{i+1,j}$	
$I_{i-1,j}$	$I_{i,j}$	$I_{i+1,j}$	
	$I_{i,j-1}$		

$$\|\nabla I_{i,j}\| \approx D_{i,j} = \sqrt{(I_{i+1,j} - I_{i,j})^2 + (I_{i,j+1} - I_{i,j})^2}$$

The most straightforward is the forward difference method. Here, the gradient at a pixel (i,j) is calculated from the difference between it and the next pixel forwards as shown, where  $I(i,j)$  is the luminance for pixel i, j,  $D_x$  is the horizontal gradient and  $D_y$  the vertical gradient.

# Gradient Computation



## Backward Difference

$$\frac{I(i, j) - I(i-h, j)}{h} \approx I'(i, j)$$

$$\|\nabla I_{i,j}\| \approx \begin{bmatrix} (I_{i,j}) - (I_{i-1,j}) \\ (I_{i,j}) - (I_{i,j-1}) \end{bmatrix}$$

		$I_{i,j+1}$	
$I_{i-1,j}$	$I_{i,j}$	$I_{i+1,j}$	
	$I_{i,j-1}$		

An obvious variant on this, referred to as the backward difference method, is to use the previous pixel (e.g.,  $I(i-1,j)$ ) rather than the next one. In both cases, it is common to calculate the mean gradient magnitude at a given location from the horizontal and vertical components although one may choose to keep the vertical and horizontal gradients separate.

# Gradient Computation

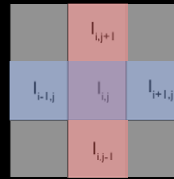


## Central Difference

### Taylor Series again

$$\nabla I(x, y) \approx \begin{bmatrix} \frac{I(i+h, j) - I(i-h, j)}{2h} \\ \frac{I(i, j+h) - I(i, j-h)}{2h} \end{bmatrix}$$

$$\nabla I_{i,j} = \begin{bmatrix} \frac{1}{2}(I_{i+1,j} - I_{i-1,j}) \\ \frac{1}{2}(I_{i,j+1} - I_{i,j-1}) \end{bmatrix}$$



A slightly more robust method is to calculate the central differences by considering the pixels before and after the pixel in question. As a larger neighborhood is considered, this approach results in a better approximation.

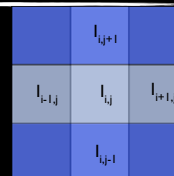
# Gradient Computation



## Sobel Operator

$$I_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

$$I_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

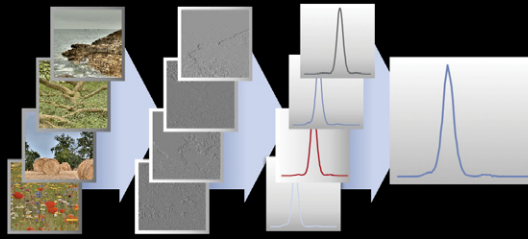


$$||\nabla I(i, j)|| = D_{i,j} = \sqrt{I_x(i, j)^2 + I_y(i, j)^2}$$

Computationally expensive, but nearly isotropic

An even larger neighborhood can be taken into account by using the Sobel operator which also looks at the diagonal pixels. This is nearly isotropic but is more computationally expensive than the simpler approaches.

# Gradient Distribution

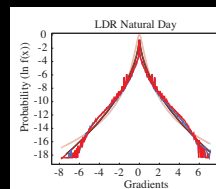


To compute the gradient distribution for an image ensemble, the individual distributions for each image need to be computed first and then averaged. One may choose to consider horizontal and vertical gradients separately or average them.

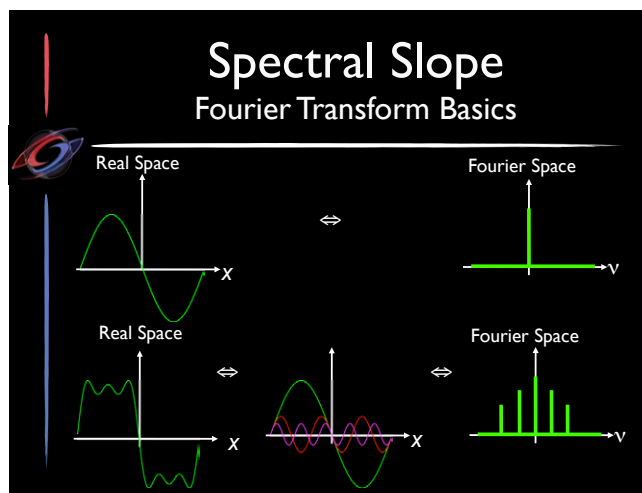
# Gradient Distribution



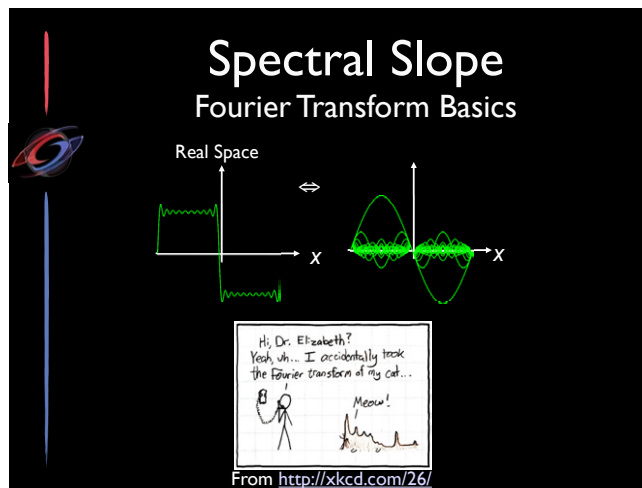
- Sharp peak at zero:
  - The world is mostly homogeneous surfaces
- Rapid fall off:
  - Some edges, mostly low contrast
- Symmetrical fall off:
  - Surfaces over same background
- Horizontal and vertical gradients tend to be similar



It has been repeatedly found that the gradient histogram has a very sharp peak at zero and then falls off very rapidly (with long tails at the higher gradients). The distribution can be modeled as  $e^{-x^\alpha}$  with  $\alpha < 1$ . The reason for the specific shape of the gradient distribution seems to be that images contain many, large surfaces which tend to be smooth and somewhat homogenous (so that the gradients tend to be near zero) along with a few high contrast edges (which will yield very high gradients). An object often finds itself over a consistent background, which means that the transition from the object's surface to the background will be similar all along the object's silhouette. This is reflected in the symmetry of the gradient distribution around the central peak.



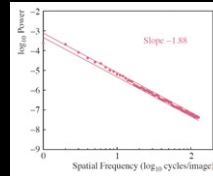
Images contain information at different spatial scales, from the large to very small. As is well known from Fourier analysis, sharp edges, can be described by a weighted sum of sinusoids, with higher frequencies being weighted less. In other words, natural images contains information at many spatial frequencies. An examination of the relative power of the different spatial frequencies reveals several interesting trends, which are so prominent, that most work on image statistics provides an analysis of the power spectrum.



# Spectral Slope

- Spectra of individual images varies
- AVERAGE spectra follow power law:  

$$A \approx \frac{1}{f^\beta}$$
- Humans most sensitive to slopes between 2.8 and 3.2



The power spectrum  $S(u,v)$  of an  $N \times N$  image is given by:

$$S(u,v) = |F(u,v)|^2 / N^2$$

where  $F$  is the Fourier transform of the image, and  $(u,v)$  are pixel indices in Fourier space. Two-dimensional frequencies can also be represented with polar coordinates  $(f,\theta)$ , using  $u = f \cos(\theta)$  and  $v = f \sin(\theta)$ .

While the spectra of individual images varies considerably (which may play a role in the rapid detection of certain scenes or objects), when the averaging over a sufficiently large number of images (and across orientation), a clear pattern arises: the lower frequencies contain the most power, with power decreasing as a function of frequency. In fact, on a log-log scale, amplitude as function of frequency lies approximately on a straight line. That is, the averaged spectrum tends to follow a power law, which can be well modeled with  $P = 1/f^\beta$ , where  $P$  is the power as function of frequency  $f$ , and  $\beta$  is spectral slope.

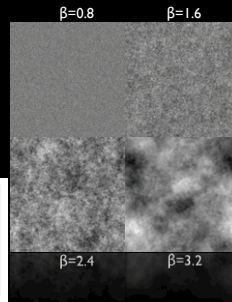
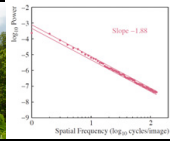
# Spectral Slope

Study	# of Images	$\beta \pm 1sd$
Burton G. J., Moorhead I. R. 1987	19	$2.1 \pm 0.24$
Dong D., Atick J. 1995	320	2.30
Dror R. O., Adelson E. H., Willsky A. S. 2001	95	2.29
Field D. J. 1987	6	2.0
Field D. J. 1993	85	2.20
Field D. J., Brady N. 1997	20	$2.20 \pm 0.28$
van Hateren J. 1992	117	$2.13 \pm 0.36$
Huang J., Mumford D. 1999	216	1.96
Parraga C. A., Brelstaff G., Troscianko T. 1998	29	$2.22 \pm 0.26$
Reinhard E., Shirley P., Troscianko T. 2001	133	$1.88 \pm 0.42$
Ruderman D., Bialek W. 1994	45	1.81
van der Schaaf A., van Hateren J. 1996	276	$1.88 \pm 0.42$
Thomson M., Foster D. 1997	82	2.38
Tolhurst D. J., Tadmory Y., Chiao T. 1992	135	$2.4 \pm 0.26$
Torralba A., Oliva A. 2003	12,000	2.08
Webster M., Mihraya E. 1997	48	2.26

A number of studies have reported values for the spectral slope for different image ensembles. While the particulars of the image ensembles vary considerably, they tend to focus on natural scenes (which usually means simply eliminating carpentered environments). The average spectral slope varies from 1.8 to 2.4, with most values clustering around 2.0.

# Spectral Slope

- Spectral slope is related to autocorrelation
- Increasing slope increases coarseness
- Self similar (fractal)



One of the most prominent facets of such a power law is that these images tend to be scale invariant. That is, they share features at every scale, which is also a hallmark of fractal images. In practice, this means that we should be able to zoom in and out of an image, or travel through a natural scene, and expect that the statistics will remain roughly constant. It is also worth noting that, according to the Wiener-Kintchine theorem, the power spectrum and the auto-correlation function form a Fourier transform pair. This means that the spectral slope of image ensembles can be interpreted as describing relations between pairs of pixels. Intuitively, the means that since since the surface of a given object tends to be rather homogenous, it is expected that neighboring pixels will be similar and that the farther apart two pixels are, the less similar they will be.

## Applications: Scene Classification

Spectral Slope differs by scene



Forest (2.15)



Close-up (2.23)



distant meadow (2.4)

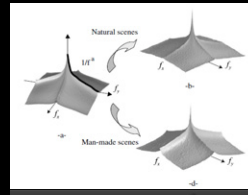
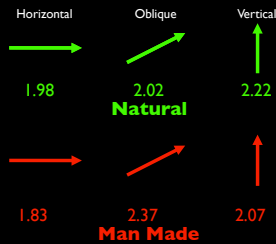
Webster & Miyahara (1997)

There is also some emerging evidence that the slope of the power spectrum is distinct for different scenes or objects. For example, Huang and Mumford examined 216 images which had been painstakingly segmented into pixels representing 11 different categories and found that although the image ensemble had an average slope of 1.96, there were systematic differences in the slopes across categories. Specifically, the slopes were 2.3, 1.8, 1.4, and 1.0 for man-made, vegetation, road, and sky pixels, respectively. Likewise, Webster and Miyahara analyzed the power spectra and rms-contrast of 48 natural scenes. They found significant differences in both spectral slope and contrast across the three scene types (2.15, 2.23, and 2.4 for the forest, close-up, and distant meadow scenes, respectively)

- HUANG J., MUMFORD D.: Image Statistics for the British Aerospace Segmented Database. Tech. rep., Division of Applied Math, Brown Univeristy, 1999
- TORRALBA A., OLIVA A.: Statistics of natural image categories. Network: Computation in Neural Systems 14 (2003)
- WEBSTER M., MIYAHARA E.: Contrast adaptation and the spatial structure of natural images. Journal of the Optical Society of America A 14 (1997)

# Applications: Scene Classification

- and orientation
- and orientation by scene



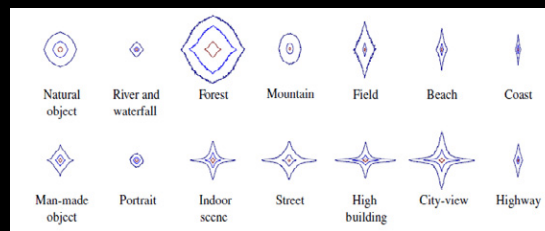
Torralba & Oliva (2003)

Furthermore, when the power spectra are not averaged over all orientations, it is clear that there is some variation as a function of angle: natural images tend to concentrate their power in horizontal and vertical angles. For example, in examining over 6000 man-made and 6000 natural scenes, Torralba and Oliva found that the slope varied as a function of both scene type and orientation (with slopes of 1.98, 2.02, and 2.22 for horizontal, oblique, and vertical angles in natural scenes and 1.83, 2.37, and 2.07 for man-made scenes). Thus, the spectral slope may be useful in object or scene discrimination. It is critical to mention that if two images have similar spectral slopes, then swap- ping their power spectra will not affect recognition as long as phase information is preserved.

- BADDELEY R. J., ABBOTT L. F., MICHAEL C. A., BOOTH C. A., SENGPIEL F., FREEMAN T., WAKEMAN E. A., ROLLS E. T.: Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society B: Biological Sciences* 264, 1389 (1997)
- OLIVA A., TORRALBA A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42, 3 (2001)
- RUDERMAN D. L.: Origins of scaling in natural images. *Vision Research* 37 (1997)
- SWITKES E., MAYER M. J., SLOAN J. A.: Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research* 18, 10 (1978)
- VAN HATEREN J., VAN DER SCHAAF A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B* 265 (1998)
- TORRALBA A., OLIVA A.: Statistics of natural image categories. *Network: Computation in Neural Systems* 14 (2003)
- BILLOCK V. A., CUNNINGHAM D. W., TSOU B. H.: What visual discrimination of fractal textures can tell us about discrimination of camouflaged targets. In *Proceedings of Human Factors Issues: In Combat Identification and in Image Analysis Using Unmanned Aerial Vehicles* (2008)
- TOLHURST D. J., TADMOR Y., CHIAO T.: Amplitude spectra of natural images. *Ophthalmic and Physiological Optics* 12 (1992)
- TADMOR Y., TOLHURST D. J.: Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research* 33, 1 (1993)

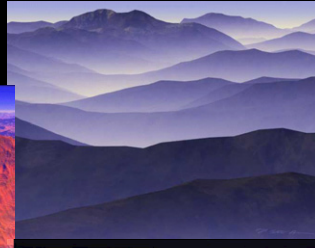
# Applications: Scene Classification

- and orientation by scene



Torralba & Oliva (2003)

## Applications: **Scene Synthesis**

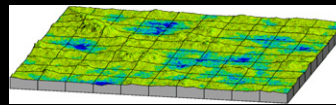
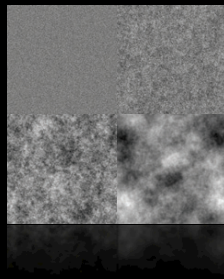


© Ken Musgrave

## Applications: **Scene Synthesis**



• Height map for terrain



Deussen



# Applications: Scene Synthesis

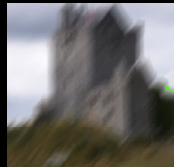
## Plant modeling



Weber

# Applications: Deblurring

Resulting photograph



?  
Real World



When taking a photograph of a scene, it is not uncommon that either the camera or an object in the scene moves. The longer the aperture is open, the more likely this is to be the case. As a result, all or part of the image will be blurred. A number of approaches for sharpening an image have been proposed. One type of approach, blind motion deconvolution, essentially treats the photograph as the result of a convolution between an unknown sharp image and an equally unknown blurring kernel. The goal is to estimate the blur kernel so that it can be deconvolved with the blurred image to yield the sharp image. Naturally this is an underspecified problem, so additional constraints are needed and recently, a number of researchers have employed natural image statistics to provide them.


# Applications:

## Blind motion Deconvolution


**Constraint:** Real world image must follow power law  
(Caron et al, 2002; Jalobeanu et al, 2002; Neelamani et al, 2004)

**Constraint:** Estimate Blur by optimizing to match real gradient distributions (Fergus et al, 2006)

Before



After



For example, Caron et al. assume that the sharp image's power spectra follows a power law distribution, while Jalobeanu et al. and Neelamani et al. use an interaction between power spectra and wavelets. Fergus et al. estimate the blur kernel by optimizing for the gradient distributions using priors derived from natural images. All of these approaches assume camera motion. That is, that there is a single blur kernel for the entire image


- ARON J. N., NAMAZI N. M., ROLLINS C. J.: Non-iterative blind data restoration by use of an extracted filter function. *Applied Optics* 41, 32 (2002)
- JALOBÉANU A., BLANC-FÉRAUD L., ZERUBIA J.: Estimation of blur and noise parameters in remote sensing. In *Proc. of Int. Conference on Acoustics, Speech and Signal Processing* (2002)
- NEELAMANI R., CHOI H., BARANIUK R.: ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *Signal Processing, IEEE Transactions on* 52, 2 (2004)
- FERGUS R., SINGH B., HERTZMANN A., ROWEIS S., FREEMAN W.: Removing camera shake from a single photograph. *ACM Transactions on Graphics (TOG)* 25 (2006)

# Applications:

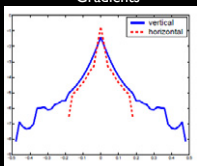
## Blind motion Deconvolution

**Constraint:** Model blurring - Gradients of blurred image are different (less high frequencies). Find non-blurred region and blur it until the gradients match (Levin et al., 2007).

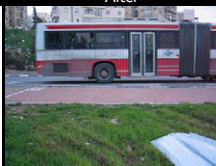
Before



Gradients



After



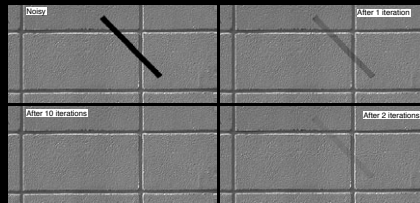
In an alternate approach, Levin use the gradient structure to find those pixels that are blurred and segment them from the rest of the image. Additionally, rather than use statistics of sharp images, of the blurring process. Specifically, they model how the the gradient distribution changes as a function of blurring to discovering the blurring kernel for a given image. One primary features of motion blurring is the attenuation of higher frequencies. This shows in the slope of the power spectra (by increasing  $\beta$ ) as well as in gradient histogram (e.g., in particular by removing the longer tails at the higher gradients). Levin attempts to recover the blur kernel by applying different blurs to the image to find the one that produces a gradient histogram that matches the blurred one. This requires a non-blurred image or image region. Using a completely different image tends not to produce reliable results (due to differences in the underlying gradient histogram). Since much terrestrial motion tends to be horizontal, the majority of the effects of motion blurring are also horizontal. Thus, the vertical gradients can under some circumstances be used to calculate the blur of the horizontal components.

- LEVIN A.: Blind motion deblurring using image statistics. *Advances in Neural Information Processing Systems* (2007)

# Applications: Image Inpainting

- Remove an unwanted object from an image
- Fill in hole by copying from elsewhere in image

Match based on spectra (and other) information  
(Hirani & Totsuka, 1996)

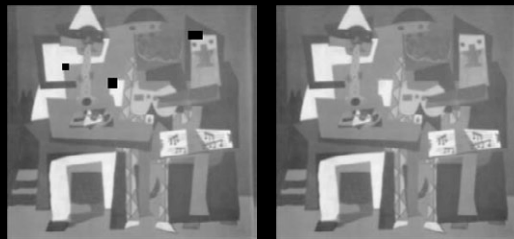


One image processing application area where image statistics have been used successfully is known as image inpainting. In order to remove an unwanted object from an image, the texture that would otherwise be behind that object need to be re-created, a task that can be particularly difficult if the holes are large. Among the various approaches, several have used image statistics to find optimal replacements for the missing texture and structure, usually by assuming that a patch similar to the missing portion can be found somewhere in the remaining image. For example, Hirani & Totsuka, use a combination of spectral and spatial information to find the replacement patch. Levin et al. select the missing section based on the gradients at the boundary region as well as to maximize the match the global gradient histogram. Shen et al. fill in the missing region by completing the gradient maps and then reconstructing the image by solving a Poisson equation.

- HIRANI A. N., TOTSUKA T.: Combining frequency and spatial domain information for fast interactive image noise removal. In SIGGRAPH '96 (1996)
- LEVIN A., ZOMET A., WEISS Y.: Learning how to inpaint from global image statistics. In Proceedings of the Ninth IEEE International Conference on Computer Vision (2003)
- SHEN J., JIN X., ZHOU C., WANG C. C. L.: Gradient based image completion by solving poisson equation. Computers and Graphics (2005)

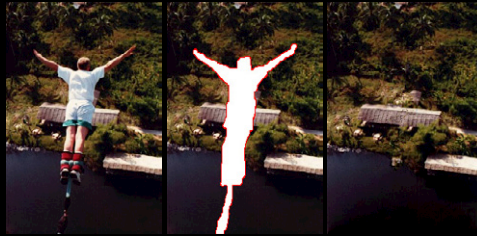
# Applications: Image Inpainting

Match gradients at boundary and maximize global gradients  
(Levin et al., 2003)



# Applications: Image Inpainting

Complete gradient maps, then reconstructing via Poisson equation  
(Shen et al., 2005)



# Higher Order Statistics

Erik Reinhard



# Types of Statistics

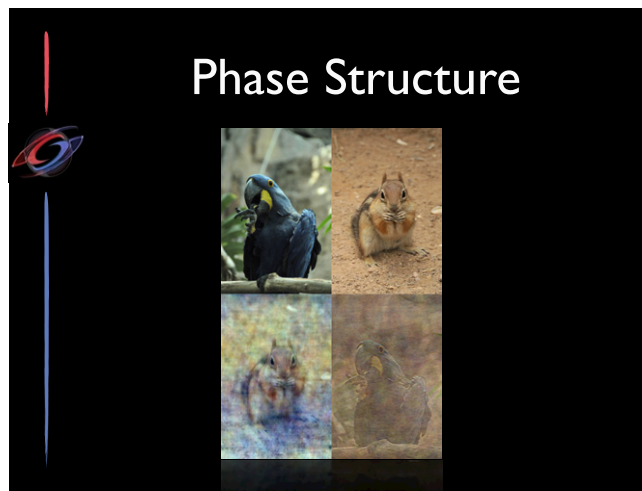
- Phase Spectra
- Wavelets
- Principal Components Analysis
- Independent Components Analysis

Whereas second order statistics take into account pairs of pixels through various transformations (in particular gradients and the amplitude spectrum), it is possible to find statistical regularities in larger groups of pixels using suitable image manipulations. The statistics that we are looking at in this section are listed on this slide.

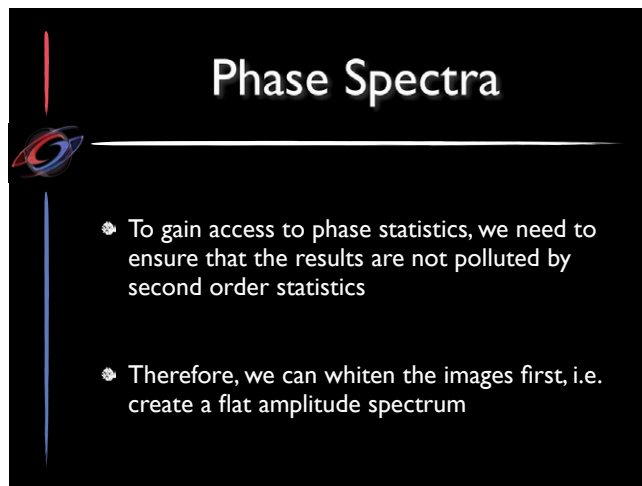
## Phase Spectra

- Amplitude spectra show great statistical regularities for natural images
- But so what? Is useful information encoded in amplitude spectra?

It can be argued that although statistical regularities are present in power spectra of natural images, much of the perceptually relevant information is encoded in phase spectra.



In this example, we have swapped the phase spectra of two images. As can be seen, much of the image structure has swapped. Thus, the image structure is largely encoded into the phase spectrum. Second order statistics such as the auto-correlation function (and therefore power spectra in Fourier space) and variance are insensitive to signal phase. They therefore do not adequately measure image structure.



- To gain access to phase statistics, we need to ensure that the results are not polluted by second order statistics
- Therefore, we can whiten the images first, i.e. create a flat amplitude spectrum

To gain access to phase information without polluting the results with first and second order information, we can whiten the images first. This amounts to adjusting the spectral slope to become flat. The auto-correlation function will therefore be zero everywhere, except at the origin.

- THOMSON M. G. A.: Visual coding and the phase structure of natural scenes. Network: Computation in Neural Systems 10 (1999)
- THOMSON M. G. A.: Beats, kurtosis and visual coding. Network: Computation in Neural Systems 12 (2001)

# Phase Spectra

- After whitening, we can compute first order statistics on resulting images
- Can be seen as computing statistics on edges
- Results show positive skew and high kurtosis, i.e. long tails
- Nonetheless, phase spectra remain somewhat of a mystery

By removing the second moment from consideration, it is now possible to compute skewness and kurtosis on the whitened signal. The whitened skew and whitened kurtosis are thus a measure of variations in the phase spectra. The results of applying this procedure to a set of natural images, leads to the conclusion that the whitened images are almost always positively skewed and are always positively kurtosed. In contrast, if the phase spectrum is randomized on the same images, the whitened skewness and kurtosis are close to zero.

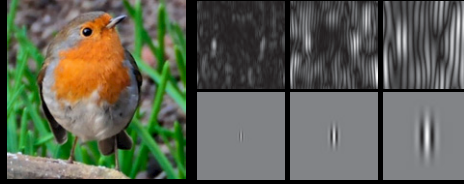
While positive skewness and kurtosis of phase spectra points in the direction of the presence of statistical regularities in the phase spectrum, these results are relatively weak and do not easily explain aspects of human vision. Furthermore, they do not appear to have found employ in any graphics related applications that we are aware of.

# Wavelets

- Phase spectra are computed over entire images
- What about spatially localized analysis?
- Wavelets do this
- They are also selective to specific orientations and scales

We have seen that second order statistics, and in particular the amplitude spectrum of image ensembles, show striking statistical regularities. The amplitude spectrum can be analyzed for different orientations, but as it is computed in Fourier space, it provides information of whole images rather than regions of images. In this section we study transforms that allow us to analyze images in local regions of space, as well as for different orientations and spatial frequencies. Such models and transforms started with Gabor who introduced the notion of time-frequency representations. It is noted that such systems are in some sense similar to how images are analyzed by portions of the human visual system.

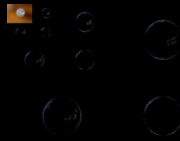
# Gabor Filters



Sinusoids weighted by Gaussians

There are several transforms that allow us to analyze images in local regions of space, as well as for different orientations and spatial frequencies. Such models and transforms started with Gabor who introduced the notion of time-frequency representations. Many wavelet transforms are self-similar, consisting of spatially localized band-pass basis functions which vary only by a dilation, translation or rotation. They have found many direct applications in image analysis and image processing.


# Haar Wavelets



Simplest type of wavelet, the Haar basis has many applications in image processing




# Wavelet Analysis



- Both phase and amplitude can be measured and correlated in a wavelet decomposition
- Surprising result: natural images are scale-invariant in both phase and amplitude

Field has shown that both phase and amplitude can be measured and correlated across scales and location. Although the amplitude spectrum has shown scale invariance, the phase structure revealed by Fourier space analysis remains somewhat of a mystery. However, by analyzing wavelets, which are simultaneously localized both in space and frequency, it can be shown that natural images are scale invariant in both phase and amplitude.

# Wavelet Analysis



- This means that the locations of edges are correlated across scales and orientation
- At higher frequencies, neighbors in frequency become more highly correlated
- At higher frequencies, neighbors in space become less correlated
- Phases in natural images contain alignment that extends over greater ranges at higher frequencies

This means that in local image regions, the location of edges are correlated across scales. At higher frequencies neighbors in frequency become more highly correlated, whereas neighbors in space become less correlated. In other words, phases in natural images contain alignment that extend over greater ranges at higher frequencies. Phase structures tend to change somewhat in orientation as well as position when moving across scales. Such behavior is well modeled with wavelets. It is therefore possible that this may help explain why mammalian visual system use wavelet-like structures in the visual cortex. It may well be that this is why wavelets are useful in a range of applications, as outlined above.

# Wavelet Analysis

- Distributions of histograms of wavelet coefficients have high kurtosis, i.e. long tails
- Can be modeled with a generalized Laplacian

It was found that the distributions of histograms of wavelet coefficients are highly non-Gaussian, showing in particular long tails, i.e. having high kurtosis. In particular, these histogram distributions can be modeled by a generalized Laplacian.

# Meaning of High Kurtosis

- Many natural image statistics end up showing high kurtosis
- This means that lots of values are small and some are large
- Effectively sparse coding

If the bandwidths of the wavelet filters are chosen to be around 1-2 octaves, that the kurtosis of the histograms of wavelet transforms is maximal. This means that under these conditions the resulting transform is maximally sparse.

# Sparse Coding

- In human vision, sparse coding is an important feature:
  - Variability of input is explained by fewer neurons
  - Metabolic efficiency
  - Minimizes wiring length
  - Increases capacity in associative memory

In clusters of neurons, sparsity is an important feature. It means that much of the variability in the input signal is explained by a small number of neurons. Such neuronal activity is metabolically efficient, minimizes wiring length, and increases capacity in associative memory

# Applications of Wavelets

- Image denoising
- Image compression
- Object detection
- Image retrieval

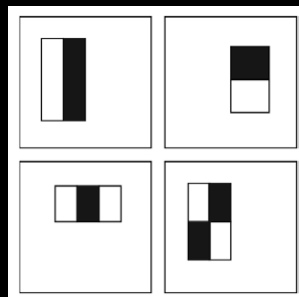
# Image Denoising



An example application of wavelet statistics is image denoising, where a Bayesian approach is employed to remove noise. Assuming that the wavelet coefficients of natural images follow a generalized Laplacian distribution, as outlined above, Gaussian image noise translates into Gaussian pollution of wavelet coefficients.

- SIMONCELLI E. P.: Bayesian denoising of visual images in the wavelet domain. In Bayesian Inference in Wavelet Based Models, Müller P., Vidakovic B., (Eds.). Springer-Verlag, New York, (1999)
- PORTILLA J., SIMONCELLI E. P.: Image denoising via adjustment of wavelet coefficient magnitude correlation. In Proceedings of the 7th International Conference on Image Processing (2000)

# Face Detection



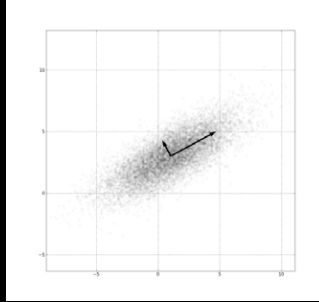
Viola & Jones use a small set of wavelet-like features to detect faces

An application area where wavelets have found extensive use is object detection. An important problem in that area is finding an effective and compact way of characterizing objects in images. Some object categories, such as faces, show surprisingly little variation in terms of overall structure and color, other types of objects however can be more complex. To overcome this issue, wavelets, and in particular Haar wavelets, can be used to provide a standardized template for characterizing objects in images. Using a training scheme, a small subset of Haar wavelets can be derived that can appropriately describe the structure of a particular class of images, such as for instance faces or pedestrians. This set of basis functions then forms the classifier that can be used to detect the object in question in new image.

More recently, a real-time face detection algorithm was proposed by Viola and Jones where features are applied on the image in a cascade fashion, ensuring that more complex classifiers are only used on instances that have been already accepted by simpler ones. Although the features used in this work are similar to Haar wavelets, some are more complex, as shown here. Image-based retrieval using the wavelet decomposition of each image has also been demonstrated by Jacobs et al. An image querying metric is developed using truncated, quantized versions of the wavelet decompositions of the images. This forms a compact signature for each image that is fast to compute and captures sufficient amount of information to allow searching.

- OREN M., PAPAGEORGIOU C., SINHA P., OSUNA E., POGGIO T.: Pedestrian detection using wavelet templates. In IEEE Conference in Computer Vision and Pattern Recognition, (1997)
- PAPAGEORGIOU C. P., OREN M., POGGIO T.: A general framework for object detection. In Sixth International Conference on Computer Vision, (1998)
- VIOLA P., JONES M.: Robust real-time object detection. International Journal of Computer Vision (2001).
- JACOBS C. E., FINKELSTEIN A., SALESIN D. H.: Fast multiresolution image querying. In SIGGRAPH '95 (1995)

# Principal Components Analysis (PCA)



One of the main properties of natural images is that pixel values are correlated. Nearby pixels tend to have similar values while these correlations tend to be less strong as the distance between pixels increases. One way to analyze these correlations is by decomposing the image data to a set of components that are as decorrelated as possible. Each of these components will then capture some particular mode of variation within the data.

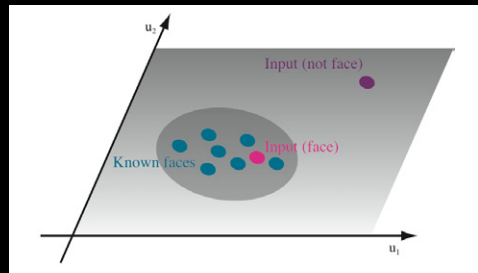
- HANCOCK P., BADDELEY R. J., SMITH L.: The principal components of natural images. Network: Computation in Neural Systems 3, 1 (1992)

## PCA

- ❖ PCA computes orthogonal components
- ❖ Axes maximize decorrelation
- ❖ Eigenvectors: directions of variation of the data
- ❖ Eigenvalues: importance of each direction
- ❖ Each component/axis captures some mode of variation of the data

Principal component analysis (PCA – also known as the Karhunen-Loève transform) is a solution commonly employed to that end. By computing the eigenvectors and corresponding eigenvalues of the covariance matrix of the image collection, a set of orthogonal, decorrelated components can be derived.

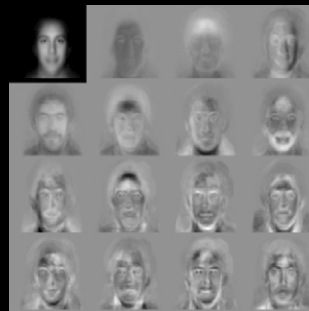
# Face Recognition (Eigenfaces)



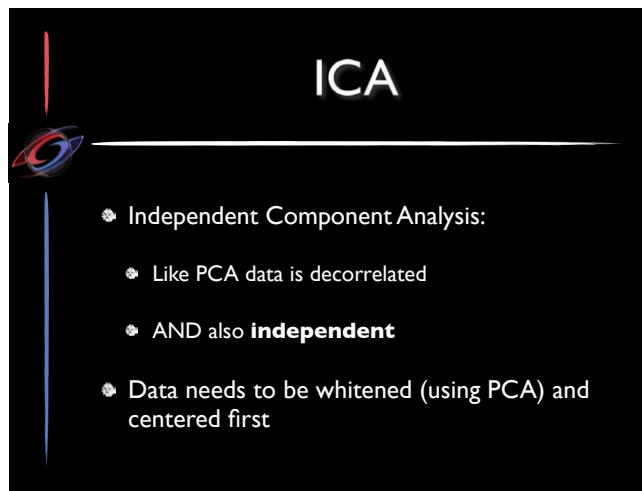
PCA can be used to reduce the dimensionality of a given set. This observation has been used in object and in particular face recognition applications successfully. The face recognition method known as Eigenfaces is one example. The database of face images is decomposed into a set of eigenvectors and eigenvalues, each of which corresponds to some dimension along which faces vary. Faces tend to have a lot of common characteristics in terms of structure, making them an ideal subject for dimensionality reduction.

- TURK M., PENTLAND A.: Eigenfaces for recognition. Journal of cognitive neuroscience 3, 1 (1991)
- TURK M., PENTLAND A.: Face recognition using eigenfaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1991)

# Face Recognition (Eigenfaces)



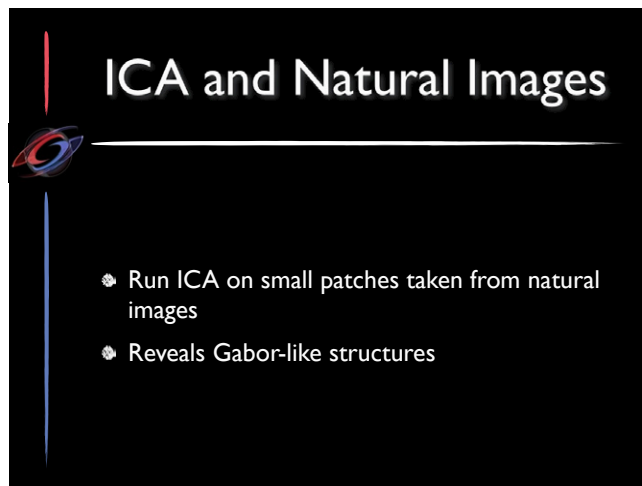
In order to match input faces to samples in the database, the new face can be projected to the computed "face space". It then becomes a simple matter of computing the distance between this new point and the database to decide whether the input is in fact a face. Recognition can then take place by finding the points closest to the input.



While Principal Components Analysis measures covariances of pixels separated by varying distances, the components that are found constitute a linear transform. This means that although the data may be decorrelated, there is no guarantee that the resulting components are in any way independent. In particular, only if the input data happens to have a Gaussian distribution, then the resulting principal components are both decorrelated and independent. As we have seen, many of the statistics of natural images that are currently known point in the direction of high kurtosis, i.e. they are highly non-Gaussian. This has given rise to the use of Independent Components Analysis (ICA), which is a technique to separate a multivariate signal into a set of components that will sum to the original signal under the assumption that the input is non-Gaussian. Like PCA, ICA is a special case of blind source separation, which is a general collection of techniques that try to separate a set of signals from a mixed signal without prior knowledge of the source signals or the mixing process.

ICA algorithms tend to require a set of preprocessing steps to make the problem more tractable. In particular, the data need to be centred and whitened. Also, data reduction is often applied. The whitening can be achieved by running the PCA algorithm first. By keeping only the first  $n$  components, data reduction is achieved. It ensures that only those components are used that are meaningful to the problem being solved. Moreover, it speeds-up the computations required to determine independent components. Several algorithms are known, including infomax (Bell & Sejnowski), extended infomax (Lee et al.), fastICA (Hyvärinen), Jade (Cardoso), kernel ICA (Bach & Jordan) and RADICAL (Learned-Miller & Fisher). Although the implementations vary, their goals are the same: represent multivariate data as a sum of components in such a way that the components reveal the underlying structure that gave rise to the observed signal.

- COMON P.: Independent component analysis: A new concept. *Signal Processing* 36, 3 (1994)
- HYVÄRINEN A., KARHUNEN J., OJA E.: *Independent Component Analysis*. John Wiley & Sons, (2001)
- ACHARYYA R.: A New Approach for Blind Source Separation of Convolutional Sources: Wavelet Based Separation Using Shrinkage Function. (2008)
- BELL A. J., SEJNOWSKI T. J.: An information- maximization approach to blind separation and blind deconvolution. *Neural Computation* 7, 6 (1995)
- LEE T.-W., GIROLAMI M., SEJNOWSKI T. J.: Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation* 11, 2 (1999)
- HYVÄRINEN A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* 10, 3 (1999)
- CARDOSO J. F.: High-order contrasts for independent component analysis. *Neural Computation* 11, 1 (1999)
- BACH F. R., JORDAN M. I.: Kernel independent component analysis. *Journal of Machine Learning Research* 3 (2002)
- LEARNED-MILLER E. G., FISHER III J. W.: ICA using spacings estimates of entropy. *Journal of Machine Learning Research* 4 (2003)

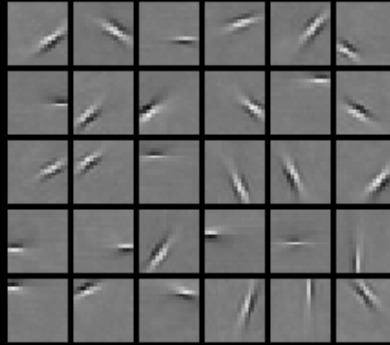


One could ask what are the independent components in the context of natural images. It could be argued that these are the objects that comprise a scene. It is exactly these that provide the structure that gives rise to the pixel array that is analyzed by computer algorithms as well as by the human visual system. Both PCA and ICA, however, are generative models which are based on linear superposition. Thus, they cannot take into account translation and rotation of objects.

However, it has been found that small image patches of natural image ensembles can be analyzed using ICA, revealing structures that resemble those found in the human visual system. In particular, this technique yields Gabor-like patches — elongated structures that are localized in space, orientation and frequency. Their size, aspect ratio, spatial frequency bandwidth, receptive field length and orientation tuning bandwidth are similar to those measured in cortical cells. These results lend credence to the argument that the human visual system appears to represent natural images with independent variables, each having a highly kurtotic distribution that leads to a metabolically efficient sparse coding.

- OLSHAUSEN B. A., FIELD D. J.: Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research* 37 (1997)
- BELL A. J., SEJNOWSKI T. J.: The 'independent components' of natural scenes are edge filters. *Vision Research* 37 (1997)
- VAN HATEREN J., VAN DER SCHAAF A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B* 265 (1998)

# ICA and Natural Images



# ICA and Natural Images



- Such structures have been found in cortical cells
- This means that human vision may represent images with independent variables, each having a highly kurtotic distribution that leads to a metabolically efficient sparse coding



# Color Statistics

Erik Reinhard

## Light Transduction

$$L_o(\lambda) = L_e(\lambda) + \int_{\Omega} L_i(\lambda) f_r(\lambda) \cos(\Theta) d\omega$$

$$L = \int_{\lambda} L_o(\lambda) \bar{l}(\lambda) d\lambda$$

$$M = \int_{\lambda} L_o(\lambda) \bar{m}(\lambda) d\lambda$$

$$S = \int_{\lambda} L_o(\lambda) \bar{s}(\lambda) d\lambda$$

Images are formed as light that is transduced to electric signals. This happens both in the retina as well as in electronic image sensors. Before that, light is emitted and usually reflected several times off different surfaces before it reaches a sensor. This process is modeled by the rendering equation. This equation models the behavior of light at a surface: light  $L_i$  impinging on a surface point from all directions  $\Omega$  is reflected into an outgoing direction of interest, weighted by the bi-directional reflectance distribution function (BRDF)  $f_r$ . The amount of light  $L_o$  going into that direction is then the sum of the reflected light and the light emitted by the surface point (which is zero unless the surface is a light source). Here, we have specifically made all relevant terms dependent on wavelength  $\lambda$  to indicate that this behavior happens at all wavelengths, including all visible wavelengths. This means that  $L_o(\lambda)$  can be seen as a spectral distribution.

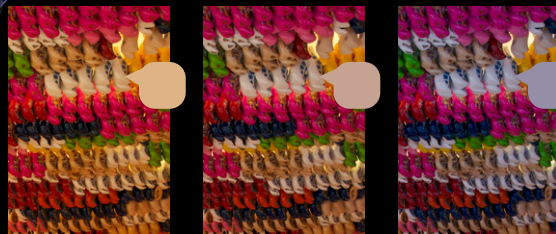
Under photopic lighting conditions, the human visual system contains three cone types, each with a different peak sensitivity. Thus, each cone type integrates the incident light according to a different weight function. A spectral distribution therefore gives rise to a triple of numbers (L, M, S) which represent the signal that is passed on from the photoreceptors to the remainder of the human visual system. Such triples are called tristimulus values.

# Implications

- Metamerism: different spectra integrate to the same cone responses, and are therefore perceived identically
- This allows us to build color displays, for instance
- Color statistics can be collected on tristimulus values, rather than color spectra

An important implication of this behavior is that different spectral distributions can integrate to the same tristimulus values. Such spectral distributions are then necessarily perceived to be identical, which is known as metamerism. It is the reason that we are able to build display devices with three primaries that are not the same as the responsivities of the cones: tristimulus values in one color space can be converted to tristimulus values in another color space, and although they do not represent the same spectral distributions, through metamerism they lead to the same percept. Further, rather than study spectral distributions, we can limit ourselves to the study of tristimulus values.

# Color Constancy



- Humans can discount the color of the illumination

Humans tend to effortlessly discount the illumination of their environments. In other words, the color of the light sources in a particular viewing environment are to a large extent ignored, so that human vision is fairly good at detecting surface colors. This phenomenon is known as color constancy. Computational models that aim to achieve the same are known as white balancing algorithm. They intend to remove color casts from images due to illumination.

# Color Constancy

- ✿ Cannot be computed analytically from retinal input; it is an under-constrained problem
- ✿ Human vision makes statistical assumptions

As light is reflected off surfaces according to the rendering equation, reconstructing either surface reflectance and/or illumination is an under-constrained problem. This means that white balancing can only be achieved by means of statistical assumptions.

# Statistical Assumptions

- ✿ Grey world:
  - ✿ Spectrum of light sources usually off-white
  - ✿ Average BRDF of a scene often close to grey
  - ✿ Average color of an image yields estimate of dominant illuminant

In color spaces such as the aforementioned LMS space, equal values of the three components denote achromatic colors. One way to achieve such neutral colors is to start with an equal energy spectrum, i.e. a spectral distribution which has the same value  $L_0$  for each wavelength  $\lambda$ . This could happen if a scene is illuminated by an equal-energy light source, a source that emits the same energy at all wavelengths. If the BRDF then also reflects all wavelengths equally, a neutral color would be achieved. In practice, a scene is illuminated by only one or at most a few light sources, with an emission spectrum that is off-white. The BRDFs in a scene are much more variable. However, a surprising finding is that when a collection of BRDFs are averaged, the result ends up being a distribution function that is close to grey. This is known as the grey-world assumption.

- REINHARD E., KHAN E. A., AKYÜZ A. O., JOHNSON G. M.: Color Imaging: Fundamentals and Applications. A K Peters, Wellesley, 2008



The implication of this finding is that if we were to average the colors of all pixels in an image, and the grey-world assumption holds, the average tristimulus value is a good indicator of the color of the light source. Of course, when averaged over natural image ensembles, the grey world assumption does tend to hold. In single images this assumption may or may not hold. In particular, if the surface colors in the scene are biased towards some specific saturated color, the average reflectance will not tend toward grey.



The grey world assumption is often used to aid white balancing. After all, if we know the color of the illuminant, then we can correct all pixels by simply dividing all pixels by the average image value. The grey-world assumption is a statistical argument that is necessary to perform white balancing on images in the absence of further information about the illuminant, given that white balancing is by itself an underconstrained problem.

# Possible Fixes

- Exclude most saturated pixels from average
- Optionally: convert to CIELAB
  - Compute 2D histogram on  $a^*$  and  $b^*$  channels
  - Spread of histogram and distance to origin determine if color cast is likely due to illumination or reflectance

Note that we have glossed over the fact that this procedure is best applied in a perceptual color space, such as LMS, thereby mimicking chromatic adaptation processes that occur in the human visual system. If an image is given in a different color space, most likely the sRGB space, the image should first be converted to LMS. The approximation of the illuminant can be improved by excluding the most saturated pixels from the estimation. Alternatively, the image can be subjected to further statistical analysis to determine if the color distribution is due to colored surfaces or due to a colored light source. The image can be first converted to the CIELAB color opponent space. Ignoring the lightest and darkest pixels as they do not contribute to a reliable estimate, the remaining pixels are used to compute a two-dimensional histogram  $F(a,b)$  on the two chromatic channels  $a$  and  $b$ .

In each channel, the chromatic distributions are modeled using the mean and variances of the histogram projections onto the  $a$  and  $b$  axes. In CIELAB neutral colors lie around the  $(a,b) = (0,0)$  point. To assess the strength of the cast, we need to measure how far the histogram lies from this point. If the spread of the histogram is small, and lies far away from the origin, the image is likely to be dominated by strong reflectances rather than illumination.

- FUNT B., BARNARD K., MARTIN L.: Is machine colour constancy good enough? In Proceedings of the 5th European Conference on Computer Vision (1998)
- ADAMS J., PARULSKI K., SPAULDING K.: Color processing in digital cameras. IEEE Micro 18, 6 (1998),
- GASPARINI F., SCHETTINI R.: Color balancing of digital photos using simple image statistics. Pattern Recognition 37 (2004)

# White Patch Algorithm

- Assume that lightest patches in the scene are neutral in color
- Their color therefore represents the illuminant

It was found that while grey-world algorithms work well on some images, alternate solutions such as the white-patch algorithm (Land) perform better on texture-rich images. The white-patch algorithm assumes that the lightest pixels in an image depict a surface with neutral reflectance, so that its color represents the illuminant. Both the grey-world and white patch algorithms are special instances of the Minkowski norm.

- LAND E.: The retinex theory of color vision. Scientific American 237, 6 (1977)
- FINLAYSON G., TREZZI E.: Shades of gray and colour constancy. In Proceedings of the Twelfth Color Imaging Conference (2004)

# Grey-Edge Assumption

- The difference between two colored pixels tends to evaluate to grey

A further generalized assumption can be made about images, which is that the average difference between two pixels evaluates to grey. This is known as the grey-edge assumption.

- VAN DE WEIJER J., GEVERS T.: Color constancy based on the grey-edge hypothesis. In Proceedings of the International Conference on Image Processing (2005)
- GIJSENIJ A., GEVERS T.: Color constancy using natural image statistics. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2007)

# Algorithm Selection

- Different white balancing algorithms tend to work best on specific types of images
- Can therefore collect statistics on the image pixels and select an appropriate algorithm based on the outcome
- **Weibull distribution** is shown to be indicative

Having noted that many color constancy and white balancing algorithms exist, with none of them universally applicable, Gijsenij and Gevers use natural image statistics to classify an image, and then use this classification to select the most appropriate white balancing algorithm for the image at hand. In particular, they use the finding that the distribution of edge responses in an image can be modeled by means of a **Weibull distribution**. It is now possible to fit Weibull parameters to the derivatives of a large number of images, and correlate their values with the white balancing algorithm that performs best for each image. The parameter space tends to form clusters where a specific algorithm tends to produce the most accurate result. This means that the Weibull distribution can be used to select the most appropriate white balancing algorithm for each image individually.

- GIJSENIJ A., GEVERS T.: Color constancy using natural image statistics. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2007)
- GEUSEBROEK J., SMEULDERS A.: A six-stimulus theory for stochastic texture. International Journal of Computer Vision 62, 1-2 (2005)

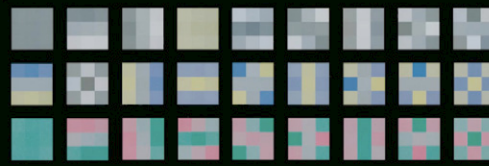
# A Further Implication

## Remember

- For grey values, in RGB (as well as LMS and similar color spaces) we have  $R=G=B$
- If values average to grey, then in RGB-like color spaces strong correlations exist between channels

If the average surface reflectance is grey, and light sources tend towards white, then in the LMS color space and similar RGB-like spaces, values in one color channel tend to be good predictors of values in the other two color channels. In other words, if we find a high value for a red pixel, then chances are that the green and blue values are high as well. This means that the color channels in such color spaces are highly correlated.

# Statistical Decorrelation



Principal axes of 3x3 chromatic patches  
(Ruderman et al. 1998)

An interesting experiment was conducted whereby a set of spectral images was converted to log LMS space, before being subjected to Principal Components Analysis (PCA) (Ruderman et al.). The logarithm was taken as a data-conditioning measure. Applied in this manner, PCA rotates the axes to the point where they are maximally decorrelated. The resulting color space is termed  $L\alpha\beta$  where  $L$  is a light-ness channel, and  $\alpha$  and  $\beta$  are color opponent channels. By starting in LMS cone space, the rotation yields a new color space which surprisingly corresponds closely to the color opponent space found in the ganglion cells (these are the cells that transport the visual signal from the retina through the optic nerve to the lateral geniculate nucleus, thought to be a relay station in the brain).

Color opponent spaces are characterized by a single achromatic channel, typically encoding luminance or light-ness, and two chromatic axes which can be thought of as spanning red-green and yellow-blue (although the axes do not precisely correspond to these perceptual hues). The chromatic axes can have positive and negative values; a positive value can for instance denote the degree of redness, whereas a negative value in the same channel would denote the degree of greenness. A consequence of color opponency is that we are not able to simultaneously perceive an object to have green and red hues. Although we may describe objects as reddish-blue or yellowish-green, we never describe them as reddish-green. The same holds for yellow and blue.

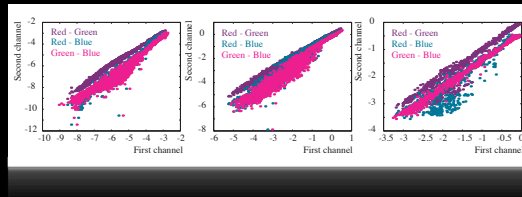
- RUDERMAN D. L., CRONIN T. W., CHIAO C.: Statistics of cone responses to natural images: Implications for visual coding. Journal of the Optical Society of America A 15, 8 (1998)

# Statistical Decorrelation



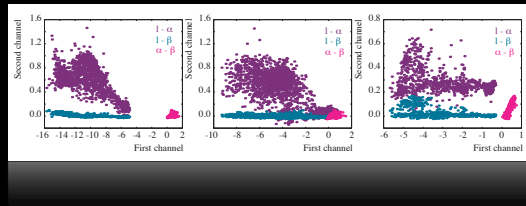
For 2000 random samples drawn from each of the 3 images shown earlier, their distribution is plotted in the next two slides. The point clouds form more or less diagonal lines in RGB space when pairs of channels are plotted against each other, showing that the three color channels in RGB space are almost completely correlated for these images. This is not the case for the same pixels plotted in  $L\alpha\beta$  space.

## Correlations in RGB/LMS





# Color Opponent Space



# Statistical Decorrelation

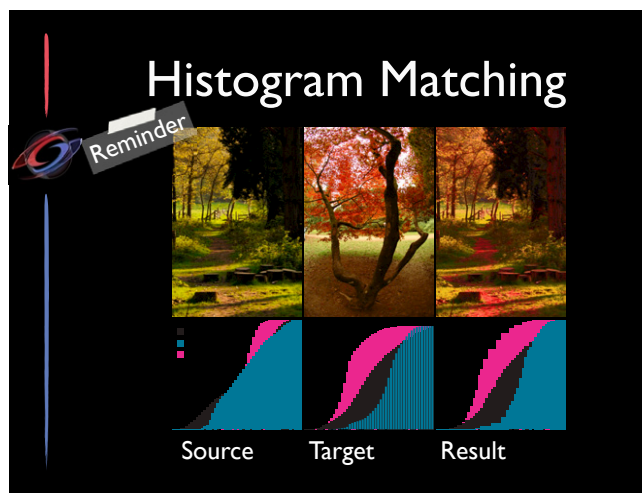
## • $L\alpha\beta$ color space

- decorrelated - values of pixels in one channel do not predict the values in another
- L - luminance
- $\alpha, \beta$  - opponent channels

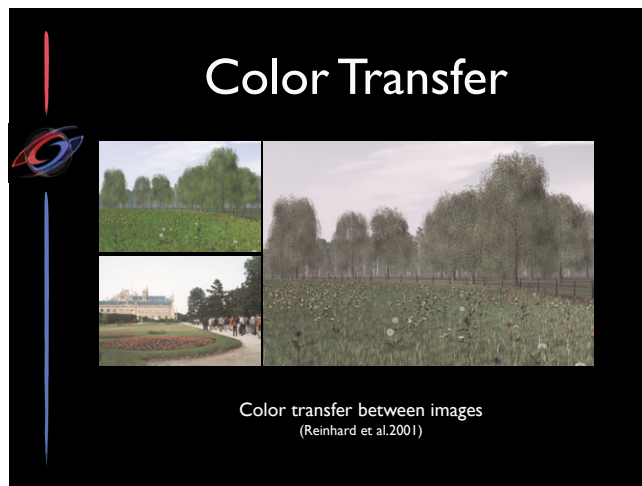


The color opponency of the  $L\alpha\beta$  space is demonstrated here, where the image is decomposed into its separate channels. The image representing the  $\alpha$  channel has the  $\beta$  channel reset to 0 and vice versa. We have retained the luminance variation here for the purpose of visualization. The image showing the luminance channel only was created by setting both the  $\alpha$  and  $\beta$  channels to zero. The fact that natural images can be transformed to a decorrelated color space, which coincides with a color decomposition occurring in the human visual system, points to a beautiful adaptation of human vision to its natural input.

It also points to several applications that are possible due to the fact that opponent spaces are decorrelated. In particular, we highlight color transfer, an algorithm that attempts to transfer the look and feel of one image to another on the basis of its color content. Due to the three-dimensional nature of color spaces, this is in the general sense a complicated three-dimensional problem. However, by converting to a color opponent space, the data in each channel will be decorrelated from the other channels. As a first example, the histogram projection onto the  $a$  and  $b$  axes yields meaningful results as the CIELAB space used here is an example of a color opponent space.



Perhaps one of the most direct applications of the use of decorrelated color spaces is color transfer between images. Earlier, we mentioned that the shape of the distribution could be matched between two images, a process known as histogram matching. Although this is a simple way to transfer the color palette between two images, the results are often less than ideal.



Many different techniques exist that aim to transfer some statistical properties of one image to another. One of the simplest approaches was proposed by Reinhard et al. where the mean and standard deviations of the three channels are matched. To make this process straightforward, images can be converted to a decorrelated color space, such as the  $L\alpha\beta$  space proposed by Ruderman. In this space, properties of all the pixels can be computed separately in each of the three axis. Doing this for both a source and a target image yields a set of factors and terms that can be used to transform the pixel data in one image to match the appearance of the other image.

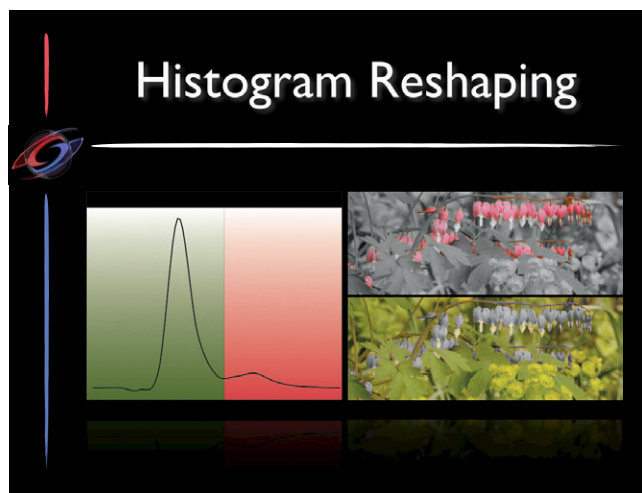
- E REINHARD, M ASHIKHMIN, B GOOCH, P SHIRLEY: Color transfer between images. IEEE Computer Graphics and Applications (2001)



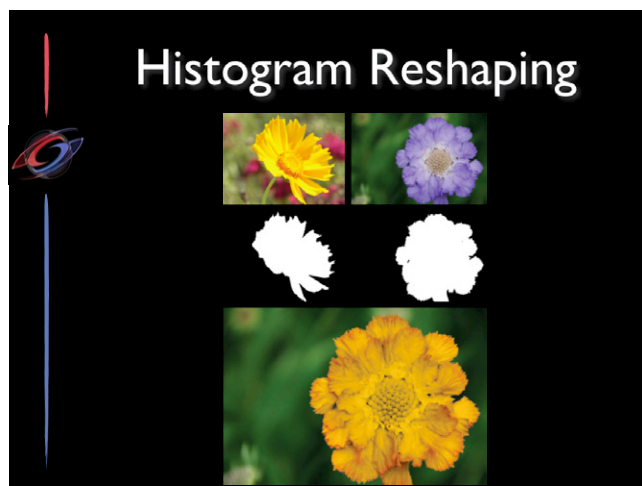
Of course using a decorrelated color space such as Lab cannot guarantee that every image will in fact be perfectly decorrelated. To rectify that, statistical analysis similar to what Ruderman did can be performed on the actual input images to derive image-specific color spaces, or the color palette transfer can be handled as a 3D problem instead of 3 independent 1D problems.

- ABADPOUR A., KASAEI A.: A fast and efficient fuzzy color transfer method. Proceedings of the IEEE Symposium on Signal Processing and Information Technology (2004)
- GRUNDLAND M., DODGSON N.A.: Color histogram specification by histogram warping. Proceedings of SPIE (2005)
- NEUMANN L., NEUMANN A.: Color style transfer techniques using hue, lightness and saturation histogram matching. Computational Aesthetics 2005 (2005) p
- PITIE F., KOKARAM R., DAHYOT R.: N-dimensional probability density function transfer and its application to colour transfer. International Conference on Computer Vision (ICCV'05), Beijing (2005)
- SENANAYAKE C.R., ALEXANDER D.C.: Colour Transfer by Feature Based Histogram Registration. British Machine Vision Conference (2007)
- XIAO X., MA L.: Color transfer in correlated color space. Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications (2006) pp. 309





- POULI T., REINHARD E.: Progressive Histogram Reshaping for Creative Color Transfer and Tone Reproduction, NPAR (2010)



# Histogram Reshaping



## Conclusions

# Statistics



- There are many ways to transform images, after which we can compute statistics
- When we transform images according to how we think the human visual system operates, we end up with highly kurtotic and sometimes independent representations
- Sparse coding is good for human vision, and probably good for solving engineering problems

# Applications



- Many applications already known
  - Object detection
  - Compression
  - Deblurring
  - Inpainting
  - Color transfer
  - etc.

# Applications

- Hopefully, as our knowledge of our environment increases, there will be many more to come
- Graphics, computer vision and image processing are prime areas of research that we think may benefit from natural image statistics

**Questions?**

# Color Imaging

## Fundamentals and Applications



Erik Reinhard  
Erum Arif Khan  
Ahmet Oğuz Akyüz  
Garrett M. Johnson

The following section is an extract from the book 'Color Imaging: Fundamentals and Applications' by Erik Reinhard, Erum Arif Khan, Ahmet Oğuz Akyüz and Garrett M. Johnson, published by AK Peters in 2008.



# 1 Natural Image Statistics

Images of natural scenes tend to exhibit statistical regularities, which are important for the understanding of the human visual system that has evolved to interpret such natural images. In addition, these properties may prove valuable in image-processing applications.

If all greyscale images that consist of  $1000 \times 1000$  pixels are considered, then the space that these images occupy would be a million-dimensional space. Most of these images never appear in our world: natural images form a sparse subset of all possible images. In recent years, the statistics of natural images have been studied to understand how properties of natural images influence the human visual system. In order to assess invariance in natural images, usually a large set of calibrated images are collected into an *ensemble* and statistics are computed on these ensembles, rather than on individual images.

Natural image statistics can be characterized by their order. In particular, first-order, second-order, and higher-order statistics are distinguished [34]:

**First-order statistics** treat each pixel independently, so that, for example, the distribution of intensities encountered in natural images can be estimated.

**Second-order statistics** measure the dependencies between pairs of pixels, which are usually expressed in terms of the power spectrum (see Section 1.2).

**Higher-order statistics** are used to extract properties of natural scenes which can not be modeled using first- and second-order statistics. These include, for example, lines and edges.

Each of these categories is discussed in some detail in the following sections.

## 1.1 First-Order Statistics

Several simple regularities have been established regarding the statistical distribution of pixel values; the gray-world assumption is often quoted in this respect. Averaging all pixel values in a photograph tends to yield a color that is closely related to the blackbody temperature of the dominant illuminant [29]. The integrated color tends to cluster around a blackbody temperature of 4800 K.

For outdoor scenes, the average image color remains relatively constant from 30 minutes after dawn until within 30 minutes of dusk. The correlated color temperature then increases, giving an overall bluer color.<sup>1</sup>

The average image color tends to be distributed around the color of the dominant scene illuminant according to a Gaussian distribution, if measured in log space [21, 22]. The eigenvectors of the covariance matrix of this distribution follow the spectral responsivities of the capture device, which for a typical color negative film are [22]:

$$L = \frac{1}{\sqrt{3}} (\log(R) + \log(G) + \log(B)), \quad (1.1a)$$

$$s = \frac{1}{\sqrt{2}} (\log(R) - \log(B)), \quad (1.1b)$$

$$t = \frac{1}{\sqrt{6}} (\log(R) - 2\log(G) + \log(B)). \quad (1.1c)$$

For a large image database, the standard deviations in these three channels are found to be 0.273, 0.065, and 0.030. Note that this space is similar to the  $L\alpha\beta$  color space. Although that space is derived from LMS rather than RGB, both operate in log space, and both have the same channel weights. The coefficients applied to

---

<sup>1</sup>Despite the red colors observed at sunrise and sunset, most of the environment remains illuminated by a blue skylight, which increases in importance as the sun loses its strength.

each of the channels are identical as well, with the only difference being that the role of the green and blue channels has been swapped.

## 1.2 Second-Order Statistics

The most remarkable and salient natural image statistic that has been discovered so far is that the slope of the power spectrum tends to be close to 2. The power spectrum of an  $M \times M$  image is computed as

$$S(u, v) = \frac{|F(u, v)|^2}{M^2}, \quad (1.2)$$

where  $F$  is the Fourier transform of the image. By representing the two-dimensional frequencies  $u$  and  $v$  in polar coordinates ( $u = f \cos \phi$  and  $v = f \sin \phi$ ) and averaging over all directions  $\phi$  and all images in the image ensemble, it is found that on log-log scale amplitude as a function of frequency,  $f$  lies approximately on a straight line [6, 10, 32–34]. This means that spectral power as a function of spatial frequency behaves according to a power law function. Moreover, fitting a line through the data points yields a slope  $\alpha$  of approximately 2 for natural images. Although this spectral slope varies subtly between different studies [6, 9, 11, 32, 40], and with the exception of low frequencies [20], it appears to be extremely robust against distortions and transformations. It is therefore concluded that this spectral behavior is a consequence of the images themselves, rather than of particular methods of camera calibration or exact computation of the spectral slope. We denote this behavior by

$$S(f) \propto \frac{A}{f^\alpha} = \frac{A}{f^{2-\eta}}, \quad (1.3)$$

where  $A$  is a constant determining overall image contrast,  $\alpha$  is the spectral slope, and  $\eta$  is its deviation from 2. However, the exact value of the spectral slope depends somewhat on the type of scenes that make up the ensembles. Most interest of the natural image statistics community is in scenes containing mostly trees and shrubs. Some studies show that the spectral slope for scenes containing man-made objects is slightly different [41, 42]. Even in natural scenes, the statistics vary, dependent on what is predominantly in the images. The second-order statistics for sky are, for example, very different from those of trees.

One of the ways in which this becomes apparent is when the power spectra are not circularly averaged, but when the log average power is plotted against angle. For natural image ensembles, all angles show more or less straight power spectra, although most power is concentrated along horizontal and vertical directions [33, 34] (see also Figure 3). The horizon and the presence of tree trunks are said to be factors, although this behavior also occurs in man-made environments.

The power spectrum is related to the auto-correlation function through the Wiener-Khintchine theorem, which states that the auto-correlation function and the power spectrum form a Fourier transform pair [25]. Hence, power spectral behavior can be equivalently understood in terms of correlations between pairs of pixel intensities.

A related image statistic is contrast, normally defined as the standard deviation of all pixel intensities divided by the mean intensity ( $\sigma/\mu$ ). This measure can either be computed directly from the image data, or it can be derived from the power spectrum through Parseval's theorem [34]:

$$\frac{\sigma^2}{\mu^2} = \sum_{(u,v)} S(u, v). \quad (1.4)$$

This particular contrast computation can be modified to compute contrast in different frequency bands. Frequency-conscious variance can then be thresholded, yielding a measure which can detect blur [12]. This is useful as lack of contrast can also be caused by the absence of sufficient detail in a sharp image.

The above second-order statistics are usually collected for luminance images only, as luminance is believed to carry the greatest amount of information. However, chrominance channels are shown to exhibit similar spectral behavior [27], and therefore, all subsequent qualitative arguments are expected to be true for color as well.

The fact that the power spectral behavior of natural images, when plotted in log-log space, yields a straight line with a slope of around -2 is particularly important, since recent unrelated studies have found that in image interpretation tasks, the human visual system performs best when the images conform to the above second-order statistic. In one such study, images of a car and a bull were morphed into each other, with varying distances between the images in the sequence [28]. Different sequences were generated with modified spectral slopes. The minimum distance at which participants could still distinguish consecutive images in each morph sequence was then measured. This distance was found to be smallest when the spectral slope of the images in the morph sequence was close to 2. Deviation of the spectral slope in either direction resulted in a deteriorated performance to distinguish between morphed images.

In a different study, the effect of spectral slope on the detection of mirror symmetry in images was assessed [31]. Here, white noise patterns with varying degrees of vertical symmetry were created and subsequently filtered to alter the spectral slope. An experiment, in which participants had to detect if symmetry was present, revealed that performance was optimal for images with a spectral slope of 2.

These studies confirm the hypothesis that the HVS is tuned to natural images. In fact, the process of whitening (i.e., flattening the power spectrum to produce a slope  $\alpha$  of 0) is consistent with psychophysical measurements [1], which indicates that the HVS expects to see images with a  $1/f^2$  power spectrum.

### 1.3 Power Spectra

In this section, we show how the power spectrum of an ensemble of images can be established. We used 133 images from the van Hateren database of natural images [15]. The images in this ensemble are available as luminance images. Example images are shown in Figure 1.

For this image ensemble, the power spectrum was computed and the spectral slope was estimated. The power spectrum computation proceeds as follows.<sup>2</sup>

A window of  $512 \times 512$  pixels was cut out of the middle of the image upon which further processing was applied. Then, the weighted mean intensity  $\mu$  was subtracted to avoid leakage from the DC-component of the image, where  $\mu$  is defined as

$$\mu = \frac{\sum_{(x,y)} L(x,y)w(x,y)}{\sum_{(x,y)} w(x,y)}. \quad (1.5)$$

Next, the images were prefiltered to avoid boundary effects. This is accomplished by applying a circular Kaiser-Bessel window function (with parameter  $\alpha = 2$ ) to the image [14]:

$$w(x,y) = \frac{I_0\left(\pi\alpha\sqrt{1.0 - \left(\frac{x^2+y^2}{(N/2)^2}\right)}\right)}{I_0(\pi\alpha)}, \quad 0 \leq \sqrt{x^2+y^2} \leq \frac{N}{2}. \quad (1.6)$$

Here,  $I_0$  is the modified zero-order Bessel function of the first kind and  $N$  is the window size (512 pixels). In addition, this weight function was normalized by letting

$$\sum_{(x,y)} w(x,y)^2 = 1. \quad (1.7)$$

---

<sup>2</sup>This calculation closely follows the method described in [34].



Figure 1: Example images drawn from the van Hateren database [15]. Images courtesy of Hans van Hateren.

This windowing function was chosen for its near-optimal trade-off between side-lobe level, main-lobe width, and computability [14]. The resulting images were then compressed using the Fourier transform:

$$F(u, v) = \sum_{(x, y)} \frac{L(x, y) - \mu}{\mu} e^{2\pi i(ux + vy)}. \quad (1.8)$$

Finally, the power spectrum was computed as per (1.2) and the resulting data points plotted. Although frequencies up to 256 cycles per image are computed, only the 127 lowest frequencies were used to estimate the spectral slope. Higher frequencies may suffer from aliasing, noise, and low modulation transfer [34]. The estimation of the spectral slope was performed by fitting a straight line through the logarithm of these data points as a function of the logarithm of  $1/f$ . This method was chosen over other slope estimation techniques, such as the Hill estimator [16] and the scaling method [8], to maintain compatibility with [34]. In addition, the number of data points (127 frequencies) is insufficient for the scaling method, which requires at least 1,000 data points to yield reasonably reliable estimates.

With this method, second-order statistics were extracted. The 1.87 spectral slope reported for the van Hateren database was confirmed (we found 1.88 for our subset of 133 images). The deviations from this value for the artificial image ensemble are depicted in Figure 2. The angular distribution of power tends to show peaks near horizontal and vertical angles (Figure 3). Finally, the distribution of spectral slopes for the 133 images in this ensemble is shown in Figure 4.

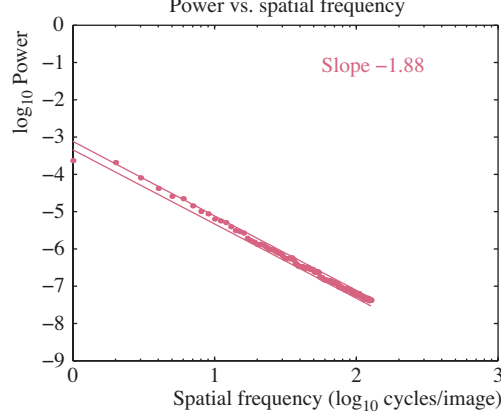


Figure 2: Spectral slope for the image ensemble. The double lines indicate  $\pm 2$  standard deviations for each ensemble.

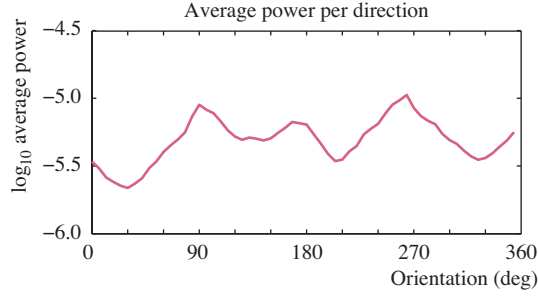


Figure 3: Log average power as function of angle.

## 1.4 Higher-Order Statistics

One of the disadvantages of using amplitude information in the frequency domain is that phase information is completely discarded, thus ignoring the position of edges and objects. For this reason, higher-order statistics have been applied to natural image ensembles. The simplest global  $n$ th-order statistics that may capture phase structure are the third and fourth moments, commonly referred to as skew and kurtosis [25]:

$$s = \frac{E\{x^3\}}{E\{x^2\}^{3/2}}, \quad (1.9a)$$

$$k = \frac{E\{x^4\}}{E\{x^2\}^2} - 3. \quad (1.9b)$$

These dimensionless measures are, by definition, zero for pure Gaussian distributions. Skew can be interpreted as an indicator of the difference between the mean and the median of a data set. Kurtosis is based on the size of a distribution's tails relative to a Gaussian distribution. A positive value, associated with long tails in the distribution of intensities, is usually associated with natural image ensembles. This is for example evident when plotting log-contrast histograms, which plot the probability of a particular contrast value appearing. These plots are typically non-Gaussian with positive kurtosis [33].

Thomson [39] has pointed out that for kurtosis to be meaningful for natural image analysis, the data should

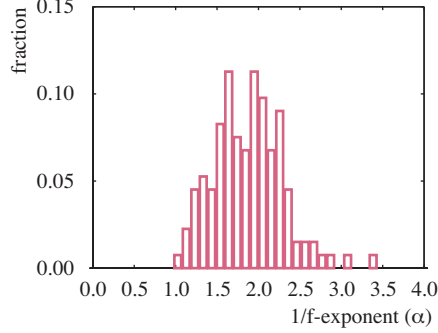


Figure 4: A histogram of spectral slopes, derived from an image ensemble of 133 images taken from the van Hateren database.

be decorrelated, or whitened, prior to any computation of higher-order statistics. This can be accomplished by flattening the power spectrum, which ensures that second-order statistics do not also capture higher-order statistics. Thus, skew and kurtosis become measures of variations in the phase spectra. Regularities are found when these measures are applied to the pixel histogram of an image ensemble [38, 39]. This appears to indicate that the HVS may exploit higher-order phase structure after whitening the image. Understanding the phase structure is therefore an important avenue of research in the field of natural image statistics. However, it appears that so far it is not yet possible to attach meaning to the exact appearance of higher-order statistics.

## 1.5 Gradient Field Statistics

Rather than study the statistical properties of pixels, we can also look at derived quantities, such as gradients. The gradient field of an image is a vector-valued quantity  $\nabla L$ , defined as:

$$\nabla L = (L(x+1, y) - L(x, y), L(x, y+1) - L(x, y)). \quad (1.10)$$

Natural images typically have strong discontinuities in otherwise relatively smooth regions. This means that large image areas have only small gradients, with larger gradients happening more sparsely. This means that the components of the gradient field will show a characteristic distribution, which can be observed by computing its histogram. The gradient distribution of a set of four example images, shown in Figure 5, is plotted in Figure 6. The images used here were captured in high dynamic range (see the following chapter), and therefore contain much larger gradients than encountered in conventional 8-bit images. In addition, these images are linear by construction, having a gamma of 1.0. Conventional images are likely to have gamma correction applied to them, and this may affect the distribution of gradients in images.

## 1.6 Response Properties of Cortical Cells

One of the reasons to study the statistics of natural images is to understand how the HVS codes these images. Because natural images are not completely random, they incorporate a significant amount of redundancy. The HVS may represent such scenes using a sparse set of active elements. These elements will then be largely independent.

Here, it is assumed that an image can be represented as a linear superposition of basis functions. Efficiently encoding images now involves finding a set of basis functions that spans image space and ensures that the coefficients across an image ensemble are statistically as independent as possible [26]. The resulting basis functions (or filters) can then be compared to the response properties of cortical cells in order to explain





Figure 5: Example images used for computing gradient distributions, which are shown in Figure 6. In reading order: Abbaye de Beauport, Paimpol, France (July, 2007), Combours, France (July, 2007), Fougères, France (June, 2007), Kermario, Carnac, France (July, 2007).

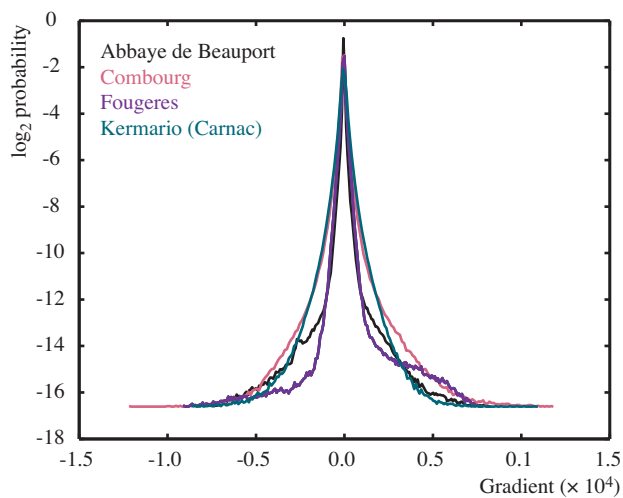


Figure 6: The probability distribution of gradients for the four images depicted in Figure 5.

the early stages of human visual processing. This has resulted in a number of different representations for basis functions, including principal components analysis (PCA) [2, 13], independent components analysis (ICA) [3, 4, 15], Gabor basis functions [10, 11], and wavelets [11, 17, 26].

The PCA algorithm finds a set of basis functions that maximally decorrelates pairs of coefficients; this is achieved by computing the eigenvalues of the covariance matrix (between pairs of pixel intensities). The corresponding eigenvectors represent a set of orthogonal coefficients, whereby the eigenvector with the largest associated eigenvalue accounts for the largest part of the covariance. By using only the first few coefficients, it is possible to encode an image with a large reduction in free parameters [10]. If stationarity is assumed across all images in an ensemble and between all regions in the images, i.e., the image statistics are the same everywhere, then PCA produces coefficients that are similar to the Fourier transform [5]. Indeed, PCA is strictly a second-order method assuming Gaussian signals.

Unfortunately, decorrelation does not guarantee independence. Also, intensities in natural images do not necessarily have a Gaussian distribution, and therefore PCA yields basis functions that do not capture higher-order structure very well [26]. In particular, the basis functions tend to be orientation and frequency sensitive, but with global extent. This is in contrast to cells in the human primary cortex which are spatially localized (as well as being localized in frequency and orientation).

In contrast to PCA, independent components analysis constitutes a transformation resulting in basis functions that are non-orthogonal, localized in space, and selective for frequency and orientation. They aim to extract higher-order information from images [3, 4]. ICA finds basis functions that are as independent as possible [18]. To avoid second-order statistics from influencing the result of ICA, the data is usually first decorrelated (also called whitened), for example using a PCA algorithm. Filters can then be found that produce extrema of the kurtosis [34]. A kurtotic amplitude distribution is produced by cortical simple cells, leading to sparse coding. Hence, ICA is believed to be a better model than PCA for the output of simple cortical cells.

The receptive fields of simple cells in the mammalian striate cortex are localized in space, oriented, and bandpass. They are therefore similar to the basis functions of wavelet transforms [11, 26]. For natural images, strong correlations between wavelet coefficients at neighboring spatial locations, orientations, and scales, have been shown using conditional histograms of the coefficients' log magnitudes [36]. These results were successfully used to synthesize textures [30, 37] and to denoise images [35].

## 1.7 Dimensionality of Reflectance and Daylight Spectra

The reflectance spectra of large numbers of objects have been analyzed for the purpose of determining their dimensionality. We have so far assumed that a spectral distribution is sampled at regular intervals, for instance from 380 to 800 with a 5 or 10 nm spacing. This leads to a representation of each spectrum by 43 or so numbers.

Values in between sample points can be approximated, for instance, by linear interpolation. However, it is also possible to use higher-order basis functions and, thereby, perhaps reduce the number of samples. The question then becomes which basis functions should be used, and how many are required.

It has been found that the eigenvalues of the covariance matrix decay very fast, leading to only a few principal components that account for most of the variance [7]. Similar results have been found for daylight spectra [19], suggesting that spectral distributions can be represented in a low-dimensional space, for instance having no more than six or seven dimensions [23, 24].



## References

- [1] J J Atick and N A Redlich. What does the retina know about natural scenes? *Neural Computation*, 4:196–210, 1992.
- [2] R J Baddeley and P J B Hancock. A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings of the Royal Society of London B*, 246:219–223, 1991.
- [3] Anthony J Bell and Terrence J Sejnowski. Edges are the 'independent components' of natural scenes. *Advances in Neural Information Processing Systems*, 9, 1996.
- [4] Anthony J Bell and Terrence J Sejnowski. The independent components of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.
- [5] T Bossomaier and A W Snyder. Why spatial frequency processing in the visual cortex? *Vision Research*, 26:1307–1309, 1986.
- [6] G J Burton and Ian R Moorhead. Color and spatial structure in natural scenes. *Applied Optics*, 26(1):157–170, 1987.
- [7] J Cohen. Dependency of the spectral reflectance curves of the munsell color chips. *Psychonomic Science*, 1:369–370, 1964.
- [8] Mark E Crovella and Murad S Taqqu. Estimating the heavy tail index from scaling properties. *Methodology and Computing in Applied Probability*, 1(1):55–79, 1999.
- [9] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–358, 1995.
- [10] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12):2379–2394, 1987.
- [11] David J Field. Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In M Farge, J C R Hunt, and J C Vassilicos, editors, *Wavelets, fractals and Fourier transforms*, pages 151–193. Clarendon Press, Oxford, 1993.
- [12] David J Field and Nuala Brady. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, 37(23):3367–3383, 1997.
- [13] Peter J B Hancock, Roland J Baddeley, and Leslie S Smith. The principle components of natural images. *Network*, 3:61–70, 1992.
- [14] Frederic J Harris. On the use of windows for harmonic analysis with the discrete fourier transform. *Proceedings of the IEEE*, 66(1):51–84, 1978.
- [15] H van Hateren and A van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B*, 265:359–366, 1998.
- [16] Bruce M Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975.
- [17] Jarmo Hurri, Aapo Hyvärinen, and Erkki Oja. Wavelets and natural image statistics. In *Proceedings of the 10<sup>th</sup> Scandinavian Conference on Image Analysis*, pages 13–18, June 1997.

- [18] Aapo Hyvärinen. Survey on independent components analysis. *Neural Computing Surveys*, 2:94–128, 1999.
- [19] Deane B Judd, David L MacAdam, and Günther Wyszecki. Spectral distribution of typical light as a function of correlated color temperature. *Journal of the Optical Society of America*, 54(8):1031–1040, 1964.
- [20] Michael S Langer. Large-scale failures of  $f^{-\alpha}$  scaling in natural image spectra. *Journal of the Optical Society of America A*, 17(1):28–33, 2000.
- [21] Hsien-Che Lee. Internet color imaging. In *Proceedings of the SPIE*, volume 3080, pages 122–135, 2000.
- [22] Hsien-Che Lee. *Introduction to color imaging science*. Cambridge University Press, Cambridge, 2005.
- [23] Laurence T Maloney. Evaluation of linear models of surface spectral reflectance with small number of parameters. *Journal of the Optical Society of America A*, 3:1673–1683, 1986.
- [24] Laurence T Maloney and Brian A Wandell. Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3(1):29–33, 1986.
- [25] Chrysostomos L Nikias and Athina P Petropulu. *Higher-order spectra analysis*. Signal Processing Series. Prentice Hall, 1993.
- [26] B A Olshausen and D J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [27] C A Párraga, G Brelstaff, T Troscianko, and I R Moorhead. Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, 15(3):563–569, 1998.
- [28] C A Párraga, T Troscianko, and D J Tolhurst. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology*, 10(1):35–38, 2000.
- [29] F H G Pitt and E W H Selwyn. Colour of outdoor photographic objects. *The Photographic Journal*, 78:115–121, 1938.
- [30] Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.
- [31] Stéphane J M Rainville and Frederick A A Kingdom. Spatial-scale contribution to the detection of mirror symmetry in fractal noise. *Journal of the Optical Society of America A*, 16(9):2112–2123, 1999.
- [32] D L Ruderman and W Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, 1994.
- [33] Daniel L Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5(4):517–548, 1997.
- [34] A van der Schaaf. *Natural image statistics and visual processing*. PhD thesis, Rijksuniversiteit Groningen, The Netherlands, March 1998.
- [35] Eero P Simoncelli. Bayesian denoising of visual images in the wavelet domain. In P Müller and B Vidakovic, editors, *Bayesian Inference in Wavelet Based Models*, volume 141 of *Lecture Notes in Statistics*, pages 291–308, New York, 1999. Springer-Verlag.

- [36] Eero P Simoncelli. Modelling the joint statistics of images in the wavelet domain. In *Proceedings of the 44<sup>th</sup> SPIE Annual Meeting*, volume 3813, pages 188–195, July 1999.
- [37] Eero P Simoncelli and Javier Portilla. Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proceedings of the 5<sup>th</sup> International Conference on Image Processing*, October 1998.
- [38] Mitchell G A Thomson. Higher-order structure in natural scenes. *Journal of the Optical Society of America A*, 16(7):1549–1553, 1999.
- [39] Mitchell G A Thomson. Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems*, 10(2):123–132, 1999.
- [40] D J Tolhurst, Y Tadmor, and T Chiao. Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12:229–232, 1992.
- [41] Christian Ziegeus and Elmar W Lang. Statistics of natural and urban images. In *Proceedings of the 7<sup>th</sup> International Conference on Artificial Neural Networks*, volume 1327, pages 219–224, Berlin, 1997. Springer-Verlag.
- [42] Christian Ziegeus and Elmar W Lang. Statistical invariances in artificial, natural and urban images. *Zeitschrift für Naturforschung A*, 53a(12):1009–1021, 1998.

## Index

Auto-correlation function, 3

Average image color, 2

Blur detection, 3

Color

average image color, 2

correlated color temperature, 2

Contrast

natural image statistics, 3

Correlated color temperature

outdoor environment, 2

Ensemble, 2

Gray-world assumption, 2

ICA, 9

Image

ensemble, 2

Independent components analysis (ICA), 9

Kaiser-Bessel window, 4

Kurtosis, 6, 9

Natural image statistics, 2

$1/f$  failure, 3

$1/f$  statistic, 3

and cortical cells, 9

chrominance, 4

contrast, 3

correlated color temperature, 2

dimensionality of spectra, 9

first-order, 2

higher-order, 2, 6

image ensemble, 2

kurtosis, 6

power spectrum, 3

second-order, 2

skew, 6

stationarity, 9

whitening, 4

Parseval's theorem, 3

PCA, 9

Power spectrum, 3

Principal components analysis (PCA), 9

Reflectance

dimensionality of spectra, 9

Skew, 6

Whitening, 4, 9

Wiener-Khintchine theorem, 3