Evaluating Models of Natural Image Patches

Chris Williams

Neural Information Processing
School of Informatics, University of Edinburgh

January 15, 2018

Outline

- Evaluating Models
- Comparing Whitening and ICA Models
- Spherically Symmetric Distribution
- Lp-spherical Distributions

Evaluating Models

- The natural way to compare models is in terms of the expected log likelihood

$$L = \mathbb{E}[\log p(\mathbf{u}|M)]$$
$$\simeq \frac{1}{n}\sum_{i=1}^{n} \log p(\mathbf{u}_i|M)$$

- $KL(p_{true}||p_M)$ argument shows that log likelihood is highest for correct generative model
- Avoid overfitting issues by using a separate test set to evaluate the expectation
- Eichhorn, Sinz and Bethge (2009) compute the Average Log Loss

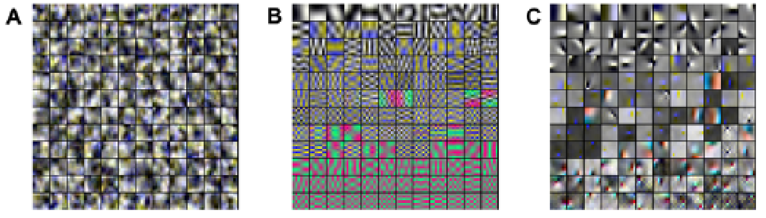
$$ALL = \frac{1}{D}\mathbb{E}[-\log p(\mathbf{u}|M)]$$

where D is the number of (colour) pixels in the patch.
Units: bits/component

Comparing Whitening and ICA Models

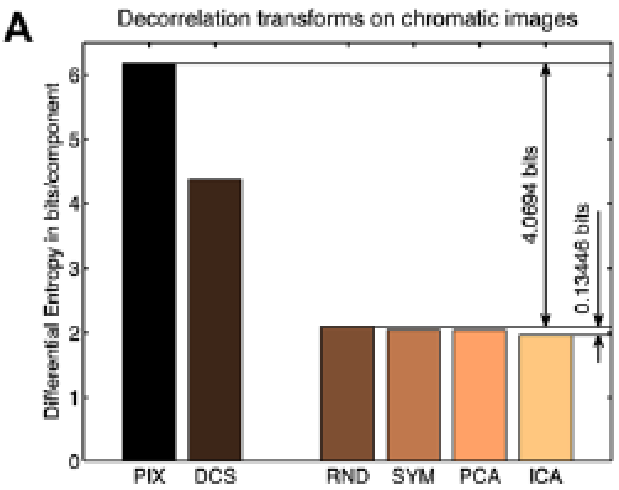
Eichhorn, Sinz and Bethge (2009)

- Recall that ICA basis can be thought of as first whitening, then a rotation in the whitened space
- Compare 4 bases: RND (random in the whitened space), SYM (=ZCA basis), PCA and ICA
- Model for $\mathbf{v} = W\mathbf{u}$ is factorized, they fit a generalized Gaussian to each of the marginals v_i, $i = 1 \ldots, D$



A = RND, B = PCA, C = ICA basis

Figure: Eichhorn, Sinz and Bethge (2009)



- DCS = separation of DC component
- Notice the small differences between RND, SYM, PCA and ICA
- Spherically symmetric distribution (SSD) is much better, at 1.67 bits/component (cf 1.78 for ICA)

Spherically Symmetric Distribution

$$p(\mathbf{u}) \propto f(\mathbf{u}^T \Sigma^{-1} \mathbf{u})$$

- In general the density has elliptical contours
- If $f(z) = \exp(-z)$ then this is a Gaussian
- Model applies more generally, e.g. multivariate Student-t (heavy tails).
- Whitening transformation $\mathbf{v} = W\mathbf{u}$
- Spherical model is a function of $|\mathbf{v}|^2$ s.t. $\Sigma^{-1} = W^T W$
- Method is called *radial Gaussianization* (Lyu & Simoncelli, 2008; Sinz & Bethge, 2008); we first transform with W to get a spherical model, then perform a nonlinear transformation in $r = |\mathbf{v}|$
- Can also approximate this e.g. with a mixture of several Gaussians with same (zero) mean but different scaling of the covariance.

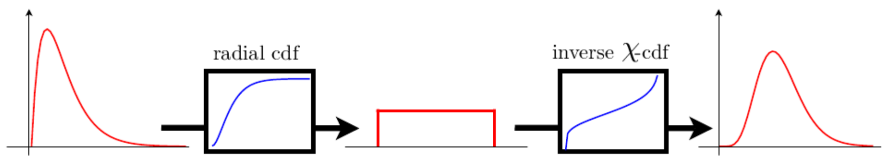


Figure credit: Matthias Bethge

- The SSD model is a better model for image patches than ICA
- However, as it is radially symmetric, it does not prefer the ICA basis over RND, PCA etc. So there seems to be no reason why there should be Gabor-style filters ...
- Radial Gaussianization (RG) has a similar effect to contrast gain control (or divisive normalization, DN)

$$g(r) = \frac{r}{\sqrt{b + cr^2}}$$

- Results in Lyu & Simoncelli (2008) show that RG is superior to DN for image patch modelling

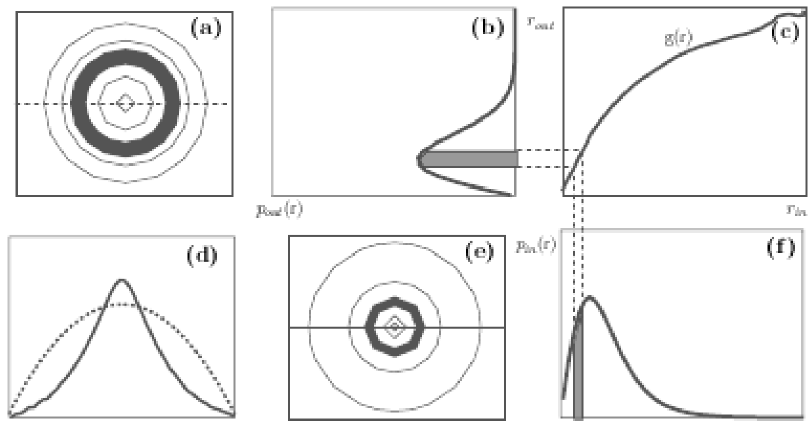


Figure 3: Radial gaussianization procedure, illustrated for two-dimensional variables. Joint densities of (a) a spherical gaussian and (b) a nongaussian SSD (multivariate Student's t). Plotted levels are chosen such that a spherical gaussian has equal-spaced contours. (b,f) Radial marginal densities of the joint gaussian and SSD densities in a,e, respectively. Shaded regions correspond to shaded annuli. (c) Radial map of the RG transform. (d) Log marginal densities of the joint gaussian (dashed line) and SSD (solid line) densities.

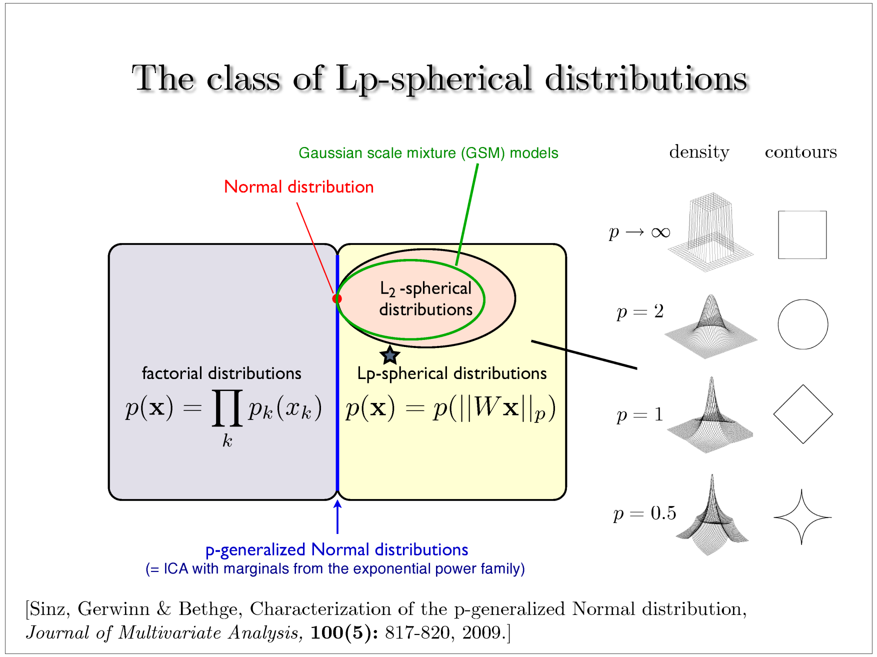
Figure credit: [Lyu and Simoncelli 2009]

Lp-spherical Distributions

- Consider L_p spherical distributions, $p(\mathbf{u}) = p(||W\mathbf{u}||_p)$
- L_p norm

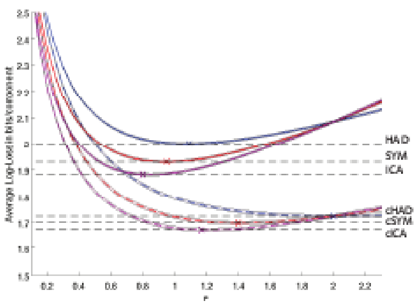
$$||\mathbf{x}||_p = (\sum_{i=1}^{D} |x_i|^p)^{1/p}$$

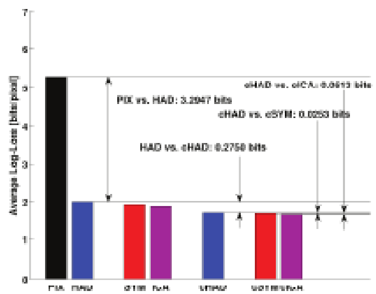
strictly only a norm for $p \geq 1$



Slide credit: Matthias Bethge

Results for Lp-spherical Distributions





Sinz and Bethge (2008)

- HAD basis = Hadamard (similar to RND)
- For $p = 2$ all models are invariant to a rotation of basis
- Focus on the lower lines (top ones are for a p-generalized Normal distribution)
- Results show that lower ALL can be obtained for $p < 2$

- Gabor-type filters (ICA basis) are superior to SYM and HAD bases
- However, this effect is weak: the contribution relative to cHAD is less than 2% in redundancy reduction
- Sinz and Bethge's conclusion: “orientation selectivity is not crucial for redundancy reduction, while contrast gain control may play a more important rôle

	Model	bandpass filtering	orientation selectivity	divisive normalization	complex cell pooling	"deep"-ness	references
Linear models	PCA/ Whitening	✓					Pentland (1984), Srinivasan et al (1982), Field (1987), Atick (1990), Ruderman & Bialek (1994),
Linear models	ICA	✓	✓				Olshausen & Field (1996) Bell & Sejnowski (1997)
ν-spherical models	L_2-spherical model	✓		✓			Zetzsche et al (1999), Schwarz & Simoncelli (2001), Lyu & Simoncelli (2008) Sinz & Bethge (2008)
ν-spherical models	L_p-spherical model	✓	✓	✓			Köster & Hyvarinen (2007) Sinz & Bethge (2009)
ν-spherical models	L_p -nested model	✓	✓	✓	✓	(✓)	Sinz et al (2010) Sinz & Bethge (in press)
	Multilayer ICA	✓	✓			✓	Chen & Gopinath (2000) Hosseini and Bethge (2009)
	Deep Belief Nets (DBN)	✓	✓			✓	Hinton et al (2006)
	Mixture of GSMs	✓	✓	✓	✓		Lee & Lewicki (2002), Karklin & Lewicki (2003+5+8), Bethge & Hosseini (2007)

Slide credit: Matthias Bethge

	Model	bandpass filtering	orientation selectivity	divisive normalization	complex cell pooling	"deep"-ness	references	***redundancy reduction*** [bits/pixel]
Linear models	PCA/ Whitening	✓					Bethge, JOSA A, 2006 Eichhorn, Sinz, Bethge, PLOS CB, 2009	2,7
Linear models	ICA	✓	✓				Bethge, JOSA A, 2006 Eichhorn, Sinz, Bethge, PLOS CB, 2009	2,92
ν-spherical models	L_2-spherical model	✓		✓			Eichhorn, Sinz, Bethge, PLOS CB, 2009 Lyu&Simoncelli, NIPS 2008 Sinz&Bethge, NIPS, 2008	3,05
ν-spherical models	L_p-spherical model	✓	✓	✓			Sinz&Bethge, NIPS, 2009	3,17
ν-spherical models	L_p -nested model	✓	✓	✓	✓	(✓)	Sinz et al, NIPS, 2010	3,2
	Multilayer ICA	✓	✓			✓	Hosseini et al, 2009	3,0
	Deep Belief Nets (DBN)	✓	✓			✓	Theis, Bachelor thesis, 2010	2,9
	Mixture of GSMs	✓	✓	✓	✓		Hosseini et al, patent (pending) 2007	3,3

Slide credit: Matthias Bethge

- Note the technical difficulty in evaluating the ALL for some models (e.g. Karklin and Lewicki, ISA, DBN etc)
- The Bethge and Hosseini reference is a patent (WO/2009/146933, published 10.12.2009)
- Basically a mixture of GSMs. It works by
  - assigning an image patch to a specific class
  - transforming the image patch, with a pre-determined class-specific transformation function
  - coding and quantizing the transformed coefficients
- Mixture of GSMs can be seen as an overcomplete model

References

- M. Bethge and R. Hosseini Patent (WO/2009/146933, published 10.12.2009) Method and Device for Image Compression
- J. Eichhorn, F. Sinz and M. Bethge. Natural Image Coding in V1: How Much Use Is Orientation Selectivity? PLoS Computational Biology 5(4) e1000336 (2009)
- S. Lyu and E. Simoncelli. Nonlinear Extraction of Independent Components of Natural Images Using Radial Gaussianization. Neural Computation 21 1485-1519 (2008)
- F. Sinz and M. Bethge. The Conjoint Effect of Divisive Normalization and Orientation Selectivity on Redundancy Reduction. NIPS*2008 (2008)