Statistical Invariances in Artificial, Natural, and Urban Images

Christian Ziegaus and Elmar W. Lang

Institute of Biophysics, University of Regensburg, D-93040 Regensburg

To answer the question about the way our visual system processes images it has to work with every day, it is necessary to investigate the statistical structure of these pictures. For this purpose we investigated several ensembles of artificial and real-world greyscale images to find different invariance properties: translation invariance by determining an average pair-correlation function, scale invariance by investigating the power spectrum and the coarse graining of the images, and a new hierarchical invariance recently proposed [D. L. Ruderman, Network **5**, 517 (1994)]. The results of our work indicated that the assumption of translational invariance can be taken for granted. Our results concerning the scale invariance are qualitatively the same as those found by Ruderman and others. The deviations of the distributions of the logarithmically transformed images from a Gaussian distribution cannot be seen as clearly as stated by Ruderman. This results from the fact that for a correct determination of the deviations the non-linear transformation must be considered. Depending on the preprocessing of the images the results concerning the hierarchical invariance differed widely. It seems that this new invariance can be confirmed only for logarithmically transformed images.

Key words: Natural Images; Urban Images; Image Statistics; Invariance Properties of Images; Image Preprocessing.

1. Introduction

The development of the mamalian visual system is strongly dependent on early visual stimulation [1] and thus is influenced by the statistics of its environment. Hence the visual system is expected to be optimally adapted to the statistics of natural images it deals with. Natural images are far from random and contain distinctive features and particular types of structures. In order to understand the way visual information is processed and the coding strategy employed within the primary visual cortex the statistical properties of visual input patterns are of primary interest. Early stages of information processing seem to encode only simple features like oriented edges, lines or bars and disparity. They thereby represent visual information in a less redundant and more efficient way. Whether efficient coding means mere redundancy reduction [2, 3], mutual information maximization [4], reconstruction fidelity [5] or sparseness of coding [6] is still in dispute and is currently explored by many groups. However, the proper coding strategy certainly depends on the statistical properties of the stimulus patterns occuring in a natural or urban environment.

There have been several recent investigations of the statistical properties of natural images [7, 8, 9, 10]. Since there is no way to collect enough data to fully characterize an image environment, these studies seek to identify a simple underlying structure or invariance property in the image probability distribution. One such symmetry frequently assumed is translational invariance. Whether it will always hold with natural images is not clear, however. Undoubtedly their most robust statistical property is an invariance to scale [8]. Recently, evidence has been presented supporting the notion of a hierarchical invariance in natural scenes. It relates to the conversion of exponential histograms to Gaussian distributions via local non-linear transformations.

In this investigation we further explore these invariance properties of images taken in natural and urban environments and compare them to an artificial reference ensemble constructed to have strictly Gaussian statistics.

2. Image Ensembles and Preprocessing

Image ensembles have been gathered from natural as well as urban environments. Photographs were taken with a photo camera (Minolta Dynax 7000i,

35-70 mm autofocus) with fixed aperture ($f/5.6$) and variable exposure time.

Processed photographs then have been digitized with a Dia-Scanner ScanMaker 35t at a resolution of 622 dpi or alternatively with the scanner HP ScanJet IIcx using 150 dpi resolution. Both scanners coded the image brightness to eight bits (256 grey levels).

Most images, however, have been sampled (30 s for each scene) with a CCD video camera (JVC Compact VHS-Camcorder GR-AX48, 4-50 mm) with fixed focal length (natural scenes 50 mm, urban scenes 4 mm) and variable aperture. Of each scene five images were directly fed into a computer with a framegrabber (Cortex CX100, Stemmer PC-Syteme GmbH, Kehlheim, Germany). Image brightness again is encoded with 256 grey levels.

From these digitized images with variable sizes smaller ones with a constant size of 256×256 pixels were chosen and formed the various image ensembles from which the image statistics were extracted. Depending on the environment, three natural ensembles (NAT1-3) and one urban ensemble (ZIV) have been formed. In addition, an artificial ensemble (K2) has been considered for comparative purposes. These ensembles will be described in the following.

2.1. The Artificial Ensemble K2

Fourier images with $1/|k|$-amplitude spectra and random phases were generated on a computer and inverse Fourier-transformed into the spatial domain to generate an image which is perfectly scale invariant, hence is characterized by a power spectrum $S(|k|) \propto 1/|k|^2$. To this end a square grid in Fourier space has been chosen with M grid points along any of the two dimensions. Each pixel value represented one Fourier coefficient according to

$$F(k) = \begin{cases} = \dfrac{1}{|k|}\exp(i\phi) = \dfrac{1}{\sqrt{k_x^2 + k_y^2}}\exp(i\phi) \\ \qquad \text{if } k \neq 0 \\ 0 \quad \text{if } k = 0 \end{cases} \tag{1}$$

$$k_{x,y} = \frac{2\pi n_{x,y}}{M\Delta} \text{ with } n_{x,y} = -\frac{M}{2}\cdots\frac{M}{2}, \ \phi \in [0, 2\pi],$$

where $\Delta = 1°$/pixel is the pixel distance in degrees. The phase angles ϕ have been chosen at random from the interval given. With $F(-k) = F^\star(k)$ the Fourier

transform into the spatial domain resulted in an image with real valued pixel intensities

$$I(r) = \frac{1}{\pi}\int_0^\infty \frac{1}{|k|}\cos(-k\,r + \phi)\,\mathrm{d}\,k \tag{2}$$

and a maximal and minimal intensity given by

$$I_{\max} = \sum_{k_x \neq 0}\sum_{k_y \neq 0}\frac{1}{\sqrt{k_x^2 + k_y^2}} \text{ and } I_{\max} = -I_{\min}. \tag{3}$$

The distribution of these pixel intensities $I(x)$ is Gaussian [11] with mean 0 (given by $F(k = 0) = 0$) and standard deviation σ_I given by Parceval's theorem [7, 12]

$$\sigma_I^2 = \sum_{k_x}\sum_{k_y}\left|F(k_x, k_y)\right|^2 = \sum_{k_x \neq 0}\sum_{k_y \neq 0}\frac{1}{k_x^2 + k_y^2}. \tag{4}$$

After determining for each Fourier-transformed image the corresponding maximal and minimal pixel intensities, the latter have been rescaled linearly to span the interval $[0, \ldots, 255]$ according to

$$I(r) = 255.0\frac{I(r) - I_{\min}}{I_{\max} - I_{\min}} = 127.5\left(1 + \frac{I(r)}{I_{\max}}\right). \tag{5}$$

This transformation results in positive pixel intensities only and provides the same range of pixel intensity values for all digital images to be considered later. Figure 1 presents an example of the images thus obtained.

The actual spread of the rescaled pixel intensities can be characterized by the relation (see Table 1 for comparison)

$$\frac{\sigma^2}{\langle I_{\max} - I_{\min}\rangle} \approx \frac{M}{\sqrt{2}}, \ \langle I_{\max}\rangle \approx \frac{\sigma^2}{\sqrt{2}M}. \tag{6}$$

Table 1. Calculated image size $M_{\exp}$ using (6). $\langle I_{\max} - I_{\min}\rangle$ has been determined from 5000 ($M = 512, 256$), 10000 ($M = 128, 64$) and 50000 ($M = 32, 16, 8$) artificial images.

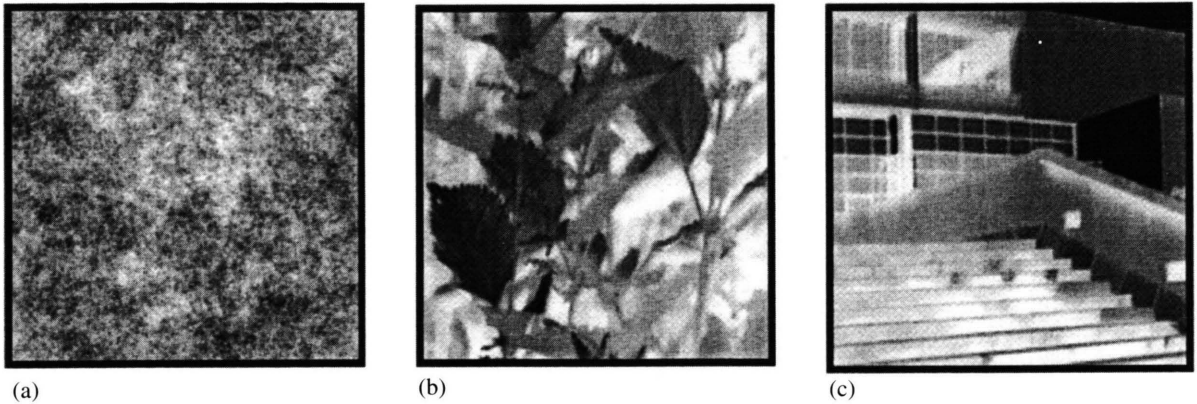| Image size M	$\langle I_{\max} - I_{\min}\rangle$	$M_{\exp}$
512	27863.1072	507.1677
256	12242.5424	255.5978
128	5296.9640	128.6350
64	2235.6402	64.9087
32	913.3623	34.3713
16	356.2586	16.5934
8	128.6634	8.3948

(a) (b) (c)

Fig. 1. Examples of the images used in this study. Figure (a) shows an example of the artificial image ensemble K2. It has been obtained by inverse Fourier–transformation of images in frequency space with a $1/|k|^2$–power spectrum and random phase. They resemble the images examined by Field, though they were obtained by multiplying the amplitude spectrum of random images with $1/|k|$ and transforming the resulting images back to the spacial domain. In Fig. (b) an example of the natural ensemble NAT3 representative for all examined natural images can be found. This image has been obtained with a CCD camera within a forest and represents averages over five single frames (see text for details). Figure (c) is an example for the images of the urban image ensemble ZIV. They have been obtained with a CCD camera at different places at the campus of the University of Regensburg. Grabbing five pictures and averaging over these images results in the example shown here.

The distribution of pixel intensities $I(x)$ after rescaling with I_{max} as calculated using (5) is thus given by

$$P(I) = \frac{\sigma_I}{255 M \sqrt{\pi}} \exp \left\{ - \left(\frac{\sigma_I}{2M} \right)^2 \left(\frac{I}{127.5} - 1 \right)^2 \right\}. \tag{7}$$

2.2. The Natural Ensemble NAT1

Photographs have been taken on a sunny winter day around a little lake, with focal lengths 35 mm (NAT135) and 70 mm (NAT170), respectively and aperture $f/5.6$. The pictures have been digitized with a Dia-Scanner (ScanMaker 35t) with 622 dpi resolution. Then 96 images with size 256×256 pixels each were extracted randomly from these digitized images. It is to be noted that due to an internal transfer function the scanner did not represent the whole range of grey levels properly and produced pixel histograms with frequent gaps. This led to artefacts in the statistical evaluation and renders the results obtained of limited value only.

2.3. The Natural Ensemble NAT2

Photographs have been taken on a sunny day in April at the border and within a forest, with focal lengths 35 mm (NAT235) and 70 mm (NAT270) and aperture $f/5.6$. The pictures have been digitized with a scanner (HP ScanJet IIxc) with 150 dpi resolution, and a total of 102 images with size 256×256 pixels each were extracted at random then.

2.4. The Natural Ensemble NAT3

Various scenes have been taken within a forest on a windy day in June with a CCD video camera with fixed focal length (50 mm) and variable aperture. From every scene filmed for 30 s five single images have been gathered and directly fed into the computer with a framegrabber. From these individual frames an average image has been constructed, and then 100 smaller sized images (256×256 pixels each) have been extracted at random. One example of these images can be found in Figure 1.

2.5. The Urban Ensemble ZIV

Scenes have been filmed with a video camera during August around the campus of the University of Regensburg, with focal length 4 mm and variable aperture. The small focal length has been chosen to have enough structure in the images and to avoid large unstructured surfaces. Also any curvilinear image distortions introduced by the optics of the camera could

Table 2. Summary of the used focal lengths for certain ensembles and the resulting pixel distances acccoring to (8).

Focal length [mm]	Ensemble	Pixel distance [degree]
4	ZIV	0.0877
50	NAT3	0.0083
35	NAT1/NAT2	0.065
70	NAT1/NAT2	0.034

be avoided thereby. Naturally occuring contours like trees have been carefully avoided, too. The scenes thus contain only man-made structures as may be seen from Figure 1. Again 97 images of size 256×256 pixels have been extracted for later analysis.

2.6. Estimation of Pixel Distances

Pixel distances will be expressed in degrees of angle a due to large variations in the size of the visual field with changing focal length. A reference line of known length b has been placed a known distance d apart and has been photographed or filmed. With increasing focal length the number of pixels p increases, hence the angular range per pixel decreases. Table 2 summarizes the pixel distances estimated according to

$$a = 2p^{-1} \arctan\left(\frac{b}{2d}\right). \tag{8}$$

3. Ensemble Statistics

To evaluate the image statistics and to discover possible invariance properties, raw pixel intensities have been transformed first either linearly or non-linearly to difference or log-contrast intensities, respectively. According to the Weber-Fechner-law [13] the sensational strength of any sensory input variable is generally proportional to the logarithm of the stimulus intensity. Hence any pixel intensity has been transformed to a logarithmic contrast intensity

$$\Phi(x) \equiv L(x) = \ln\left(\frac{I(x)}{I_0}\right) \tag{9}$$

with I_0 chosen to result in an intensity distribution with zero average, i. e. $\sum_x L(x) = 0$. This logarithmic contrast function better mimics the behaviour of the visual system, which also focuses on contrast rather than intensity differences [6, 14]. Further any inhomogenities concerning the light flux falling onto

a surface are thereby equalized largely [15 - 17]. Besides this non-linear transformation the pixel intensities have been linearly transformed, too, to yield difference intensities with zero average

$$\Phi(x) \equiv D(x) = I(x) - \langle I \rangle$$

$$\text{with} \quad \langle I \rangle = \frac{1}{N_x} \sum_x I(x), \tag{10}$$

with N_x the number of pixels in the image. We will refer to these linearly transformed pixel intensities as difference intensities henceforth. All digitized images have been treated in both ways, and some results will depend on the data treatment as will be demonstrated shortly.

3.1. Translational Invariance

Almost all investigations so far generally assumed translational invariance of the images considered, though this is by no means trivial as landscape images will always have sky in their upper part, for example. If images are indeed translationally invariant, then any two-point-correlation function $A_x(y) = \langle \Phi(x) \Phi(x + y) \rangle$ of pixel intensities at points x and y will depend on their relative distance only and not on x itself. The average has to be taken over all images comprising the ensemble under investigation. Besides calculating, for different starting points x, the second order correlation function $A_x(y)$ of pixel intensities, whether expressed as logarithmic contrast values or as difference intensities, an average correlation function has been determined also according to the relation

$$\bar{A}(y) = \frac{1}{N_x} \sum_x A_x(y) = \frac{1}{N_x} \sum_x \langle \Phi(x) \Phi(x + y) \rangle, \tag{11}$$

which gives the mean over the number of pixels contained in any given image. Whereas any comparison of correlation functions centered at individual origins x is difficult because of the unavoidable noise and because of an insufficient number of members of the ensemble, the averaged correlation functions provide a much smoother measure of translational invariance. The results do not depend on whether difference intensities or logarithmic contrast values are used to calculate the intensity correlation function. Hence only the results where the difference intensities have been used are shown in this section.
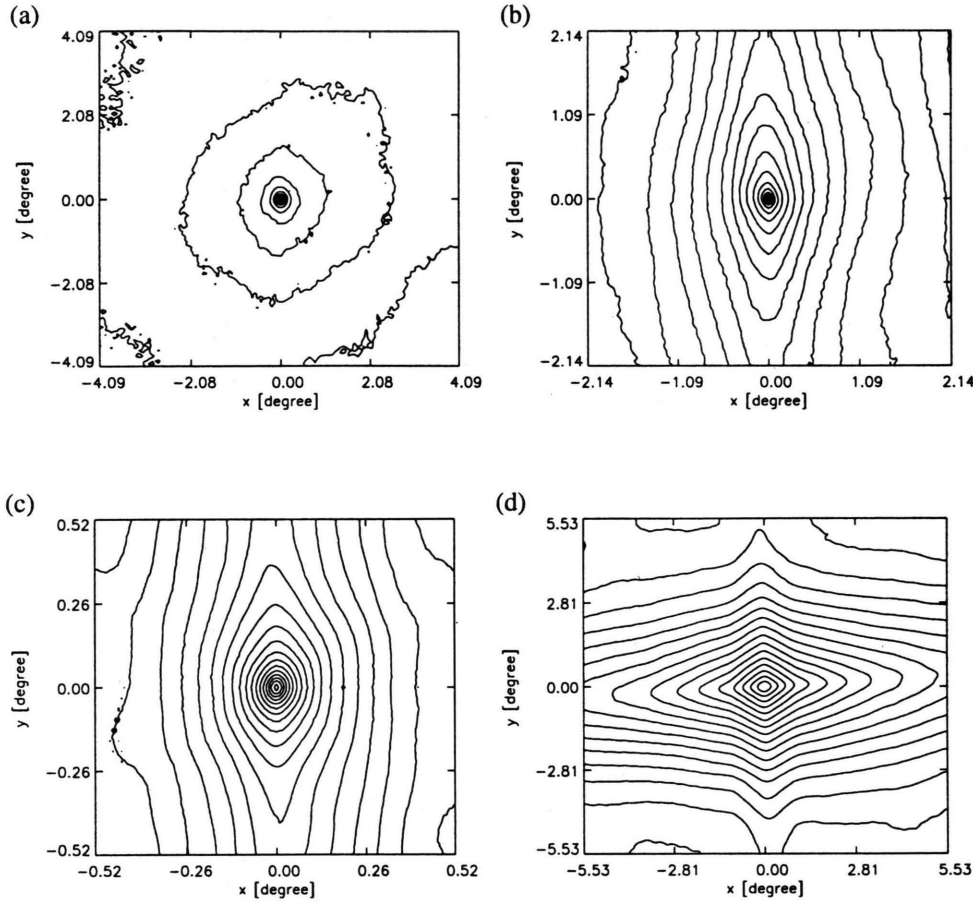
(a)

(b)

(c)

(d)

Fig. 2. Average correlation function $\bar{A}(y)$ for the ensembles NAT1 (a), NAT2 (b), NAT3 (c) and ZIV (d) for the difference intensities. Contours are shown at equal intervals of correlation. It can be seen that vertical correlations increase from NAT1 to NAT3, which can be related to the increasing number of vertical structures contained within the images. In contrast, maximal correlations can be found in the horizontal direction reflecting the large horizontal structures in the pictures of the ensemble ZIV.

The average correlation functions of the various image ensembles are compared in Fig. 2 for the natural images and for the images from an urban environment. Though seemingly translationally invariant, rotational isotropy is increasingly lost as more and more vertical structures (stemming from trees largely) dominate the natural images, whereas the average correlation function from images of the urban environments reflect the predominance of horizontal contours.

3.2. Scale Invariance

Since Deriugin [18] investigated invariance properties of television images, several researchers demonstrated the scale invariance of natural images [7 - 9, 19 - 21]. If scale invariance prevails, the power spectrum $S_\Phi(k)$ of the intensity distribution, given by

$$
\begin{aligned}
S_\Phi(k) &= \int dy \ \exp(-i k \, y)\langle \Phi(x)\Phi(x+y)\rangle \\
&= \left\langle \left| \int dy \ \exp(-i k \, y)\Phi(y)\right|^2 \right\rangle
\end{aligned}
\tag{12}
$$

should, after averaging over all orientations, scale like

$$
S_\Phi(|k|) \propto \frac{1}{|k|^{2-\eta}}
\tag{13}
$$

with $|k|$ the modulus of the spatial frequency and $\eta \neq 0$ an anomalous exponent. With a nonclassical

exponent, pixel intensities need to be renormalized upon changing the scale according to

$$\Phi(\boldsymbol{x}) \rightarrow c^\nu \Phi(c\boldsymbol{x}) \tag{14}$$

with c the scale factor and ν a universal exponent independent of the physical property studied. Image statistics thus become independent of the lens' focal length used, i. e. they are independent of the distance of observation.

After preprocessing, the images sampled via photo camera or video camera are discretized versions of the environment portrayed. To calculate the power spectrum, the substitutions

$$\boldsymbol{x} \quad \rightarrow \quad x_{mn} = (ma, na), \tag{15}$$

$$\Phi(\boldsymbol{x}) \quad \rightarrow \quad \Phi_{mn} = \phi(ma, na) \tag{16}$$

had to be made, resulting in a discretized version of the power spectrum given by

$$S_{\text{est}}(k_{rs}) = \left(\frac{a}{2\pi M^2}\right)^2 \frac{1}{N_i} \tag{17}$$

$$\cdot \sum_{i=1}^{N_i} \left| \sum_{m,n=0}^{M-1} \phi_{mn}^i W_{mn} \exp(-2\pi i k_{rs} x_{mn}) \right|^2,$$

with a the pixel distance measured in degrees, N_i the number of images in the ensemble, $0 \le m, n \le 255$ the pixel index, $k_{rs} = \left(\frac{2\pi r}{Ma}, \frac{2\pi s}{Ma}\right)$, $M = 256$, $-\frac{M}{2} \le r, s \le \frac{M}{2}$ the discretized spatial frequency and M the number of pixels in each spatial direction. W_{mn} represents a window function, the detailed form of which is not too important, to alleviate the artefacts, a two-dimensional Fourier transform of a square image will inevitably exhibit. A two-dimensional Bartlett window is used for convenience [8, 22].

Scale invariance must show up in any arbitrarily chosen physical property of the image ensemble, but it does not tell the form of the stationary distribution from which the $\Phi(\boldsymbol{x})$ are drawn. Both aspects may be further investigated through the process of *coarse graining* [23], which replaces an $N \times N$ block of pixel intensities by their average according to

$$\Phi_N = \frac{1}{N^2} \sum_{m,n=1}^{N} \phi_{mn}. \tag{18}$$

If the underlying probability distribution $P_N(\Phi)$ of the scaling fields with scaling variable N is indeed scale invariant, it must not change upon changing N except for some renormalization of the scaling fields (14) in case of an anomalous exponent $\eta \neq 0$ [24]. In order to compare pixel histograms of different N, the scaling fields should be normalized with their rms value. With $\Phi(\boldsymbol{x}) \equiv D(\boldsymbol{x})$ representing difference intensities and normally distributed pixel intensities, one would have a Gaussian distribution according to

$$P(\tilde{D}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\tilde{D}^2\right)$$

$$\text{with} \quad \tilde{D} = \frac{D}{D_N^{RMS}}. \tag{19}$$

In case of non-linearly transformed log-contrast intensities $\Phi(\boldsymbol{x}) \equiv L(\boldsymbol{x})$, the correspondingly transformed and normalized distribution with zero average becomes

$$P(L) = \frac{\sigma_I \sigma_L}{255 M \sqrt{\pi}} \exp(\sigma_L L + \bar{L})$$

$$\cdot \exp\left\{-\left(\frac{\sigma_I}{2M}\right)^2 \left(\frac{\exp(\sigma_L L + \bar{L})}{127.5} - 1\right)^2\right\},$$

$$\sigma_L^2 = \int L^2 P(L)\,\mathrm{d}L, \tag{20}$$

$$\bar{L} = \ln(I_0) = \int_0^{255} \ln(I)\,P(I)\,\mathrm{d}I$$

$$= \int_{-\infty}^{\ln(255)} L\,P(L)\,\mathrm{d}L.$$

Except for investigating the distribution of difference intensities or the logarithmic contrast values it is advantageous also to consider the distribution of local gradients $G \approx |\nabla \Phi|$ in the images. If the $\Phi(\boldsymbol{x}) \equiv D(\boldsymbol{x})$ are normally distributed and scale invariant, a Rayleigh distribution of local gradients would result [25, 26]:

$$P(\tilde{G}) = \frac{\pi}{2}\tilde{G} \exp\left[-\frac{\pi}{4}\tilde{G}^2\right] \text{ with } \tilde{G} = \frac{G}{\bar{G}_N}. \tag{21}$$

The corresponding distribution of local gradients $G = \sqrt{G_x^2 + G_y^2}$ in case of logarithmically transformed pixel intensities $\Phi(\boldsymbol{x}) \equiv L(\boldsymbol{x})$, as well as the average local gradient $\bar{G} = \int \mathrm{d}G_x \int \mathrm{d}G_y G \cdot P(G_x, G_y)$ must be obtained numerically assuming factorization of the joint probability density
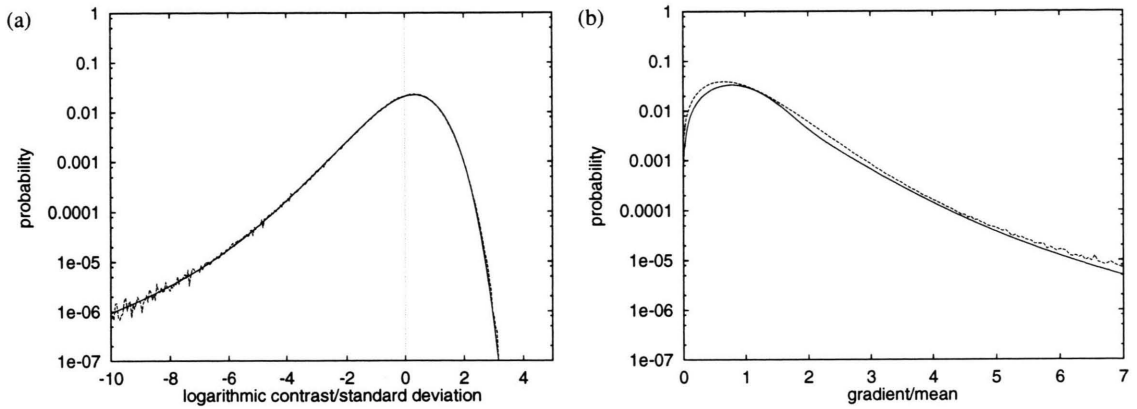
Fig. 3. Comparison between the theoretically derived and the experimentally determined distribution of the non–linearly transformed pixel intensities (a) and corresponding local gradients (b). A very good match between the distributions can be seen. The distribution of gradients does not match exactly due to the numerical calculations. The parameters according to (20) are: $\sigma_I = 1487.499$, $\sigma_L = 0.268$, $\bar{L} = 4.815$.
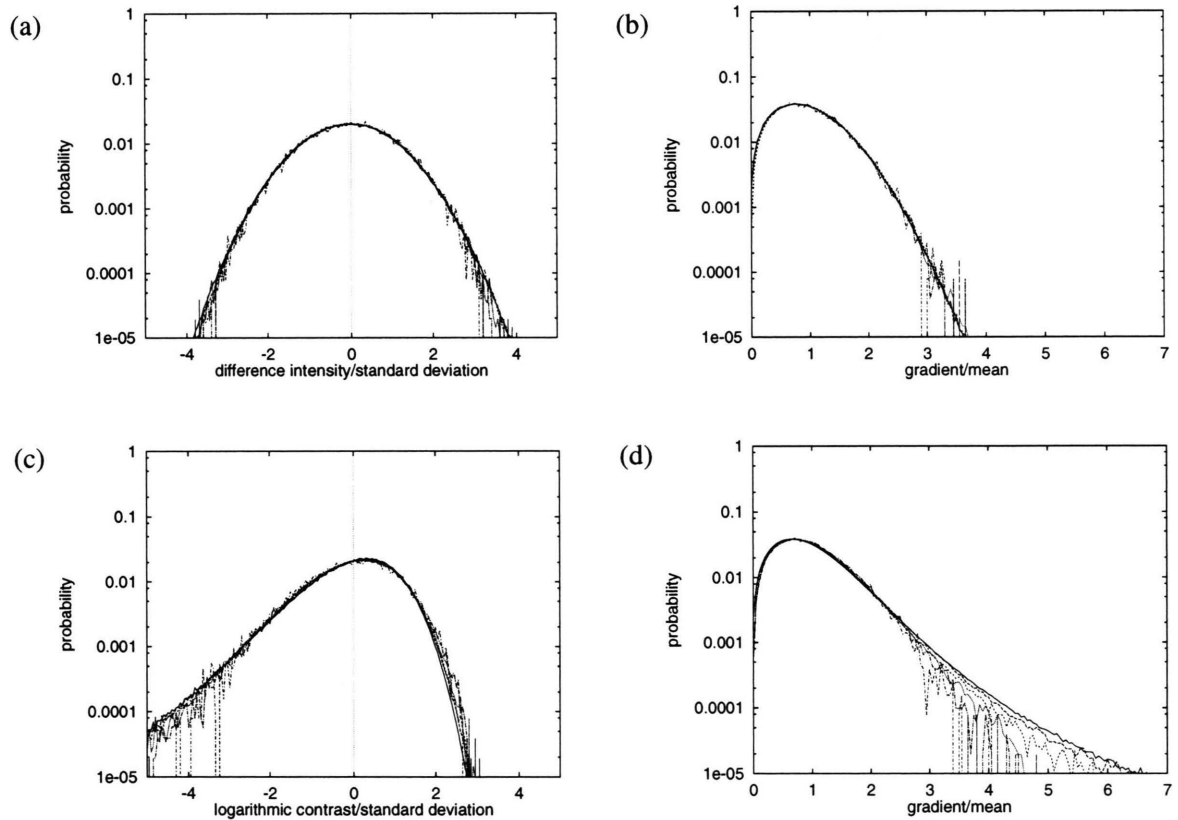

Fig. 4. Scaling of the distributions of difference intensities (a, b), and logarithmic contrast values (c, d) with the corresponding distributions of gradients (b, d) over scales $N = 1, 2, 4, 8, 16$, and 32 for the artificial ensemble K2. The histogram for the difference intensities resembles almost perfectly a Gaussian or Rayleigh distribution, respectively. The deviation of the distributions of logarithmic contrast values from a Gaussian or Rayleigh–distribution must be attributed to the non–linear preprocessing. Hence, to determine differences from Gaussian behaviour when examining natural and urban ensembles, one must not use a Gaussian or Rayleigh–distribution for comparison.

$P\left(G_x, G_y\right) = P\left(G_x\right) P\left(G_y\right)$ with $P\left(G_{x,y}\right)$ chosen according to (20). The resulting theoretical distribution functions are compared to the experimental results of the artificial image ensemble K2 in Figure 3.

3.2.1. Statistics of the artificial ensemble K2

As a reference, the statistics of the artificial ensemble K2 is presented in Fig. 4 for the linearly and the non-linearly transformed pixel intensities. The related power spectrum, calculated via the Fourier transform of the correlation function according to

$$S_\Phi\left(k\right) = \int d y \, \exp\left(-i k \, r\right) \langle \Phi\left(x\right) \Phi\left(x + y\right) \rangle,$$

(22)

corroborates the scale invariance property of the images forming the ensemble K2. Hence, $S\left(|k|\right) \propto |k|^\alpha$ yields with $\alpha = -1.951 \pm 0.052$ the best estimate in the least squares sense. The small deviation of the exponent from the theoretical value $\alpha = 2$ is due to the large scatter of the data at high wave numbers. Also the pixel intensities, having been normalized to 256 greylevels, exhibited some rounding errors leading to small deviations from the exact scaling behaviour, too. Because of the large scatter at high wave numbers, the power spectra of natural and urban image ensembles are not to be obtained directly via a Fourier transform of the related second order correlation functions of pixel intensities. The resulting power spectra are averages over all orientations, of course. A coarse graining of the pixel intensities has been performed for the scales $N = 1, 2, 4, 8, 16,$ and 32. Whereas the linearly transformed pixel intensities show the expected Gaussian and Rayleigh distributions, the non-linearly transformed pixel values show characteristic deviations due to the corresponding non-linear transformation of the probability density function, as can be seen from Fig. 4 for the linearly and the non-linearly transformed intensities. Hence all logarithmically transformed pixel contrast values of the various image ensembles investigated have to be compared to these transformed probability distributions to judge any characteristic deviations of natural image histograms from random image histograms. The orientationally averaged power spectra, however, seem to be robust against any of these transformations.

Table 3. Summary of the anomalous exponent η according to (13).

Ensemble	Transformation	Slope $-2 + \eta$	η
K2	linear	-1.969 ± 0.002	0.031
K2	non–linear	-1.933 ± 0.002	0.067
NAT235	linear	-1.967 ± 0.019	0.034
NAT235	non–linear	-2.041 ± 0.015	-0.041
NAT270	linear	-2.215 ± 0.016	-0.215
NAT270	non–linear	-2.215 ± 0.018	-0.215
NAT3	linear	-2.319 ± 0.054	-0.319
NAT3	non–linear	-2.369 ± 0.062	-0.369
ZIV	linear	-2.634 ± 0.052	-0.634
ZIV	non–linear	-2.665 ± 0.054	-0.665

3.2.2. Statistics of the natural ensembles NAT

Again the power spectra show an increasing asymmetry in going from NAT1 to NAT3 due to the increasing predominance of vertical contours in these images, though the results are not shown here. The spectra show an approximate scaling over more than five orders of magnitude with small anomalous exponents collected in Table 3. Results are also fairly robust against any transformation of the pixel intensities. The pixel histograms show characteristic deviations from the corresponding distributions of the K2 ensemble. Since we want to show that if pixel intensities are logarithmically transformed the deviation from Gaussian behaviour cannot be seen as clearly as stated by Ruderman [8], we will show only the distributions of non-linearly transformed pixel intensities.

Figure 5 shows the histograms for the non-linearly transformed pixel intensities of the ensemble NAT3. The averaging procedure used to obtain these images, however, had no noticable effect on the histograms. It is our finding that any kind of smoothing applied by averaging over a certain number of images or by convolving the images using a Gauss-kernel improves the results of the coarse graining procedure since such operations lead to smoother pixel histograms for the linearly as well as for the non-linearly transformed pixel intensities.

Most remarkable is the almost exponential distribution for high logarithmic contrast values and gradients as can be seen clearly from Figure 5.

3.2.3. Statistics of the urban ensemble ZIV

The images of this ensemble contain man-made structures only with prominent horizontal and vertical contours. This is reflected in the power spectrum
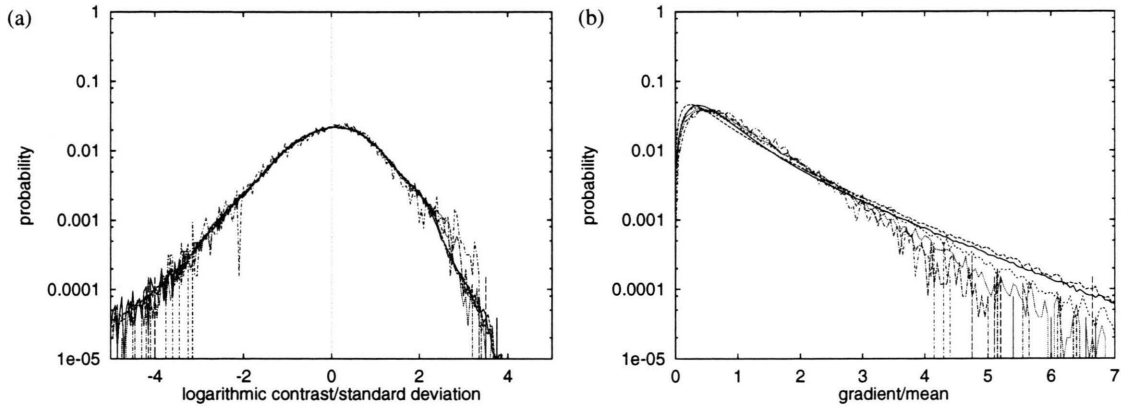
Fig. 5. Scaling of the distributions of logarithmic contrast values (a) and gradients (b) over scales $N = 1, 2, 4, 8, 16$, and 32 for the natural ensemble NAT3.
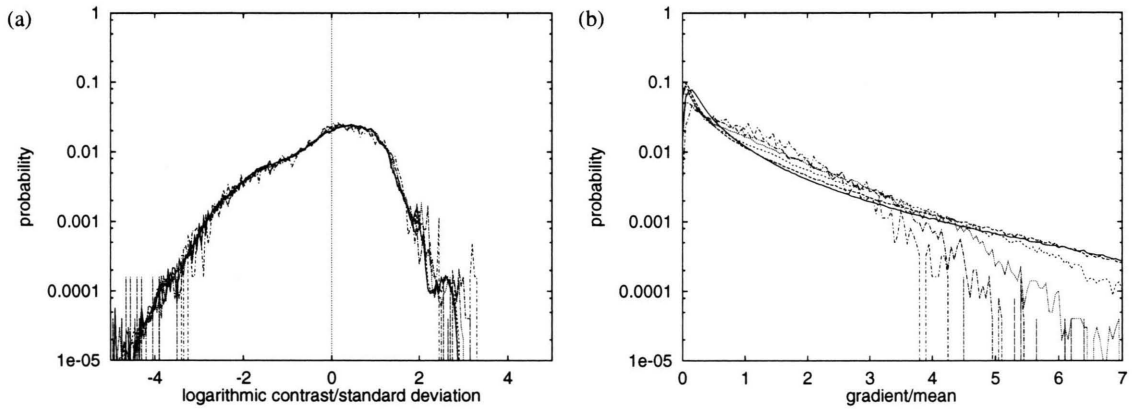

Fig. 6. Scaling of the distributions of logarithmic contrast values (a) and gradients (b) over scales $N = 1, 2, 4, 8, 16$, and 32 for the urban ensemble ZIV.

which is very asymmetric along the k_x and k_y direction. It is also of interest that the anomalous scaling exponent η of the orientationally averaged power spectrum shows the largest anomalous exponent of all image ensembles considered in this study (cf. Table 3). Still scaling is confirmed over almost five orders of magnitude.

The histogram of the linearly transformed pixel intensities shows an almost perfect coincidence with a Gaussian distribution with small deviations for very low pixel intensities only. To the contrary the histogram of local gradients exhibits pronounced deviations from a Rayleigh distribution at very small and high gradients and also strong suppression for intermediate values. Again an almost exponential probability distribution is obtained experimentally over most of the range of local gradients encountered. The

decay of the exponential is much weaker than in case of the natural ensemble NAT3, i. e. high local gradients are much more probable in urban than in natural environments. The histograms for the logarithmically transformed intensities and the corresponding gradients are shown in Figure 6.

3.2.4. Comparison of logarithmical contrast distributions

Figure 7 shows the distributions of the logarithmically transformed pixel intensities for all ensembles together with the theoretically derived distribution and the histograms of local gradients.

The distributions of the logarithmically transformed pixel intensities show for all ensembles only small differences at high log contrast values ($I \gg I_0$)
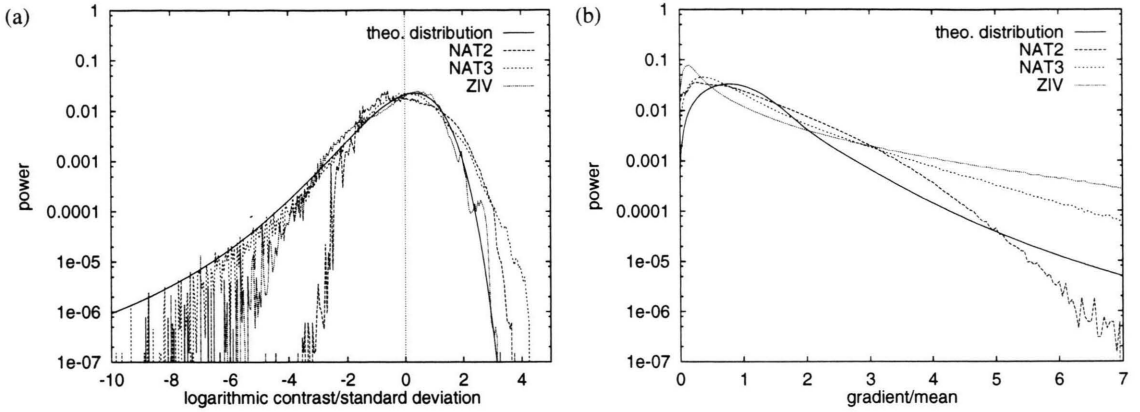
Fig. 7. Scaling of the distributions of logarithmic contrast values (a) and gradients (b) for the real–world ensembles NAT2/NAT3/ZIV and the theoretically derived histograms from Figure 3.

when compared to the theoretically distribution. For low values ($I \ll I_0$) the ensemble NAT2 exhibits the greatest deviations, whereas the other ensembles again resemble the theoretical histogram rather well. The images of the ensembles NAT3 and ZIV comprise a greater range of grey values, since the photographs underlying the ensemble NAT2 have been quite dark. Again this demonstrates the dependence of the results concerning the coarse graining on the method used to construct the image ensembles.

The comparison of the experimentally obtained distributions of local gradients of the ensembles NAT3 and ZIV with the theoretical histogram shows an increasing probability of very small and very large gradients in going from the theoretical distribution to NAT3 and ZIV. This is due to the increasingly pronounced edges and large unstructured areas in the latter images. Ensemble NAT2 shows similar results for small gradients, but results become unreliable for large local gradients. Since the range of grey values in these images is quite small, the occurence of large gradients will be underestimated in the corresponding probability distribution.

3.3. Hierarchical Invariance

Recently Ruderman [8] discussed the possibility of a hierarchical invariance of natural images. It is related to the observation that simple linear filtering of logarithmic contrasts produces exponential histograms much like those observed experimentally. If these exponential tails are due to a superposition of many distributions with widely differing variances,

one may try to find a local non-linear transformation which can turn the distributions to Gaussians. The transformation proposed by Ruderman [8] amounts to calculating

$$\Psi(x) = \frac{\Phi(x) - \bar{\Phi}(x)}{\sigma(x)} \tag{23}$$

with $\bar{\Phi}(x)$ the average pixel intensity (whether linearly or non-linearly transformed) within a block of size $N \times N$ pixels and $\sigma(x)$ the standard deviation of pixel intensity fluctuations within the block. This process enlarges regions of low local contrast and reduces regions of large local contrast. Besides these *variance modified images* one may as well consider so called *variance images*. These have been constructed in the present investigation in two different ways. One may either substitute any pixel intensity by the related variance of pixel intensity fluctuations within block $N \times N$ without thereby changing the image size or one may substitute the whole block of pixel intensities by the variance of their intensity fluctuations. We will refer to both procedures as variance images *without* and *with block substitution*, respectively. The pixel intensities of these variance images may be transformed in the same way as the original images. The interesting observation is that these variance images seem to exhibit similar histograms as do the original images. This observation asks for the possibility to iterate this procedure and to its possible outcome. In doing so one may select the block size according to the smallest possible kurtosis of the distribution of pixel intensities, as Gaussian distributions are charac-
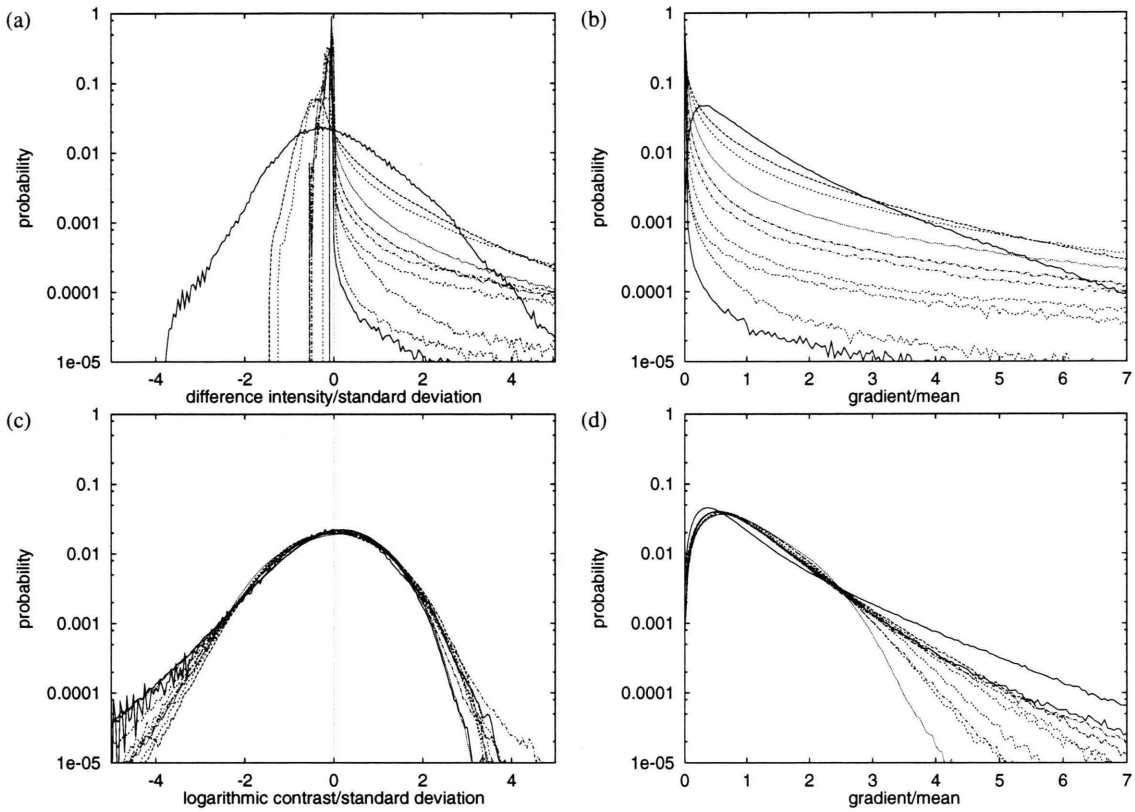
Fig. 8. Distributions of the linearly (a, b) and non–linearly (c, d) transformed pixel intensities (a, c) and the corresponding local gradients (b, d) for ensemble NAT3 and the iteration procedure *without* block substitution. For each transformation ten steps with 100 images have been performed.

terized by vanishing kurtosis (fourth moment of the distribution). Hence the resulting histograms would resemble Gaussian distributions as closely as possible. Variance modified as well as variance images have been constructed for both the artificial ensemble K2 to provide a reference system and the ensemble NAT3 of natural images. The results for the other natural ensembles NAT1/2 and the urban ensemble ZIV are qualitatively the same as for ensemble NAT3 and are not shown here.

3.3.1. The artificial ensemble K2

Ten iterations with 100 images of the ensemble K2 have been performed to obtain variance modified images of block sizes 3×3 to 19×19 pixels. For every iteration the block size resulting in the lowest kurtosis has been chosen. If a hierarchical invariance prevails, all histograms should exhibit a similar shape, which

was clearly not observed with linearly transformed pixel intensities but seems to hold in case of non-linearly transformed pixel intensities (log contrasts). The calculations for the artificial ensemble have been performed for comparative purposes.

3.3.2. The natural ensemble NAT3

Again ten iterations with 100 images of the ensemble NAT3 have been performed to obtain variance modified images of block sizes 3×3 to 19×19 pixels and the block size resulting in the lowest kurtosis has been chosen during every iteration. The resulting histograms for linearly transformed pixel intensities are shown in Fig. 8 without and in Fig. 9 with block substitutions. Large deviations from Gaussian statistics are found in case of linearly transformed pixel intensities which do not support the hierarchical invariance hypothesis. As with the ensemble K2, the
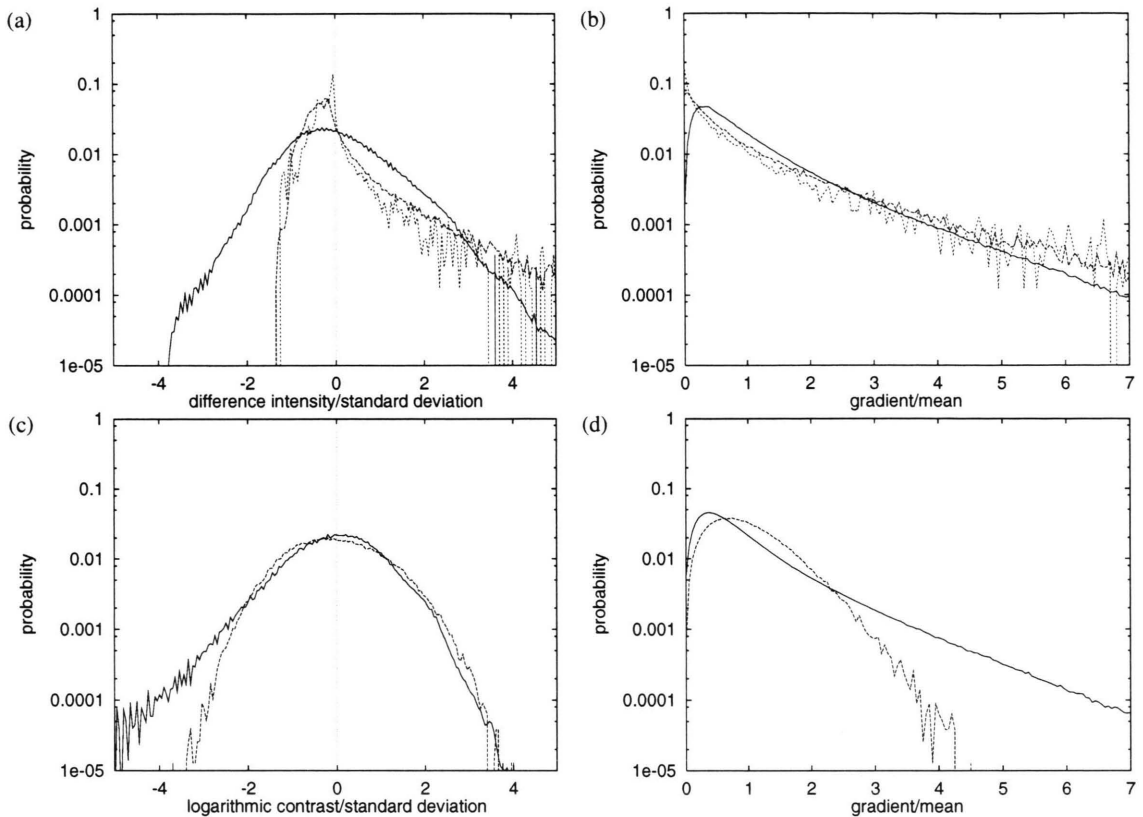
Fig. 9. Distributions of the linearly (a, b) and non–linearly (c, d) transformed pixel intensities (a, c) and the corresponding local gradients (b, d) for ensemble NAT3 and the iteration procedure *with* block substitution. As with ensemble K2 only a few steps where possible before the images became too small for further calculations.

logarithmic contrast histograms do indeed approach Gaussian statistics with increasing number of iterations, hence supporting the notion of a hierarchical invariance. Contrary to scale invariance the result does depend strongly on the way the raw pixel intensities are transformed.

4. Conclusions

Natural images are far from random and occupy an infinitesimally small volume in the space of all possible images. The information contained in natural images is thus highly redundant and asks for a more convenient encoding scheme than simply expressing them on a pixel-by-pixel basis. It is the statistical structure of these images which determines the most suitable encoding algorithm according to some well defined optimality criterium. Since there is no way to collect enough data to fully characterize an image

environment one seeks to identify possible symmetries and invariance properties of the underlying image probability distribution.

Recent investigations of the statistical nature of natural images having generally assumed a translational invariance of such images, have demonstrated an invariance to scale and have proposed a new hierarchical invariance in these images [7, 6, 8].

We have investigated all three invariance properties in ensembles of natural and urban images and, further, constructed an artificial ensemble of random images with purely Gaussian statistics. Although one might not expect translational invariance to hold in non-random images in general – sky will always be in the upper part of any natural or urban image, for example – we could provide strong evidence in favour of translational invariance in the image ensembles using a smoothing procedure to calculate second order correlation functions of pixel intensities. This

helped alleviate the unavoidable scatter in the correlation functions due to insufficient ensemble averaging. The calculated two point correlation functions were mostly highly asymmetric, thus reflecting prominent contours in the underlying images. This is most obvious in comparing the ensembles NAT3 and ZIV.

The artificial ensemble K2 proved useful to explore the characteristics of any deviations of histograms of pixel intensities and local gradients in non-random images from Gaussian statistics. This is important, as we have shown that any non-linear preprocessing of raw pixel intensities renders Gaussian and Rayleigh distributions inappropriate to compare with the corresponding pixel and local gradient histograms of natural and urban images. After appropriate coordinate transformations, the resulting distributions of logarithmically transformed pixel intensities and related distributions of local gradients exhibited less dramatic deviations from Gaussian and Rayleigh distributions, though the general conclusions of earlier investigations [8] remain still valid. Concerning the hierarchical invariance of natural images recently discussed by Ruderman [8] we persued the proposed iterative scheme of replacing pixel intensities by the variance of the intensity distribution within a pixel block of size $N \times N$. This has been done in two different ways whereby the size of the image has been reduced according to the appropriate block size or not. With block substitution the iteration could be performed only two or three times, whereas without block substitution as many as ten iterations could be performed. Contrary to results reported by Ruderman [8] the outcome of this iterative procedure depended strongly on the preprocessing of the raw pixel intensities. With logarithmic contrast pixel intensities, indications of a hierarchical invariance could indeed be found while with difference intensities no such invariance was observed. Contrary to scale invariance, these results are not robust against any non-linear transformation of the pixel intensities and renders this invariance related to the hierarchical structure of natural images less generally valid. As the visual information processing system of mamals seems to rely on logarithmic contrast intensities, the proposed variance normalization invariance may still have relevance to the possibility to Gaussianize natural images via this iterative procedure. Exploring natural image statistics further is clearly essential to a fundamental understanding of visual information processing.

Acknowledgement

[1] C. B. Blackmore and G. Cooper, Nature, London **228**, 477 (1970).
[2] F. Attneave, Psychol. Rev. **61**, 183 (1954).
[3] H. B. Barlow, Possible principles underlying the transformation of sensory messages, Sensory Communication ed W. A. Rosenblith. MIT Press, Cambridge MA 1961.
[4] R. Linsker, Neural Comp. **1**, 402 (1989).
[5] D. L. Ruderman, Network **5**, 147 (1994).
[6] D. J. Field, Neural Comp. **6**, 559 (1994).
[7] D. J. Field, J. Opt. Soc. Amer. A **4**, 2379 (1987).
[8] D. L. Ruderman, Network **5**, 517 (1994).
[9] D. L Ruderman and W. Bialek, Phys. Rev. Lett. **73**, 814 (1994).
[10] R. J. Baddeley and P. J. B. Hancock, Proc. Roy. Soc. London B **246**, 219 (1991).
[11] H.-O. Peitgen and D. Saupe, The Science of Fractal Images, Springer, Berlin 1988.
[12] R. Bracewell, The Fourier transform and its applications, McGraw-Hill, New York 1965.
[13] V. F. Nesteruk and N. N. Porfireva, Opt. Spect. **29**, 606 (1970).
[14] R. M. Shapley and P. Lennie, Ann. Rev. Neurosci. **8**, 547 (1985).
[15] C. Fyfe and R. J. Baddeley, Network **6**, 333 (1995).
[16] P. Haberäcker, Digitale Bildverarbeitung: Grundlagen und Anwendungen, Hanser, Wien 1983.
[17] J. C. Russ, The image processing handbook, CRC Press, Boca Raton, New York 1995.
[18] N. G. Deriugin, Telecom. **1**, 1 (1956).
[19] G. K. Burton and I. R. Moorhead, Appl. Opt. **26**, 157 (1987).
[20] C. J. StC. Webber, Network **2**, 169 (1991).
[21] D. J. Tolhurst, Y. Tadmor, and T. Chao, Opht. Physiol. Opt. **12**, 229 (1992).
[22] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, Numerical Recipes in C, University Press, Cambridge MA 1990.
[23] L. P. Kadanoff, Physics **2**, 263 (1966).
[24] J. Cardy, Scaling and Renormalisation in Statistical Physics, University Press, Cambridge 1996.
[25] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, 3rd edition 1991.
[26] C. W. Gardiner, Handbook of Stochastic Methods, Springer, Berlin 1983.