Higher Order Image Pyramids*

Joshua Gluckman

Polytechnic University,
Dept. of Computer and Information Science,
Brooklyn, New York

Abstract. The scale invariant property of an ensemble of natural images is examined which motivates a new early visual representation termed the higher order pyramid. The representation is a non-linear generalization of the Laplacian pyramid and is tuned to the type of scale invariance exhibited by natural imagery as opposed to other scale invariant images such as 1/f correlated noise and the step edge. The transformation of an image to a higher order pyramid is simple to compute and straightforward to invert. Because the representation is invertible it is shown that the higher order pyramid can be truncated and quantized with little loss of visual quality. Images coded in this representation have much less redundancy than the raw image pixels and decorrelating transformations such as the Laplacian pyramid. This is demonstrated by showing statistical independence between pairs of coefficients. Because the representation is tuned to the ensemble redundancies the coefficients of the higher order pyramid are more efficient at capturing the variation within the ensemble which leads too improved matching results. This is demonstrated on two recognition tasks, face recognition with illumination changes and object recognition which viewpoint changes.

1 Introduction

It has been argued for some time that the problem of early vision is to determine the most efficient method for representing images. Early formulations of this theory were given by Atteneave and Barlow who argued for redundancy reduction through the efficient coding hypothesis [2] [3]. The rational is rooted in information theory: if later visual processes are to perform probabilistic inference the input should be as statistically independent as possible The implication for early vision is that images should be coded to match the expected input, the ensemble of natural images (for examples see Fig. 1). Since, considerable advances have been made in understanding this ensemble by studying the statistics of natural images and arguably the most important feature to emerge is scale invariance (see [4] for a recent review). While multi-resolution representations such as the Laplacian, steerable, and wavelet pyramids are motivated by scale invariance they are not necessarily tuned to the type of scale invariance found in images.

In this paper we present a novel method for representing images which is motivated by the type of scale invariance found in the ensemble of natural images. This representation is a non-linear transformation based on the Laplacian

* This work was supported by a NSF ITR award no. IIS-0219078.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part II, LNCS 3952, pp. 308–320, 2006.
© Springer-Verlag Berlin Heidelberg 2006

Fig. 1. Examples of natural images from the Van Hatteren database [1]

pyramid of Burt and Adelson [5] and is referred to as the higher order image pyramid. Because the transformation is invertible it is shown that the higher order pyramid can be quantized and truncated with little loss of visual quality. Once an image is coded the bits of the higher order pyramid exhibit far less redundancy than the raw image pixels and linear transformations such as the Laplacian pyramid. This is shown by demonstrating that the mutual information between pairs of coefficients is near zero.

Tuning a representation to the ensemble redundancies implies the representation efficiently captures the variation within the ensemble. Capturing this variation leads to improvements in recognition tasks. This is demonstrated by showing improved matching using the higher order pyramid on an object recognition task with varying viewpoint and a face recognition task with varying illumination.

2 Scale Invariance

The first statistical evidence of scale invariance was found by Field who discovered a $1/f$ power law in the amplitude spectra of natural images indicating that spatial correlations are scale invariant [6]. Further evidence was found by Ruderman and by Zhu and Mumford who demonstrated that histograms of derivative filtered images are consistent across scale indicating scaling in the higher order statistics [7] [8]. Although the presence of scale invariance in images is clear, no method exists for fully taking advantage of this property. One reason for this is the lack of a simple model that suggests how to represent the type of scale invariance found in images. Next, we develop such a model.

The type of scale invariance seen in images is examined by considering the behavior of each particular image as the scale is reduced. Once particular images are considered the range of intensities is fixed, hence the moments are finite. Then scale invariance appears in the expected value of the moments of the distribution of intensity as scale is reduced which is

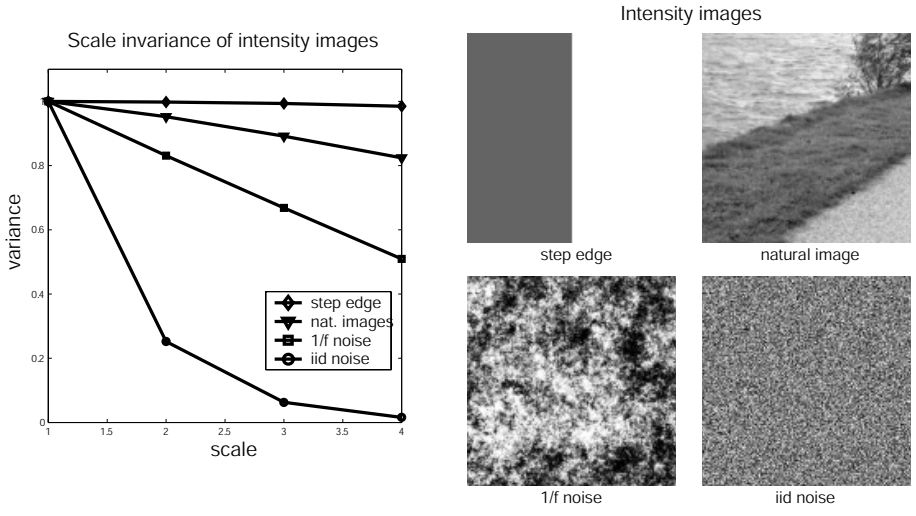

Fig. 2. The variance in intensity as a function of scale (the number of times the images are downsampled) for a pure step edge, a set of natural images, $1/f$ noise, and iid noise. The step edge, $1/f$ noise, and natural images all exhibit scale invariance indicated by a slow linear decay.

$$E\{|I(x/s)|^k\}, \tag{1}$$

where $I(x)$ is an image, s is the spatial scale, and the expectation E is taken over an ensemble of images. We assume the mean intensity is subtracted from each image so k is a central moment.

Fig. 2 shows the result of estimating (1) for $k = 2$ using 40 images selected from the van Hateren database [1]. A scale change of $I(x)$ is implemented by smoothing and downsampling. Each image is repeatedly blurred, downsampled and a variance ($k = 2$) is estimated. For each scale the variances of all 40 images are averaged. The plot shows the reduction in variance as a function of scale and compares the behavior of images to that of a pure step edge, $1/f$ noise, and iid noise. All plots are normalized to 1 at the first scale. The step edge, the images, and $1/f$ noise all exhibit scale invariance because the variance falls linearly. The same result is found for the higher moments implying

$$E\{|I(x/s)|^k\} = E\{|I(x/(2s))|^k\} + K. \tag{2}$$

The actual value of the constant K is unimportant. However, for images K lies somewhere between that of a step edge and $1/f$ noise,

$$K_{1/f} < K_{\text{images}} < K_{\text{step}}. \tag{3}$$

What differentiates images is the *rate* at which the moments decay with scale.

Multi-resolution representations such as the Laplacian pyramid are motivated by the fact that images exhibit scale invariance. These representations code

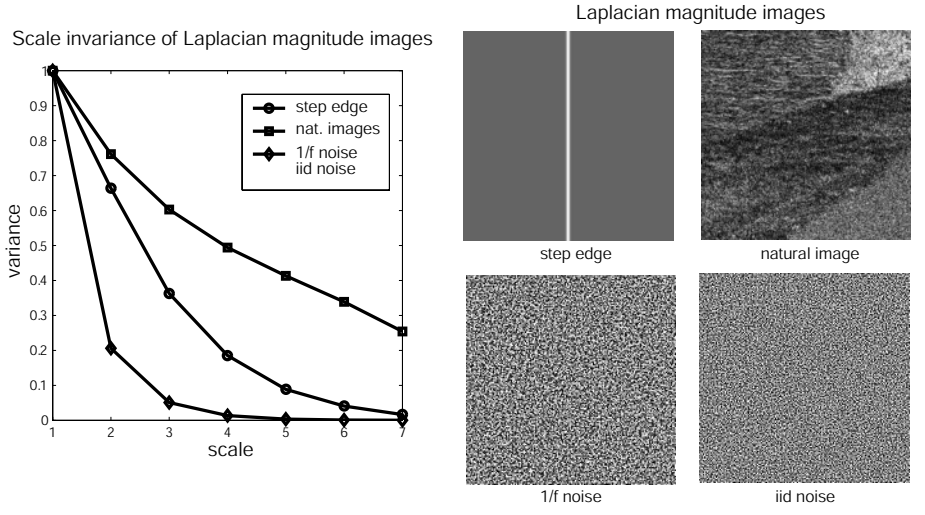

Fig. 3. The variance of the Laplacian magnitude as a function of scale for a pure step edge, a set of natural images, and $1/f$ noise. Only the natural images are scale invariant.

$$L(x/s) = I(x/s) - I(x/(2s)),$$

which is the prediction error between scales. However, such representations do not explicitly differentiate between step edges, $1/f$ noise, and natural images. To describe the type of scale invariance that images exhibit, L must be examined.

Although L, the levels of a Laplacian pyramid, are no longer scale invariant according to eq. (2) they still resemble images which suggests a hidden scale invariance. Indeed, if a new image (the Laplacian magnitude) is defined

$$I^{(2)} = \log |L| \tag{4}$$

which is independent of the sign then scale invariance reappears. The log function is used to reduce the sensitivity of the statistics to the extreme values. The fact that redundancy exists in the magnitude is well known and has been documented by Simoncelli et al [9] [10] [11]. Here it is shown that the form of the redundancy is scale invariance. Next, the same experiment is repeated except using $I^{(2)}$ for each of the 40 natural images, the step edge, the $1/f$ noise, and the iid noise.

Fig. 3 shows the result of the experiment. Now the behavior of images is very different. The Laplacian magnitude of a step edge is a line edge and the Laplacian magnitude of $1/f$ noise is similar to iid noise neither of which have scale invariance. For natural images the Laplacian magnitude appears scale invariant. Furthermore, if we start from a lower scale and include the data from Fig. 2 the same pattern emerges as in the first experiment (see Fig. 4). Thus $I^{(2)}$ for natural images behaves like I meaning both (2) and (3) hold. Therefore, we define the following sequence

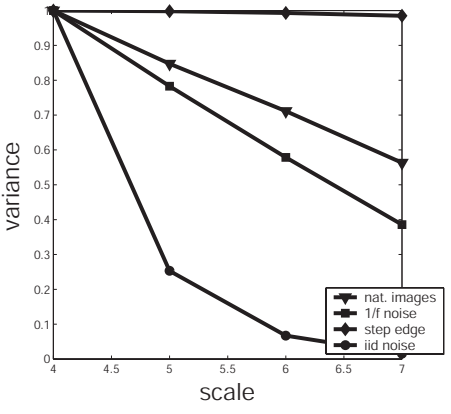

Fig. 4. The plot for the natural images from Fig. 3 is combined with the plot for the other images from Fig. 2. The rate of decay for the Laplacian magnitude of the natural images is similar to that of the intensity images.

$$I^{(1)}(x/s) = I(x/s)$$
$$I^{(n)}(x/s) = \log|I^{(n-1)}(x/s) - I^{(n-1)}(x/(2s))|,$$

and assume that if (2) and (3) are true for $I^{(n)}$ they are true for $I^{(n+1)}$. Just as the Laplacian pyramid is a representation that assumes I is scale invariant, the higher order pyramid is a representation that assumes $I^{(1)}, I^{(2)}, I^{(3)} \ldots$ are scale invariant. The implication is that the higher order pyramid is tuned to the type of scale invariance of natural images and not that of other scale invariant images such as $1/f$ noise and the step edge. In the next section the construction of the higher order image pyramid is described.

3 Higher Order Pyramids

We begin with a description of the Laplacian pyramid on which higher order pyramids are based. The Laplacian pyramid is motivated by the scale invariance of $I^{(1)}$ and only codes the information at each scale that is not predicted by the lower scale. The transformation $\mathcal{L}$ constructs a Laplacian pyramid for an image I and is defined by the set,

$$\mathcal{L}(I) = \{L_1, L_2, \ldots, L_s, G\}, \tag{5}$$

where $L_i = I(x/2^{i-1}) - I(x/2^i)$ are levels of the pyramid. L_1 is the finest scale, L_s is the coarsest scale and G is the lowpass residual. Using $\mathcal{L}_i$ to refer to the i^{th} member of the set $\mathcal{L}$, it is straightforward to invert $\mathcal{L}$ by

$$I = \mathcal{L}^{-1} = \sum_{i=1}^{s+1} \mathcal{L}_i \tag{6}$$

where upsampling and convolution are implicit (for details see [5]).

A higher order pyramid is motivated by the scale invariance of $I^{(1)}, I^{(2)}, I^{(3)} \ldots$ where the prediction error is $L_i/|L_i|$ which are the sign bits. However, the sign operation introduces aliasing and is not rotationally invariant which is why scale invariance does not appear until the lower scales (see the abscissa axis of Fig. 3). Rather than coding the sign bits we code

$$\mathcal{C}(I) = \{L_1(x)/M_1(x/2), \ldots, L_s(x)/M_s(x/2), G\} \tag{7}$$

where $M_i = |L_i|$. The operation $L_i(x)/M_i(x/2)$ is accomplished by dividing a level of the Laplacian pyramid by its magnitude after blurring, downsampling, and upsampling. Thus, $\mathcal{C}$ constructs a pyramid on the input image I and divides each level by a reduced resolution version of the magnitude at that level. The result is a contrast normalized Laplacian pyramid. The higher order pyramid differs from other contrast normalized image representations by explicitly representing the magnitudes as well via additional pyramids [7] [12]. The higher order pyramid is recursively defined by the set

$$\mathcal{H}^{(1)}(I) = \{\mathcal{C}(I), \{\log M_1(x/2), \ldots, \log M_s(x/2)\}\} \tag{8}$$

$$\mathcal{H}^{(n)}(I) = \{\mathcal{C}(I), \{\mathcal{H}^{(n-1)}(\log M_1(x/2)), \ldots, \mathcal{H}^{(n-1)}(\log M_s(x/2))\}\}. \tag{9}$$

Because this is a doubly recursive representation there are two notions of depth, s the number of times the scale is reduced when building the pyramids and n the number of times pyramids are built on the images $I^{(1)}, I^{(2)}, I^{(3)} \ldots$. We refer to n as the order of the pyramid. A first order pyramid, $\mathcal{H}^{(1)}$, is a contrast normalized Laplacian pyramid along with $I^{(2)}$ the residual magnitude images. A second order pyramid, $\mathcal{H}^{(2)}$, constructs a first order pyramid on each of the residual magnitudes and keeps the $I^{(3)}$ residuals . For higher orders this process is repeated. Fig. 5 demonstrates a higher order pyramid built up to order 3. Each time the order is increased the residual magnitudes decrease in scale by a factor of 2. It is straightforward to show that the total number of coefficients is twice the number of pixels. Hence, the representation is twice overcomplete.

$\mathcal{H}$ is a set with two elements each of which is a set. The notation $\mathcal{H}_{1,i}$ refers to the i^{th} member of the first set and $\mathcal{H}_{2,i}$ refers to the i^{th} member of the second set. Using this notation it is straightforward to invert the pyramid by

$$\mathcal{H}^{(n)^{-1}} = \sum_{i=1}^{s} \mathcal{H}_{1,i} \exp\left(\mathcal{H}_{2,i}^{-1}\right) + \mathcal{H}_{1,s+1}, \tag{10}$$

$$\mathcal{H}^{(1)^{-1}} = \sum_{i=1}^{s} \mathcal{H}_{1,i} \exp\left(\mathcal{H}_{2,i}\right) + \mathcal{H}_{1,s+1}, \tag{11}$$

which recursively collapses all of the pyramids and multiplies back in the magnitudes. Like the Laplacian pyramid, there is no loss of information when inverting the higher order pyramid.

For any order, the pyramid can be inverted without the residual magnitudes which we refer to as a truncated pyramid. This is done by redefining eq. 11 to

$$\mathcal{H}^{(1)^{-1}} = \sum_{i=1}^{s} \mathcal{H}_{1,i} + \mathcal{H}_{1,s+1}. \tag{12}$$

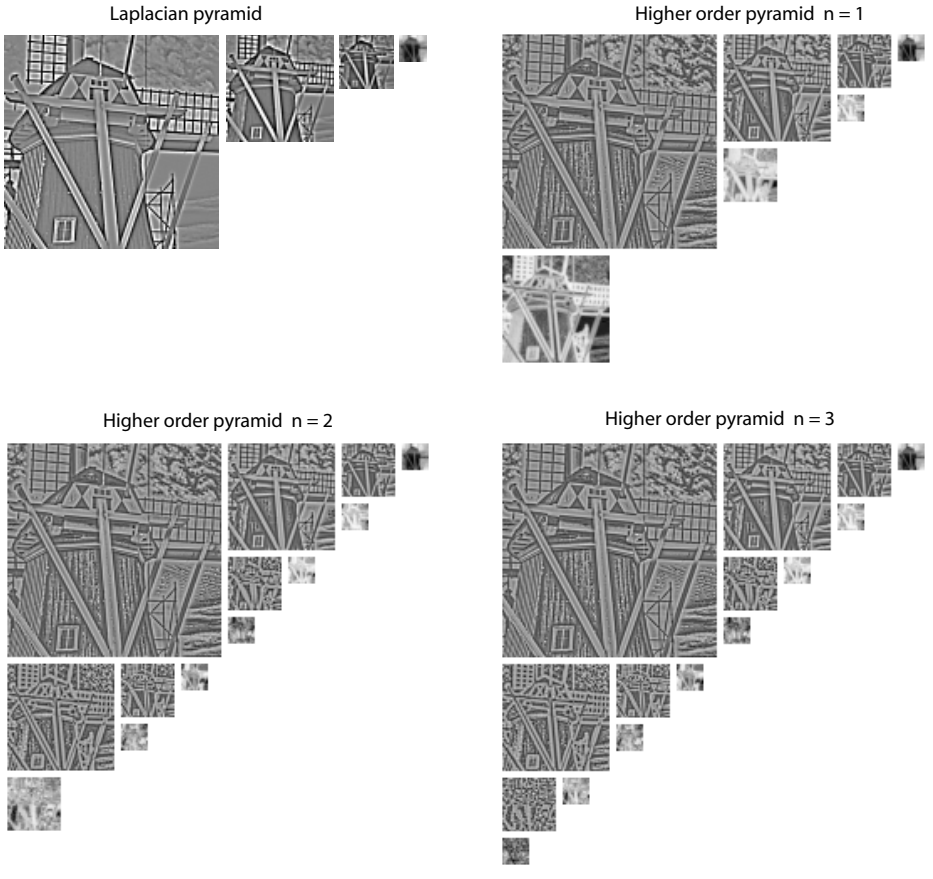

Fig. 5. A higher order pyramid is built from a Laplacian pyramid which in this example is three levels deep along with a lowpass residual. A first order pyramid, $n = 1$, is a contrast normalized pyramid with a residual magnitude shown below each level. A second order pyramid, $n = 2$, builds first order pyramids on the magnitudes that are larger than the lowpass residual. A third order pyramid repeats the process.

Although the result is a loss of information, the visual quality of the reconstructed images rapidly improves as n increases. The transformation is also stable with respect to quantization in that visual quality also rapidly improves with the size of the quantization bins. Fig. 6 demonstrates examples of reconstructed images from truncated pyramids with coefficients quantized to integer values resulting in an average entropy of 2.2 bits per coefficient.

4 Experiments

Three experiments are performed. The first is application independent and measures the degree to which pairs of coefficients are independent. The reason for

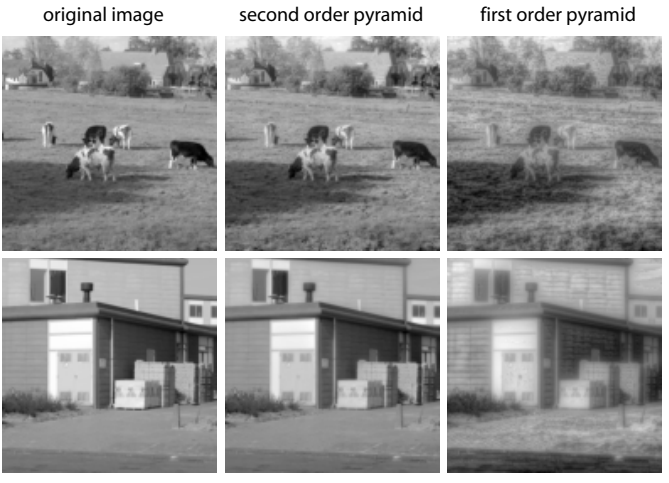

Fig. 6. The images are examples of inverting the higher order pyramid after truncating (removing the residual magnitudes) and quantizing the coefficients. The residual magnitude becomes significantly less important as the order increases. Furthermore, the transformation is stable with respect to quantization of the coefficients.

this test is that the efficiency of a representation is reflected by the lack of redundancy or independence in the coefficients of the representation. The second and third experiments are recognition tests. A good representation should lead to improved matching. The rational is that a less redundant representation is a better space in which to perform matching in the presence of uncertainty. In our experiments two types of uncertainty are considered, face recognition with illumination changes and object recognition with viewpoint changes. In all of our experiments the Laplacian and higher order pyramids are implemented using a 7-tap binomial filter and image borders are handled via reflection.

4.1 Redundancy

To test for redundancy the joint distribution for a pair of coefficients is measured from a set of 40 natural images. For each image, statistics are gathered for all pairs of coefficients displaced by 5 pixel positions in the diagonal direction. This is done for the raw image pixels, for a level in the Laplacian pyramid, and for a level in the higher order pyramid. A distance of 5 is chosen to ensure the coefficients are not within the support of the convolution filter.

The distributions are displayed as conditional distributions in Fig. 7 where intensity represents the likelihood of y conditioned on the value of x. The Laplacian pyramid has the familiar bowtie shape documented by Simoncelli [9]. This is indicative of the fact that the magnitudes are correlated. Meanwhile the coefficients of the higher order pyramid are clearly independent because the likelihood of x is the same for all values of y. This is also reflected by the mutual information shown below each plot. Similar results are found for different spatial offsets

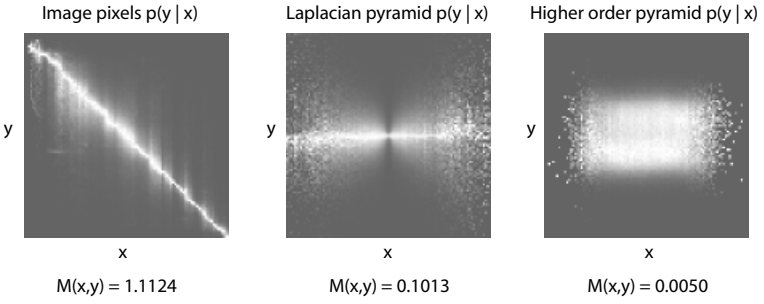

Fig. 7. Conditional distributions for pairs of coefficients at the same spatial offset. In the pixel domain the coefficients are correlated. In the Laplacian pyramid they are decorrelated but still dependent because the variance of y scales with x. In the higher order pyramid they are independent. Below each graph is M the measured mutual information between the coefficients for each representation.

both within and between different levels of the higher order pyramid. Independence has also been demonstrated after applying contrast normalization [12]. However because the higher order pyramid is invertible we can be sure that the independence does not arise from destroying information. It is important to note that correlations exist for neighboring pairs of coefficients within and between levels however these are not due to redundancy in natural images but arise from the convolution. Hence, they exist even for noise images.

4.2 Face Recognition with Illumination Changes

Using the Yale illumination face database and experimental setup reported in [13], we test the ability of higher order pyramids to perform face recognition in the presence of illumination changes. The database contains 10 faces each under 64 different illuminants. They are grouped into 5 subsets according to the severity of the direction of illumination. Example are shown in Fig. 8(a). Each face is cropped and downsampled to a resolution of 96×80 pixels. We compared the matching performance of the representations shown in Fig. 8(b). The pyramids are constructed 3 levels deep and the lowpass residual and the residual magnitude images are not used. Because the levels of the pyramids are subsampled each level is weighted by $4^{(s-1)}$ so that all levels contribute equally. Then each representation is treated as a vector of coefficients and matching is done with normalized cross correlation. One image of each persons face is used as the training image and the remaining 630 images are used as test images. For all representations log intensities are used.

The results are reported in Fig. 9 and demonstrate a clear advantage for the higher order pyramid as compared to the Laplacian pyramid and the raw image pixels. For further comparison the results from Chen et al. are also included [13]. It should be pointed out that all of the other methods with the exception ofGradient Angle use all of the images in Subset 1 and 2 for training.

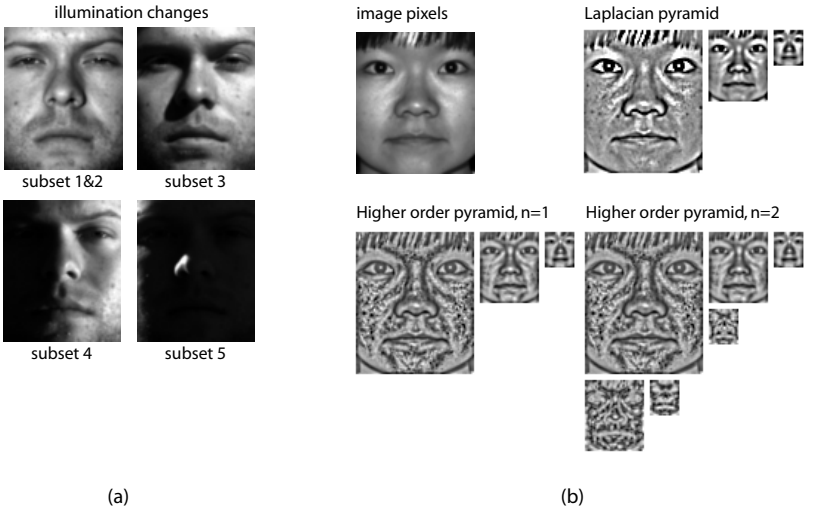

Fig. 8. (a) Examples from the Yale illumination face database. Recognition is done using normalized cross correlation with the representations in (b).

Furthermore, Subset 5, the one with the most sever illumination directions, is not used in any of the other experiments.

The Gradient Angle method learns, from a database of images, an illumination insensitive weighting of the magnitude of the gradient. Because the magnitudeis

Face recognition error rate (%) vs. Illumination				
Method	Subset 1&2	Subset 3	Subset 4	Subset 5
Eigenfaces	0	16.7	69.3	–
Linear subspace	0	1.7	12.9	–
Cones-attached	0	0.8	9.3	–
Gradient Angle	0	0.0	1.4	–
Cones-cast	0	0.0	0.0	–
Image pixels	0.6	15.0	47.9	46.3
Laplacian pyramid	0	1.7	10.0	24.2
Higher order pyramid, n=1	0	0	8.6	11.6
Higher order pyramid, n=2	0	0	0	0

Fig. 9. Face recognition error rates. The methods in **bold** are experiments we performed using normalized cross correlation and the other results are from [13]. The images are grouped into subsets according to the severity of the illuminant direction. Subset 5 is the most severe and is not included in the results reported by the other methods. Out of 630 test images the second order pyramid had no errors.

sensitive to illumination the Gradient Angle method is a form of contrast normalization thus bears a resemblance to the first order pyramid. However, the perfect performance of the second order pyramid indicates that representing the magnitude (in a less redundant space) rather than normalizing the magnitude is important.

4.3 Object Recognition with Varying Viewpoint

Using the recently available Amsterdam object database [14], we test the ability of higher order pyramids to perform recognition in the presence of viewpoint changes. This is a large database containing 1000 small objects. Each object is imaged on a rotational stage from 72 positions at 5° increments. We used the grey scale dataset at quarter resolution which was further downsampled to a resolution of 72×96 pixels. In addition log intensities are used. Examples of some of the objects are shown in Fig. 10.

For each object, N evenly spaced training views are used. Each test image is matched between all training views. Again we tested the performance of matching using the image pixels, the Laplacian pyramid, and the higher order pyramid where each representation is treated as a vector of weighted coefficients. The depth of the pyramids is the same as the face recognition experiment except the lowpass residual images are included this time. In the face recognition experiment the lowpass residual is too dependent on illumination. Furthermore, sum of squared distances (SSD) is used rather than normalized correlation due to the use of the lowpass images. One important difference between the object images and the face images is the background. In order to avoid the background biasing the results the mean vector is computed for the entire set of training views. This mean vector is then subtracted from each training vector and from each test vector prior to computing the SSD.

Due to the size of the database 10 trials are performed on subsets of 100 randomly selected objects. For each trial N training images are used for each of the 100 chosen objects and 5000 test images are randomly chosen from among

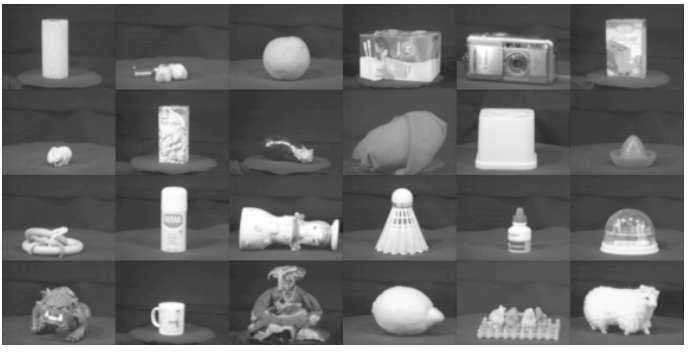

Fig. 10. Example of objects from the Amsterdam library of object images (ALOI)

the 7200 images of the chosen objects. Below are the recognition error rates averaged over the 10 trials for $N = 6, 8, 12$. In all cases the second order pyramid outperforms the other representations.

Object recognition error rate (%) vs. number of training views			
Representation	6 views	8 views	12 views
Image pixels	6.90	3.72	1.56
Laplacian pyramid	5.83	2.81	1.01
Higher order pyramid, n=1	4.10	2.08	0.63
Higher order pyramid, n=2	3.74	1.50	0.31

5 Discussion

The scale invariant properties of natural images leads to the higher order pyramid as an early visual representation which is simple to compute and straightforward to invert. Images coded in the space of higher order pyramids exhibit far less redundancy than the raw image pixels and decorrelating transformations such as the Laplacian pyramid. This is demonstrated by showing the independence between pairs of coefficients and improved matching in two recognition experiments. The recognition experiments are only intended to demonstrate the potential for redundancy reduction as a goal for the representation of visual patterns and the higher order pyramid as a step in this direction. The real benefit is the further processing of the higher order pyramid and the incorporation of learning techniques.

It is interesting to note that the representation is not oriented. Because a Laplacian pyramid is used to construct the higher order pyramid the basis functions are all circularly symmetric Gaussian functions albeit combined in a non-linear way. This might seem strange given the wide spread use of oriented filters such as Gabor functions and steerable derivatives throughout computer vision. In addition, oriented receptive fields are commonplace in computational models of human vision. Furthermore, it is well known that the optimal linear basis in which to represent natural images is made up of oriented basis functions [6] [15] [16]. If we assume that scale invariance as defined in this paper is a natural description of images then we must ask why use oriented basis functions. We briefly speculate on some possible answers. (1) The model proposed here does not fully capture the scale invariance of images. (2) Other statistical properties of images that are not implied by scale invariance are more important. (3) Oriented basis functions are simply the best way to represent scale invariance when restricted to a linear framework. However, in a non-linear framework they may no longer be needed.

References

1. van Hateren, J.: Independent component filters of natural images compared with simple cells. Proc. of the R. Stat. Soc. of London B **265** (1998) 359–366
2. Attneave, F.: Some informational aspects of visual perception. Psych. Rev. **61** (1954) 183–193
3. Barlow, H.: Possible principles underlying the transformaiton of senosry messages. In: Sensory Commuincation. MIT Press (1961)
4. Zhu, S.: Statistical modeling and coneptualization of visual patterns. IEEE Trans. on Pattern Analaysis and Machine Intelligence **25** (2003) 691–712
5. Burt, P., Adelson, E.: The laplacian pyramid as a compact image code. IEEE Trans. on Communications. **31** (1983) 532–540
6. Field, D.: Relations between the statistics of natural images and the response properties of cortical cells. J. of the Opt. Soc. of Am. A **4** (1987) 2379–
7. Ruderman, D.: The statistics of natural images. Network **5** (1993) 477–500
8. Zhu, S., Mumford, D.: Prior learning and gibbs reactionn-diffusion. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 1236–
9. Simoncelli, E., Adelson, E.: Noise removal via bayesian wavelet coding. In: Int'l Conf. on Image Processing. (1996) 379–382
10. Simoncelli, E., Freeman, W.: The steerable pyramid: A flexible architecture for multi-scale derivative computation. In: Int'l Conf. on Image Processing. (1995)
11. Buccigrossi, R., Simoncelli, E.: Image compression via joint statistical characterization in the wavelet domain. IEEE Trans. on Image Processing **8** (1999) 1688–
12. Schwartz, O., Simoncelli, E.: Natural signal statistics and sensory gain control. Nature Neuroscience **4** (2001) 819–825
13. Chen, H., Belhumeur, P., Jacobs, D.: In search of illumination invariants. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. (2000)
14. Geusebroek, J., Burghouts, G., Smeulders, A.: The amsterdam library of object images. Int'l J. Computer Vision **61** (2005) 103–112
15. Olshausen, B., Field, D.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381** (1996) 607–609
16. Bell, A., Sejnowski, T.: The independent components of natural scenes are edge filters. Vision Research **37** (1997) 3327–38