Zipf's Law in Image Coding Schemes

Michael Crosier and Lewis D Griffin
Computer Science
University College London
London, WC1E 6BT, UK
m.crosier@cs.ucl.ac.uk

Abstract

Zipf's Law describes a power law between the frequency and rank of an event which is observed, most famously, in word distributions in human languages [1]; where it has been argued [2] that it represents an optimally efficient communication code. From an analogy between image coding schemes and natural language, we hypothesize that conforming to Zipf's Law is a necessary property of a good image description code, i.e. one that provides optimal local descriptors for vision tasks such as object categorization and recognition.

Following the analogy between images and documents, we have developed an 'alphabet' of Basic Image Features (BIFs) [3]. This alphabet allows us to label each pixel, at each scale, with one of a small number of letters. 'Words' in this approach are then fixed square patches of BIF 'letters'.

We test whether the frequency statistics of BIF words obeys Zipf's law, and find that as we vary certain parameters the Zipfness varies systematically. Remarkably, we find that for certain parameter settings we get extremely Zipf behaviour. Moreover the encoding at these settings looks optimal to eyeball measure.

1 Introduction

1.1 Local descriptors

There has recently been a great deal of interest in using local descriptors (e.g. [4], [5]; see [6] for an overview and evaluation) to perform vision tasks such as object categorization [7] and recognition [4].

We are interested in the properties of such local cues which make them suited for these types of tasks. We argue that the graded rarity of descriptors in natural images is an important factor in their effectiveness: If a cue appears in all images, regardless of content, it will be a poor discriminator of objects; as it will if it only ever appears in a particular view of some object instance. Here we explore the frequency statistics of cues from different image coding schemes across natural images.

1.2 Zipf's Law

If counts of word occurrences in a corpus of natural language are sorted into descending order, a power law is observed between the rank r and frequency p of a word so that $p \propto r^{-\alpha}$ with α close to unity. Named after George Zipf, who first noted this relationship for English text [1], Zipf's Law has been empirically observed for a wide range of both human and non-human languages, for example Russian [8] and the whistles of bottlenose dolphins [9].

Motivated by results [2] showing that, under certain assumptions, an information source must obey Zipf's Law in order to ensure optimal efficiency of communication; and by an analogy between image coding schemes and natural language; we hypothesize that conforming to Zipf's Law is a necessary property of a good image description code.

1.3 Coding schemes

Following the analogy between images and documents, we have developed an 'alphabet' of Basic Image Features (BIFs) [3]. This alphabet allows us to label each pixel, at each scale, with one of a small number of letters. At present we have two-, three- and six-letter alphabets based on different sets of Gaussian derivative filters (see figure 1). 'Words' in this approach are fixed square patches of BIF 'letters'.

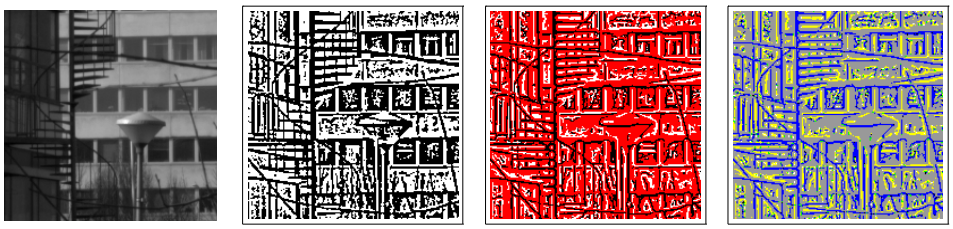


Figure 1: A section of an image from the van Hateren collection [10], encoded respectively with 2, 3 and 6 label BIF schemes.

The BIF coding schemes correspond to a partitioning of filter response space [3]: In the 2-letter scheme, we take the sign of a Laplacean of Gaussian filter (using [11]). For the three letter alphabet, we compare the absolute value of the Laplacean with the gradient magnitude, taking the sign of the Laplacean if $\frac{\sigma^2}{4}\left(I_{xx}+I_{yy}\right)^2 > \left(I_x^2+I_y^2\right)$ where I_x etc. are image derivatives computed at scale σ using Gaussian derivative filters, and a third label otherwise. In both schemes, points are then subsampled at a distance of Δ pixels.

1.4 Testing the 'Zipf Hypothesis'

In this paper we explore the degree to which the frequency statistics of 'words' in natural and synthetic images, encoded with our 2-label ('2-BIFs') and 3-label ('3-BIFs') systems, adhere to Zipf's Law.

We present results which, for certain parameter settings, exhibit strikingly Zipf-like behaviour. Further, these settings seem to correspond well to those which we would choose by eyeball measure. The work presented here will form a basis for testing our 'Zipf Hypothesis' through employing a range of the codes explored here in vision tasks such as object categorization.

2 Methods & Results

2.1 Counting

We counted occurrences of overlapping square patterns ('words'), of sizes 1x1 up to 10x10 pixels, in images which had been encoded with one of our BIF schemes. These counts were then pooled together (so a typical pixel was counted in one 1x1 patch; four 2x2 patches; nine 3x3 patches etc.) and ranked in order of decreasing frequency.

For each size of pattern, approximately 14.6 million patches were drawn evenly from a set of 100 encoded images. For a given area of image, it is possible to extract one more $(d-1)\,\mathrm{x}\,(d-1)$ patch than $d\,\mathrm{x}\,d$ patches in each direction and so we normalised the image area to ensure that the same number of patches were counted of each size.

In order to economise on memory expenditure by reducing the number of patterns recorded, our counting algorithm takes a cutoff value for the least number of occurrences of some pattern across an ensemble of images which we consider to be of interest. This has the effect of cropping the number of ranked patterns in our results, but causes no bias in the distribution of those ranks which are recorded. Indeed, in the absence of such an explicit cutoff its value would default to 1: no pattern will be counted which does not occur at least once in some image. We then proceed iteratively as follows: For each patch size $d\,\mathrm{x}\,d$, we scan through each encoded image, selecting a patch at each position. We extract its four $(d-1)\,\mathrm{x}\,(d-1)$ subpatches and check to see that these have all been recorded sufficiently often (for otherwise it is impossible that the parent patch will occur frequently enough); if so, we increment the count for that patch. Most patterns will occur less often than their smaller subpatches and so, after scanning through all $d\,\mathrm{x}\,d$ patches in all images, we scan through our table of counts and remove references to any pattern which does not occur at least *cutoff* times; before moving on to larger patches.

We consider only patches up to 10x10 pixels for computational reasons. There is a possibility that this could introduce a small bias into our results. However, we argue that since, out of the approximately 300000 patterns in each experiment which occur more frequently than our cutoff, only around 1% of these are of size 10x10; that the number of 11x11 or larger patterns which we are excluding from our results will be vanishingly small.

Four sets of images were considered, one natural and three synthetic:

- Natural images from the van Hateren database [10], containing varied views of countryside, forests, lakes and buildings.
- Phase randomised noise images, based on the power spectra of these same natural images.
- Images generated using a dead leaves model [12]. A large number of opaque discs with randomly sampled positions, radii and luminances are 'stacked' so

that nearer discs occlude those further away. These images capture the occlusive nature as well as approximate scale invariance of natural images.

- 'Coin-flip' random images. Each pixel of these was selected independently from either 2 labels (where the images were to be compared with 2-BIF coded images) or 3 labels (for comparison with 3-BIF coded images). The outcome probabilities for each label were determined by the relative frequencies of that label in the sets of encoded natural images. Thus these images are comparable not with the grey-scale images upon which we employ our coding schemes; but with the 2- or 3-label results of these encodings.

2.2 Zipfness measure

On a log-log plot of frequency vs. rank, Zipf's Law is characterised by a straight line with slope -1. From the one-dimensional space of all such lines we select the best-fitting, and use as our Zipfness measure the weighted sum of squared errors between this and the data, normalised by the log of the number of ranks considered, i.e.

$$\frac{\min_{\beta} \sum_{r=1}^{R} \frac{1}{r} \left(\log f_r - (\beta - \log r)\right)^2}{\log R}, \tag{1}$$

where r is rank, f_r frequency and R the total number of distinct patterns whose occurrences have been counted.

The weighting ensures that equally sized parts of the log-log plot contribute equally to the measure. Normalisation is necessary to compensate for the fact that, for differing parameters, different numbers of patterns appear sufficiently often to be recorded by our counting algorithm.

2.3 Results

We calculated the frequency distributions of patches for each of our four image types coded using either a 2- or 3-label BIF scheme, over a range of parameters.

Because of the well-known scale invariance of natural images (and its replication in our synthesized images) one would expect that for a given ratio between filter scale σ and spatial sampling Δ, varying σ or Δ should make no significant difference to the distribution of frequencies: this is indeed what we find. We therefore define a new parameter $k = \sigma / \Delta$ to be our single degree of freedom in each coding scheme, and study how varying k affects the frequency statistics of BIF words for each BIF alphabet and image type.

Figures 2 and 3 show how frequency distributions vary over k: for small k (i.e. very sparse subsampling), pixels in the encoded images will be essentially independent (see fig. 5, bottom-right) and so the distribution of patches of a given size will be determined almost solely by the relative frequencies of each label. For our 2-label

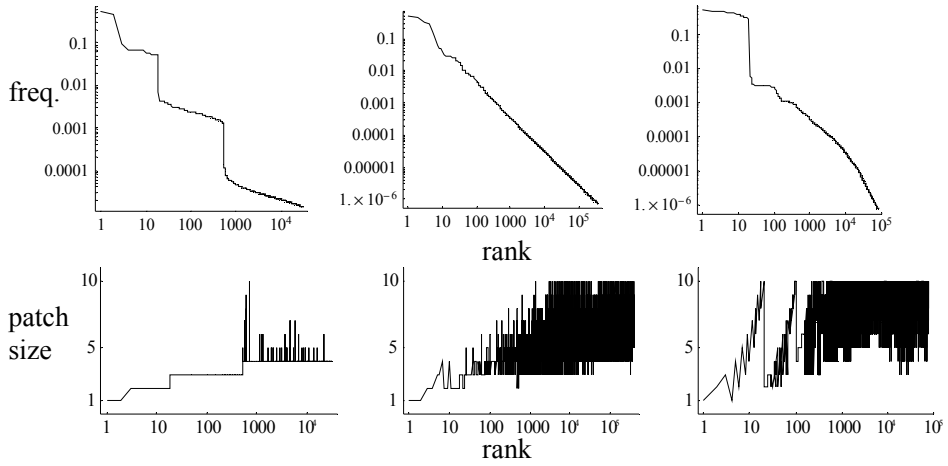


Figure 2. Top: Log-log plots of frequency of occurrence against rank for natural images encoded with 2-BIFs. From left to right, the ratio k between filter scale and spatial sampling takes the values 0.2, 1.5 and 12. Bottom: The size of patches at each rank in the same results. Similar frequency distributions are observed for phase randomised noise and dead leaves model images, although patch size distributions vary slightly: in particular, for intermediate values of k the results for dead leaves images tend to show peaks in patch size for low ranking patterns, caused by large areas of zero contrast in some of the original images. This is explored further in figure 8.

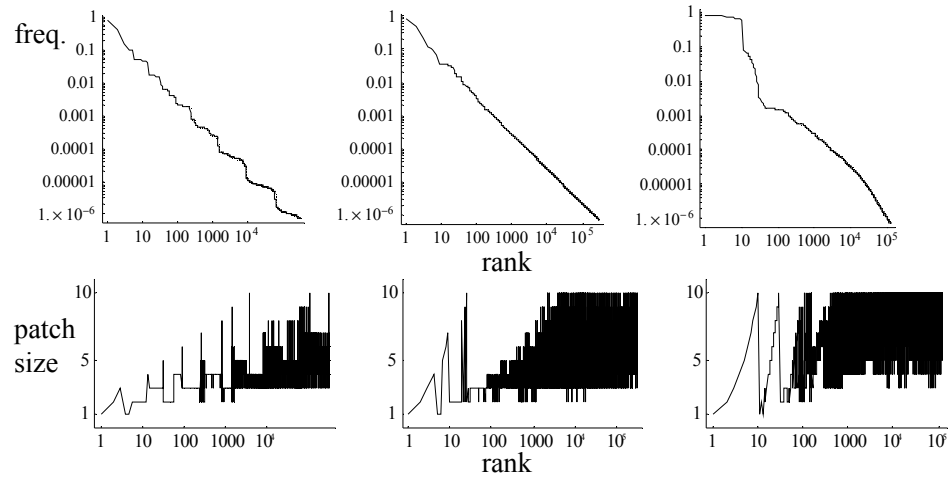


Figure 3:As figure 2, but for natural images encoded with 3-BIFs. From left to right, k takes the values 0.2, 1.0 and 12. The main difference from figure 2 is in the results for small k (i.e. undersampled coding) where the rank-frequency plot is less step-like. This is a consequence of the relative frequencies of the labels in our 3 label system being less symmetric than those for 2-BIFs, resulting (as shown in figure 5) in large areas of the same (red) label in the encoded images.

system, each label occurs with approximately the same frequency and so patches of each size are close to being uniformly distributed, resulting in the stepped plot seen in figure 2. In figure 3 (3-BIFs), this effect is less pronounced due to the uneven label frequencies in 3-BIF encoded images; but is still present.

For large k, on the other hand, encoded images will be characterised by large connected areas of the same label (see fig. 5, bottom-left). Most patches observed, of all sizes, will therefore be single-label patterns, and any moderately complex pattern will be very rare (see fig. 6). This accounts for the high initial frequency and consequent steep dropoff seen in figures 2 and 3.

Whilst encodings for small k are undersampled; and for large k oversampled; those for a reasonably broad mid-range of k-values appear 'just right': as in the top-right of figure 5. Remarkably, this seemingly optimal mid-range corresponds quite well to those distributions which closely follow Zipf's Law, as in the middle parts of figures 2 and 3.

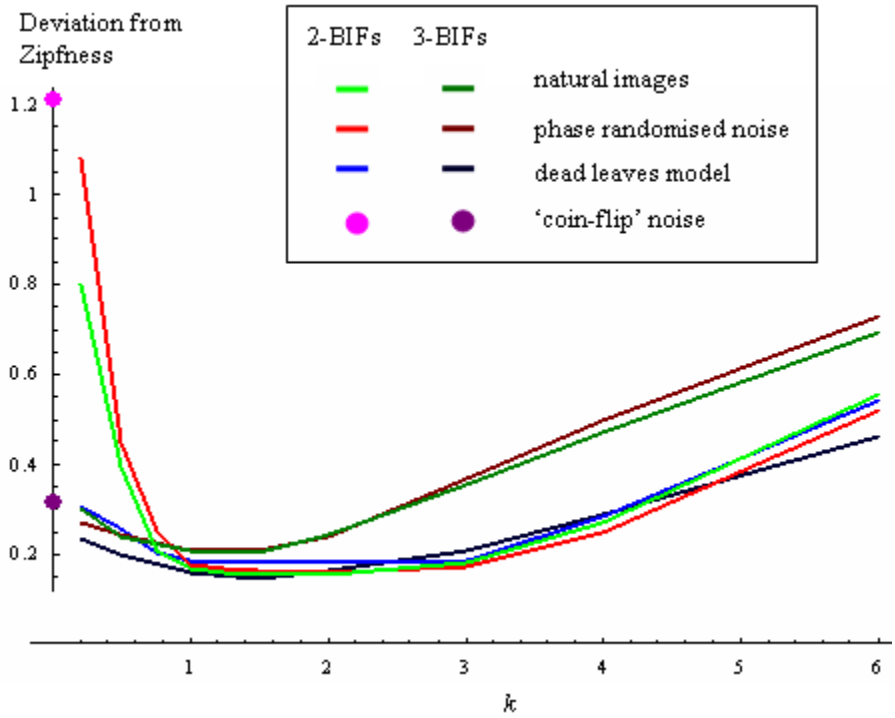


Figure 4: Variation of 'Zipf-dissimilarity' score with changing k for all image types and coding schemes. The 'coin-flip' images were not encoded in the same way as the other image types, and so do not vary over a parameter k. They are shown here as having $k = 0$ since they represent a logical limiting value for the other encoding schemes as $k \to 0$: As k decreases, encoded pixels are sampled increasingly sparsely and so become more independent of their neighbours. 'Coin-flip' noise models an encoded image in which the pixels are all independent.

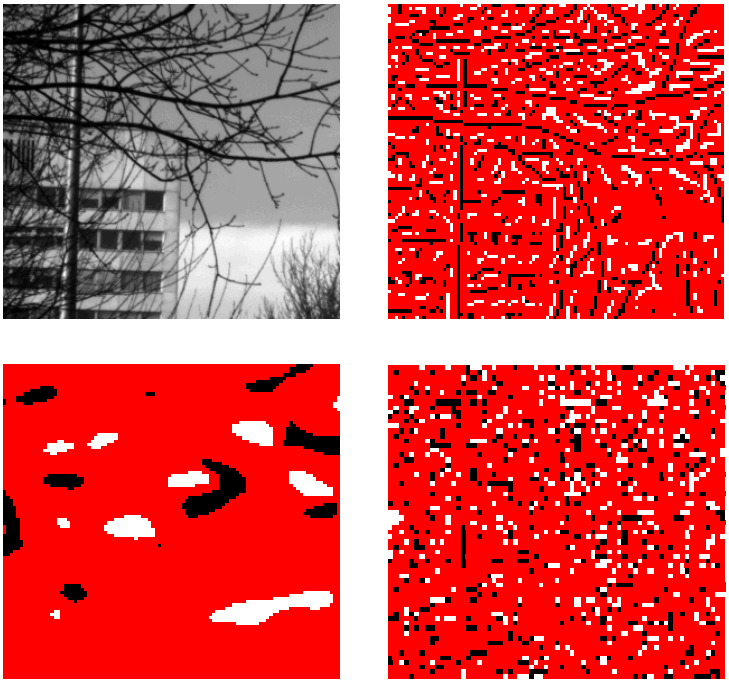


Figure 5: A section of a natural image from the van Hateren database, and 3-BIF encodings of the same image part with different values of k (anticlockwise from bottom-left: k =6.0, 0.2, 1.0. Recall that $k = \sigma/\Delta$, the ratio between filter scale and spatial sampling: In these examples, Δ was kept fixed and σ varied to produce the stated values of k while allowing the section of image to be encoded into the same number of pixels). The top-right encoding ($k = 1.0$) clearly displays more of the original image's structure. Comparison with the results in figure 4 shows that codings with the lowest Zipf-dissimilarity scores also appear optimal to eyeball measure.

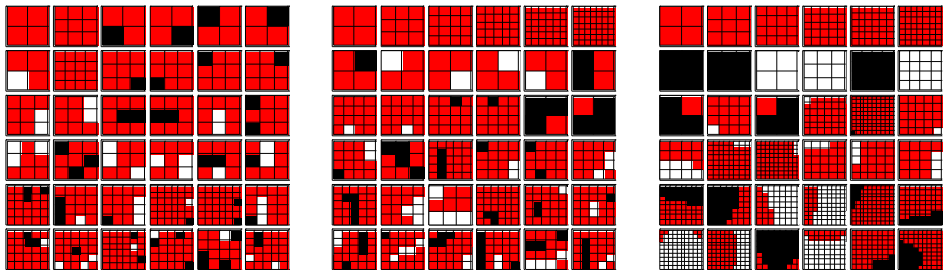
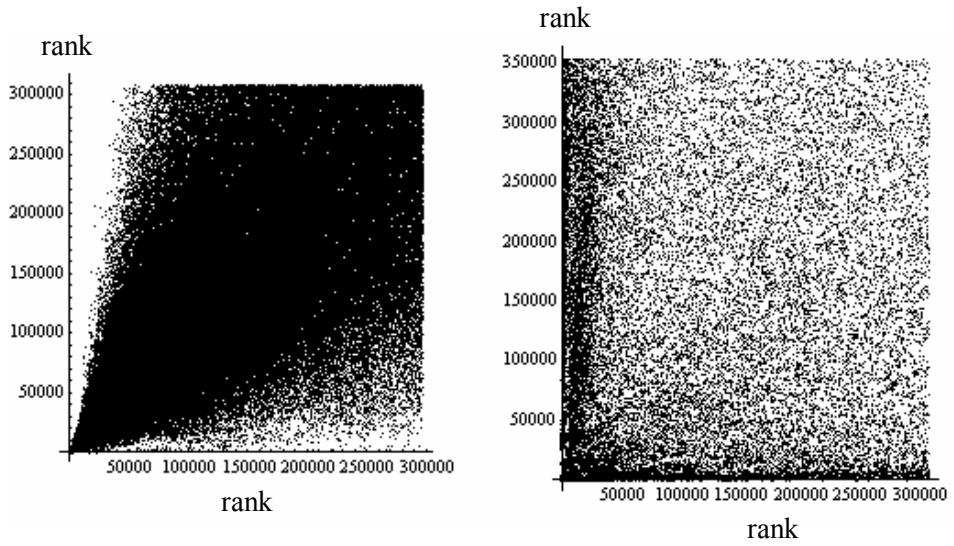


Figure 6: Examples of 3-BIF 'words' occurring in natural images for different values of k (left to right: k =0.2, 1.0, 12). The top row of each set shows the 6 most popular patches; the second row the 11th to 16th most frequent; the third row the 101th to 106th most frequent; and so on to the 6 patches ranked around 100000 in frequency of occurrence.



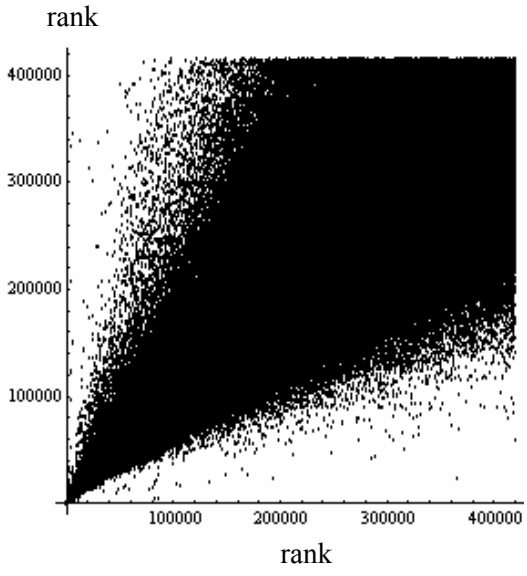


Figure 7: Stability of pattern rank. Top left: The rank of patterns occurring in both natural and phase randomised noise images encoded with 2-BIFs with $k = 1.0$ (correlation coefficient ρ =0.68). Top right: The rank of patterns occurring in natural images encoded with 2-BIFs, first with $k = 1.0$ and then $k = 4.0$ (correlation coefficient ρ =0.06). Right: The rank of patterns occurring in two different sets of natural images, both encoded with 2-BIFs with $k = 2.0$ (correlation coefficient ρ =0.84). Far greater rank-stability is seen between different types of images encoded in the same way than between the same images encoded with different ratios (k) between filter scale and spatial sampling.

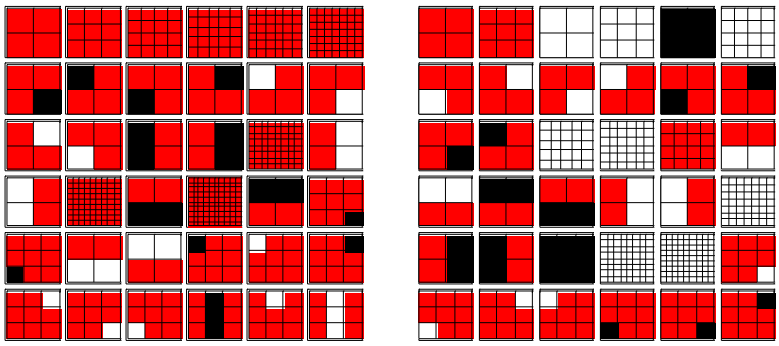


Figure 8: As fig. 6, but showing on the left patterns occurring in natural images encoded with 3-BIFs and $k = 1.0$; and on the right patterns occurring in dead leaves model images encoded with 3-BIFs and $k = 1.0$. The results for phase randomised noise images are very similar to those for natural images. The large connected areas of white or black labels are caused by 3-BIF-encoding relatively large areas of zero-contrast (which are very seldom seen in natural or phase randomised images) in the dead leaves images. This differing behaviour is mirrored in figure 4, where the results for 3-BIF encoded dead leaves model images appear more Zipf-like than results for natural images or phase randomised noise

This 'mid-range' is made more explicit in figure 4, which shows the full range of Zipfness scores for all the image types and encoding schemes studied. For both 2- and 3-label systems, there is very good correlation between the scores for natural and phase randomised noise images, and as expected (see caption of fig. 4) both of these seem to be tending towards the results for 'coin-flip' noise as $k \to 0$. Images encoded with 2-BIFs appear to achieve optimal Zipfness around $k = 1.5$; and those encoded with 3-BIFs around $k = 1.0$.

Figure 4 shows that the frequencies of occurrence of patterns in natural and phase randomised noise images (and, for 2-BIF coding, dead leaves model images) encoded in the same way are distributed in a similar fashion. Figure 7 expands on this by showing (left-hand side) that the same patterns occur at reasonably similar ranks in these distributions. By contrast (fig. 7, right-hand side), when the same images are encoded with different parameters there is almost no correlation between the ranks of individual patterns in the two codes.

3 Conclusions and Further Work

We have tested whether the frequency statistics of BIF words obeys Zipf's law. We find that as we vary the ratio between filter scale and spatial sampling the Zipfness varies systematically. We have found that for certain parameter settings we get extremely Zipf behaviour. Moreover the encoding at these settings looks optimal to eyeball measure.

We are currently testing whether Zipfness is a predictor of performance of the code when it is used for the task of texture classification.

Acknowledgements

EPSRC-funded project ‘Basic Image Features’ EP/D030978/1.

References

[1] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*: Addison-Wesley, 1949.

[2] B. Mandelbrot, "Contribution a la the´orie mathe´matique des jeux de communication.," in *Publications de l’institut de statistique de l’universite´ de Paris*, vol. 2, 1952, pp. 1–124.

[3] L. D. Griffin and M. Lillholm, "Feature category systems for 2nd order local image structure induced by natural image statistics and otherwise.," presented at SPIE 6492(09):1-11, 2007.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," presented at Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, 1999.

[5] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns," in *Computer Vision - ECCV 2000*, vol. Volume 1842/2000, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, 2000, pp. 404-420.

[6] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 1615-1630, 2005.

[7] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," presented at Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, 2003.

[8] A. Gelbukh and G. Sidorov, "Zipf and Heaps Laws' Coefficients Depend on Language," in *Computational Linguistics and Intelligent Text Processing : Second International Conference, CICLing 2001, Mexico-City, Mexico, February 18-24, 2001, Proceedings:*, 2001, pp. 332.

[9] B. McCowan, S. F. Hanser, and L. R. Doyle, "Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires," *Animal Behaviour*, vol. 57, pp. 409-419, 1999.

[10] J. 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 B: Biological Sciences*, vol. 265, pp. 359-366, 1998.

[11] B. M. ter Haar Romeny, *Front-End Vision and Multi-Scale Image Analysis*: Kluwer Academic Publishers, 2003.

[12] A. B. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *International Journal of Computer Vision*, vol. 41, pp. 35-59, 2001.