Quantifying Texture Scale

Bouremoum S, Protonotarios ED and Griffin LD
University College London - CoMPLEX & Computer Science
s.bouremoum.12@ucl.ac.uk

Keywords
Texture, scale, length-scale, perception, vision, psychophysics, basic image features.

Introduction

Quantitative descriptors that capture texture properties effectively have applications in biomedicine, geoscience and materials science. Ground-truth measurements, against which descriptors can be calibrated and their accuracy can be assessed, are required. These can be hard to obtain for properties such as order and scale for which accepted definitions do not exist. Protonotarios *et al* (2014) described an algorithm to quantify the order of point patterns. The algorithm was shown to be in accordance with an interval scale for human perception of order obtained through a pairwise comparison psychophysical experiment. We report progress on a similar approach for texture scale.

Textures vary from regular to stochastic. There is a clear mathematical definition of a repeat unit for exactly regular textures (Schattschneider, 1978), and the size of the repeat unit is a natural definition of scale for such textures. The concept of a repeat unit can be extended to near-regular textures by treating them as noisy and/or deformed departures from regular textures (Liu et al, 2002). For example, Ardizzone *et al* propose a method to detect the scale of regular and near-regular textures. They begin by converting the texture into a point map using keypoint detection and then estimate texture scale based on a model for how many points should be expected to occur in windows of different sizes. Their model is not expected to work for less regular textures and is not calibrated against perceptual data. We are unaware of any algorithms designed for less regular textures.

We propose an algorithm for the quantification of texture scale which is consistent with human perception and works for all levels of texture regularity. First, through a psychophysical experiment we demonstrate that a perceptual interval scale for texture scale exists consistently within and between human subjects. Next, we use the method presented in Protonotarios *et al* (2014) to obtain an estimate of perceptual texture scale for the a texture dataset. We then use these estimates as ground-truth values in the development and validation of an algorithm for the automated quantification of texture scale.

Perceptual Length-Scale

We carried out two psychophysical experiments, one to investigate whether humans have a consistent perception of texture scale, and one to compute an estimate of perceptual texture scale.

Window Selection Task

Nine subjects participated in a window-selection task in which they selected “the smallest window through which they had a good idea of what the entire texture looks like”. The interface enforced upright square windows that could be positioned anywhere within the texture. Each subject performed the task twice on each of the 112 images in the Brodatz dataset (Brodatz, 1966). We analysed log window sizes.

Investigating individual subjects’ consistency, we found that one of the nine subjects was much less consistent than everyone else and performed the task quicker. We discarded this subject’s poor quality responses. Amongst the remaining, the mean intra-subject correlation was 0.82 suggesting that subjects have a consistent criterion for their selections. Investigating consistency between subjects, we found a mean inter-subject correlation of 0.73. Albeit smaller than the intra-subject correlation, it supports the claim that there is a common perception of texture scale between subjects.

Although correlated, we observed systematic linear differences between subjects' responses. For this reason, subjects' responses were aligned before averaging. Figure 1 shows some typical mean responses.

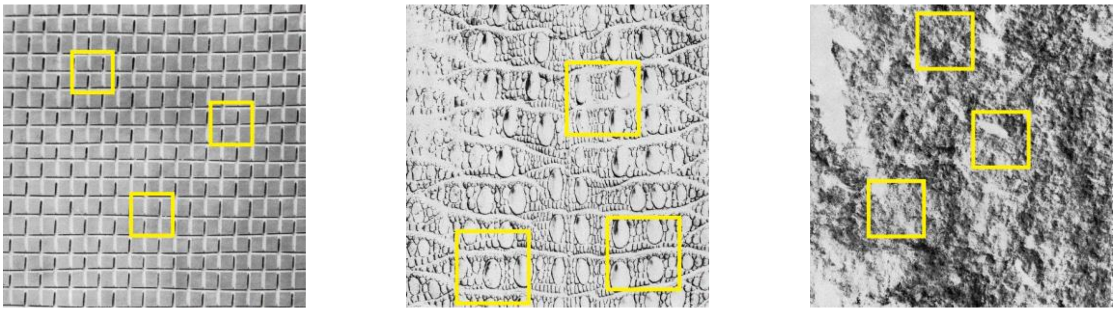

Figure 1. Sample mean responses to window selection task for textures of varying regularity in the Brodatz dataset

Pairwise Comparison Task

Forty participants were presented with pairs of textures and instructed to 'select the one that has the largest length-scale, i.e. the scale over which the texture appears consistently'. Responses were used to construct a perceptual length-scale using the methods from Protonotarios *et al* (2014).

In brief, each texture is assumed to have a single true value on a perceptual interval-scale. Each perception is a noisy realization of the true value. When comparing two textures, the noisy realizations are compared and the observer reports which is largest. If the noise for each perception is an i.i.d. Gaussian (Thurstone, 1927), then there exists a monotonic preference function, which takes the form of a cumulative Gaussian distribution function mapping the signed difference between the true values, to the probability that the one will be preferred to the other. In the perceptual scales derived with this method, equal intervals correspond to equal discrimination performances.

We fit the model to the 9500 collected comparisons by likelihood maximization. The resulting interval scale is invariant under multiplication so we report distance in just-noticeable difference (jnd) units, the distance between two stimuli such that an observer will have a 75% chance of correctly ordering them (Torgerson, 1958). The perceptual length-scale extended over 11.2 jnds for the 112 Brodatz textures.

We assessed the quality of the interval scale model by computing a deviance statistic. The deviance is the difference between the log-likelihood of the dataset given the ML model, and the log-likelihood of the dataset given the saturated model. The saturated model specifies a separate ML probability for the outcome of each trial (Wichmann & Hill, 2001). The goodness-of-fit of the model is assessed by comparing the empirical deviance to the distribution of deviances that result from Monte Carlo generated datasets from the ML model. The empirical deviance, 4246, falls in the 95% interval of acceptable deviances for 10,000 repetitions, [3954, 4390], thus the interval-scale model is accepted.

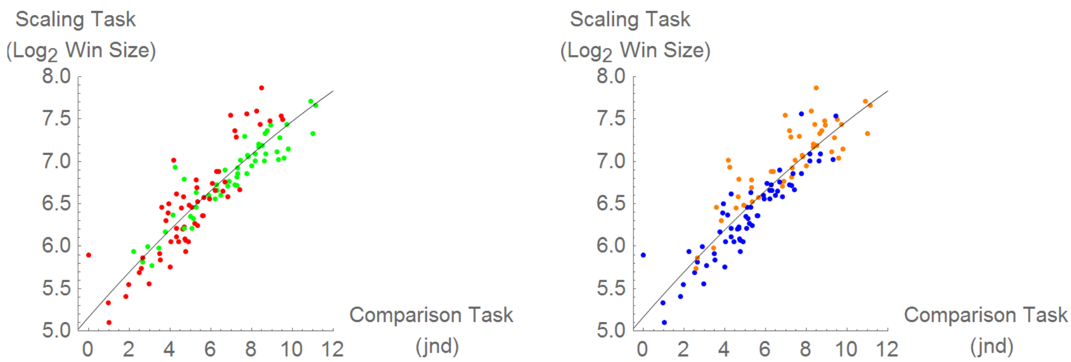

Figure 2. Scatter-plots of window selection task *vs.* pairwise comparison task with quadratic best-fit lines.
Left: points coloured according to texture regularity. Green: highly stochastic textures. Red: all other textures.
Right: points coloured according to texture isotropy. Blue: isotropic textures. Orange: anisotropic textures.

Analysis & Comparison of Perceptual Datasets

The results of the two experiments are well correlated ($\rho = 0.90$), but there are some small but systematic differences as shown in figure 2. Subjects select relatively smaller windows for stochastic and isotropic textures compared to their judgements in the pairwise comparison experiment. The data does not allow us to conclude that either task is superior to the other. However *(i)* it was clear that subjects prefer the pairwise task, *(ii)* ceiling effects for very large or small textures are likely to be lessened with pairwise comparisons, and *(iii)* no alignment is needed for this comparison as it is for direct scaling. Henceforth we use the estimate scales from the pairwise comparison dataset as our ground-truth.

Algorithmic Quantification of Texture Scale

We use Basic Image Features (BIF) columns (Griffin *et al*, 2009) as our texture representation. BIFs encode the second order local structure of the image and have been proved to be very effective in texture (Crosier & Griffin, 2010) and object (Griffin *et al,* 2013) recognition. With BIFs, pixels are classified according to approximate local symmetry based on responses to derivative of Gaussian (DtG) filters. For each scale of the underlying DtG filters, pixels are classified into seven classes. We use filters of scale $\sigma = \{0.5., 0.71, 1.00\}$ pixels. Thus each pixel gets a classification into one of 243 ($= 7^3$) texton classes. A sample from a texture is then characterised by the 243-bin histogram of the pixel labels. We use the Bhattacharyya distance to compute distance between BIF histograms.

For each texture and a range of window sizes we calculate the mean distance between the BIF histograms for a large number of non-overlapping window pairs. We obtain curves of mean-histogram distance *vs.* window size. Figure 3 shows two typical curves for two distinct Brodatz textures. The curves for all textures look similar; in all cases monotonically decreasing with increasing window size.

There are several plausible ways to get a window scale from these curves. We only report the one we found most effective. We introduce a single parameter, η, governing the percentage height between the bottom and top values of a curve at which we set the prediction threshold for all textures. The estimated texture scale is the estimated window size at which the histogram difference attains the threshold level. The dashed lines in figure 3 illustrates how to obtain a window size prediction for two distinct textures.

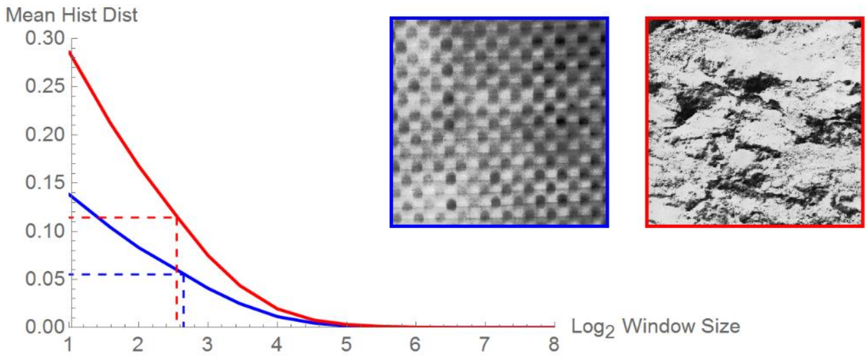

Figure 3. Histogram distance curves for two Brodatz textures & illustration of the scale estimate when $\eta = 40\%$

The readings from curves are mapped to the perceptual length-scale values through a tuned monotonic quadratic function. The process is repeated for a range of cut-off values η. We choose the η minimising the root mean square error (RMSE) between the transformed predictions and the perceptual scale.

Results & Discussion

The value of η resulting in the smallest RMSE is $\eta = 1.5\%$. With this value of η, the RMSE between predicted texture scales and the values estimated from the pairwise comparison data for all the images in the Brodatz dataset is 1.4 jnds. Looking at textures for which our algorithm gives poor predictions, we find that if we remove 10 out 112 textures and refit to the remaining, the RMSE drops to 1.0 jnd. The 10 removed textures show some commonalities. In particular, six of them are textile images. We note that the RMSE values of 1.4 and 1.0 take in to account the number of parameters in the algorithm.

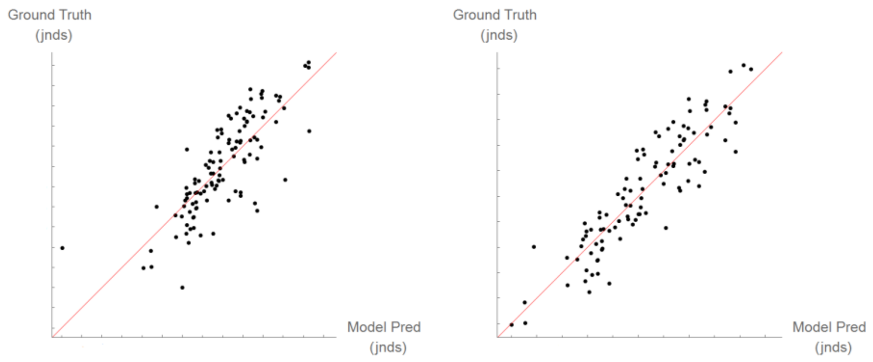

Figure 4. Scatter plots of ground-truth *vs* model prediction. Left: All Brodatz images. Right: Excluding 10 images.

Conclusion

Explaining to subjects that texture scale is the size at which the texture appears consistently, we collected two datasets from psychophysical experiments. The first (direct scaling) supports the existence of a consistent and shared perception of texture scale between subjects. The second (pairwise comparison) enabled us to construct an absolute interval-scale for texture scale which we use as ground-truth for validation of a texture scale estimation algorithm. The algorithm is based on BIF histogram texture descriptors and has a single parameter controlling a cut-off threshold and three parameters from a tuned quadratic function. On the complete Brodatz dataset, our model has an adequate RMSE error of 1.4 jnds. Removing a small proportion of the data results in a much improved RMSE of 1.0 jnd. The RMSE should be compared to the extent of the scale interval which is 11.2 jnds for the Brodatz. In future work we intend to improve our prediction algorithm, perform an analysis of the textures for which the model underperforms and investigate further the effect of the task on subjects' perception of texture scale.

References

Protonotarios E., Baum B., Johnston A., Hunter G. L. and Griffin L. D. (2014) 'An absolute interval scale of order for point patterns', *Journal of The Royal Society Interface,* 11(99), 20140342.

Schattschneider. D. (1978) 'The plane symmetry groups: their recognition and notation', *American mathematical monthly*, 439-450.

Liu Y., Tsin Y. and Lin. W. C. (2005) 'The promise and perils of near-regular texture', *International Journal of Computer Vision* 62. 1-2, 145-159

Brodatz P. (1966) 'Textures: a photographic album for artists and designers', *Dover Publications*.

Griffin L. D., Lillholm M., Crosier M. and Sande J. (2009) 'Basic image features (bifs) arising from approximate symmetry type' in *Scale Space and Variational Methods Computer Vision*, 5567, 343-355.

Griffin L. D., Wahab M. H. and Newell A.J. (2013) 'Distributional learning of appearance', *PLOS one*, 8(2):e58074.

Crosier M. and Griffin L. D. (2010) 'Using basic image features for texture classification', *International Journal of Computer Vision* 88(3), 447-460.

Ardizzone E., Bruno A. and Mazzola G (2013) 'Scale detection via keypoint density maps in regular or near-regular textures', *Pattern Recognition Letters* 34.16, 2071-2078.

Thurstone L. L. (1927) 'A law of comparative judgement', *Psychology Review,* 34, 273-286.

Torgerson W. S. (1958) 'Theory and methods of scaling", Wiley.

Wichmann F. A. and Hill N. J (2001) 'The psychometric function: I. Fitting, sampling and goodness of fit', *Perceptual Psychophysics* 63.8, 1293-1313.