Towards Self-Organized Feature Maps from Gabor Filter Responses

James Orwell[1], Ramón Turnes[2], María J. Carreira[2]
Diego Cabello[2], James Boyce[1] and John F. Haddon[3]

[1]*Image Processing Group, Kings College London*
[2]*Dept. Electrónica e Computación, Univ. Santiago*
[3]*Defense Research Agency, Farnborough*
e-mail: jo@physig3.ph.kcl.ac.uk, elreal@usc.es

Abstract

In this paper we develop a system initially proposed by Lampinen[3] to extract perceptual primitives in successive stages. Several orientations and scales of Gabor Filter are used to form a high dimensional space in which the original data is represented. The Self Organizing Map is used to find a useful sub-space representation. We investigate its use *hierachically*, first to represent the orientations at a particular scale, and then again, using the outputs from the various scales. We investigate different representations of the filter responses, and show that a useful feature map needs training data biased toward perceptually significant data. High amplitude filter responses serve this purpose. Histograms from such feature maps are used to classify image data: features and whole objects - using the Learning Vector Quantization method. Preliminary experiments, investigating the dynamics of the process, confirm Lampinen's claim of an architecture tolerant to distortions of noise, clutter, and small rotations and scalings. Experiments are presented on synthetic, aerial infra-red, and indoor scene imagery.

key words: perceptual grouping, Gabor filters, pattern recognition, SOM, LVQ.

1 Introduction

This research is motivated by the goal of a hierachical, self-organized representation of image data, in order to recognize objects. There is evidence that relations between features are an important cue for pattern recognition tasks and an explicit representation of such relations exist in nature and would seem useful in artificial systems. These relations include proximity, parallelness, symmetry and closure. There is a tension between the need for an explicit formulation of these relations, and the desire to develop an autonomous system which learns from the input data: ideally, the usefulness of the relations would also be learned. The Self Organizing Map is a suitable paradigm for researching such systems. Its output is a two-dimensional manifold through the data space and can thus be easily visualized. A hierachy of maps can thus be constructed, enabling a succession of transformations of the input data, each stage of which being available for inspection.

This paper describes the Gabor filters used to create the initial data, and the manner of representation and sampling which provides useful feature maps. A two layer hierarchy has been implemented [3], grouping data from first different orientations, and then different scales. Images classified using these first and second layer maps are presented. We discuss ways of making

 This work was part funded by the DRA (Farnborough), under Research Contract SF/U740

the sampling strategy optimal in an information theoretic sense, and bridging the gap between supervised and unsupervised methods of feature grouping.

2 Extraction of features: Gabor filters

Gabor filters were introduced by Daugman[1][2]. They are two-dimensional filters sensitive to different orientations and frequencies. There is evidence that visual cortical cells in mammals closely resemeble this configuration [6]; importantly, their conjoint resolution in image and frequency space is optimized. Gabor filters are sinsoidals softened by a gaussian[3]:

$$\psi(f,\theta,x,y) = exp(i(f_x x + f_y y) - \frac{f^2(x^2+y^2)}{2\sigma^2}) \tag{1}$$

with $f_x = f cos\theta$ and $f_y = f sen\theta$ where f is the central frequency of the pass-band filter, θ is the spatial orientation and σ is the fundamental width of the filters. We quantize f and θ in $f_k = \frac{\pi}{2^k}$ and $\theta_t = \frac{\pi}{8}t$ such that the coefficients k and t determine the frequency and orientations. We have considered eight different orientations ($t = 0, 1...7$) and up to three frequencies ($k = 0, 1, 2$). Lee [9] assets that, to form a tight frame, a wavelet family of Gabor filters need between two and three frequency steps per octave; thus more frequencies will be required to have a complete representation of the input data.

We convolve each input image $I(x,y)$ with the kernels $G^f_\theta(x,y)$ to obtain filter outputs $H^k_t(x,y)$:

$$G^f_\theta(x,y) = \psi(\theta, f, x - \frac{k_w}{2}, y - \frac{k_h}{2}) \tag{2}$$

where k_w and k_h are the width and height of the kernels respectively, chosen to be 7, 15, and 25 pixels for $k = 0, 1, 2$ respectively.

Once the images have been filtered we create a data file for each image and considered frequency such $H^k(x,y) : H^k_0(x,y), H^k_1(x,y), H^k_2(x,y) \dots H^k_7(x,y)$. The *amplitude* of a gaborjet will refer to the modulus of the above 8-vector.

2.1 Presenting the Gabor Data to the Map

Several considerations mitigate against presenting this data to the map for training straight away. First, the statistics of the output vary greatly with respect to the scale of the filter used: convolution with a low frequency filter will, in general, produce a greater range of values: commensurability across scales would be useful. Second, the input images had a variability of dynamic range which must be accommodated, either before or after filtering. Third, it is by no means obvious that the original filter output data is in its most suitable form for clustering, using the Euclidean distance metric the Self Organizing Map employs. The statistics of the amptitudes of response are leptokurtic at the top end (*i.e.* a relatively large tail), which may hinder clustering of similar strong edge features, since, relative to more moderate amplitudes, they can have a very large variance. These considerations suggest the following scheme. We wish to innure the gaborjets against variability from image to image: thus we define the *response* of the gaborjet to have zero mean and unit variance, over a given image, by subtracting from each amplitude the image mean amplitude, and dividing by the image variance. This normalizes the amplitude statistics across images. This does have the consequence that the response from a particular patch is dependent upon statistics elsewhere in the image. This also occurs in biological systems, and could later be corrected by a foveation, just as one shields one's eyes from the sun. Next, to correct for leptokurtic tendancies in the input data, the response was bounded to the range {0,1} by passing it through a *arctangent* sigmoid function. The response was added, as an extra (ninth) component, to the data input to the map. Two alternatives schemes to represent the first eight components were investigated: First, we normalize these values to unity, such that they lie surface of an

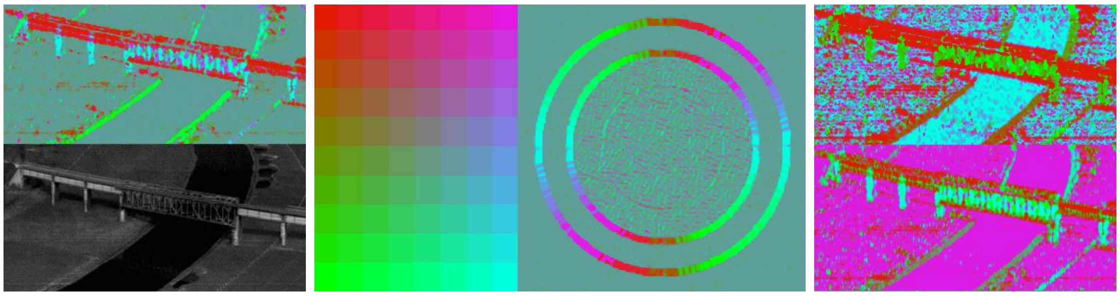

Figure 1: Labelling scheme for a Self-Organized Map. Each pixel's winning neuron is labelled with a unique colour. The matrix shows the colour allocation. The top left FLIR image, and the artificial circle image, are labelled with a map trained with high bias ($n = 10$). The top and bottom right images had biases of $n = 1$ and $n = 5$ respectively. Bottom Left is the original image.

eight-dimensional hypersphere. No information is lost since the response component enables a full reconstruction. Second, we scale the components such that their modulus is equal to the response. There is now some redundancy of coding, but this is offset by the utility of rendering the response explicitly in the input (and thus trained codebook) representation. Theoretically, the differences between these two techniques are interesting: two gaborjets, differing only in response, suffer a far larger separation in the latter representation, since the difference is coded in all dimensions. In the former representation the only separation is in the response component. This should imply different grouping characteristics; but in practice any differences were negligable, compared to variations in other parameters.

3 Training the Maps

The Self Organizing Map is an unsupervised neural network. Input data are not labelled and the target of the organized map is to partition them into similar feature sets. For more details see [4].

3.1 Sampling the Input Data

The lattice of neurons self-organizes into a configuration that models the distribution of the input data. Thus, densely populated areas in the input space attract a large number of neurons. In this application, *i.e.* the generation of a useful feature map, there is no *a priori* reason to suppose the input distribution will generate such a map. Indeed, the bulk of typical image data will be 'uninteresting', *i.e.* lacking in salient features. Training a map on this data will generate a map of which the majority of neurons will model this part. Our hypothesis is that biasing the sampling of training data towards vectors with a high response will lead to a more useful map. This needs the lemma that saliency is correlated with areas of high response: intuitively this is so, and is a common assumption among workers in the field[7].

Such a map will be more useful for two reasons. first, 'uninteresting' areas will be assigned to a smaller number of neurons, simplifying the labelling of the image by eliminating irrelevent variations of classification. Second, a greater proportion of the map will be assigned to the high response vectors, which implies the classification is more articulate about the salient features.

Two methods of biased sampling were implemented. First, a weight was assigned to the vector, which was a function of the response: the number of times a training pattern was presented to the map was proportional to its weight. Second, a sample was selected from the image set with a probability dependent upon its response. Specifically, we determined that the probability of

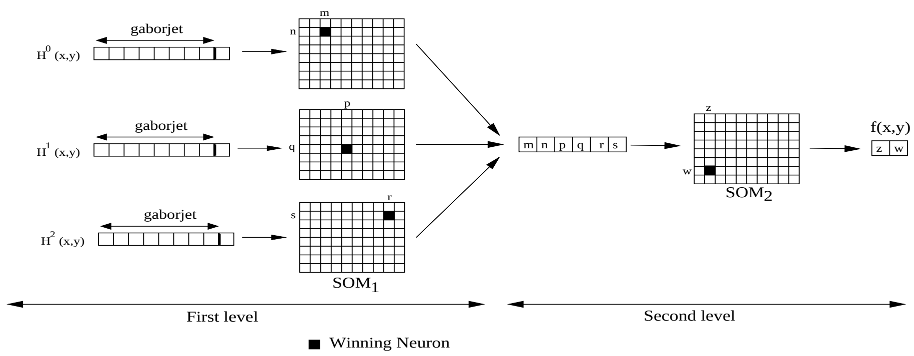

Figure 2: Level 1 and 2 maps and result for a point(x,y) from the original image

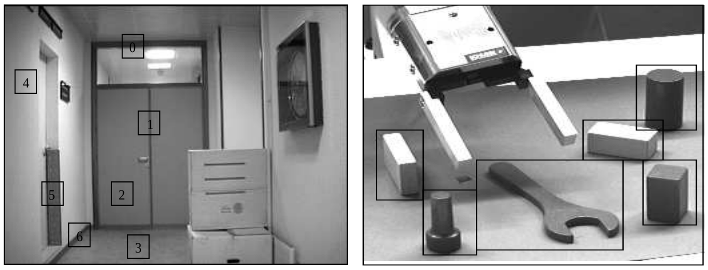

Figure 3: Object and sub-object features used to test the feature maps.

selection was propotional to the nth power of the response, where n is a parameter to be fixed. The former is more computationally efficient; the latter allowed us to explicity compare the response histogram of the input data with the ninth component of the codebook vectors. There is little theoretical difference between these two methods.

The characteristics of the map, and thus of the image labelled with the neuron co-ordinates, are highly dependent on the sampling distribution: see fig 2.1.

3.2 First level

In the first level we have three maps of NxM neurons, such we consider each of the frequencies independently. The training vectors for this level are the nine-component gaborjets, as described in section 2; see fig 2.

We have considered three different sets of grey-scale images:

1. Images captured by two cameras situated over a mobile robot. The robot was moved in straight trajectory and the distance between all the images is almost the same.

2. A publicly available set of images [1], showing objects in an indoor scene in different lighting conditions and with varying clutter.

3. A selection of images from five different aerial sequences of Forward Looking Infra-Red (FLIR) imagery.

To train the map, the package developed at Helsinki University was used [2]. The authors acknowledge this help, and recommend the package to others. Three principle parameters needed

[1] http://www.informatik.ini-kiel-de./-jpa/images.html

[2] http://nucleus.hut.fi/nnrc/nnrc-programs.html

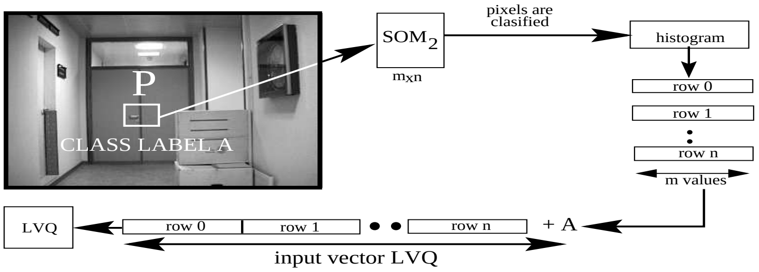

Figure 4: LVQ classification

to be set: the training length, the learning rate, and the neighbourhood size. Experiments were performed on a range of parameter settings, and two criteria were used to fix these at appropriate levels: the quantization error of the map, measured against a validation set; and the classification performance of higher processing, using a map with particular parameters. However it must be stressed that, in the opinion of the authors, particular parameter settings were not critical to the performance of the system: there was a 'broad minimum' to the configuration it was not difficult to find.

4 Second level Map

Once the first level has been trained and the input patterns classified we train the map of the second level. As figure 2 shows, the outputs of the first level are joined to make an input vector to the second level SOM. The only difference here is that now we have a 6-dimensional input space, and the units of the space are in lattice neurons. The same type of map is used to those of the first level. Analysis of the second layer codebook vectors is a more difficult excercise, since the neuron weights refer to more abstract entities: simultaneously firing first layer neurons, rather than the Gabor filter responses to which the first layer weights refered. Thus, visual inspection of labelled images must play a more dominant role in diagnosing the behaviour of the self-organizing system. To train the second layer, the data must also be sampled. In the second layer we have three responses to consider, for any one data sample. It was reasoned that the response from the lowest frequency (*i.e.* coarsest scale) should be used to pilot the sampling stategy, since this varies across the image most smoothly: we plan to change this to sample as a function of the *maximum* response across scales, for a given pixel.

Following Lampinen ([3]), we compile histograms of neuron responses, over some local neighbourhood. It is hypothesized that the histograms of these responses, summed over the neighbourhood, are distortion tolerant statistics. This hypothesis is tested in the next section.

5 Classification of Neighbourhood Histograms

Object and sub-object patterns are labelled (see Fig 3), and classified using the Learning vector Quantization Algorithm [5].

We select patches from the input images (see Fig 4) and calculate the winning neurons for all the pixels of the patch: a two-dimensional histogram. Example histograms can be seen in fig.6, showing the the histograms of horizontal (left) and vertical (right) lines for two different patches.

To compare histograms extracted from divers patch size, we enforce unit area on the histogram. The histogram is treated by the classifier as a one-dimendional histogram (Fig 4). The classifier was tested on a validation set, unseen during the training.

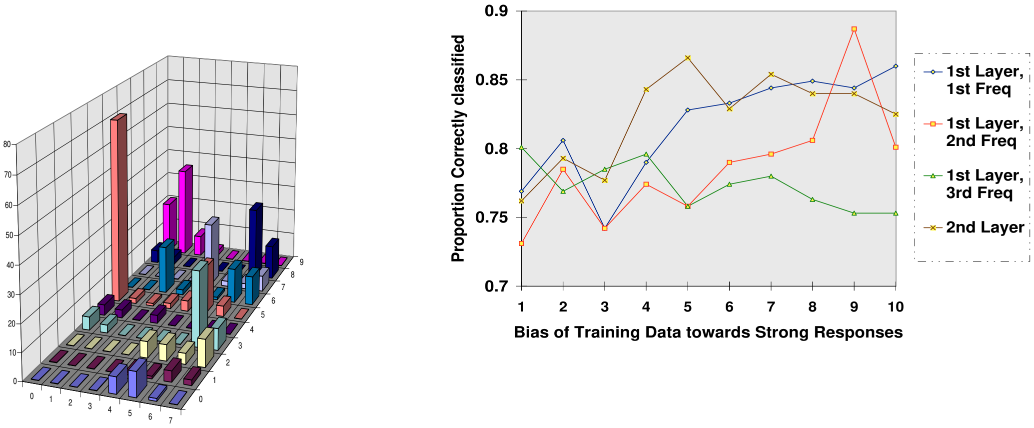

Figure 5: Example histogram of the neuron responses; classification performance as a function of the bias towartds high response of the (FLIR) sampling data.

6 Results

The classification experiments presented here are on the three types of imagery listed above. In the robot camera sequence, seven different classes were labelled (see fig 3.2 left). In the corridor sequence, manually labelled objects were used; ussee fig 3.2.

For each of the sets, we showed the global average accuracy in the classification and the class We see in table I that the average accuracy is high (around 90%) and horizontal, vertical and diagonal lines have an even higher accuracy (around 98 %). With the second set of images (see table II) we see that the size of the maps is not a critical parameter to obtain a good classification of the different classes from figure 3.2(right).

Fig 6 shows the classification accuracy as a function of the bias towards the high response data, using the infra red sequences. The accuracy increases with the training bias in all but the 3rd (coarse) frequency first layer map.

Table I:

Map size	A_v	A_0	A_1	A_2	A_3	A_4	A_5	A_6
8x10	91.29	94.58	96.72	92.72	98.52	71.82	94.12	92.77
4x5	92.43	94.74	98.80	91.82	95.82	74.29	96.30	94.17

Table II:

Map size	A_v	A_1	A_2	A_3	A_4	A_5	A_6
5x6	98.41	89.82	100	100	100	100	100
4x5	100	100	100	100	100	100	100
3x4	100	100	100	100	100	100	100
5x2	100	100	100	100	100	100	100
2x3	86.92	91.41	100	85.86	100	73.61	96.97

7 Discussion and Future Work.

The feature map is most useful when heavily biased towards high response Gabor filter data. This state of affairs might be considered unsatisfactory, since this bias is manually determined: it is an external observation that high-response outputs deserve a disproportionate share of lattice. One principled way in which we might obviate such measures is to demand that, rather than consider *probabilty* density (which we have distorted), we consider *information* density as our 'driving force' (see Plumbley [8]) behind the distribution of the neurons. That is, we attempt to equalize the information which each neuron transmits to a higher level. In such a way the necessary bias might be invoked automatically.

Alternatively, we can view the bias as blurring the distinction between supervised and unsupervised methods. Traditional supervised methods are cumbersome and unpopular partly because they require extensive ground-truthing. Biasing the training set is a easy way to incorporate knowledge into a feature map. We propose that higher levels of feature mapping, representing complex, possibly non-local collections of features, could be similarly piloted by such judicious selection of training samples. Thus, the task of perceptual grouping – representing salient groups of features – could be achieved in similar manner by training a self-organizing system on and only on such related features. At this low-level stage, almost any image could be used to train the map – 'as long as similar elementary features like edges are present in both' (Lampinen[3]: higher abstractions will require more specific imagery, and, though technically unsupervised, will become more akin to a loosely supervised architecture).

8 Conclusions

We have shown that Gabor features can be grouped by a hierachically to create a feature map that is perceptually useful, smoothly varying with input data, intrinsically parallel, and achieves massive dimensionality reduction. The most critical parameter in the dynamics of the process is the sampling strategy. At present this strategy is arbitrarily designated, yet we hope to achieve a principled method for its determination. Although the temporal costs are high on a serial computer, the process is intrinsically parallel, and achieves encouraging results on the preliminary quantitative classification experiments.

References

[1] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency and orientation optimized by 2D visual cortical filters", *J. Opt. Soc.Am. (A)*, vol. 2, no. 7, pp1160-1169, 1985.

[2] J. Daugman, "Complete discrete 2D Gabor transformations by neural networks for image analysis and compression", *IEEE Trans. ASSSP*, vol. 36, no. 7, pp 1169-1179, Jul 1988.

[3] J. Lampinen and E. Oja, "Distortion Tolerant Pattern Recognition Based on Self-Organizing Feature Extraction", *IEEE Transactions on Neural Networks*, vol. 6, no. 3, May 1995

[4] T. Kohonen, "Self-organized formation of topollogically correct feature maps", *Biological Cybernetics*, vol.43, pp 59-69,1982.

[5] T. Kohonen, "Improved versions of learning vector quantization",*Proceedings of the International Joint Conference of Neural Networks*, pp 545-550, San Diego, Junio 1990.

[6] D.H. Hubel and T.N Weisal, "Functional Architecture of the Macaque Monkey Visual Cortex", *Proc. Royal Soc. B, London*, vol. 198, pp 1-59, 1978.

[7] A. Shashua and S. Ullman, "Structural Saliency". In *Proc. IEEE Int. Conf. Comp. Vision*, Dec 1988, pp.321-327

[8] M. D. Plumbley, "Towards an Information Density Measure for Neural Feature Maps" Awaiting publication. M.Plumbley@kcl.ac.uk.

[9] Tai Sing Lee, "Image Representation Using 2D Gabor Wavelets" *IEEE Patt. Anal. and Machine Int*, Vol 18, No.10 October 1996