Classification in High Resolution Images with Multiple Classifiers

A. Monadjemi and B. T. Thomas and M. Mirmehdi
Department of Computer Science
University of Bristol
Bristol, BS8 1UB, England
email: {monadjem,barry,majid}@cs.bris.ac.uk

Abstract

We examine the use of high frequency features in high resolution images and demonstrate how they can increase texture classification accuracy when used in combination with lower frequency features. We used eight features, four low frequency and four high frequency, derived from patches of 4032 × 2688 images. Furthermore, we experiment with both single and multiple classifiers to illustrate the effectiveness of such a combination. Outcomes of classification tests on outdoor scene patches are presented and discussed.

Keywords: Image Classification, High Resolution images, High Frequency Analysis, Multiple classifiers.

1 Introduction

In recent years, our research group has developed a neural network based system for classifying images of typical outdoor scenes to an area accuracy of approximately 90% [1]. The system was trained with features extracted from segmented regions of an extensive image database. Texture information was represented using Gabor filters. A common problem confronting this approach is that many regions in typical outdoor scenes are too small to allow a significant range of spatial frequency to be included in the feature set.

This paper presents a pilot study designed to establish if high resolution images would provide a sufficient increase in texture information to justify the extra computational complexity. Our approach is to train a classifier to distinguish between four object classes using frequency information alone. Patches extracted from high resolution images are used in the training. The patch size is representative of the spatial scale of objects typical in the low-resolution database of outdoor scenes used in [1]. They are large enough, however, at this higher resolution to allow classification performance to be assessed over a significant range of higher spatial frequencies.

High frequency information in images can be commonly associated with edges and noise [2]. We are not concerned, at least directly, with edge information. Instead, we are examining overall texture which will embody both edge and noise in ’homogeneous’ regions as well as any fine or coarse resolution characteristics. Fine resolution textures, such as roads and pavements, are particularly rich in high frequency information content.

Some researchers disregard higher frequencies (HF) since the power spectra of natural images show an exponential decay with frequency and in most cases image acquisition is made through a conventional optical system that filters out the very high frequencies [3]. Nevertheless, we show in this study that higher frequency information extracted from high resolution images can improve classification performance. It must be emphasized that in the high resolution images we have used for our experiments in this paper, what we refer to as *lower frequencies* (LF) correspond to the highest frequencies found in normal lower resolution (e.g. 512 × 512) images.

Classification using a single classifier can lead to inefficiently trained and/or complex classifiers. Multiple classifiers are gaining much popularity [4] and can be designed such that extra classification steps can be carried out only when necessary. We perform our experiments using a single classifier at first and then illustrate the improvements gained by using two different arrangements of a multiple classifier system. In one arrangement the final decision is based on a weighted average of two separate classifiers while in the other arrangement, a secondary classifier is only used when we do not have enough confidence in the decision of the primary classifier. We refer to these two different combinations of classifiers as W_AVG and PRI_SEC.

Wang et al. [5] have shown that considerable classification improvement can be achieved when the feature set can be divided into separate subsets depending on their classification power and characteristics. This maps very well in terms of our use of LF and HF features and hence motivates our employment of multiple classifiers.

Following a brief literature review in the next section, we present the low frequency and high frequency features used in our experiments in Section 3. Then, in Sections 4 and 5 we report on our classification tests using both single and multiple classifiers. Summary and conclusions are provided in Section 6.

2 Background and Literature Review

Work on the analysis of HFs in higher resolution images in particular is scarce. Many high resolution image processing applications have focused on astronomy or remote sensing, where the objects sought are usually very small and imaging facilities are excellent in quality. For example, Legault [6] presented the requirements of a high resolution imaging system for astronomical image processing for which the Modulation Transfer Function was considred as a criterion for high frequency response measurement. Also, Myint [7] utilized wavelet transforms as a multi-band approach to analyse textures in high resolution multi-spectral remote-sensed images. He demonstrated the essential role of the spatial resolution factor, which could be interpreted as the importance of higher frequency data, in dealing with fine resolution textures. Freeman and Pasztor [8] used probabilistic models and Markov Random Fields for the estimation and retrieval of higher frequency information, eliminated during the typical recording procedure of video frames, to produce higher resolution pictures for HDTV systems.

Multiple classifiers have recently become an active topic of investigation within the computer vision and pattern recognition communities [4]. Multiple classifier classification can be simply defined as using more than one classifier, feature set or both and combining their outputs by employing a particular *Combination Rule* to obtain the final answer. The most important motivations for using multiple classifiers are the advantages gained in simplifying the complexities of classification problems and amplification of the strengths of each individual classifier or feature set in the global classification procedure [9]. Multiple classification has recently been applied in many diverse applications [10, 11, 12]. For example, Wan and Fraser have implemented a multi-classification scheme for different complex remote sensing classification problems [12]. Cappelli et al. have used it in fingerprint classification [10], and Jiang et al. have developed a combined classifier system for grammar guided sentence recognition [11].

In cases of deploying different feature sets, as performed in this work, multiple classifier systems can only improve classification performance when the features are sufficiently different for each classifier [5]. Therefore, the increased performance using a combination of LF-tuned and HF-tuned classifiers can be interpreted as a measure of relative independence between the two feature groups. In other words, if there is no definite gain in classification using both LF and HF features compared to using LF features only, then quite clearly the HF features are shown to not convey any extra classification information over the LF features.

3 Feature Selection

In this work, we wish to classify our images into 4 categories: trees, pavements, cars, and roads. The input images were 256×256 patches obtained from high resolution 4032×2688 pixel images of outdoor scenes. The total number of patches in our dataset was 166. This was divided into a training set of 114, a validation set of 20, and a test set of 32 patches. Figure 1 shows two typical examples in each class of our input patches.

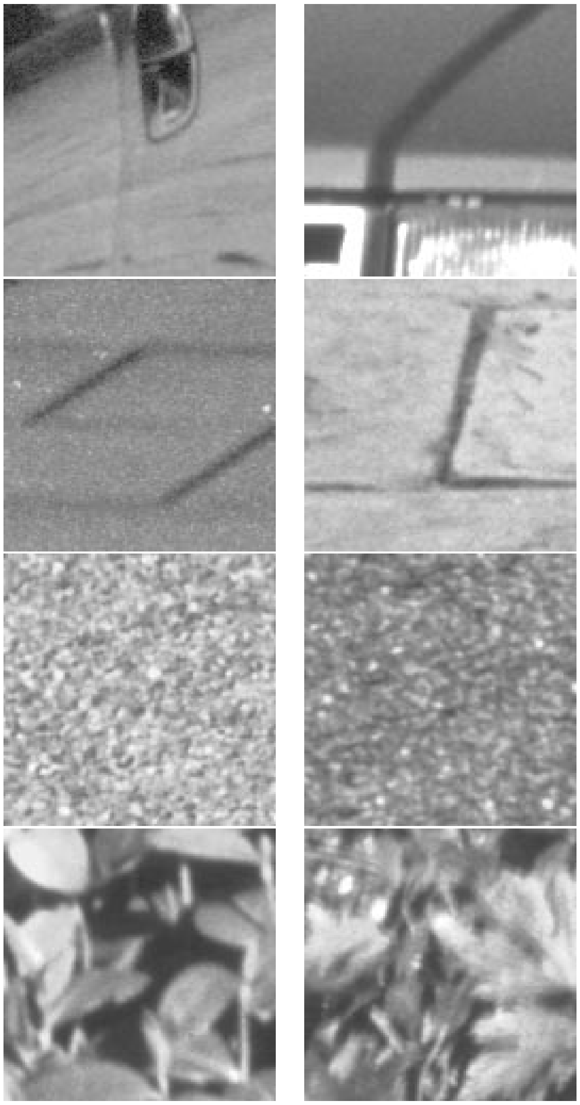

Figure 1: Eight sample images of four classes, from top row to bottom: car, pavement, road, tree.

The use of Gabor filters as a tool for the analysis of textures is now a common idea, mostly due to their ability to extract information in different spatial frequency ranges and orientations. Daugman [13] originally proposed this family of 2-D filters in the 1980s as a building block for interpreting the orientation-selective and spatial-frequency-selective receptive field properties of neurons in the human visual cortex, and as useful operators for practical computational image processing. Many works have applied Gabor filters for texture classification tasks. For example, Fogel and Sagi have discussed their application to biological texture discrimination [14], Jain and Farrokhnia have shown their advantages in unsupervised texture classification [15], and more recently, Mirmehdi and Perissamy have used them in a content-based image retrieval application for extracting features from colour textures [16]. The Gabor filter in the frequency domain is:

$$g(u,v) = e^{-\pi(\frac{u_p^2}{\sigma_x^2}+\frac{v_p^2}{\sigma_y^2})} . e^{-2\pi j(x_0 u + y_0 v)} \tag{1}$$

where, $u_p = (u - \omega_x) * cos(\theta) + (v - \omega_y) * sin(\theta)$, and $v_p = -(u - \omega_x) * sin(\theta) + (v - \omega_y) * cos(\theta)$, are the rotated/displaced coordinates in the frequency plan, ω_x and ω_y are filter central frequencies (modulation factors) in x and y directions, θ is filter orientation parameter, σ_x and σ_y are filter standard deviations in x and y directions, and x_0 and y_0 are horizontal and vertical displacements in the spatial domain. We keep $x_0 = 0$, $y_0 = 0$, and set $\omega_x = \omega_y$, and $\sigma_x = \sigma_y$ in all the experiments. As we intended to compare both low and high frequency features, two different Gabor filter banks with different central frequencies and bandwidths in the frequency domain were considered:

$$LF\ Filter \Rightarrow G(\omega_1, \sigma_1, \theta_1), \quad where \quad \omega_1 = 25,\ \sigma_1 = 10 \tag{2}$$

$$HF\ Filter \Rightarrow G(\omega_2, \sigma_2, \theta_2), \quad where \quad \omega_2 = 100,\ \sigma_2 = 40 \tag{3}$$

and $\theta_1 = \theta_2 = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}$. Thus, each filter bank contains four filters with the same central frequency and standard deviation but different orientations. Figure 2 shows the filter banks in the frequency domain, where the four inner peaks are the lower frequency filters ($\omega_1 = 25$) and the four larger, outer peaks are the higher frequency ones ($\omega_2 = 100$). The values for $(\omega_1, \omega_2, \sigma_1, \sigma_2)$ were chosen to reflect the coverage and density of the corresponding features in the frequency space which is also consistent with past applications of Gabor filters. More specifically, as Table 1 shows, the selected LF central frequency ($\omega_1 = 25$) are close to maximum frequencies in lower resolution images such as a 512×512 image, and the HF central frequency ($\omega_2 = 100$) is considerably further away from the highest analysable frequency for patches in lower resolution images. Therefore, we would like to consider if using HF features obtained from high resolution images would improve classification performance. We will examine this in the next two sections.

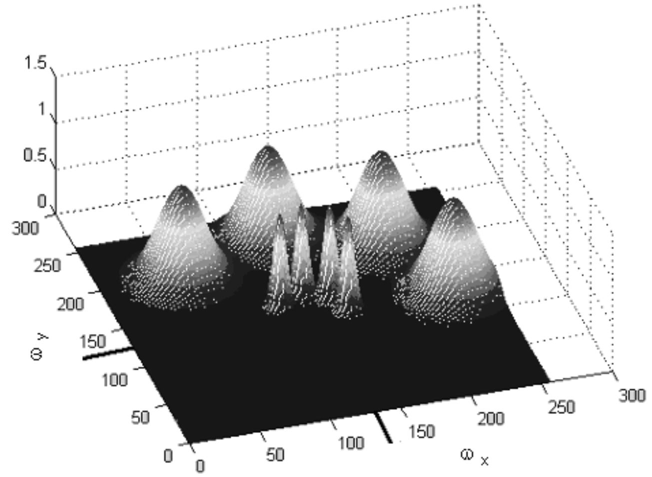

Figure 2: Applied Gabor filter bank, 4 lower frequency with $\omega = 25$ and $\sigma = 10$ (inner), and 4 higher with $\omega = 100$ and $\sigma = 40$ (outer). Orientations were $\theta = 0, \pi/4, \pi/2, 3\pi/4$.

4 Single Classifier Tests

To determine the classification performance of LF and HF features two series of experiments were performed. In the first series, we used a single classifier to determine the role of LF and HF features individually and when used together. In the second series of experiments, discussed in Section 5, classification was carried out by employing two different combining algorithms using a multiple classifier.

Original Size	**Patch Size**	**Max. Analysable Freq. in Patch**
4032×2688	256	$256/2 = 128$
1024×1024	$256 \times 1024/2688 = 97.52$	$97.52/2 = 48.76$
512×512	$256 \times 512/2688 = 48.76$	$48.76/2 = 24.38$

Table 1: The highest analysable frequency in different resolutions.

For classification we employed a back-propagation neural network with one hidden layer. We experimented with many different numbers of hidden layer nodes, but only the optimum results, where each classifier showed the best performances overall, are reported. To measure the classification error, we used the Sum of Square Errors, SSE, as the difference between the ground truth G (i.e. the *expected* outputs of classifiers), and the network classification C (i.e. the *real* outputs of the classifiers) across N classes:

$$SSE = \sum_{i=1}^{N} (G_i - C_i)^2 \tag{4}$$

Another metric used for classification evaluation, was *Classification Error*, CE, showing the percentage of wrong class assignments across the complete labelled test set.

Table 2 represents the classification results using different single classifiers. As expected, a single LF classifier with SSE and classification error (CE) 6.56 and 15.6% respectively reported better results than a single HF classifier with SSE at 15.21 and CE at 43.7%. However, a single classifier using all eight LF and HF features together, provided the best performance with SSE at 6.40 and CE at 9.3%. This resulted in 3 incorrect class assignments out of a total of 32 test samples. The improvement in the LF/HF classifier demonstrates the usefulness of HF features for scene classification. More details of these tests can be found in [17].

Classifier	**SSE**	**CE**	**CE (%)**
LF ($\omega_1 = 25$)	6.56	5	15.6%
HF ($\omega_2 = 100$)	15.21	14	43.7%
Both LF/HF	6.40	3	9.3%

Table 2: Single classifier results

5 Multiple Classifier Tests

In the second series of experiments, for continuity and comparison, again neural network-based multiple classification schemes were employed. Neural networks have shown great ability as general function approximators, especially when the function to be approximated is nonanalytic and complex, for instance a classifier with a number of unordered inputs. However, there are still some problems in their training and consequently classification performances, which have a direct relation with level of complexity and size of the network [9].

In this work, we used a multiple classifier system consisting of two individual classifiers, one for LF features and one for HF features. Their outputs were then combined through a *combine module* to make a final decision. We examined two different *combination algorithms* for the combine module. In the first, here called W_AVG, an unknown input sample was presented to both classifiers. Having obtained the respective responses, the combine module computed their weighted average, using two weighting factors called k_1 and $k_2 = 1 - k_1$, and then selected the class with the minimum SSE. In the second, here called PRI_SEC, a primary/secondary scheme was applied in which the LF classifier was used as the main primary classifier. If this classifier was certain of its classification (i.e. the primary classifier's SSE was less than a given threshold, δ_1) then the secondary classifier was not required. Otherwise, the secondary HF classifier was deployed to aid in the final classification if it demonstrated high certainty in its decision (i.e. if the secondary SSE was less than another threshold, δ_2). If the secondary classifier did not demonstrate high certainty, then the primary classifier's decision was accepted anyway as the final classification. All other conditions were the same as with the previous experiments. Figure 3 illustrates the combining algorithms used.

To find out the optimum values of the combine module parameters, (averaging weight k_1 for W_AVG scheme and certainty/error thresholds δ_1 and δ_2 for PRI_SEC scheme), a linear search algorithm of the parameter spaces was developed. This measured the performance of classification on a validation set of 20 samples, changed the parameters and iterated the classification test for new parameters and continued this procedure for all possible values of parameters and finally returned the parameters with minimum error. In our test, the parameter space was $[0,1]$ for k_1 (and $k_2 = 1 - k_1$), and $[0,0.4]$ for δ_1 and δ_2. The optimum parameters selected were:

$$\begin{cases} W\_AVG: & k_1^{opt} = 0.750 \ (LF), \ k_2^{opt} = 0.250 \ (HF) \\ PRI\_SEC: & \delta_1^{opt} = 0.012 \ (LF), \ \delta_2^{opt} = 0.027 \ (HF) \end{cases}$$

The optimum weights for the W_AVG scheme show the higher importance attached to the LF features. However, the role of the HF features is still significant and the improved performance of the multiple classifier system depends on the combination of both LF and HF features. In the PRI_SEC scheme, the optimum error threshold δ_1^{opt} is smaller than δ_2^{opt} and this demonstrates that the system considers a higher level of acceptable error for HF features, i.e. the combine module considers LF features to be more reliable than HFs.

Classifier	**SSE**	**CE**	**CE (%)**
W_AVG	4.77	3	9.3%
PRI_SEC	4.52	3	9.3%

Table 3: Multiple classifier results

Table 3 presents the results of applying the W_AVG and PRI_SEC methods. Both combinations in our schemes gave more accurate classifications compared to a single classifier. The W_AVG combining method had an SSE of 4.77 compared to 6.40 for the LF/HF single classifier, demonstrating an improvement of 25.5%. Similarly, the PRI_SEC method had an SSE of 4.52, demonstrating an improvement of 29.4%. While the PRI_SEC combining rule performed better than the W_AVG, both had the same CE classification errors (i.e. 3) as each other, as well as the LF/HF single classifier. The CE classification error is a measure of the final, crisp decision for the texture classification for our network and hence the SSE should be considered as a more suitable measure of the performance of the classifiers.

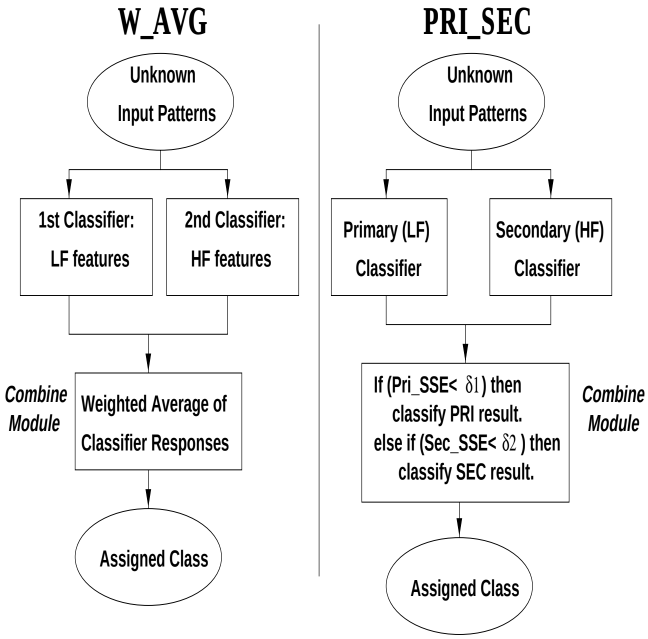

Figure 3: The W_AVG and PRI_SEC multiple classifier algorithms including their combine modules.

6 Conclusion

Our experiments have demonstrated the importance and practicality of high frequency features as an additional aid in texture analysis in *high resolution* images. High frequency features are not used in texture analysis when dealing with fairly low resolution images. In fact the frequencies used are generally low to medium frequencies. We have shown that the performance of a single classifier can increase considerably if both lower and high frequencies are used together. It is worth reminding the reader that the lower frequencies we refer to in this paper correspond to the highest frequencies present in lower resolution images.

Furthermore, we demonstrated that the performance achieved can be enhanced by using multiple classifiers. We examined two methods of combining classifier decisions, W_AVG and PRI_SEC, to decrease the overall single classifier error by 25.5% and 29.4% respectively. Multiple classifiers can also be used for further feature evaluation. Although here we expected more performance from our LF features [17], in other applications the relative performances of the various features may be not very clear. Thus, multiple classifiers can be exploited to determine the reliability and importance of different subsets of features in a particular pattern classification task. This in turn can be useful in refining or ranking of features according to their efficiency.

Acknowledgement

The first author, Mr. Amirhassan Monadjemi, is on a scholarship funded by the Iranian Ministry of Science, Research and Technology, (IMSRT).

References

[1] M. Everingham, B. Thomas, T. Troscianko, and D. Easty. Neural-network virtual reality mobility aid for the severely visually impaired. In *Proceedings of 2nd European Conference on Disability, Virtual Reality and Associated Technologies*, pages 183–192, 1998.

[2] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing Analysis and Machine Vision*. International Thomson Computer Press, 1996.

[3] O. Nestares, R. Navarro, J. Portilla, and A. Tabernero. Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *Journal of Electronic Imaging*, 7(1):166–173, 1998.

[4] J. Kittler and F. Roli. Foreword on multiple classifier systems. In *Multiple Classifier Systems*, pages 5–6. Springer-Verlag, 2000.

[5] W. Wang, P. Jones, and D. Partridge. Diversity between neural networks and decision trees for building multiple classifier systems. In *Multiple Classifier Systems*, pages 240–249. Springer-Verlag, 2000.

[6] T. Legault. What is a MTF curve? http://perso.club-internet.fr/legault/mtf.html.

[7] S. Myint. Image texture analysis with high-resolution multi-spectral image data using Wavelet transforms. In *University Consortium for Geographic Information Science (UCGIS) Annual Assembly and Summer Retreat*, pages 121–130, 2000.

[8] W. Freeman and E. Pasztor. Markov networks for super-resolution. In *Proceedings of the 34th Annual Conference on Information Sciences and Systems (CISS 2000)*, pages 60–67. Department of Electrical Engineering, Princeton University, 2000.

[9] T. Ho. Complexity of classification problems and comparative advantages of combined classifiers. In *Multiple Classifier Systems*, pages 97–106. Springer-Verlag, 2000.

[10] R. Cappelli, D. Maio, and D. Maltoni. Combining fingerprint classifiers. In *Multiple Classifier Systems*, pages 351–361. Springer-Verlag, 2000.

[11] X. Jiang, K. Yu, and H. Bunke. Classifier combination dor grammar-guided sentence recognition. In *Multiple Classifier Systems*, pages 383–392. Springer-Verlag, 2000.

[12] W. Wan and D. Fraser. A multiple self-organizing mao scheme for remote sensing. In *Multiple Classifier Systems*, pages 300–309. Springer-Verlag, 2000.

[13] J. Daugman. Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustic Speech and Signal Processing*, 36(7):1169–1179, 1988.

[14] I. Fogel and D. Sagi. Gabor filters as texture discriminator. *Biological Cybernetics*, 61:102–113, 1989.

[15] A. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[16] M. Mirmehdi and R. Perissamy. CBIR with perceptual region features. In T Cootes and C Taylor, editors, *Proceedings of the 12th British Machine Vision Conference*, pages 511–520. BMVA Press, September 2001.

[17] A. Monadjemi, B. Thomas, and M. Mirmehdi. Experiments on high resolution images towards outdoor scene classification. In *Proceedings of the 7th Computer Vision Winter Workshop, CVWW 2002*, volume 1, pages 325–334, 2002.