Text and Picture Segmentation by the Distribution Analysis of Wavelet Coefficients *

Jia Li
EE Department
Stanford Univ., CA 94305
jiali@isl.stanford.edu

Robert M. Gray
EE Department
Stanford Univ., CA 94305
rmgray@stanford.edu

Abstract

This paper presents an algorithm to segment text and picture in an image using two features based on the statistical distribution of the wavelet coefficients in high frequency bands. The algorithm breaks the image into blocks and classifies every block as background, text or picture according to the two features. The block size is variable so that the segmentation can be accurate at the boundary of two types and avoids misclassifying due to over-localized region analysis.

1 Introduction

Statistical classification is an important topic in image processing. Classification, which helps to interpret an image, can also be incorporated with other image processing to improve performance. One well-known example is image compression. For training-based image compression algorithms, such as vector quantization [1], a codebook is optimally designed under the assumption that the data to be quantized is statistically consistent with the training data. Hence, different quantizers are required for different types of data. If source data is a mixture of several statistical types, classification is often applied before quantization to decide which quantizer to use [1].

A particularly interesting type of classification is the segmentation of pictures and text. By pictures, we mean continuous-tone images such as photographs. By text, we mean normal text, tables and graphs. One application is the World Wide Web. With the exponentially increasing popularity of the World Wide Web, web information is handled more and more frequently. Since most web pages are a mixture of text and pictures, the segmentation of the two is preferred in many kinds of processing.

*This work was supported by the National Science Foundation under NSF Grant No. MIP-931190 and by a gift from Hewlett-Packard, Inc.

This paper presents an efficient algorithm to segment text and pictures using wavelet transforms [2]. Wavelet transforms have played important roles in classification of texture [3, 4] and abnormalities in medical images [5, 6]. An application of wavelet transforms is the formation of classification features by the statistics of wavelet coefficients. The moments of wavelet coefficients are the most commonly used [3, 4, 5, 6]. In our paper, however, we pay direct attention to the distribution pattern of wavelet coefficients and define features depending on the shape of the histogram of wavelet coefficients. The classification is block-based, i.e., an image is broken into blocks and every block is classified separately. The block size is variable and content-based, however, so that the segmentation can be accurate at the boundaries of two types and avoid misclassification due to over-localized region analysis.

In section 2, we define the two features used in our classification. Section 3 presents the segmentation algorithm and section 4 shows the result.

2 Classification Features

It has been observed that, for pictures, the wavelet coefficients in the high frequency bands, i.e., LH, HL and HH bands [7], tend to follow a Laplacian distribution. Although this approximation is controversial in some applications, it will be seen to work quite well as a means of distinguishing continuous tone images from text by means of the goodness of fit. As an example, the histograms of coefficients in the LH band for one picture image and one text image are plotted in Fig. 1. Another important difference to note is the continuity of the observed distribution. The histogram of the coefficients of the text image suggests that the values are concentrated on a few discrete values, but the histogram for the picture image shows much better continuity of distribution. However, we point out that in practice the histograms of the two types are

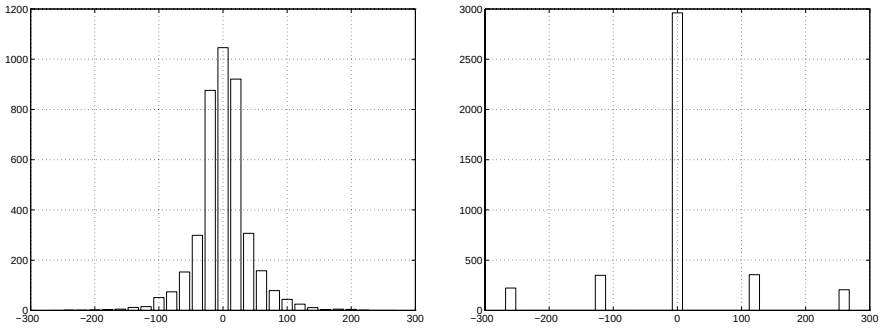

Figure 1: The histograms for the wavelet coefficients in LH band. left: picture image, right: text image.

usually not as clean as those of the examples shown here, especially for the text class. Since text class includes artificial graphs, the perfect concentration at a few values, which appeared in Fig. 1, is often not present.

To measure the goodness of match between the observed distribution and the Laplacian distribution, we use the χ^2 test [8], normalized by the sample size N, denoted by $\bar{\chi}^2$. Suppose that the data range is divided into k bins (as shown in Fig. 1), f_i is the relative frequency of bin i, and F_i is the probability of bin i according to the Laplacian distribution, then we can calculate $\bar{\chi}^2$ by

$$\bar{\chi}^2 = \chi^2/N = \sum_{i=1}^{k}(f_i - F_i)^2/F_i \quad .$$

We then define a criterion, denoted by L, to measure the likelihood of the wavelet coefficients being composed by highly concentrated values. For example, the right panel of Fig. 1 shows that the data only lies in the five far apart categories, indicating a high L. The efficiency of L estimating the likelihood of the wavelet coefficients having a highly discrete distribution depends on how concentrated peak values can be found, and how robust the identification of these values is to local fluctuations.

To calculate L, we partition the data range into a set of 'zones'. The histogram in a zone should have a concentrated peak value and vanish towards the ends of the zone. Omitting the slight difference at the ends of the data range, we specify a zone $[t_i, t_{i+1}], i = 0, 1, ..., r-1$ as follows:

1. $\exists t^* \in (t_i, t_{i+1})$, s.t., $h(t^*)$ is the maximum value of the histogram for $t \in [t_i, t_{i+1}]$, where $h(t)$ denotes the histogram.
2. $h(t_i)/h(t^*) < \delta$, and $h(t_{i+1})/h(t^*) < \delta$, where δ is a threshold.
3. $h(t_i) \leq h(t)$, for all $t \in (t_i, t^*)$; $h(t_{i+1}) < h(t)$, for all $t \in (t^*, t_{i+1})$
4. There does not exist $[\tau, \tau'] \subset [t_i, t_{i+1}]$ and $[\tau, \tau'] \neq [t_i, t_{i+1}]$, so that $[\tau, \tau']$ satisfies the above three conditions.

For each zone $[t_i, t_{i+1}], i = 0, 1, ..., r-1$, we define a concentration level β_i. A high value of β_i indicates that the data in this zone are tightly distributed around the peak value t^*. We evaluate β_i by the percentage of the data in a narrow neighborhood of t^* based on the total number of data in the zone. Then, we define L as a weighted sum of the concentration levels β_i, where the weight is the probability of the data lying in a zone.

Although more sophisticated classification methods, e.g., CART [9], can be applied to two dimension data $\bar{\chi}^2$ and L, for simplicity and speed, we combine the two features into one function, denoted by $\Delta(\bar{\chi}^2, L)$, and we classify based only on this function. The definition of $\Delta(\bar{\chi}^2, L)$ is a sum of two nondecreasing piecewise linear functions depending on $\bar{\chi}^2$ and L respectively, i.e.,

$$\Delta(\bar{\chi}^2, L) \quad = \quad \frac{\Delta_1(\bar{\chi}^2) + \Delta_2(L)}{2}$$

where

$$\Delta_1(\bar{\chi}^2) \quad = \quad \begin{cases} 0 & \bar{\chi}^2 < \theta_1 \\ (\bar{\chi}^2 - \theta_1) \cdot \frac{1}{\theta_2 - \theta 1} & \theta_1 \leq \bar{\chi}^2 < \theta_2 \\ 1 & \bar{\chi}^2 \geq \theta_2 \end{cases}$$

and

$$\Delta_2(L) \quad = \quad \begin{cases} 0 & L < \theta_1' \\ (L - \theta_1') \cdot \frac{1}{\theta_2' - \theta_1'} & \theta_1' \leq L < \theta_2' \\ 1 & L \geq \theta_2' \end{cases}$$

The parameters $\theta_1, \theta_2, \theta_1'$ and θ_2' are chosen according to the statistics of the data. However, experiments showed that the classification result is not very sensitive to these parameters.

3 The Algorithm

In our segmentation algorithm, we have three classes: background, text, and picture. The classification is block-based, i.e., the image is divided into blocks and every block is identified as one class. However, the block size is variable in the process of classification so that accurate segmentation can be achieved at the boundaries of several types and misclassification due to over-localizing can be avoided.

We transform the blocks using a Haar wavelet to one level and use the wavelet coefficients in the LH and HL bands to classify the block. If the variance of the data is zero, the block is classified as background. Otherwise, $\Delta(\bar{\chi}^2, L)$ is evaluated. If $\Delta(\bar{\chi}^2, L)$ approaches extreme values, i.e., it strongly suggests the block is either pure text or pure picture, then the block is classified to the specific class. Otherwise, the block is marked as undetermined. It is then subdivided into four subblocks and classification is done for the subblocks separately. If the feature value of a subblock strongly suggests the type of the subblock, it will be classified. It is highly possible that the subblock has no sharp feature values because of its small sample size, even if it contains a single class. We then make use of context information to help classifying such subblocks. In context based classification, we first check the image types of the surrounding blocks. If the block has adjacent text and picture blocks, i.e., the block is at the boundary of two types, the classification can not be helped by merging surrounding blocks since the merged big block will contain mixed types. In this case, we will classify only according to the subblock's feature value. The feature space is divided into two regions, text and picture, with no undetermined region. If the block does not have adjacent text and picture blocks simultaneously, it is reasonable to guess that the cause of the ambiguous feature is small sample size. Consequently, we merge the surrounding blocks to form a big block and classify based on the big block. The class of the merged block is taken as the class of the undetermined block.

The algorithm is presented in the list below.

1. (a) Transform the image by the Haar wavelet.
   (b) Divide the image into square blocks of size 32×32. For every block $B_j, j = 1, ..., J$, form the sequence of data $\mathbf{v}_j = \{x_1, ..., x_{16\times16\times2}\}$, where $x_1, ..., x_{16\times16}$ are the wavelet coefficients for this block in the LH band and $x_{16\times16+1}, ..., x_{16\times16\times2}$ are the wavelet coefficients for this block in the HL band.
   (c) Let $1 \to j$.
2. For block B_j, calculate the variance σ_j^2 of the data sequence.
3. If $\sigma_j^2 = 0$,
   classify the block B_j as background, $j+1 \to j$, go back to 2.
   Otherwise:
   (a) Calculate $\Delta_j(\bar{\chi}^2, L)$ for $\mathbf{v}_j$
   (b) If $\Delta_j(\bar{\chi}^2, L) < T_1$, where T_1 is a chosen threshold, classify the block B_j as picture, go to 3e.
   (c) If $\Delta_j(\bar{\chi}^2, L) > T_2$, where T_2 is a chosen threshold, classify the block B_j as text, go to 3e.
   (d) If $T_1 < \Delta_j(\bar{\chi}^2, L) < T_2$
      i. Divide block B_j into four equal size subblocks. Then the sequence of data $\mathbf{v}_j$ is divided into four subsequences: $\mathbf{v}_{j,1}$, $\mathbf{v}_{j,2}$, $\mathbf{v}_{j,3}$, and $\mathbf{v}_{j,4}$. Every subsequence is a sequence of the wavelet coefficients for one of four subblocks.
      ii. For every subblock, calculate $\Delta_{j,l}(\bar{\chi}^2, L), l = 1, ..., 4$.
      If $\Delta_{j,l}(\bar{\chi}^2, L) < T_1'$, classify the subblock as picture, go to 3e.
      If $\Delta_{j,l}(\bar{\chi}^2, L) > T_2'$, classify the subblock as text, go to 3e.
      If $T_1' < \Delta_{j,l}(\bar{\chi}^2, L) < T_2'$
         A. If B_j has both adjacent text blocks and adjacent picture blocks, classify $B_{j,l}$ as picture if $\Delta_{j,l}(\bar{\chi}^2, L) < T_0'$, and classify $B_{j,l}$ as text otherwise, go to 3e.
         B. Form a larger block $\bar{B}_j$ by merging the 8 blocks surrounding B_j. Denote the surrounding eight blocks as $B_{j_k}, k = 1, ..., 8$.
         C. If B_{j_k} is not classified as background, it is merged into $\bar{B}_j$; otherwise, it is not merged into $\bar{B}_j$. Use the wavelet coefficients of $\bar{B}_j$ to form a larger sequence of data $\bar{\mathbf{v}}_j$. In fact, $\bar{\mathbf{v}}_j$ is the combination of $\mathbf{v}_j$ and $\mathbf{v}_{j_k}$, where B_{j_k} is merged into $\bar{B}_j$.
         D. Calculate $\bar{\Delta}_j(\bar{\chi}^2, L)$ for $\bar{\mathbf{v}}_j$.
         If $\bar{\Delta}_j(\bar{\chi}^2, L) < T_0$, the subblock $B_{j,l}$ is classified as picture; otherwise, the subblock $B_{j,l}$ is classified as text.
   (e) Set $j+1 \to j$, go back to 2.

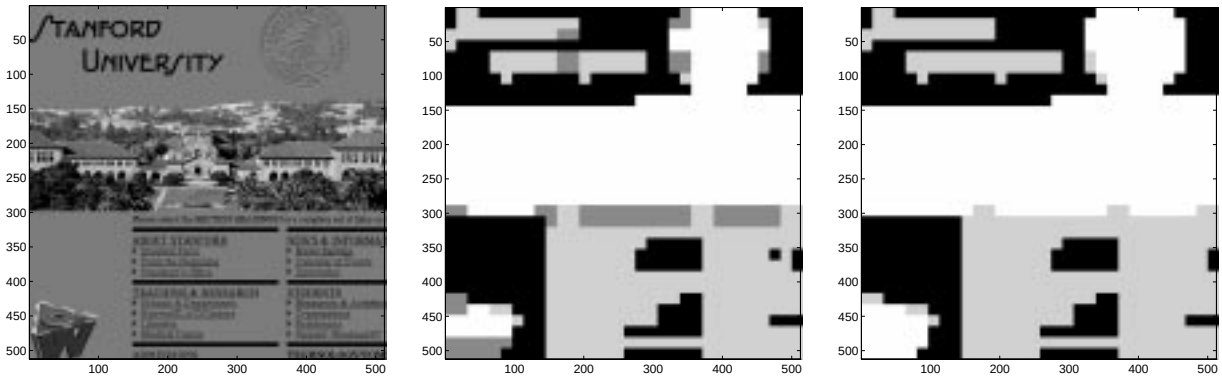

Figure 2: One sample image and its classification results. The image in the middle is the intermediate result with undetermined class blocks, and the image on the right is the final result. White: picture, light gray: text, dark gray: undetermined, black: background.

Image ID	1	2	3	4	5	6
Pe (%)	0.49	0.76	2.69	0.28	0.15	0.15

Table 1: Ratios of classification errors with respect to human labeling for 6 sample images

The parameters T_1, T_1', T_2, T_2', and T_0, T_0' are thresholds to be chosen. In our algorithm, we chose the following values: $T_1 = T_1' = 0.15, T_2 = T_2' = 0.85, T_0 = T_0' = 0.6$.

4 Results

In this section, we give a classification example. The original image and the classification results are shown in Fig. 2. The classified image in the middle of Fig. 2 shows the intermediate result of the algorithm, i.e., the result before undetermined blocks being subdivided or merged with surrounding blocks. Thus, there are four classes in the image: background, text, picture, and undetermined class. As we can see, the undetermined blocks are the mixtures of multiple types, located at the boundaries of several types. The blocks containing a single class are correctly classified as background, text or picture. This media result shows that the feature function is efficient in distinguishing text from picture and detecting the mixtures of the two types, a critical property for accurate segmentation at the boundary of the two types. The classified image on the right side of Fig. 2 shows the final result of the algorithm. Clearly shown in Fig. 2, the final result provides a good classification for the image. As the starting block size is significantly large, we avoid misclassification inside the region of any type. The large block size also makes the algorithm fast. At the boundary area, as smaller block sizes are used, the classification accuracy is not limited by the original block size.

We also applied the algorithm on a set of 6 images with size around 1650×1275. By comparing the results with the human classified images, we got the misclassification rates shown in Table 1. As we can see, except for one image with error rate 2.69%, all the other images are classified with error rates lower than 1%.

References

[1] Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.

[2] O. Rioul and M. Vetterli, "Wavelets and Signal Processing," *IEEE Signal Processing Magazine,* vol. 8, no. 4, pp. 14-38, Oct 1991.

[3] Michael Unser, "Texture Classification and Segmentation Using Wavelet Frames," *IEEE Transactions on Image Processing,* vol. 4, no. 11, pp. 1549-1560, Nov. 1995.

[4] Loum, G., Provent, P., Lemoine, J. and Petit, E., "A New Method for Texture Classification Based on Wavelet Transform," *Proceedings of Third International Symposium on Time-Frequency and Time-Scale Analysis,* pp. 29-32, June 1996.

[5] A.P. Dhawan, Y. Chitre, C. Kaiser-Bonasso and M. Moskowitz, "Analysis of Mammographic Microcalcifications Using Gray-level Image Structure

Features," *IEEE Transactions on Medical Imaging,* vol. 15, no. 3, pp. 246-259, June 1996.

[6] Weaver, J.B., Healy, D.M., Jr., Nagy, H., Poplack, S.P., Lu, J., Sauerland, T. and Langdon, D., "Classification of Masses in Digitized Mammograms with Features in the Wavelet Transform Domain," *Proceedings of the SPIE - Wavelet Applications,* vol. 2242, pp. 704-710, April 1994.

[7] Martin Vetterli and Jelena Kovacevic, *Wavelets and Subband Coding*, chapter 7, Prentice-Hall Inc., 1995.

[8] G. W. Snedecor and W. G. Cochran, *Statistical Methods,* Iowa State University Press, Ames, Iowa, 1989.

[9] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees,* Wadsworth, 1983.