Multiresolution Hadamard Representation and Its Applications to Document Image Analysis

Kung-Hao Liang, Fu Chang, Tzu-Ming Tan, and Wen-Liang Hwang

Institute of Information Science,

Academia Sinica, Nankang,

Taipei, TAIWAN, R.O.C.

Correspondence to K. H. Liang, Fax: 886-2-2782-4814, E-mail: khliang@iis.sinica.edu.tw

Abstract

Document image analysis is a critical technology in the construction of modern digital libraries. In document image analysis, *a priori* knowledge of the structure of signals (i.e. strokes) is required. In this paper, a special class of the wavelet transform, referred to as the multiresolution Hadamard representation (MHR), is proposed. The basis of MHR provides a stroke model, by which the strokes in a Chinese character can be extracted easily. This multiresolution analysis provides a means to reveal the size/scale of a character. Applications on the extraction of half-tone pictures, the determination of scales for characters and the dynamic binarization of images using the multiresolution Hadamard representation are also presented.

1 Introduction

Document image analysis (DIA) comprises techniques that extract useful information from image form of documents. It is a well-known fact that documents contain objects (e.g. characters) that can vary in size but look similar in shape. Moreover, objects of the same size are usually grouped together in well-segmented blocks. Things of such nature are certainly ideal for multiresolution techniques to apply [1, 2, 3, 5].

A choice of kernels (basis) is essential to multiresolution analysis [5]. To suite for the goal of this paper, we propose a set of kernels, consisting of two functions that are associated with Hadamard transforms (Figure 1). Both of them are useful for expressing certain aspects of objects that have approximately the same width.

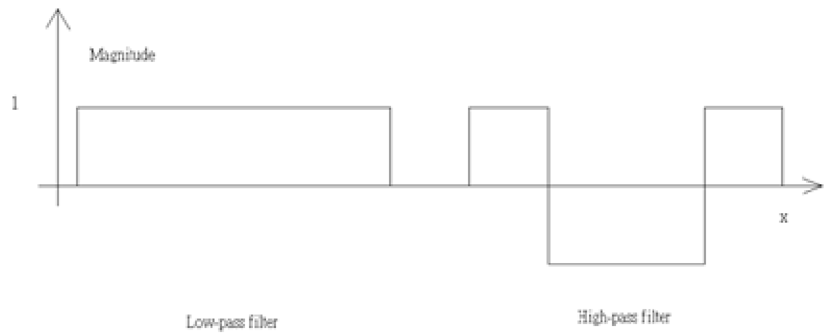


Figure 1: The kernel functions proposed in this article, including a low-passed filter (left) and a high-passed filter (right).

It has been a common practice in document analysis to extract features (e.g. run length) that are associated with the width of objects. However, many such practices assume that the processed images are binary (black/white). Hadamard kernels, on the other hand, can be directly applied to gray-scaled images. Images of such nature preserve a lot more information than the derived black/white images and serve as better source data for various operations to work on [4, 6]. All applications in this article are assumed to deal with gray-scaled images.

2 Multiresolution Hadamard Representation

In this article, multiresolution Hadamard representation (MHR) is proposed as a general framework for DIA. It is based on 2-D dyadic wavelet representation, where a low-pass filter LF and a high-pass filter HF are applied, in both horizontal and vertical directions, to the transformed objects obtained at previous level [2]. The results comprise 4 different channels for each level of transform: low-passed (L), horizontal (H), vertical (V) and diagonal (D) channels (see Figure 2a). They are defined recursively as follows.

$$L_{j+1} = L_j * LF_{(horizontal)} * LF_{(vertical)},$$

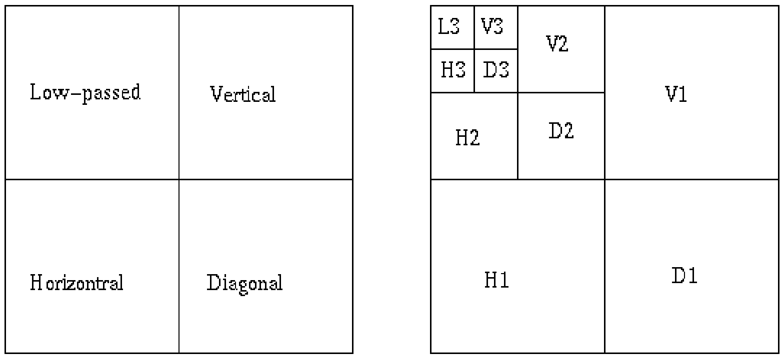
$$H_{j+1} = L_j * LF_{(horizontal)} * HF_{(vertical)},$$

$$V_{j+1} = L_j * HF_{(horizontal)} * LF_{(vertical)},$$

$$D_{j+1} = L_j * HF_{(horizontal)} * HF_{(vertical)},$$

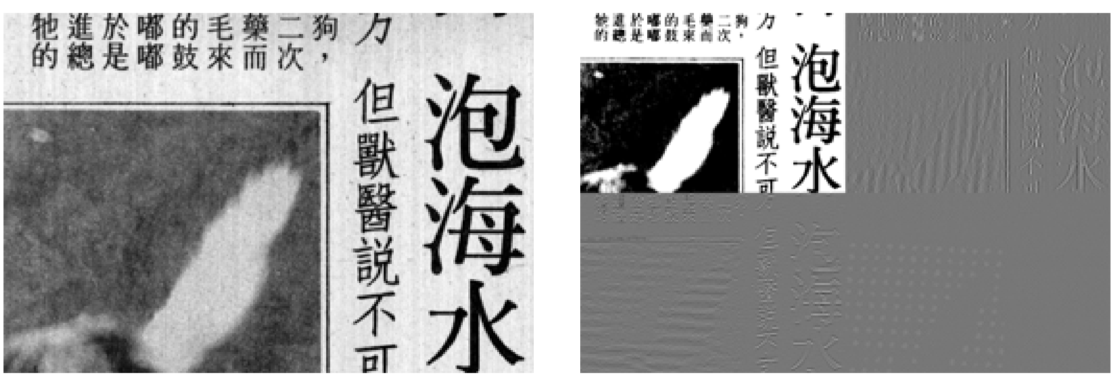
where subscript j indicates the level of transform, and * is the sign for convolution. Note that when $j = 0$, L_j denotes the original image. On the other hand, when LF and HF filters are applied to the j^{th} level, they are scaled by the factor 2^j. Figure 2b shows the example of a three-level transform.

In MHR, two Hadamard coefficients [1,1,1,1] and [1, -1, -1, 1] serve as LF and HF filter respectively (see Figure 1). In the H and V channels, LF and HF are applied alternatively but in the direction orthogonal to each other. Consequently, a vertical stroke would react strongly in the vertical channel and a horizontal stroke in the horizontal channel. A diagonal stroke would react in both channels. Its reaction in each channel is less vigorous than horizontal or vertical strokes in their corresponding channel, but the combined strength in the two channels is still comparable to either of them (see the upper right and lower left part of Figure 3b).



(a) (b)

Figure 2: (a) The four channels of a one-level wavelet transform. (b) Three-level wavelet transform.



(a) (b)

Figure 3: (a) A newspaper article. (b) The output of wavelet transform at level j = 1.

3 Applications

We use the following three applications to demonstrate the effectiveness of MHR.

3.1 Extraction of half-tone pictures

While characters show up their strengths in the H and V channels, we observe that half-tone pictures react strongly in the D channel. Half-tone pictures (as contained in Figure 3a) are generally produced in newspapers in such a way that the tone of the pictures is produced by small black dots with regular spacing between them. In the lower right part of Figure 3b, we see that a significant regular pattern appears in the location of the picture. This pattern, known as the Moire pattern [7], is certainly a good clue for the existence of half-tone pictures.

3.2 Scale determination for characters

Characters of different widths differentiate themselves by the reaction strengths at different levels of transform. The general rule being that: a character of width W has strongest response at the level j whose canonical size ($= 2^j$) is closest to W among all other levels. We refer to this 2^j as the scale of the character. Scale can be determined by comparing the measure $abs(Hj - Vj)$ among all j, where $abs(\cdot)$ stands for the absolute magnitude. We take the difference, rather than the sum, of Hj and Vj, so as to avoid the possibility of confusing a horizontal (vertical) stroke with a very thick vertical (horizontal) stroke. Figure 3a comprises characters with 3 different sizes. The normalized measures of these characters of different sizes are shown in Table 1. The results demonstrate that the maximum measure of a particular character size, represented by the boldface, occurs at the appropriate level of the transform.

Table 1: The transformed values at different levels for characters of different sizes.

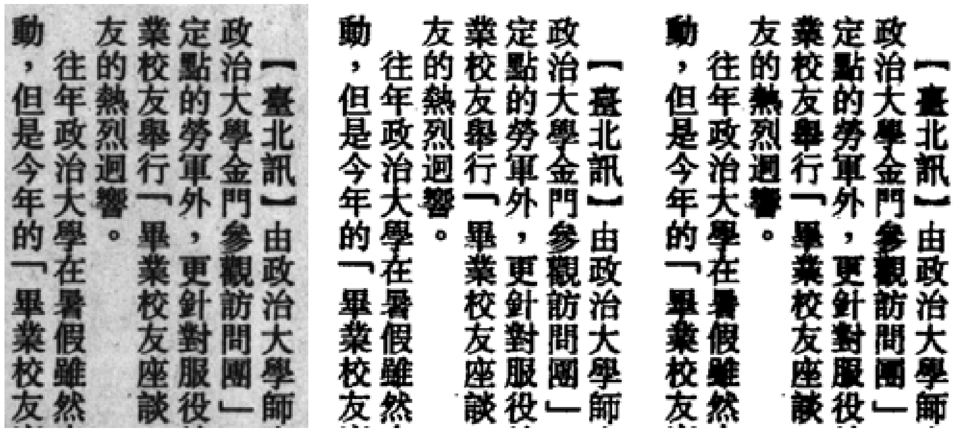
	Level 1	Level 2	Level 3
Large characters	2.75	3.55	**5.62**
Medium characters	3.40	**4.80**	4.37
Small characters	**5.25**	4.93	3.28

3.3 Window-based dynamic binarization

Scale information, among other things, is very useful for window-based dynamic binarization. The idea is the following. For each pixel P locating within or near a character, we set a window centered at P, whose size is determined by the width, and therefore the scale, of the given character. The important observation being that, within such a window, there is a fairly good mixture of background and foreground pixels. By taking advantage of this fact, we can make accurate determination of the (dark/bright) tone of pixel P. The result derived by the technique that implements the above idea is shown in Figure 4b. We also compare it with a binary image obtained by a static thresholding method (Figure 4c). In a separate study, we fed both images to a commercial character-recognition software and found that the former image gains 20 precent higher recognition rate than the latter.

4 Conclusions

In this paper, the multiresolution Hadamard representation is proposed as a general framework for document image analysis. The basis of MHR is designed as a stroke model to extract strokes in several scales. The result shows that a high-quality binary document image can be obtained using this representation. In the future, the potential of Multiresolution Hadamard Representation on



(a) (b) (c)

Figure 4: (a) Gray-scaled image. (b) Binary image derived from window-based dynamic binarization method. (c) Binary image obtained by a static thresholding method.

the task of optical character recognition will be investigated.

References

[1] K. Etemad, D. Doermann and R. Chellappa, "Multiscale segmentation of unstructured document pages using soft decision integration," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-19, no.1, pp.92-96, 1997.

[2] S. G. Mallat, "A theory of multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-11, no.7, pp.674-693, 1989.

[3] M. D. Swanson and A. H. Tewfik, "A binary wavelet decomposition of binary images," *IEEE Trans. Image Process.,* vol. IP-5, no.12, pp.1637-1650, 1996.

[4] Y. Y. Tang, H. Ma, J. Liu, B. Fa and D. Xi, "Multiresolution analysis in extraction of reference lines from documents with gray level background," *IEEE Trans. Patt. Anal. Machine Intell.,*

vol. PAMI-19, no.8, pp.921-926, 1997.

[5] M.Vetterli and J.Kovacevic, *Wavwlets and subband coding,* Prentice-Hall, New Jersey, 1995.

[6] L. Wang and T. Pavlidis, "Direct gray-scale extraction of features for character recognition," *IEEE Trans. Patt. Anal. Machine Intell.,* vol. PAMI-15, no.10, pp.1053-1067, 1993.

[7] J. C. Y. Yang and W. H. Tsai, "Document image segmentation and quality improvement by Moire pattern analysis," in *Proc. 7th Optical character recognition and document analysis workshop,* Academia Sinica, Taiwan, pp. 307-318, December, 1997.