Journal of Southeast University (English Edition) Vol. 30, No. 1, pp. 27 − 31 Mar. 2014 ISSN 1003—7985

Wavelet transform and gradient direction based feature extraction method for off-line handwritten Tibetan letter recognition

Huang Heming[1,2] Da Feipeng[1] Han Xiaoxu[3]

([1] School of Automation, Southeast University, Nanjing 210096, China)
([2] School of Computer Science, Qinghai Normal University, Xining 810008, China)
([3] Department of Computer and Information Science, Fordham University, New York 10458, USA)

Abstract: To improve the recognition accuracy of off-line handwritten Tibetan characters, the local gradient direction histograms based on the wavelet transform are proposed as the recognition features. First, for a Tibetan character sample image, the first level approximation component of the Haar wavelet transform is calculated. Secondly, the approximation component is partitioned into several equal-sized zones. Finally, the gradient direction histograms of each zone are calculated, and the local direction histograms of the approximation component are considered as the features of the character sample image. The proposed method is tested on the recently developed off-line Tibetan handwritten character sample database. The experimental results demonstrate the effectiveness and efficiency of the proposed feature extraction method. Furthermore, compared with the detail components, the approximation component contributes more to the recognition accuracy.

Key words: pattern recognition; wavelet transform; gradient direction; Tibetan; handwritten character

doi: 10.3969/j.issn.1003 −7985.2014.01.006

The study of English and Chinese character recognition has a history of more than half a century and many effective techniques have been developed for such recognition stages as pre-processing, feature extraction, classification, and post-processing. However, the study on Tibetan character recognition has just been begun since the last decade. The recognition of printed Tibetan characters has been studied first; the reason may be that it is comparatively easier[1-4]. Recently, the recognition of on-line handwritten Tibetan characters has been studied by a few researchers[5-6]. As far as off-line handwritten Tibetan character recognition is concerned, a database of off-line handwritten Tibetan character samples, THCDB, is introduced by Huang et al[7]. A sparse representation-based classification algorithm is also proposed by them[8].

The main contribution of this paper is that the gradient direction histograms based on the wavelet transform are proposed as the features of off-line handwritten Tibetan character recognition. Concretely, to a character image, the first level approximation component of the Haar wavelet transform is calculated first; and then, it is partitioned into several equal-sized zones; thirdly, the local gradient direction histograms of each zone are calculated; and finally, all the histograms are considered as the features of the character sample image. The experimental results on the THCDB show that the proposed method of combining the wavelet transform and the gradient direction histogram leads to a competitive feature extractor.

1 Database and Pre-Processing

We confine ourselves to the recognition of the 30 Tibetan consonants, namely, ཀ, ཁ, ག, ང, ཅ, ཆ, ཇ, ཉ, ཏ, ཐ, ད, ན, པ, ཕ, བ, མ, ཙ, ཚ, ཛ, ཝ, ཞ, ཟ, འ, ཡ, ར, ལ, ཤ, ས, ཧ, and ཨ. All samples of these consonants come from the THCDB sample database, and the sample number of each character class ranges from 264 to 318. To each character class, the 80% samples are used for training while the remaining 20% samples are used for testing.

To each sample image for either training or testing, the necessary pre-processes such as segmentation, size normalization, de-noising, and orientation correction are implemented to facilitate the feature extraction process and improve the classification accuracy. More details about these pre-processes can be seen in Ref. [7].

2 Proposed Feature Extraction Method

To a character image, a single level wavelet transform produces one approximation component and three detail components. Among these, the approximation component has the advantage of conserving the main information of the original image. In addition, the local gradient direction histograms of an image have the advantages of simple implementation and invariance to the stroke-width variation. The local gradient direction histograms of the approximation component make full use of the advantages of both aspects and, therefore, they are proposed as the features of the Tibetan character sample image.

Received 2013-08-01.

Biographies: Huang Heming (1969—), male, graduate; Da Feipeng (corresponding author), male, doctor, professor, dafp@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 60963016), the National Social Science Foundation of China (No. 17BXW037).

Citation: Huang Heming, Da Feipeng, Han Xiaoxu. Wavelet transform and gradient direction based feature extraction method for off-line handwritten Tibetan letter recognition[J]. Journal of Southeast University (English Edition), 2014, 30(1): 27 −31. [doi: 10.3969/j.issn.1003 − 7985.2014.01.006]

2.1 Calculation of the first level approximation component

Wavelet transform is a powerful technique in many areas[9]. For a given image A_0, as shown in Fig. 1, a single level wavelet transform produces one approximation component A_1 and three detail components H_1, V_1, and D_1. The approximation component keeps the low frequency information of the original image while the three detail components reflect the high frequency information of the original image in horizontal, vertical, and diagonal directions, respectively.

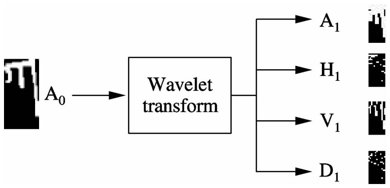
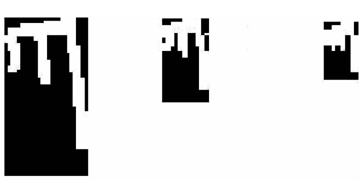


Fig. 1 The first level wavelet transform of a sample of handwritten Tibetan letter ཀ

There are many different wavelet families, such as Daubechies wavelets, Mexican hat wavelets, and Morlet wavelets. Among these, the Haar wavelet, a special case of Daubechies wavelets, has such advantages as conceptual simplicity, high speed, and memory efficiency[10]. Furthermore, the disadvantage of the Haar wavelet is that it is not continuous, and, therefore, not differentiable. However, this property becomes an advantage for the analysis of signals with sudden transitions such as character images that have many sharp edges.

To handwritten Tibetan character samples, the first level approximation component of the Haar wavelet transform conserves the main information of the original image while the higher level approximation component degenerates severely. Fig. 2 shows the top two level approximation components of letters ཀ and ད, respectively. It can be seen that the second level approximation components (see the third images of Fig. 2(a) and (b)) become so vague that it is difficult to identify them. Therefore, in our system, the feature extraction is based on the first level approximation component.



(a)

(b)

Fig. 2 Top two level approximation components of Tibetan letters ཀ and ད produced by Haar wavelet transform. (a) A sample image of letter ཀ and its first two level approximation components; (b) A sample image of letter ད and its first two level approximation components

2.2 Partitioning approximation component into equal-sized zones

Characters contain strokes, and the directions of the strokes have significant effects on the distinguishing of various characters. For a long time, stroke direction has been considered in the stroke analysis of character recognition.

For a statistical recognition based on feature vector representation, character samples are represented as the vectors of direction statistics. To realize this, the stroke direction angle is divided into a fixed number of ranges, and the number of stroke segments in each angle range is regarded as a feature value. The set of numbers of directional segments thus forms a histogram, called direction histogram. To enhance the differentiation ability, the histogram for the local zones of the character image is often calculated. In our experiments, the local direction histograms are calculated by decomposing the gradient vector at each pixel of the local zone to some standard directions.

2.3 Calculation of local gradient direction histogram

To an image $A(x, y)$, the gradient vector $(\partial A/\partial x, \partial A/\partial y)$ at pixel (x, y) is computed by

$$\left.\begin{aligned}\partial A/\partial x = {} & A(x+1, y-1) + 2A(x+1, y) + A(x+1, y+1) - \\ & A(x-1, y-1) - 2A(x-1, y) - A(x-1, y+1) \\ \partial A/\partial y = {} & A(x-1, y+1) + 2A(x, y+1) + A(x+1, y+1) - \\ & A(x-1, y-1) - 2A(x, y-1) - A(x+1, y-1)\end{aligned}\right\} \tag{1}$$

To calculate the features of image $A(x, y)$, a gradient vector is generally decomposed to some standard directions. Eight standard directions are usually specified and each of them is denoted as $A_i(x, y)$, $i = 0, 1, \ldots, 7$, respectively.

A gradient vector of arbitrary direction is decomposed into two constituents coinciding with the two adjacent standard directions. If we use l_1 and l_2 to represent the constituent lengths of two adjacent standard directions, the corresponding two direction planes are updated with $A_1(x, y) = A_1(x, y) + l_1$ and $A_2(x, y) = A_2(x, y) + l_2$, respectively. The direction planes are completed by separating the gradient vectors at all pixels. However, for the sake of recognition accuracy and computing efficiency, the eight direction planes are usually merged into four by $A_i(x, y) = A_i(x, y) + A_{i+4}(x, y)$, $i = 0, 1, 2, 3$. The local gradient direction histograms calculated in this way have two advantages: simple implementation and invariance to stroke-width variations[11].

The process of the proposed feature extraction method is shown in Fig. 3. To each sample (see Fig. 3(a)), the approximation component of a single level Haar wavelet transform is obtained first (see Fig. 3(b)). Then, it is partitioned into several equal-sized zones (see Fig. 3(c)). Thirdly, the local gradient direction histograms of all the zones are calculated (see Fig. 3(d)). Finally, all the histograms are considered as the feature values of the character image.

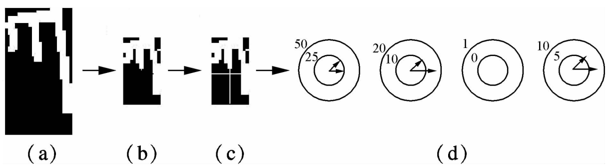


Fig. 3 Process of the proposed feature extraction method. (a) A sample image of Tibetan letter ཀ; (b) The first level approximation component produced by the Haar wavelet transform; (c) The approximation component partitioned into four equal-sized zones; (d) Four local gradient direction histograms corresponding to the four equal-sized zones

2.4 Analysis of feature vector dimension

If the width and height of a character image are denoted as w_0 and h_0, respectively, the width and height of the first level approximation component becomes $w_0/2$ and $h_0/2$, because of column-wise and row-wise down-sampling. Therefore, if the width and height of each zone are denoted as w_z and h_z, respectively, the total number of histograms is

$$4\frac{w_0/2}{w_z}\frac{h_0/2}{h_z}=\frac{w_0 h_0}{w_z h_z} \tag{2}$$

This number is also the dimension of the feature vector. It can be seen that the dimension of the feature vector is determined by the size of the sample image and that of the partitioned zone.

Since the sizes of all sample images are normalized to 48×24 in the pre-processing stage, the dimension of the feature vector is determined by

$$\frac{48\times24}{w_z h_z} \tag{3}$$

3 Modified Quadratic Discriminant Function

The quadratic discriminant function (QDF) is a popular statistical classifier and it is obtained under the assumption of the multivariate Gaussian density for each class. Up to the present, three versions of QDF, namely MQDF1, MQDF2, and MQDF3, have been developed. Comparing with the QDF, MQDF2 has been proved to be more effective due to its higher performance and less computation time. The MQDF2 is defined as

$$g_2(\boldsymbol{x},\omega_i)=\sum_{j=1}^{k}\frac{1}{\lambda_{ij}}[\boldsymbol{\varphi}_{ij}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu}_i)]^2+\frac{1}{\delta_i}r_i(\boldsymbol{x})+\sum_{j=1}^{k}\log\lambda_{ij}+(d-k)\log\delta_i \tag{4}$$

where λ_{ij} and $\boldsymbol{\varphi}_{ij}$ denote the j-th eigenvalue and its corresponding eigenvector of the covariance matrix $\boldsymbol{\Sigma}_i$, respectively; k denotes the number of principal eigenvalues; δ_i is a constant; $r_i(\boldsymbol{x})$ represents the residual of subspace projection.

Liu et al. improved the performance of the QDF in another way[11]. They combined the principle of regularized discriminant analysis (RDA) with MQDF2 by smoothing the covariance matrix of each class with the identity matrix, that is

$$\hat{\boldsymbol{\Sigma}}_i=(1-\gamma)\boldsymbol{\Sigma}_i+\gamma\sigma_i^2 \tag{5}$$

where $\sigma_i^2=\mathrm{tr}(\boldsymbol{\Sigma}_i)/d$ and $0<\gamma<1$. Furthermore, they replaced each minor eigenvalue with the average of all minor eigenvalues. The QDF modified in this way is abbreviated as MQDF3.

The MQDF3 has the advantages of high computation effectiveness and remarkable performance. Therefore, the MQDF3 is employed as the classification function of our recognition system.

Based on the abundant experiments, the number of principal eigenvalues k in Eq. (4) and the value of the regularization parameter γ in Eq. (5) are set to be 50 and 0.2, respectively.

4 Experiments

The implementation environment of our experiments is as follows. The processor is Intel Core 2 Duo CPU (E6550, 2.33 GHz), and the RAM is 2.00 GB DDR2. The operating system is MS XP professional SP3, and the programming platform is Matlab 2007a.

4.1 Evaluation of size of equal-sized zone

In this experiment, the optimal size of each equal-sized zone is evaluated by fixing k to 50. The recognition rates vs. the sizes of the zone are listed in Tab. 1. The dimensions of the feature vector that influence the recognition time are also listed in Tab. 1.

It can be seen that the best recognition rate 97.13% is reached when $w_z=2$ and $h_z=2$, and the average recognition time of a test sample is 0.137 0 s.

In a word, the optimal values for the parameters of our recognition system are that k is 50 and the size of each zone is 2×2. Under this circumstance, the dimension of

Tab. 1 Recognition rates vs. the sizes of equal-sized zones

Zone size	Dimension	Recognition rate/%	Time/ms
2 × 2	288	97.13	137.0
3 × 3	128	96.79	69.0
4 × 4	72	95.30	58.8
6 × 6	32	88.64	51.8
12 × 12	8	60.76	47.7
3 × 2	192	96.73	86.9
4 × 2	144	96.67	73.9
6 × 2	96	95.18	63.2
8 × 2	72	93.92	59.3
12 × 2	48	89.44	54.2
4 × 3	96	95.98	62.9
6 × 3	64	94.43	57.6
8 × 3	48	93.98	54.3
12 × 3	32	87.49	51.7
6 × 4	48	91.51	54.3
8 × 4	36	90.59	52.4
12 × 4	24	82.67	50.6
8 × 6	12	86.29	50.4

the feature vector is 288.

4.2 Comparison with the previous method

The recognition accuracy of the proposed feature extraction method is compared with that of the method in Ref. [9]. The features in Ref. [9] are extracted as follows. After the same pre-processing, a character image is directly partitioned into several equal-sized zones without wavelet transforms, and, for each zone, four local gradient direction histograms are calculated. There are totally $(48 \times 24)/(w_z \times h_z)$ values and they are considered as the features of each sample image.

The discriminant function MQDF3 is also employed for classification. Similarly, the size of equal-sized zones affects the recognition rate and the dimension of the feature vector potentially, as shown in Tab. 2. It can be seen that the optimal recognition rate 95.17% is reached as the size of a zone is 6 × 6, which is 1.96% higher than that of the method presented in Ref. [9].

Tab. 2 Recognition rates vs. the sizes of equal-sized zones

Zone size	Dimension	Recognition rate/%	Zone size	Dimension	Recognition rate/%
2 × 2	1152	93.75	4 × 3	384	94.43
3 × 3	512	94.25	6 × 3	256	94.76
4 × 4	288	94.84	8 × 3	192	93.98
6 × 6	128	95.17	12 × 3	128	91.51
3 × 2	768	94.15	6 × 4	192	93.92
4 × 2	576	93.92	8 × 4	144	89.44
6 × 2	384	93.92	8 × 6	96	93.92

4.3 Performance comparison for different components of wavelet transforms

The feature extraction method proposed in this paper is based on the approximation component that maintains the main information of the original image; the three detail components are also helpful to achieve good inter-class variance and distinguish similar characters since they reflect the high frequency information in horizontal, vertical, and diagonal directions, respectively. Therefore, the following experiments explore the contribution of detail components to the recognition accuracy.

For simplicity, let A, H, V, and D stand for the methods that calculate the local gradient direction histograms on the approximation component, horizontal detail component, vertical detail component, and diagonal detail component, respectively; A + H for the method that concatenate the feature vectors of methods A and H; the meaning of others, such as H + V + D, are similar.

The experiments are divided into three groups. The first group contains the methods that extract the features just from the detail component, i. e. the methods H, V, D, and H + V + D. The second group contains just method A, the proposed method. The third group contains the methods A + H, A + V, A + D, and A + H + V + D, which extract the features from both the approximation component and the detail component. For each method, the dimension of the feature vector, the recognition rate, and the recognition time are listed in Tab. 3.

Tab. 3 Feature vector dimension, recognition rate, and recognition time of the nine methods

Method	Dimension	Recognition rate/%	Time/ms
H	288	91.68	82.5
V	288	89.50	82.0
D	288	68.39	83.4
H + V + D	864	95.12	514.2
A	288	97.13	84.1
A + H	576	97.30	247.6
A + V	576	97.36	252.2
A + D	576	97.30	251.6
A + H + V + D	1152	97.36	723.5

The following two conclusions can be obtained from Tab. 3. First, the recognition rate of the proposed method is at least 2.01% higher than that of the methods of the first group, while the recognition time is roughly equal. Secondly, the recognition rate of the third group is at most 0.23% higher than that of the proposed method at the cost of three or more times the recognition time.

Overall, considering both accuracy and speed, the proposed feature extraction method is more powerful than those methods that are related to the detail components.

5 Conclusion

In this paper, the local gradient direction histograms based on the approximation component of the Haar wavelet transform are proposed as the features of a character image. With the proposed feature extraction method, a best recognition rate of 97.13% is reached for the recognition of off-line handwritten Tibetan consonants, which is 1.96% higher than that of the state-of-the-art method.

It demonstrates that the proposed method is effective.

Compared with the approximation component, the detail component contributes less in improving the recognition accuracy. In addition, the combination of the approximation component with one or more detail components improves the recognition rate slightly at the cost of too much more time. Therefore, considering both accuracy and speed, the proposed feature extraction method is more powerful than those methods based on the detail components.

References

[1] Huang H M, Da F P. General structure based collation of Tibetan syllables [J]. *Journal of Computational Information System*, 2010, **6**(5):1693 −1703.

[2] Wang H J, Zhao N Y, Deng G Y. A stroke segment extraction algorithm for Tibetan character recognition [J]. *Journal of Chinese Information Processing*, 2001, **15**(4): 41 −46. (in Chinese)

[3] Li Y Z, Wang Y L, Liu Z Z. Study on printed Tibetan character recognition technology [J]. *Journal of Nanjing University*: *Natural Sciences Edition*, 2012, **48**(1): 55 − 62. (in Chinese)

[4] Ngodrup, Zhao D C. Research on wooden blocked Tibetan character segmentation based on drop penetration algorithm [C]//*Chinese Conference on Pattern Recognition*. Chongqing, China, 2010: 84 −88.

[5] Liang B, Wang W L, Qian J J. Application of hidden Markov model in on-line recognition of handwritten Tibetan characters [J]. *Journal of Microelectronics and Computer*, 2009, **26**(4): 98 −101.

[6] Ma L L, Liu H D, Wu J. MRG-OHTC database for online handwritten Tibetan character recognition [C]//*International Conference on Document Analysis and Recognition*. Beijing, China, 2011: 207 −211.

[7] Huang H M, Da F P. A database for off-line handwritten Tibetan character recognition [J]. *Journal of Computational Information System*, 2012, **9**(18): 5987 −5993.

[8] Huang H M, Da F P. Sparse representation-based classification algorithm for optical Tibetan character recognition [J]. *Optik-International Journal for Light and Electron Optics*, 2014, **125**(3):1034 −1037.

[9] Huang H M, Da F P. Wavelet and moments based offline handwritten Tibetan character recognition[J]. *Journal of Information and Computational Science*, 2013, **10**(6): 1855 −1859.

[10] Raviraj P, Sanavallah M Y. The modified 2D-Haar wavelet transformation in image compression [J]. *Middle-East Journal of Scientific Research*, 2007, **2**(2): 73 −78.

[11] Liu C L. Normalization-cooperated gradient feature extraction for handwritten character recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(8):1465 −1469.

基于小波变换和梯度方向的脱机手写藏文字符特征提取方法

黄鹤鸣[1,2]　达飞鹏[1]　韩晓旭[3]

([1] 东南大学自动化学院,南京 210096)
([2] 青海师范大学计算机学院,西宁 810008)
([3] 福坦莫大学计算机与信息科学系,纽约 10458)

摘要:为了提高脱机手写藏文字符的识别效果,提出了一种在小波变换基础上计算局部梯度方向直方图的特征提取方法. 首先,对一个脱机手写藏文字符样本图像进行一次 Haar 小波变换,得到相应的一级近似分量;然后,将这个一级近似分量划分成几个等尺寸的子区域;最后,计算每个等尺寸子区域的局部梯度方向直方图, 并将所有子区域的全部局部梯度方向直方图的值作为该字符图片的特征. 在最近建立的脱机手写藏文字符样本数据库(THCDB)上的实验结果表明:提出的特征提取方法识别效率较高,且识别效果较好;和细节分量相比,近似分量对提高识别精度具有更大的贡献.

关键词:模式识别;小波变换;梯度方向;藏文;手写字符

中图分类号:TP391.4