# An Improved Method Based on Weighted Grid Micro-structure Feature for Text-independent Writer Recognition

Lu Xu, Xiaoqing Ding, Liangrui Peng, Xin Li
State Key Laboratory of Intelligent Technology and Systems
Department of Electronic Engineering, Tsinghua University
Beijing, P.R. China
e-mail: {xulu, dxq, plr, lixin}@ocrserv.ee.tsinghua.edu.cn

***Abstract*: Writer recognition is a very important branch of biometrics. In our previous research, a Grid Micro-structure Feature (GMSF) based text-independent and script-independent method was adopted and high performance was obtained. However, this method is sensitive to pen-width variation in practical situation. To solve this problem, an inner and inter class variances weighted high-dimensional feature matching method is proposed. The inner and inter class variances are estimated on handwriting samples with different pen-width written by different writers. Experimental results show that our method is effective.**

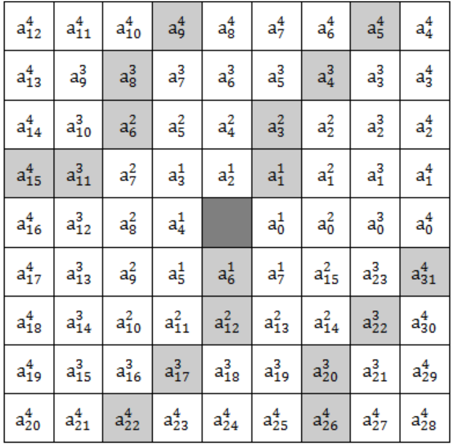
***Keywords: writer recognition; text-independent; Chinese handwriting; grid microstructure feature; pen-width; strike width; inter class variance; inner class variance.***

## I. INTRODUCTION

### A. Background and related work

Writer recognition is a very important branch of biometrics and has been developing in recent years. The current method in writer recognition or verification can be divided into two aspects: text-dependent and text-independent [1]. The text-dependent methods require the same writing scripts or characters beforehand, which is not quite similar with the actual free writing situation and the labeling is huge work. On the other hand, the text-independent method is applicable in free writing situations. The mostly used features of text-independent method [2] mainly rely on the lay-out of the passage, the character shape and geometric features, texture features and some other probability distribution function (PDF) features such as Contour-Hinge [3] and Grapheme Emission [4]. Nonetheless, these methods suffer from language constraints, algorithm complexity and other problems. Therefore some researchers are devoted to combine the two approaches and use both global features on whole texts and local features on single characters to get a higher performance [5]. The grid microstructure feature (GMSF), proposed by Li and Ding[6] is a multi-language text-independent method and is proved much more effective on eastern scripts.

In our previous work, the different writers' scripts are firstly preprocessed, only preserving the edge information. Then the edge image is covered by a $(2N+1)\times(2N+1)$ mask shown in Figure 1 and there are N layers in the mask. Each mask pixel is labeled as $a_i^L$, in which L represents the layer index, and i represents the pixel index in the L-th layer. According to this pattern, the L-th layer contains 8L pixels and L is no bigger than N. Then the mask center will traverse throughout all the edge points. (in experiment N=10, now we take N=4 as an illustration)



|  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|
| $a_{12}^4$ | $a_{11}^4$ | $a_{10}^4$ | $a_9^4$ | $a_8^4$ | $a_7^4$ | $a_6^4$ | $a_5^4$ | $a_4^4$ |
| $a_{13}^4$ | $a_9^3$ | $a_8^3$ | $a_7^3$ | $a_6^3$ | $a_5^3$ | $a_4^3$ | $a_3^3$ | $a_3^4$ |
| $a_{14}^4$ | $a_{10}^3$ | $a_6^2$ | $a_5^2$ | $a_4^2$ | $a_3^2$ | $a_2^2$ | $a_2^3$ | $a_2^4$ |
| $a_{15}^4$ | $a_{11}^3$ | $a_7^2$ | $a_3^1$ | $a_2^1$ | $a_1^1$ | $a_1^2$ | $a_1^3$ | $a_1^4$ |
| $a_{16}^4$ | $a_{12}^3$ | $a_8^2$ | $a_4^1$ | ■ | $a_0^1$ | $a_0^2$ | $a_0^3$ | $a_0^4$ |
| $a_{17}^4$ | $a_{13}^3$ | $a_9^2$ | $a_5^1$ | $a_6^1$ | $a_7^1$ | $a_{15}^2$ | $a_{23}^3$ | $a_{31}^4$ |
| $a_{18}^4$ | $a_{14}^3$ | $a_{10}^2$ | $a_{11}^2$ | $a_{12}^2$ | $a_{13}^2$ | $a_{14}^2$ | $a_{22}^3$ | $a_{30}^4$ |
| $a_{19}^4$ | $a_{15}^3$ | $a_{16}^3$ | $a_{17}^3$ | $a_{18}^3$ | $a_{19}^3$ | $a_{20}^3$ | $a_{21}^3$ | $a_{29}^4$ |
| $a_{20}^4$ | $a_{21}^4$ | $a_{22}^4$ | $a_{23}^4$ | $a_{24}^4$ | $a_{25}^4$ | $a_{26}^4$ | $a_{27}^4$ | $a_{28}^4$ |

Figure 1. Example of a 9*9 GMSF mask

During the mask's travel, if the pixel pair $< a_i^L, a_j^M >$ satisfies any of the 3 criterions:

$$(1):\begin{cases} 1 \le L = M \le N,\ i < j,\ \{i,j\} = \underset{0\le i,j\le 8L-1}{\arg}\ \min f_A(i,j) \\ a_i^L\ \textit{and}\ a_j^M\ \textit{belong to the same CC} \end{cases} \qquad (1)$$

$$(2):\begin{cases} 1 \le L = M-1 \le N-1,\ \{i,j\} = \underset{\substack{0\le i\le 8L-1 \\ 0\le j\le 8M-1}}{\arg}\ \min f_B(i,j,L,M) \\ a_i^L\ \textit{and}\ a_j^M\ \textit{belong to the same CC} \end{cases} \qquad (2)$$

$$(3):\begin{cases} 1 \le L = M-2 \le N-2,\ \{i,j\} = \underset{\substack{0\le i\le 8L-1 \\ 0\le j\le 8M-1}}{\arg}\ \min f_C(i,j,L,M) \\ a_i^L\ \textit{and}\ a_j^M\ \textit{belong to the same CC} \end{cases} \qquad (3)$$

$f_A = j - i$, $f_B$ and $f_C$ represent the distance between pixel $a_i^L$ and $a_j^M$

Then $h(i_M, j_L)$, which records the feature histogram at $(i_M, j_L)$, will increase by 1.

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.134

$$p(i_M, j_L) = \frac{h(i_M, j_L)}{\sum_{i,j,L,M} h(i_M, j_L)} \qquad (4)$$

Finally, the $p(i_M, j_L)$ distribution describes the GMSF. Therefore we can figure out that the size of the feature vector is $1 \times (4N^2 + 4N)^2$.

After we obtain the GMSF of each script, distance metrics are applied to calculate the analogy between the scripts. It is believed that scripts of one writer have the nearest distance. Besides Euclidian, Manhattan and Chi-square distance, weight of each GMSF dimension is introduced from the training category to enhance the writing style difference. The weighted distances are shown in equation (5), (6) and (7)

$$d_{WEuc}(\overline{v}_1, \overline{v}_2) = \sum_{i=1}^{N} \sqrt{\left(\frac{(v_{1i} - v_{2i})}{\sigma_i}\right)^2} \qquad (5)$$

$$d_{WMan}(\overline{v}_1, \overline{v}_2) = \sum_{i=1}^{N} \frac{|v_{1i} - v_{2i}|}{\sigma_i} \qquad (6)$$

$$d_{WChi}(\overline{v}_1, \overline{v}_2) = \sum_{i=1}^{N} \frac{(v_{1i} - v_{2i})^2}{\sigma_i (v_{1i} + v_{2i})} \qquad (7)$$

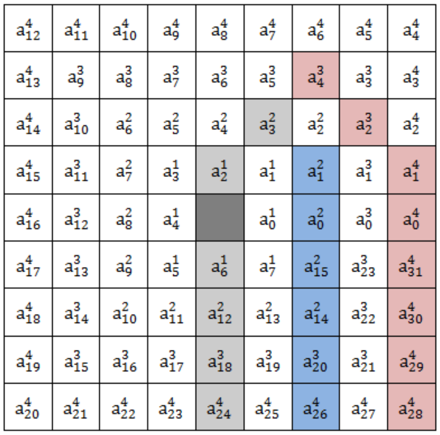
$$\sigma_i = \sqrt{\frac{1}{M-1} \sum_{k=1}^{M} (v_{ki} - \mu_i)^2} \qquad (8)$$

$\overline{v}_1, \overline{v}_2$ are feature vector, while $\sigma_i$ is the feature vector's i-th dimension standard deviation, M is samples' total number.

### B. Deficiency of GMSF

We apply the GMSF method (Only criterion 1 applied) on a 100 Chinese writers' database, trying to match the writers from testing category to training category, under the weighted Chi-square distance metric, the accuracy reaches to 94%, (if the 3 criterions work together, accuracy reaches up to 96%). However, as we examine the feature details, we find some disappointing facts, and we believe that GMSF is strike-width sensitive, which should be improved.

The essence of GMSF is to extract the script's local structure, and accumulates the local feature on the entire passage, in this process, semantic information is blurred, and singular information, which is mostly manifested in local mask, is preserved, so that GMSF is a text-independent feature. However, our multi-width strike experiments indicate that pen-width information also remains in the mask, and GMSF seems a recognizer of both writer and pen, when the pen-width component rises, GMSF cannot distinguish writers. Even though the three criterions are designed to eliminate some irrelevant strike width information by the CC constraint, we find such strike-end situation as Figure 2 shows below



| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $a^4_{12}$ | $a^4_{11}$ | $a^4_{10}$ | $a^4_9$ | $a^4_8$ | $a^4_7$ | $a^4_6$ | $a^4_5$ | $a^4_4$ |
| $a^4_{13}$ | $a^3_9$ | $a^3_8$ | $a^3_7$ | $a^3_6$ | $a^3_5$ | $a^3_4$ | $a^3_3$ | $a^4_3$ |
| $a^4_{14}$ | $a^3_{10}$ | $a^2_6$ | $a^2_5$ | $a^2_4$ | $a^2_3$ | $a^2_2$ | $a^3_2$ | $a^4_2$ |
| $a^4_{15}$ | $a^3_{11}$ | $a^2_7$ | $a^1_3$ | $a^1_2$ | $a^1_1$ | $a^2_1$ | $a^3_1$ | $a^4_1$ |
| $a^4_{16}$ | $a^3_{12}$ | $a^2_8$ | $a^1_4$ | | $a^1_0$ | $a^2_0$ | $a^3_0$ | $a^4_0$ |
| $a^4_{17}$ | $a^3_{13}$ | $a^2_9$ | $a^1_5$ | $a^1_6$ | $a^1_7$ | $a^2_{15}$ | $a^3_{23}$ | $a^4_{31}$ |
| $a^4_{18}$ | $a^3_{14}$ | $a^2_{10}$ | $a^2_{11}$ | $a^2_{12}$ | $a^2_{13}$ | $a^2_{14}$ | $a^3_{22}$ | $a^4_{30}$ |
| $a^4_{19}$ | $a^3_{15}$ | $a^3_{16}$ | $a^3_{17}$ | $a^3_{18}$ | $a^3_{19}$ | $a^3_{20}$ | $a^3_{21}$ | $a^4_{29}$ |
| $a^4_{20}$ | $a^4_{21}$ | $a^4_{22}$ | $a^4_{23}$ | $a^4_{24}$ | $a^4_{25}$ | $a^4_{26}$ | $a^4_{27}$ | $a^4_{28}$ |

Figure 2. Strike-end situation which interferes GMSF

It is a local figure of an edge image. It is assumed that there are two kinds of pen-widths in the mask. The thinner pen-width consists of the gray and blue pixels; while the wider pen-width consists of the gray and red pixels. Both of the structure and style of the handwriting are quite similar, but they greatly differ in GMSF, hence the TAR (True Acceptance Rate) will drop rapidly. The existence of such strike-end situation can hardly be eliminated only by the previous method, and they happen quite frequent as we apply the mask size as big as 21*21. So the GMSF accuracy is influenced by pen-widths of the writers.

Before we raise an example of pen-width impact on GMSF, there are some conceptions which should be defined to make the narration simpler. There are three kinds of distances in our research, the first one is the distance between the writer and himself with the same pen-width, the second is the distance between the writer and all the other writers with the same pen-width, and the third one is the distance between the writer and himself with different pen-widths. We call the first distance as the self distance, the second distance as the not-self distance, which is also inter class distance, the third one pen-width distance, which is also inner class distance. We record these distances on weighted Chi-square distance metric and calculated their average value in Table I below, from which we can obviously find out the pen-width variation impact on GMSF.

TABLE I. WRITER AND PEN-WIDTH WEIGHTED CHI-SQUARE DISTANCE COMPARISON

| Distance value<br>Distance type | Average weighted Chi-square |
|---|---|
| **Self distance (Pen-width is 0.5mm)** | 1103.5 |
| **Minimum not self distance (Pen-width is 0.5mm)** | 1667.7 |
| **Pen-width distance (Pen-width is 0.3mm and 0.5mm)** | 6261.1 |

We hope the pen-width distance is smaller than the minimum not self distance, however the experiments shows that pen-width distance already surpass the writer distance, therefore we doubt that the pen-width change may play great impact on the recognition accuracy.

## II. STATISTICAL IMPROVEMENT BY AVERAGE TEMPLATE AND INNER CLASS VARIANCE TRAINING.

### A. Obtaining average training template by morphology

Even though people write in different pen-width occasionally, we can still recognize their handwritings easily. It is believed that writing style information is an immanent feature which is independent from different pen-widths. Therefore we analyze the pen-width of some free handwriting samples, and a general conclusion is required that most common pen-width range is from 0.3mm to 1.0mm and the pen-width range is from 3 to 8 pixels on 300dpi scanned digital image. According to this information, we firstly assume that there are mainly three kinds of pen-width in our daily ordinary writing, the thin pen-width (0.3mm, 3~4 pixels), the median pen-width (0.5mm, 5~6 pixels) and the wide pen-width (1.0mm, 7~8 pixels). We can obtain the one writer's average GMSF by calculate his average template of the three pen-width GMSF as below. (see in Equation 9).

$$\overline{v_{avg}} = \frac{1}{3}(\overline{v_M} + \overline{v_T} + \overline{v_W}) \tag{9}$$

We apply this average GMSF template as the training data. As we know, $p(i_M, j_L)$ in equation (4) is actually a random vector. The average operation reduces the pen-width change and makes GMSF closer to their writing style singular feature's statistical expectation. Therefore, we believe this method can obtained better results.

Usually, if we only have one writer's median pen-width data, we can obtain his other pen-widths samples through morphology. We erode and dilate 25 writers' median pen-width samples and obtain the corresponding GMSF, then we can calculate the average GMSF template according to Eq. 9. We hope to use the generated average GMSF to stimulate the true average GMSF.

### B. Inner and inter class variances

In Equation (8), the variance of each GMSF dimension is obtained, however it is inter class variance among different writers. As in our statistic model, the inner class variance which reflects one writer's pen-width change should be considered. We adapt the distance metrics formulation to add both inter and inner class variance vector from training category.

$$d_{NeoWEuc}(\overline{v_1}, \overline{v_2}) = \sum_{i=1}^{N} \sqrt{\left( \frac{\sigma_i^{inter}(v_{1i} - v_{2i})}{\sigma_i^{inner}} \right)^2} \tag{10}$$

$$d_{NeoWMan}(\overline{v_1}, \overline{v_2}) = \sum_{i=1}^{N} \frac{\sigma_i^{inter} \,|\, v_{1i} - v_{2i} \,|}{\sigma_i^{inner}} \tag{11}$$

$$d_{NeoWChi}(\overline{v_1}, \overline{v_2}) = \sum_{i=1}^{N} \frac{\sigma_i^{inter}(v_{1i} - v_{2i})^2}{\sigma_i^{inner}(v_{1i} + v_{2i})} \tag{12}$$

In these distance metrics:

$$\sigma_i^{inter} = \sqrt{\frac{1}{M-1} \sum_{k=1}^{M} (v_{ki} - \mu_i)^2} \tag{13}$$

$\mu_i$ is the i-th mean value of different writer's GMSF.

$$\sigma_{ki}^{inner} = \sqrt{\frac{1}{2} \sum_{w=1}^{3} (v_{wi} - \hat{\mu}_i)^2} \tag{14}$$

$\hat{\mu}_i$ is the i-th mean value of the k-th writer's 3 pen-widths GMSF.

$$\sigma_i^{inner} = \frac{1}{M} \sum_{k=1}^{M} \sigma_{ki}^{inner} \tag{16}$$

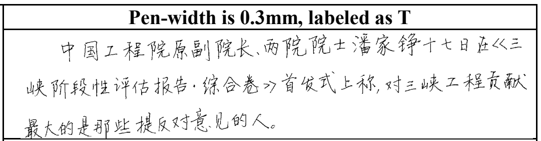
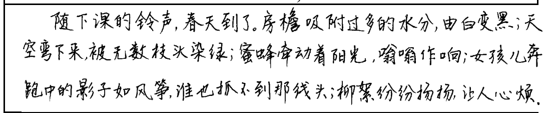
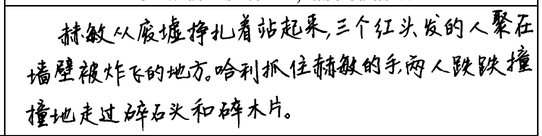
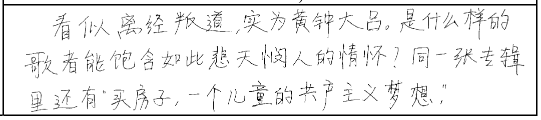
So far, we can obtain distances between testing and training category according to both previous distance metrics (Eq. 5, 6 and 7) and adapted distance metrics (Eq. 10, 11 and 12). The improved method can train the inner class variance and restrain the pen-width change on GMSF.

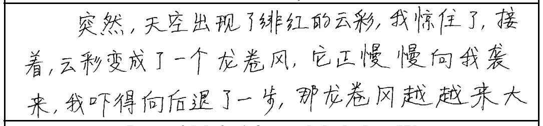
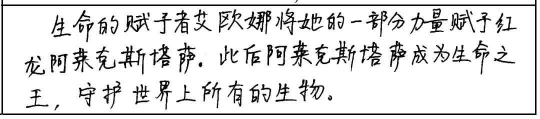
## III. MULTI-WIDTH CHINESE SCRIPT EXPERIMENT.

### A. Sample collection

We prepared a group of experiment to testify the pen-width impact of GMSF. In the experiment, 25 writers are asked to write different passages of 200-250 Chinese characters with 3 pens labeled as 0.3mm (thin), 0.5mm (median) and 1.0mm (wide) respectively. Similar data collection and sample usage instructions are referred [7]. For example, two writers' handwriting samples with three different pen-widths are shown as below (only display part of the image).

TABLE II. EXAMPLES OF DIFFERENT PEN-WIDTH SCRIPTS

| | |
|---|---|
| **No.8 Writer** | **Pen-width is 0.3mm, labeled as T** |
| |  中国工程院原副院长、两院院士潘家铮十七日在《三峡阶段性评估报告·综合卷》首发式上称，对三峡工程贡献最大的是那些提反对意见的人。 |
| | **Pen-width is 0.5mm, labeled as M** |
| |  随下课的铃声，春天到了。房檐吸附过多的水分，由白变黑；天空弯下来，被无数枝头染绿；蜜蜂牵动着阳光，嗡嗡作响；女孩儿奔跑中的影子如风筝，谁也抓不到那线头；柳絮纷纷扬扬，让人心烦. |
| | **Pen-width is 1.0mm, labeled as W** |
| |  赫敏从废墟挣扎着站起来，三个红头发的人聚在墙壁被炸飞的地方。哈利抓住赫敏的手，两人跌跌撞撞地走过碎石头和碎木片。 |
| **No.12 Writer** | **Pen-width is 0.3mm, labeled as T** |
| |  看似离经叛道，实为黄钟大吕。是什么样的歌者能饱含如此悲天悯人的情怀？同一张专辑里还有"买房子，一个儿童的共产主义梦想。" |

| | |
|---|---|
| | **Pen-width is 0.5mm, labeled as M** |
| |  突然，天空出现了绯红的云彩，我惊住了，接着，云彩变成了一个龙卷风，它正慢慢向我袭来，我吓得向后退了一步，那龙卷风越越来大 |
| | **Pen-width is 1.0mm, labeled as W** |
| |  生命的赋予者艾欧娜将她的一部分力量赋予红龙阿莱克斯塔萨。此后阿莱克斯塔萨成为生命之王，守护世界上所有的生物。 |

Now we have 75 samples in hand. These passages are manually divided in half. The top half sections belong to training category and the bottom half sections belong to testing category. Now we have 3 data sets, each set contains the training and testing category of the same pen-width. The samples in training and testing category are both labeled from 1 to 25 respectively. Then we performed the similar procedure as Li and Ding's. If the No. k test sample is recognized as the No. k training sample, it is a correct identification. The Ture Acceptance Rate (TAR) is required according to the following formula:

$$TAR = \frac{correctly\ recognized\ testing\ sample\ number}{N} \qquad (9)$$

In our experiment, N is 25, and the recognition accuracies are displayed in the tables below.

### *B. Complete training category*

If three pen-width scripts all appear in the training category, we test the thin, median and wide samples respectively, the recognition accuracy is quite high, which indicates that the writing style feature matches correctly under the same pen-width. The recognition accuracy is shown in Table III.

TABLE III. GMSF ACCURACY ON MULTI-WIDTH CHINESE SCRIPTS, MASK SIZE IS 21*21

| **Accuracy** | **Train 0.5mm+0.3mm+0.7mm** | | |
|---|---|---|---|
| **Distance metrics** | ***Test 0.5mm*** | ***Test 0.3mm*** | ***Test 1.0mm*** |
| Weighted Chi-square | 100% | 92% | 92% |

### *C. Only median pen-width training category*

We only preserve the median pen-width scripts in the training category, and the previous GMSF method shows poor performance as we expected.

TABLE IV. GMSF ACCURACY ON MULTI-WIDTH CHINESE SCRIPTS, MASK SIZE IS 21*21

| **Accuracy** | **Train0.5mm** | | |
|---|---|---|---|
| **Distance type** | ***Test 0.5mm*** | ***Test 0.3mm*** | ***Test 1.0mm*** |
| Euclidian | 100% | 16% | 28% |
| Manhattan | 100% | 24% | 8% |
| Chi-square | 100% | 8% | 4% |
| Weighted Euclidian | 100% | 4% | 8% |
| Weighted Manhattan | 100% | 28% | 8% |
| Weighted Chi-square | 100% | 16% | 8% |

From Table IV, we can obviously see that the same pen-width trial has much higher recognition accuracy, however, as the pen-width changes, the recognition accuracy drops quickly. Our suspicion is testified, and this experiment result leads to the fact that GMSF is not a robust method dealing with the pen-width change.

### *D. Average GMSF template training category*

We apply the average pen-width GMSF as training template and the recognition accuracy raises significantly as table V shows

TABLE V. GMSF ACCURACY ON MULTI-WIDTH CHINESE SCRIPTS WITH AVERAGE TRAINING GMSF, MASK SIZE IS 21*21

| **Accuracy** | **Train Average GMSF** | | |
|---|---|---|---|
| **Distance type** | ***Test 0.5mm*** | ***Test 0.3mm*** | ***Test 1.0mm*** |
| Euclidian | 84% | 60% | 56% |
| Manhattan | 88% | 56% | 60% |
| Chi-square | 80% | 60% | 52% |
| Weighted Euclidian | 96% | 64% | 72% |
| Weighted Manhattan | 96% | 48% | 60% |
| Weighted Chi-square | 96% | 72% | 68% |

Compared with Table IV, the recognition accuracies obviously increase.

### *E. Utilization of Inner and inter class variance*

According to the improved distance metrics shown in Eq. (10), (11) and (12), we update the recognition accuracy in Table VI as below.

TABLE VI. GMSF ACCURACY WITH NEW DISTANCE METRICS ON MULTI-WIDTH CHINESE SCRIPTS WITH AVERAGE TRAINING GMSF, MASK SIZE IS 21*21

| **Accuracy** | **Train Average GMSF** | | |
|---|---|---|---|
| **Distance type** | ***Test 0.5mm*** | ***Test 0.3mm*** | ***Test 1.0mm*** |
| New weighted Euclidian | 88% | 72% | 80% |
| New weighted Manhattan | 100% | 76% | 84% |
| New weighted Chi-square | 100% | 92% | 88% |

From the Table VI above, the improvement is quite significant. Because of appropriate utilization of inter class and inner class variances, we successfully reduce the pen-width impact on writers' scripts GMSF. If more samples concerning about one writer's different pen-widths scripts are collected, the inner class variances will get better trained and the recognition accuracy may be further improved.

### *F. Applicable "leave one out" experiment*

According to the experiments above, it seems that our improvement makes the current method more adaptive to the pen-width change. However, in practical situation, perhaps there is only one kind of pen-width scripts in the training category. Since our samples are not plenty enough, we perform the "leave one out" experiment. It is assumed that one writer's training script is written by median pen-width, however, his testing script is in one of the three pen-widths. We generate his thin and wide pen-width scripts and calculate the average GMSF template as his training data and constrain the pen-width variation by pre-trained inner class variance. In the "leave one out" test, recognition accuracy is shown in Table VII. (We only apply the New weighted Chi-square distance metric).

TABLE VII. "LEAVE ONE OUT" EXPERIMENT, GENERATED AVERAGE GMSF TEMPLATE AS TRAINING DATA.

| **Accuracy**<br>**Distance type** | **Train Average GMSF** | | |
|---|---|---|---|
| | ***Test 0.5mm*** | ***Test 0.3mm*** | ***Test 1.0mm*** |
| Neo weighted Chi-square | 84% | 80% | 60% |

From table VII, we can see that the performance in actual situation drops because the train average GMSF contains morphological generated scripts; however it is greatly improved than Table IV result.

## IV. CONCLUSION

In this paper, we discuss the deficiency of the current GMSF and collect samples of different pen-widths. In the following experiments, GMSF sensitivity of pen-width is verified: the recognition TAR drops rapidly when the training and testing category vary in pen-width. We also compared the writer distance and pen-width distance, and the results support the accuracy decrease. The reason of this result, strike-end, is found. Finally, statistical approaches are proposed to solve the pen-width problem, average training GMSF template is applied, inner and inter class variances are complemented to the previous distance metrics. Take new weighted Chi-square distance metric as an example, the recognition accuracy reaches up to 92% (testing category pen-width is 0.3mm) and 88% (testing category pen-width is 1.0mm), increase 76 percent and 80 percent respectively. In leave one out experiment, we stimulate the practical situation, and the recognition accuracy is improved than the previous method.

In future work, our data base will be enlarged mainly on two aspects: more writers' scripts and more scripts with different pen-widths of each one writer. We hope this data base can obtain more stable pen-width inner class variance dependent from writers' inter class variance and help understand the relationship between writer's GMSF and pen-width GMSF. Besides, the improvement on the feature selection level will be made.

## ACKNOWLEDGMENT

This work was supported by the National Basic Research Program of China (973 program) under Grant No. 2007CB311004 and the National Natural Science Foundation of China under Grant Nos. 60872086, 60933010.

## REFERENCES

[1] Plamondon. R., Lorette. G, "Automatic Signature Verification and Writer Identification- the State of the Art," Patt. Rec. 22(2), pp. 107–131 1989

[2] R. Plamondon, S. N. Srihari, "On-line and off-line handwriting recognition: a comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84, 2000.

[3] Bulacu M, Schomaker L, Vuurpijl L, "Writer identification using edge-based directional feature," Proceedings. Seventh International Conference on Document Analysis and Recognition, 2003, pp. 937-941

[4] Schomaker L, Bulacu M, Franke K, "Automatic writer identification using fragmented connected-component contours," IWFHR-9, 2004

[5] S. N. Srihari, S.-H. Cha, H. Arora *et al.*, "Individuality of handwriting," Journal of Forensic Sciences, vol. 47, no. 4, pp. 856-872, 2002.

[6] Xin Li, Xiaoqing Ding, "Writer identification of Chinese handwriting using grid microstructure feature," Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings.

[7] T. Su, T. Zhang, D. Guan. "HIT-MW dataset for offline Chinese handwritten text recognition," Proceedings of the 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR), 2006.