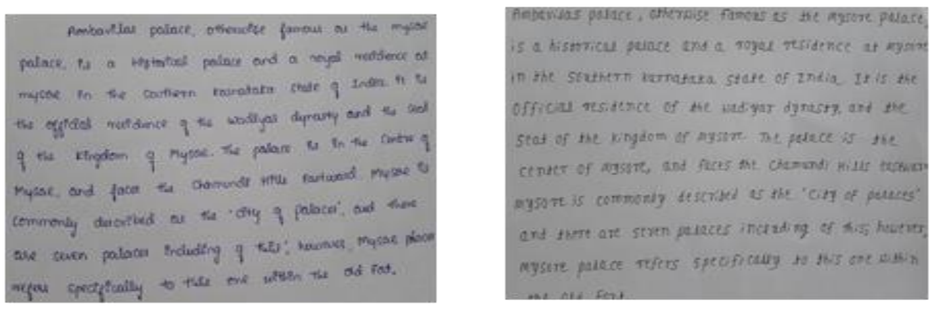
Adaptive Multi-Gradient Kernels for Handwritting based Gender Identification

[1]B. J, Navya, [2]Palaiahnakote Shivakumara, [1]G.C, Shwetha, [2]Sangheeta Roy, [1]D. S. Guru, [3]Umapada Pal and [4]Tong Lu
[1]Department of Studies in Computer Science, University of Mysore, Karnataka, India.
Email: navyagowdab.j@gmail.com,swethagc46@gmail.com, dsguruji@yahoo.com
[2]Faculty of Computer System and Information Technology, University of Malaya, Kuala Lumpur, Malaysia.
Email: shiva@um.edu.my. 2sangheetaroy@gmail.com.
[3]Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India. Email: umapada@isical.ac.in.
[4]National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China. Email: lutong@nju.edu.cn

Motivation

- Gender identification plays a vital role in forensic applications, such as document authorization, suspecting person who committed crimes etc. In addition, it can be used for studying psychological behavior of the person.
- It assists forensic investigation to identify the actual crime along with the other biometric based methods.
- However, handwriting based Gender identification is challenging due to unconstrained handwriting and individual differences in writing.



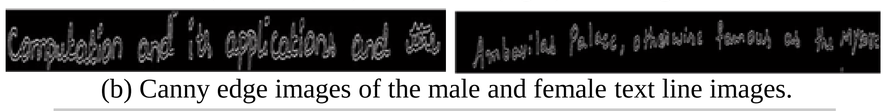
Female Male

Sample handwriting of female and male.

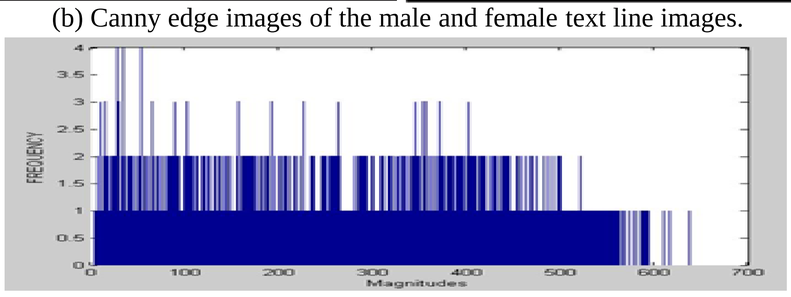
Proposed method

- We explore Adaptive Multi-Gradient (AMG) Sobel kernels based feature for gender identification in this work.
- For each segmented text lines, the proposed method finds dominant points based on gradient directional symmetry of text pixels.
- The histogram operation is performed for the AMG features extracted corresponding to dominant points of respective Sobel kernels to find the values which contributes to highest peak.
- The feature vector is formed based on value of histogram highest peaks. The proposed method finds correlation between the feature vectors between first and successive text lines in the image.
- The consistency and inconsistency based on correlation is studied for gender identification.

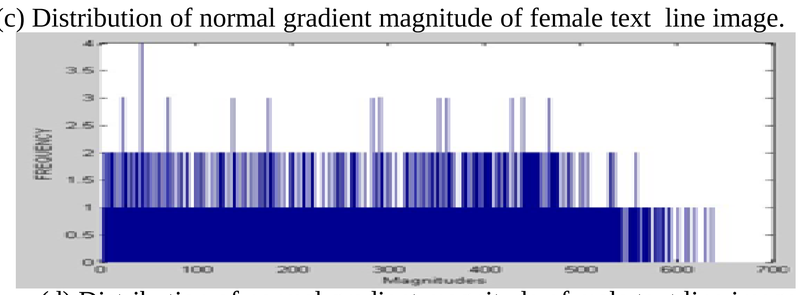
(a) Male (left) and female (right) line images



(b) Canny edge images of the male and female text line images.



(c) Distribution of normal gradient magnitude of female text line image.



(d) Distribution of normal gradient magnitude of male text line image.
Cues for gender identification with the gradient distribution of male and female line images

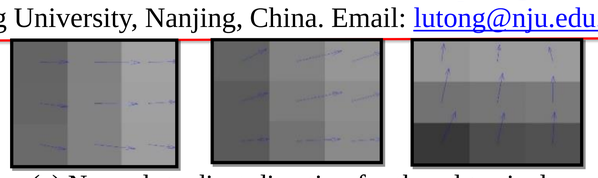
Adaptive Multi-Gradient Directional Symmetry for Dominant Pixel Detection

$S = \frac{T_{scale} S_{local}}{S_{max}}$ where S_{local} represents the Sum of Absolute Difference (SAD), and S_{max} represents the maximum SAD value.

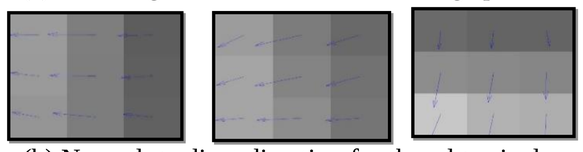
$S_{local} = SAD = \frac{1}{9}\sum_{K=-1}^{1}\sum_{L=-1}^{1}|F(i,j) - F(i+K, j+L)|$ Where $F(i,j)$is the middle pixel 3×3 window in the gray image,

S	S/2	S		S	0	-S		-S	-S/2	-S		-S	0	S
0	0	0		S/2	0	-S/2		0	0	0		-S/2	0	S/2
-S	-S/2	-S		S	0	-S		S	S/2	S		-S	0	S
-S/2	-S	0		0	S	S/2		S/2	S	0		0	-S	-S/2
-S	0	S		-S	0	S		S	0	-S		S	0	-S
0	S	S/2		-S/2	-S	0		0	-S	-S/2		S/2	S	0

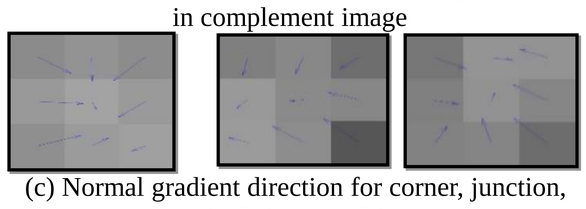
Multi Sobel directional kernels for AMG features extraction: Horizontal-Vertical, Vertical-Horizontal, South West-South East and North East-North West.



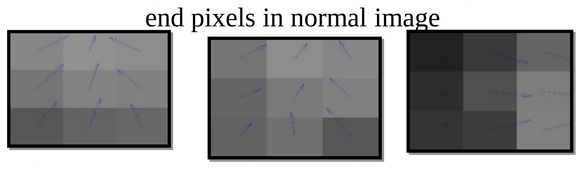
(a) Normal gradient direction for the edge pixels



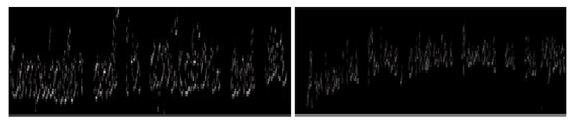
(b) Normal gradient direction for the edge pixels in complement image



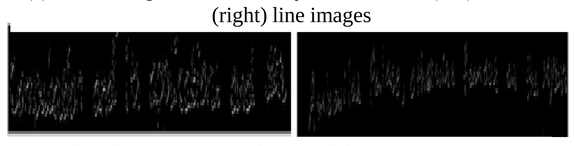
(c) Normal gradient direction for corner, junction, end pixels in normal image



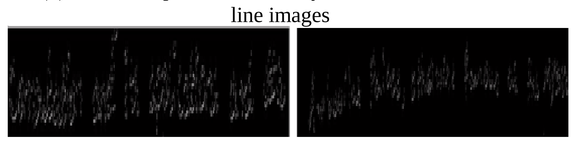
(d) Normal gradient direction for corner, junction, end pixels in complement image.
Basis for defining dominant pixels.



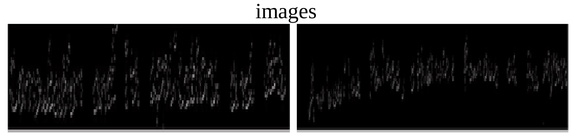
(a) Dominant pixels obtained by D1 of female (left) and male (right) line images



(b) Dominant pixels obtained by D2 of female and male line images

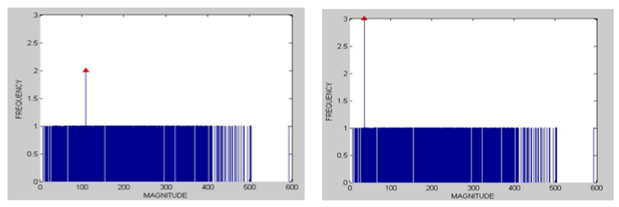


(c) Dominant pixels obtained by D3 of female and male line images

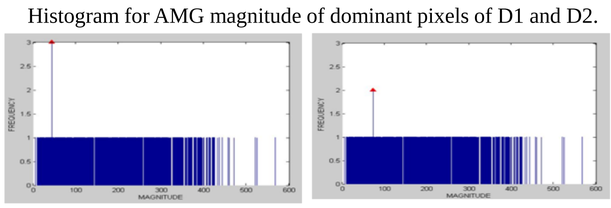


(d) Dominant pixels obtained by D4 of female and male line images
Dominant pixel detection by AMG directional symmetry.

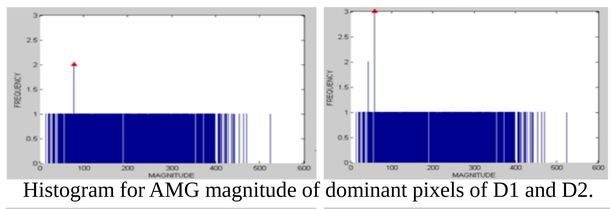
Adaptive Multi-Gradient Magnitude for Gender Classification



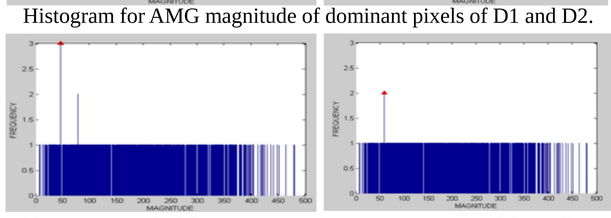
Histogram for AMG magnitude of dominant pixels of D1 and D2.



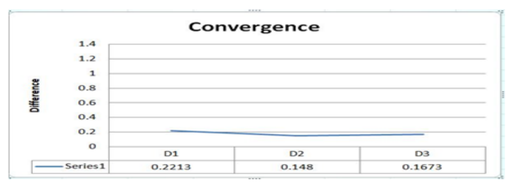
Histogram for AMG magnitude of dominant pixels of D3 and D4.
(a) Histogram for adaptive multi-gradient magnitude v/s frequencies of dominant pixels of female line image.



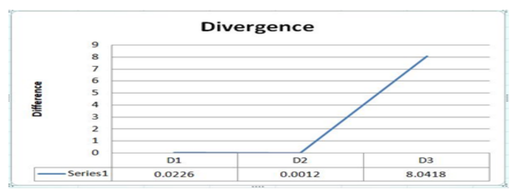
Histogram for AMG magnitude of dominant pixels of D1 and D2.



Histogram for AMG magnitude of dominant pixels of D3 and D4.
(b) Histogram for AMG magnitude vs frequencies of dominant pixels of male line image.
Illustration of choosing an AMG which contributes to the highest peak from histograms.



(a) Converging criterion for female text line image



(b) Diverging criterion for male text line image.
Studying error between correlation coefficient between first line with successive lines.

$\rho = \frac{\sum_X (G-\mu_G)(L-\mu_L)}{\sqrt{((\sum_X (G-\mu_G)^2)(\sum_X (L-\mu_L)^2))}}$ Here, ρ is correlation coefficient, G represents the ground truth vector, L represents the next vector, μ_G is the mean of ground truth vectors, μ_L denotes the mean of the next vector, and X gives the length of the vector.

Experimental Results

- We use 100 male and 100 female documents for experimentation. In addition, 990 images from QUWI, IAM-1, IAM-2 and KHATT.

Table I. Confusion matrix and classification rate (CR) of the proposed and existing methods on our and QUWI datasets

Our dataset	Proposed		Ibrahim		Bouadjenek		Bouadjenek	
	Female	Male	Female	Male	Female	Male	Female	Male
Female	90.0	10	58.9	41.1	73.9	26.1	70.3	29.7
Male	16	84.0	37.7	62.3	27.3	72.7	25.9	74.1
CR	87.0		60.6		73.3		72.2	

QUWI	Proposed		Ibrahim		Bouadjenek		Bouadjenek	
	Female	Male	Female	Male	Female	Male	Female	Male
Female	69.9	30.1	59.2	40.8	64.8	35.2	63.2	36.8
Male	29.9	70.1	45.7	54.3	32.7	67.3	34.6	65.4
CR	70.0		56.7		66.0		64.3	

Table II. Confusion matrix and classification rate (CR) of the proposed and existing methods on IAM and KHATT datasets

IAM1+IAM-2	Proposed		Ibrahim		Bouadjenek		Bouadjenek	
	Female	Male	Female	Male	Female	Male	Female	Male
Female	73.2	26.8	66.5	33.5	74.3	25.7	71.5	28.5
Male	19.9	80.1	28.3	71.7	23.9	76.1	25.2	74.8
CR	76.65		69.1		75.2		73.2	

KHATT	Proposed		Ibrahim		Bouadjenek		Bouadjenek	
	Female	Male	Female	Male	Female	Male	Female	Male
Female	74.1	25.9	68.3	31.7	72.8	27.2	72.4	27.6
Male	22.9	77.1	28.9	71.1	24.3	75.7	29.4	70.6
CR	75.6		69.7		74.2		71.5	

Conclusion

- We have proposed a novel adaptive multi-gradient feature given by multi-gradient directional Sobel kernels for gender identification based on handwriting analysis.
- Dominant pixels are detected using gradient direction symmetry in normal and complement image.
- The correlation between first and successive text lines are used for gender identification.
- The method can be extended to word level in near future.