

# SIFT 서술자를 이용한 오프라인 필기체 문자 인식 특징 추출 기법

박정국<sup>o</sup> 김경중  
세종대학교 컴퓨터공학과

Email: prfirst@sju.ac.kr<sup>o</sup>, kimkj@sejong.ac.kr

## Feature Extraction for Off-line Handwritten Character Recognition using SIFT Descriptor

Jung-Guk Park<sup>o</sup> Kyung-Joong Kim

Department of Computer Engineering, Sejong University

### 요약

본 논문에서는 SIFT(Scale Invariant Feature Transform) 기술자를 이용하여 오프라인 필기체 문자 인식을 위한 특징 추출방법을 제안한다. 제안하는 방법은 문자의 획의 방향 정보를 제공하는 특징 벡터를 추출함으로써 오프라인 문자 인식에서 성능 향상을 기대할 수 있다. 테스트를 위해 MNIST 필기체 데이터베이스와 UJI Penchar2 필기체 데이터베이스를 이용하였고, BP(backpropagation)신경망과 LDA(Linear Discriminant Analysis), SVM(Support Vector Machine) 분류기에서 성능 테스트를 하였다. 본 논문의 실험결과에서는 일반적으로 사용되는 특징추출로부터 얻어진 특징에 제안된 특징추출을 정합하여 성능향상을 보인다.

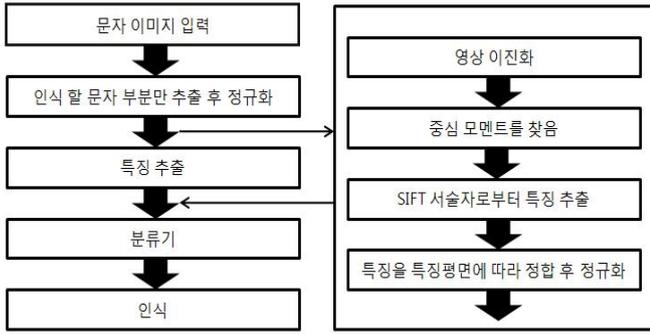
### 1. 서론

오프라인 필기체 문자 인식은 지난 50년간 패턴인식 분야에서 중요한 연구주제였다[1]. 실세계에서 무인 우편번호 인식, 수표인식, 문서 판별 등에서 사용되기 위한 연구사례가 있으며[2][3] 다양한 특징추출 방법[4][5]과 전처리[6] 및 분류기 성능향상과 함께 발전되어져 왔다. 문자 인식에서 각 문자 클래스마다 구별력이 강한 특징을 추출하는 것은 성능향상과 이어지므로 특징 추출은 중요한 부분이다. 패턴 인식문제에서도 성능을 높이기 위해서는 특징 추출과 추출된 특징을 처리하는 분류기의 성능이 주요 관심사이다.

문자 인식을 위한 특징추출 방법은 여러 각도에서 연구 중이며 S.W.Lee는 문자의 각 획의 방향을 Kirssh마스크를 이용하여 획을 4방향으로 나누어 특징 추출을 하였고[7], Dayahanker는 12방향의 특징 추출을 하였다[8]. 획의 정보를 잘 반영 할 수 있는 특징추출은 문자 인식문제를 보다 쉽게 풀 수 있는 접근방법이다. 본 논문에서는 특징 추출 부분에 대한 방법을 제안하며 그림1에 제안된 방법의 개요

를 보인다. 제안하는 방법은 D.G.Lowe[9] 가 소개한 SIFT (Scale Invariant Feature Extraction 이하 SIFT) 기술자를 문자의 정의된 중심 모멘트에 중심으로 하여 특징을 추출한다. SIFT 서술자는 지문인식[10], 사물인식[11], 문자인식[12]등에서 유용하게 사용되고 있다. 제안하는 특징 추출은 일반적으로 사용되는 문자인식을 위한 특징에 추가되어 구별력을 가진 특징을 생성하게 되는데, 이는 문자의 획의 방향에 대한 정보를 제공 할 수 있다. 제안하는 방법으로 추출된 특징은 3가지의 다른 타입의 분류기를 사용하여 기존의 특징과 비교하며 성능평가를 한다. 본 연구에서는 구별력있는 특징 추출에 중점을 두었으며 기존의 특징추출 중 일반적으로 사용되는 방법 중 문자 영상 이미지로부터 문자의 획의 정보를 충분히 반영 할 수 있는 2가지를 채택하여 제안된 방법과의 비교를 본 논문에서 보인다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 특징 추출 방법을 서술하고, 3장에서는 추출된 특징을 분류하기 위한 분류기들을 소개한다. 4장에서는 실험결과를 보이고, 5장은 본 논문의 결론을 제시한다.



제안하는 특징 추출 방법

그림 1. 문자 인식의 시스템과 본 논문의 제안한 방법

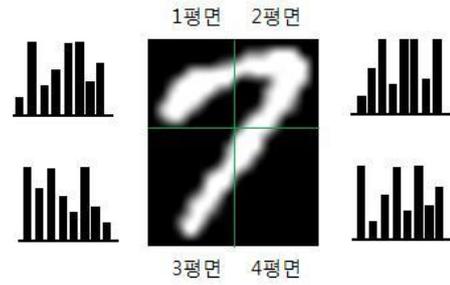


그림 2. 중심모멘트와 4분할 평면의 특징 히스토그램

## 2. 특징추출

이 절에서는 오프라인 필기체 문자 인식을 위한 특징 추출 방법을 소개한다. 제안하는 특징 추출 방법은 본 연구에서 정의된 중심모멘트를 중심으로 SIFT 서술자를 사용하여 특징 추출을 하였다. 중심모멘트는 입력된 문자 영상을 4분할 하여 4평면을 생성하며 다음과 같이 정의된다[13].

$$m_{(x,y) \in \mathbb{R}^2} x_p y_p \quad \mathbf{a} = m_{00} \quad (1)$$

$$x_c = \frac{m_{10}}{\mathbf{a}} \quad y_c = \frac{m_{01}}{\mathbf{a}}$$

$m$ 은 중심모멘트 정의이며,  $\mathbf{a}$ 는 문자의 면적을 나타내고,  $x_c, y_c$ 는 x축, y축 중심모멘트이다. 중심 모멘트를 중심으로 SIFT 서술자가 특징 추출을 하게 된다. SIFT 서술자는 다음과 같다.

$$Descriptor_{(x,y)} = [\theta_{(x,y)}, mag_{(x,y)}] \quad (2)$$

$$\theta_{(x,y)} = \tan^{-1}(I_{x,y+1} - I_{x,y-1}) / (I_{x-1,y} - I_{x+1,y})$$

$$mag_{(x,y)} = \sqrt{(I_{x+1,y} - I_{x-1,y})^2 + (I_{x,y+1} - I_{x,y-1})^2}$$

수식 2의  $Descriptor_{(x,y)}$ 는 문자 영상 이미지 픽셀 좌표  $x, y$ 에서의 SIFT 서술자를 나타내고 크기가  $mag$  이고 방향이  $\theta$ 인 벡터로 나타낼수 있다. 구해진 벡터의  $\theta$ 을 8방향으로 양자화하며, 각 픽셀의 벡터 방향  $\theta$ 에 따라서  $mag$  크기를 정합한 히스토그램을 특징으로 사용한다. 양자화된 벡터는 중심모멘트를 중심으로 4개의 특징평면으로 나누어서 정규화 하여, 한 평면당 8방향의 벡터를 가지는 4평면의 총 32개의 특징벡터를 완성한다. 제안하는 SIFT 서술자는 각 평면의 글의 획의 방향과 크기를 반영할 수 있다.

## 3. 분류기

이 절에서는 특징을 사용하여 분류를 수행 할 분류기에 대해 소개를 한다. LDA(Linear Discriminant Analysis)와 MLP(Multi-Layered Perceptron), SVM(Support Vector Machine) 분류기를 서술한다.

### A. LDA

LDA는 베이지언 학습을 이용하여 패턴 분류를 한다. 베이지언 확률은 다음과 같으며

$$P(w_i | \mathbf{x}) = \frac{p(\mathbf{x} | w_i) P(w_i)}{p(\mathbf{x})}$$

분별함수는 다음과 같다.

$$g(w_i, \mathbf{x}) = p(\mathbf{x} | w_i) P(w_i)$$

$p(\mathbf{x} | w_i)$ 는  $N(\mu, \Sigma)$ 인 정규분포를 따르며

$$p(\mathbf{x} | w_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left\{ -\frac{(\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i)}{2} \right\}$$

분별함수는 다음과 같이 얻을 수 있다[14].

$$g(\mathbf{x}, w_i) = (\mathbf{x} - \mu_i)^T \Sigma_i^{-1} (\mathbf{x} - \mu_i) + \log |P(w_i)|$$

### B. MLP

본 실험에서 사용한 Multiple Layer Perceptron 은 하나의 은닉층을 가졌으며 델타법칙으로 연결강도를 학습시킨다. 입력층은 특징벡터의 차원과 같고, 출력층의 개수를 클래스의 개수를 따른다. 어떤 입력패턴  $\mathbf{x} = [x_1, x_2, \dots, x_d]$  이고  $k$ 개의 클래스가 있을 때, 출력값  $O_k$ 는

$$O_k(\mathbf{x}) = s \left[ \sum_{j=1}^{h_n} w_{kj} s(\mathbf{v}_{j_i}^T \mathbf{x} + v_{j_0}) + w_{k0} \right]$$

$$= s\left[\sum_{j=1}^{h_n} w_{kj} \mathbf{h}_j + w_{k0}\right]$$

$s$ 는 시그모이드 함수  $s(x) = \frac{1}{1+exp^{-x}}$ 이며,  $h_n$  은 은닉층의 개수이다.  $w_{kj}$ 은 출력층과 은닉층을 연결하는 가중치이고,  $v_{ji}$ 는 은닉층과 입력층을 연결하는 가중치 값이다. 연결강도 학습에는 back-propagation(오류 역전파) 알고리즘[15]을 사용했으며, 온라인 모드로 연결값을 갱신하였다.

### C. SVM

Support Vector Machine[16]분류기는 일반적으로 이진 분류 문제를 해결하는데 사용하며, 다중 분류 문제에서는 여러개의 SVM이 필요하다. 본 연구에서는 M 분류 문제에 대해서  $\binom{M}{2}$  문제로 나누어서 해결을 하였다. 어떤 특징벡터  $\mathbf{x}$ 에 대해서 SVM은

$$f(\mathbf{x}) = \sum_{i=1}^l y_i \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b \quad (3)$$

로 나타내며,  $l$ 은 학습을 위한 특징 패턴의 개수이며,  $y_i$ 는 특징 $\mathbf{x}$ 에 대한 -1또는 1을 갖는 목표값이다.  $b$ 는 바이어스이며,  $k(x, x_i)$ 은 특징벡터의 커널 함수이며  $\Phi(\mathbf{x})$ 는 특징벡터를 확장시킨 특징공간 함수이다. 본 연구에서는

$$k(x, x_i) = \Phi(x) \cdot \Phi(x_i) \quad (4)$$

$$k(x, x_i, p) = (1 + \mathbf{x} \cdot \mathbf{x}_i)^p \quad (5)$$

두 가지 커널을 사용하였다. 수식(3)은 일반화된 선형 판별 함수로 나타 낼 수 있으며

$$\mathbf{w} = \sum_{i=1}^l y_i \alpha_i \cdot \Phi(\mathbf{x}_i)$$

$\alpha_i$ 는 아래의 최적화 문제를 풀면서 얻을 수 있다.

$$J(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2$$

을 최소화 하며

$$y_i f(x_i) \geq 1 - \xi_i, \quad \xi_i \geq 0, i = 1, \dots, l$$

을 만족한다. 이것은 2차 계획법 문제이며 듀얼문제로 변환할 수 있으며

$$W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j)$$

를 최대화 하고

$$0 \ll \alpha_i \ll C, i = 1, \dots, l, \quad \sum_{i=1}^l \alpha_i y_i = 0$$

여기에서  $C$ 는 학습하는 동안의 분류에러 임계치이다.

## 4. 실험 결과

본 연구를 위해 문자 영상의 크기를 28x28 크기로 정규화하여 Otsu[17] 방법을 이용하여 의사-이진영상으로 변환하는 간단한 전처리과정을 거치며 MNIST 데이터베이스와 Machine Learning Repository의 UJI penchar2 데이터베이스를 사용하였다.(그림3 참고) MNIST 데이터베이스는 60000개의 테스트 집합과 10000개의 테스트 집합으로 이루어져 있으며 각 클래스간 학습 집합을 균등하게 하기 위해 5000종류를 사용하였다. UJI penchar2 데이터 베이스는 필기체 영상 이미지 52개의 영문자 대문자소문자를 포함하여 14가지 스페인문자, 숫자, 21가지 각종 기호 문자 영상이미지를 포함하여, 학습 영상 이미지 80종류와 테스트 영상이미지 40종류로 이루어져있다.

분류기 매개변수 설정은 MLP의 경우 학습률은 0.2로 하였으며 MNIST 데이터베이스에 대해서는 50 epoch, UJI penchar2 데이터베이스는 100 epoch를 학습시켰다. SVM(L)은 수식(4)의 커널을 따르고 SVM(P)는 수식(5)의 커널을 따르며 커널의  $p$ 의 값은 2이며 두 SVM 모두 종료 임계치값을 0.01로 하였고 분류에러 임계치는 1로 설정하였다. 본 연구에서는 숫자 영상 이미지를 사용하여 기존의 필기체 숫자의 특징 추출 방법과 비교를 하였다.

D.Trier[4] 는 문자 영상이미지에서 특징을 추출하게 위해 Zoning technique와 Projection crossing count 방법을 소개하였고 본 실험에서는 두 가지 특징 추출을 정합하여 하나의 특징 벡터를 생성하여 특징 추출1을 만들고, 여기 특징 벡터에 제안하는 SIFT 서술자가 포함된 특징을 포함하여 특징 추출2을 생성하여 비교를 하였다. 특징 추출1에서 projection crossing count 방법으로 22차원의 벡터를 생성하고 특징벡터의 차원을 낮추고 Zoning technique 에서는 영상이미지를 7\*7로 나누어서 단일한 부분의 화소의 평균을 특징값으로 취해서 특징벡터의 차원은 총 71차원이다. 특징 추출2에서는 특징추출1에서 제안한 방법의 SIFT 서술자 특징벡터 32차원을 포함시킨다.

표1에서는 두 특징에 대해 4가지 분류기로 테스트 된 결과를 보인다. 표2와 표3에서는 MNIST데이터베이스에서 특징추출1과 특징추출2으로 추출된 특징을 SVM(L) 분류기로 학습된 결과를 보이며 제안한 특징 추출방법의 향상된 성능을 보여준다.

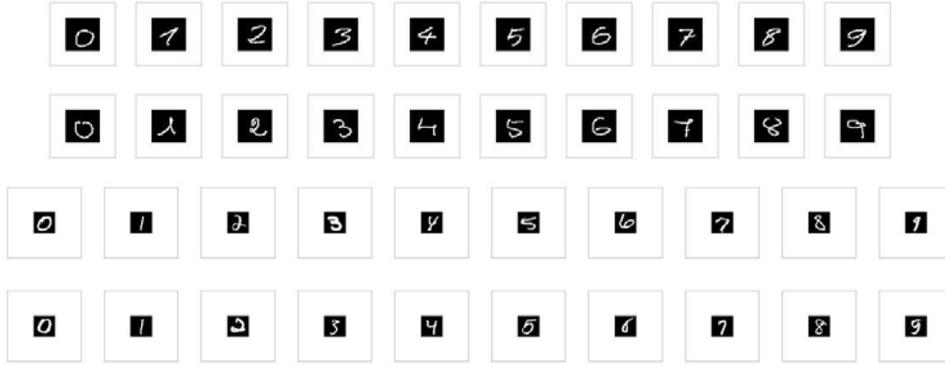


그림 3. 데이터베이스 영상이미지의 예(위 UJIPenchar2 데이터베이스 아래 MNIST 데이터베이스)

표 I  
기존의 특징 추출 방법과 제안하는 특징 추출 방법을 각 분류기를 이용한 테스트 결과

분류기	MNIST		UJI Penchar2	
	기존의 특징 추출	제안된 특징 추출	기존의 특징 추출	제안된 특징 추출
LDA	85.12%	89.00%	87.75%	89.50%
MLP	94.83%	95.04%	92.50%	94.25%
SVM(L)	92.50%	94.34%	91.75%	93.50%
SVM(P)	95.73%	96.07%	93.50%	94.75%

표 II  
SVM(L)을 사용한 MNIST 데이터 베이스 CONFUSION MATRIX(위. 특징추출1 아래.특징추출2(제안한 방법))

	0	1	2	3	4	5	6	7	8	9	accuracy
0	962	3	0	2	0	2	4	2	5	0	98.16%
1	1	962	4	3	0	9	5	0	13	0	96.20%
2	4	11	952	9	2	2	7	4	8	1	95.20%
3	2	1	13	936	1	22	0	9	10	6	93.60%
4	1	4	5	0	940	0	9	0	3	20	95.72%
5	3	8	11	26	4	816	8	2	13	1	91.47%
6	7	0	4	0	4	9	931	0	3	0	97.18%
7	0	7	8	2	11	3	0	933	4	32	93.30%
8	8	42	6	16	11	12	2	5	864	8	88.70%
9	1	217	3	14	22	6	0	16	18	917	75.53%
total											92.50

	0	1	2	3	4	5	6	7	8	9	accuracy
0	969	0	0	1	0	1	3	2	4	0	98.87%
1	1	976	7	2	2	1	4	0	6	1	97.60%
2	2	8	969	5	3	0	1	5	7	0	96.90%
3	2	0	6	955	0	21	0	7	5	4	95.50%
4	0	2	4	0	951	1	8	0	1	15	96.84%
5	2	1	4	24	1	843	6	1	9	1	94.50%
6	6	1	6	0	2	10	930	0	3	0	97.07%
7	0	6	7	0	10	1	0	950	2	24	95.00%
8	5	9	8	5	8	9	1	4	918	7	94.25%
9	2	4	2	14	16	4	0	18	221	933	76.85%
total											94.34

표 III  
SVM(L)을 이용한 UJI PENCHAR2 데이터 베이스 CONFUSION MATRIX(위. 특징추출1 아래.특징추출2(제안한 방법))

	0	1	2	3	4	5	6	7	8	9	accuracy
0	38	0	0	0	0	0	1	0	1	0	95.00%
1	2	36	0	0	2	0	0	0	0	0	90.00%
2	0	1	36	0	0	2	0	1	0	0	90.00%
3	1	0	0	39	0	0	0	0	0	0	97.50%
4	0	3	0	0	36	0	0	0	0	1	90.00%
5	0	1	0	2	0	36	0	0	1	0	90.00%
6	0	0	1	0	0	2	36	0	1	0	90.00%
7	0	0	0	0	0	0	0	38	1	1	95.00%
8	0	0	1	1	0	0	0	0	38	0	95.00%
9	1	1	2	0	0	1	0	0	1	34	85.00%
total											91.75

	0	1	2	3	4	5	6	7	8	9	accuracy
0	40	0	0	0	0	0	0	0	0	0	100%
1	1	38	0	0	0	0	0	0	0	0	95.00%
2	0	3	36	0	0	0	0	1	0	0	90.00%
3	0	0	0	40	0	0	0	0	0	0	100%
4	0	3	0	0	37	0	0	0	0	0	92.50%
5	0	0	0	1	0	36	0	0	3	0	90.00%
6	0	0	1	0	0	1	36	0	2	0	90.00%
7	0	0	0	0	0	0	0	39	0	1	97.50%
8	0	1	0	0	0	0	0	0	39	0	97.50%
9	0	1	1	2	0	2	0	0	1	33	82.50%
total											93.50

## 5. 결론

본 논문에서는 SIFT 서술자를 이용하여 획의 방향을 반영할 수 있는 특징을 추출할 수 있었고, 표1에서 보듯이 각 4가지의 분류기를 통해 제안하는 방법이 오프라인 필기체 문자 인식에서 성능을 향상시키는 것을 확인할 수 있었다. 본 연구에서는 일반적으로 높은 인식률을 위한 다양한 특징 추출방법을 연구하지 못하였지만, 기존에 사용하던 특징 추출에 사용되는 방법에 제안하는 방법을 정합하여 더 나은 특징을 추출해 낼 수 있었다. 제안하는 방법은 획의 구별력을 높혀주어 사선이나 역사선에 대해 강한 정보를 제공하여 오인식되는 샘플의 수를 낮출 수 있었고 테스트한 두 데이터베이스에 대해서 인식 향상을 보였으며, 일반적으로 오인식될 수 있는 숫자 7과 9의 혼동을 감소시켰고 비슷한 위상적 정보를 가진 획에 대한 문자의 구별력을 높힐 수 있었다.

## 감사의 글

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2010-0012876)

## 참고 문헌

- [1] Shunji Mori, Ching Y.Suen, Kazuhiro Yamamoto, "Historical Review of OCR Research and Development," Document image analysis, IEEE Computer Society Press, pp. 244-273, 1995
- [2] LeCun Y., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W. and Jackel L. D., "Backpropagation Applied to Handwritten Zip Code Recognition," Neural Computation, vol. 1, no. 4, pp. 541-551, 1989
- [3] LeCun Y., Bottou L., Bengio Y. and Haffner P., "Gradient-Based Learning Applied to Document Recognition," Intelligent Signal Processing, IEEE, pp. 306-351, 2001
- [4] Øivind Due Trier, Anil K Jain and Torfinn Taxt, "Feature Extraction Methods For Character Recognition - A Survey," Pattern Recognition, vol. 29, pp. 641-662, 1996
- [5] Cheng-Lin Liu, Kazuki Nakashima, Hiroshi Sako, Hiromichi Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques," Pattern Recognition, vol. 37, no. 2, pp. 265-279, 2004
- [6] Chen-Lin Liu, Masashi Koga, Hiroshi Sako, Hiromichi Fujisawa, "Aspect Ratio Adaptive Normalization for Handwritten Character Recognition," Lecture Note in Computer Science, vol. 1948, pp. 418-425, 2000
- [7] Seong-Whan Lee, "Off-Line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.18, no. 6, pp. 648-652, 1996
- [8] Dayashankar Singh, Sanjay Kr.Singh and Dr.Maitreyee Dutta, "Hand-Written Character Recognition using twelve Directional feature Input and Neural Network," International Journal of Computer Applications, no. 3 - Article 14 , 2010
- [9] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004
- [10] Unsang Park, Sharath Pankanti and A.K. Jain, "Fingerprint verification using SIFT features," Defense and Security Symposium, Orlando, FL, USA , vol. 6944, 2008
- [11] Koen van de Sande, Theo Gevers and Cees Snoek, "Evaluating Color Descriptors for Object and Scene Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, no. PrePrints, 2009
- [12] Zhiyi Zhang, Lianwen Jin, Kai Ding and Xue Gao, "Character-SIFT: A Novel Feature for Offline Handwritten Chinese Character Recognition," icdar, pp.763-767, 2009 10th International Conference on Document Analysis and Recognition, 2009
- [13] Sergios Theodoridis and Konstantinos Koutroumbas, Pattern Recognition 3/E, Academic Press, pp. 340-342, 2006.
- [14] Richard O. Duda , Peter E. Hart , David G. Stork, Pattern Classification 2/E, Wiley-Interscience, pp. 36-41, 2002
- [15] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning internal representations by error propagation," Neurocomputing: foundations of research, pp. 673-695, 1986
- [16] Christopher J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol.2 ,pp. 121-167, 1998
- [17] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, Man and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979