Recognition of Vehicle License Plate from Degraded Image via MA Model Identification and Stochastic Model

128564F Taishi Tamashiro

Syouhei Numata

128580H Subaru Inahuku

Background

security cameras

Vehicl

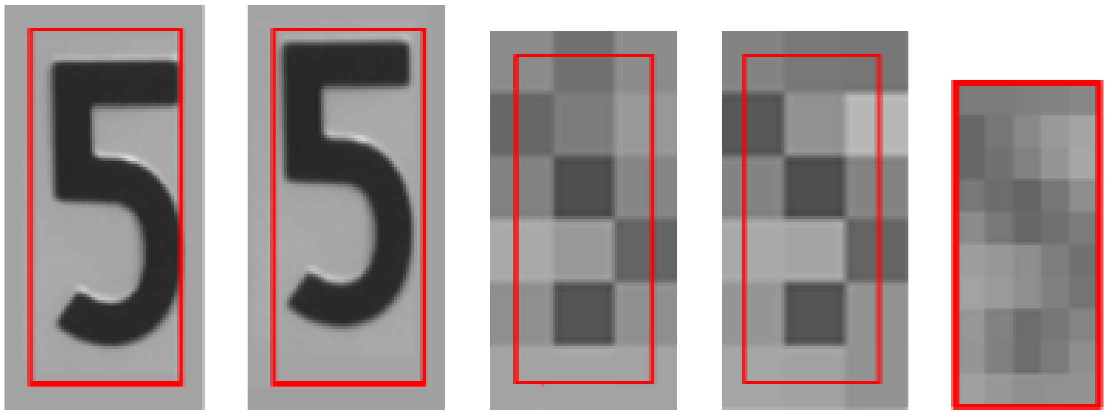
license plates

- Vehicle license plates photographed by security cameras

Recognizing vehicle license plate

A. Degradation process

- We assumed the following degradation process of which a example in Fig.2.



(a) (b) (c) (d) (e)

Fig. 2, Example of degradation process.

- (a)・・・original image
- (b)・・・location shifted image
- (c)・・・downscaled
- (d)・・・noised add
- (e)・・・position normalized image

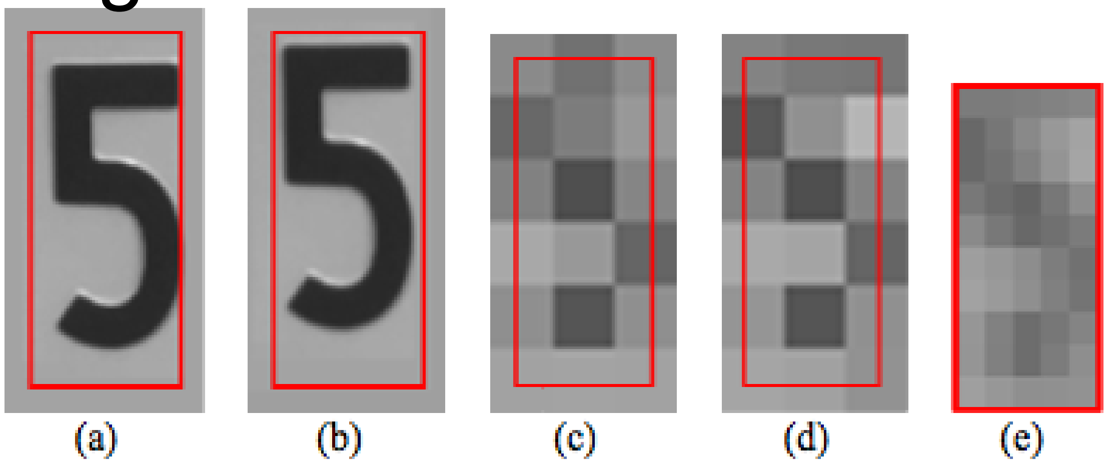


Fig. 2, Example of degradation process.

A. Degradation process

- Image degradation is caused by a Point Spread Function (PSF).
- In this paper, we pushed the PSF so as to satisfy following conditions.
- f(p,q) represents the PSF value at vertical and horizontal coordinates p and q.

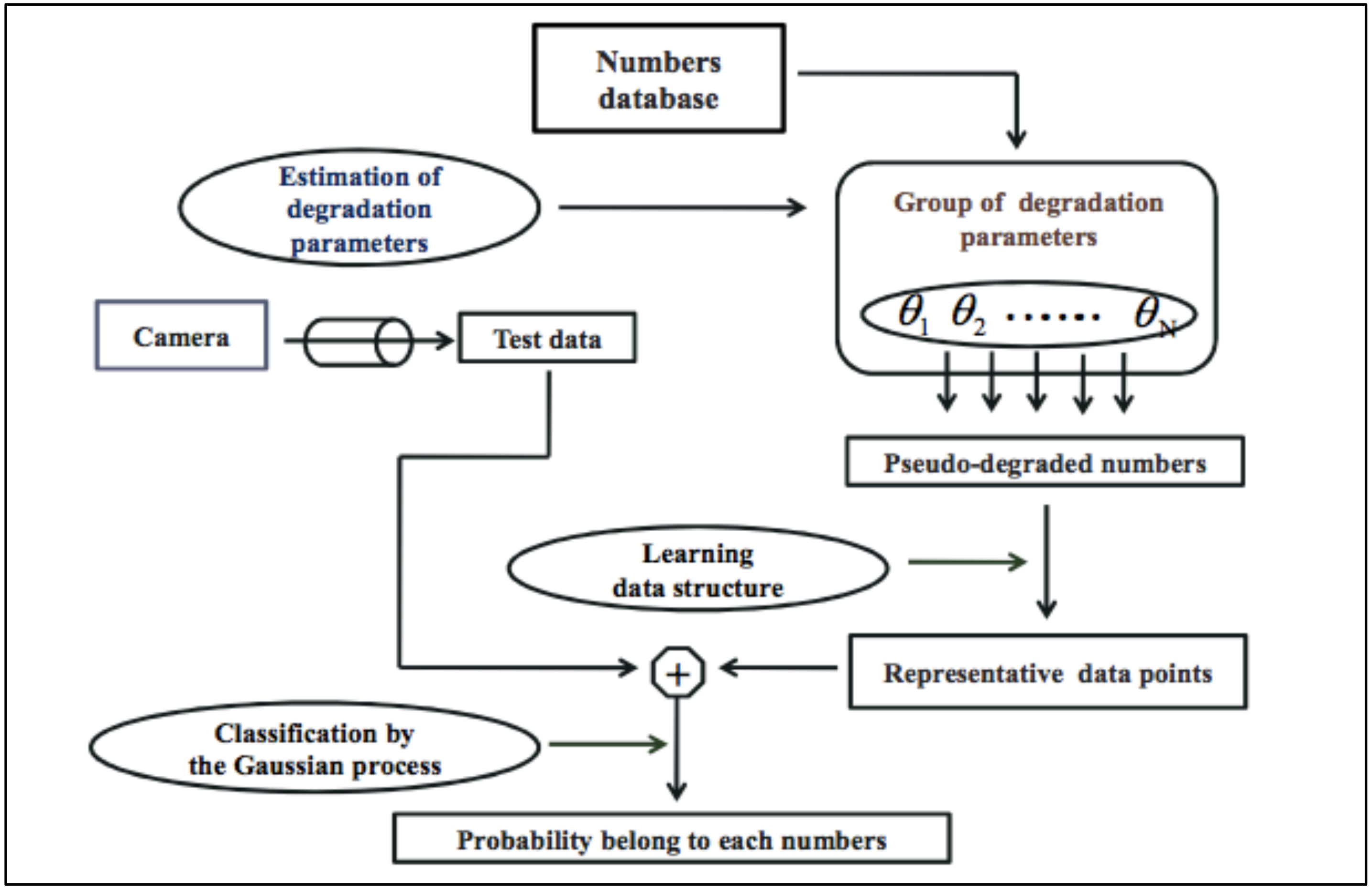
C1) $$\sum_{p=-P}^{P}\sum_{q=-Q}^{Q} f(p,q) = 1$$

$$f(p,q) \geq 0, \ (-P \leq p \leq P, -Q \leq q \leq Q)$$

A. Degradation process

- To associate above degradation process with the PSF, we prepared non-degraded image which is directly scaled down from original image to the normalized size (5×10 pixcel).
- In general, we can never solve the degradation process.
- So we attempt approximation of the PSF as parameters of MA model in the next chapter.

B.Proposed Classification system



Flow diagram of the proposed classification system

A. Classification with Gaussian processes

$N(\cdot|\cdot,\cdot)$: Gaussian distribution

$Dir(\cdot|\cdot)$: Dirichlet distribution

$W(\cdot|\cdot,\cdot)$: Wishart distribution

A^T : transpose of the matrix A

- To classify test data, we adopt Bayesian approach using the Gaussian process.
- The standard Bayesian approach is modeling test data using classifier which predicts t conditional on x.

$$P(a|\theta) = N(a|0, C) \quad \cdots (1)$$

$a = (a(x1), a(x2), \ldots, a(xn))^T$: Covariance matrix

C : Hyperparameters θ

$$P(t^*|t^u) = \int p(t^*|a(x^*))p(a(x^u)|t^u)da(x^*) \quad \cdots (2)$$

$t^U = (t(x_1^T), t(x_2^T), \ldots, t(x_n^T))^T$: Vector of classifier

$p(T|a)$: Discriminant function

$$\pi^{(j)} = p(t = j|a(x)) = \frac{exp(a^j(x))}{\Sigma_{i=1}^{M} exp(a^j(x))} \quad \cdots (3)$$

$$p(a(x^*)|t^u) = \int p(a(x^*)|a^u)p(a^u|t^u)da^u \quad \cdots (4)$$

・Considering e.q. (3)

$$p(a^u|t^u) = p(a^u)\Pi_n \Pi_i p(t|a^{(i)}(x_n)) \quad \cdots (5)$$

$$C(x_n, x_m) = exp(-\frac{1}{2\sigma^2}||x_n - x_m||) + p\sigma_{n,m} \quad \cdots (6)$$

$\delta_{n,m}$: Kronecker's delta

σ, ρ : hyperparameters

- To treat a problem with Bayesian approach, we must compute integration in e.q. (4) but the product of the soft-max function in e.q. (5) makes the integrals analytically intractable.

- Thus, we use analytic approximation based on Laplace's approximation.

- In this approach, we approximate $p(a^u|t^u)$ as the Gaussian distribution whose covariance matrix is Hessian of $-\log p(a^u|t^u)$ and mean is determined by maximizing $\log p(a^u|t^u)$ respect to a^U.

$$p(a^U|t^U) \approx N(a^U|\hat{a}, H)$$

- Hence, $\hat{a}$ is optimized in Lapalace's approximation iterative and H is following Hessian

$$H = K - W$$

K : block diagonal

$$W = diag(\pi_1^{(1)}, \dots, \pi_1^{(M)}, \dots, \pi_N^{(1)}, \dots, \pi_N^{(M)}) - \Pi\Pi^T$$

$$\Pi = diag(\pi_1^1, \dots, \pi_1^M), \dots, diag(\pi_N^1, \dots, \pi_N^M)$$

B. Learning data structure

- In previous section, we introduce classification based on the Gaussian process.
- This approach is memory based method, so all sample data are used for predicting the class of test data.

  →It takes a long times if enormous training data exist.
- Therefore, we should extract representative data points from training data.

B. Learning data structure

- We decide to analyze data structures through clustering based on variational Bayes framework.
- Assuming data structures as mixtures of the Gaussians formulation and estimation parameters of each Gaussian, we get information about data structures.
- There are many advantages in variational Bayes inference, e.g. the estimated parameters don’t fall into local solution.

B. Learning data structure

- As we adopt this approach for pretreatment of classification, we extract representative data points of each numbers by turns.
- In practice, we ignore the correlations of data vector because we thought that standard Gaussians include too many parameters to explain our datasets.
- For that reason, we used the following limited mixture of the Gaussians

$$p(\mathbf{x}) = \sum_{k=1}^{M} \pi_k N(\mu_k, \mathrm{diag}\{\alpha_k\}) \qquad (12)$$

- where { ρ0 , m0 , β0 , W0 , τ0 } are hyperparameters, respectively.

$$p(\boldsymbol{\pi}) = \mathrm{Dir}\left(\boldsymbol{\pi} \middle| \rho_0\right) \propto \prod_{k=1}^{M} \pi_k^{\rho_0 - 1} \qquad (13)$$

$$p(\boldsymbol{\mu}, \mathrm{diag}\{\boldsymbol{\alpha}\}) = \prod_{k=1}^{M} N\left(\boldsymbol{\mu}_k \middle| \mathbf{m}_0, \beta_0^{-1} \mathrm{diag}\{\boldsymbol{\alpha}_k\}\right) W(\mathrm{diag}\{\boldsymbol{\alpha}_k\} | W_0, \tau_0) \quad (14)$$

•And we set {z} as latent parameters of {x}.

•We estimate πk,μk, αk (for all k) and {z} alternately

•We regard μk as representative data points and use {μk} for training data of each number.

- As a result of integration in e.q. (4), we get $p(a(x^*)|t^U)$ as Gaussian distribution whose mean and covariance are estimated by the following equations.

$$E[a(x^*|t^U)] = k^T(t^U - \pi^U)$$

$$var[a(x^*|t^U)] = c^* - k^T(W^{-1} + K)^{-1}k$$

$$k = (d_1(x_1, x^*), \ldots, d_1(x_N, x^*), \ldots, d_M(x_1, x^*), \ldots, d_M(x_N, x^*))$$

- we generate samples of x^* from Gaussian distribution $p(a(x^*)|t^U)$ and compute $E[p(t^*)|t^U)]$ by calculating sum of the soft-max functions $p(t^*|a(x^*))$.

C. Identification of MA model

- In this section, we introduce how to apply MA model identification for degraded images. Simply we define MA model as following forms.
- where s(i,j), r(i,j) are pixel value of the image which are subtracted to turn their means to 0, {d(i,j)} are the parameters of MA model and the notation (i,j) shows the position of the image.
- s(i,j) indicates pixel of observed image and r(i,j) indicates pixel of original image

$$s(a,b) = \sum_{j=-L}^{L}\sum_{i=-L}^{L} d(i,j)r(a-i,b-j) \quad (15)$$

C. Identification of MA model

- In time-series analysis, many methods, such as the correlation analysis based, may be used to estimate parameters of the MA model by assuming that objected signals are stationary, i.e.

$$\text{A1)} \qquad E[r(a,b)] = 0, \quad E[r(a,b)r(s,t)] = 0$$

$$(\text{ except for } a = s,\ b = t\ )$$

$$E[r(a,b)r(a,b)] = \sigma^2(\text{constant}).$$

- In this case, we get more than L×L equations by shifting position in e.q. (15).

C. Identification of MA model

- Our assumption is more complex than A1. In our problem, we regard signals as non-stationary, i.e.

$$\text{A2)} \qquad E[r(a,b)] = 0, \;\; E[r(a,b)r(s,t)] = 0$$

$$(\text{ except for } a = s, b = t)$$

$$E[r(a,b)r(a,b)] = \sigma^2(a,b).$$

•In this case, correlation analysis based methods are no longer available to determine parameters because we can’t get enough equations.

C. Identification of MA model

- Therefore we adopt the method based on the gradient algorithm. Here, let us introduce new notations

$$\boldsymbol{\theta}(t) = [d(L, L), \cdots, d(-L, -L)]^T \quad (17)$$

$$\mathbf{r}(a, b) = [r(a - L, b - L), \cdots, r(a + L, b + L)]^T \quad (18)$$

$$\mathbf{v}(a, b) = [v(a - L, b - L), \cdots, v(a + L, b + L)]^T \quad (19)$$

C. Identification of MA model

- Using these notations, we get the following equation
- where e(a,b) indicates error value.

$$s(a,b) = \mathbf{r}(a,b)^T \boldsymbol{\theta}(t) + \mathbf{v}(a,b) \qquad (20)$$

$$e(a,b) = s(a,b) - \mathbf{r}(a,b)^T \boldsymbol{\theta}(t) \qquad (21)$$

C. Identification of MA model

- By shifting position, these notations are expressed as

$$W(t) = [\mathbf{e}(-N,-N),\cdots,\mathbf{e}(N,N)]^T \quad (22)$$

$$S(t) = [s(-N,-N),\cdots,s(N,N)]^T \quad (23)$$

$$V(t) = [v(-N,-N),\cdots,v(N,N)]^T \quad (24)$$

C. Identification of MA model

- where e(a,b) = [e(a-L,b-L),···, e(a+L,b+L)] and we assume N >> L. Summarizing e.q. (20) and (21), we get (25)

$$S(t) = W(t)^T \theta(t) + V(t) \qquad (25)$$

C. Identification of MA model

- We estimate θ(t) by minimizing the cost function J(θ) whose form is a quadratic criterion function,

$$J(\theta) := \|S(t) - W(t)^T \theta(t)\|^2 \qquad (26)$$

C. Identification of MA model

- using the following gradient decent method.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \delta(t)W(t)\left[R(t) - W(t)^T\hat{\theta}(t-1)\right] \quad (27)$$

- where δ(t) is the step size of iterative. Here, δ(t) must be determined to guarantee the convergence of θ. We determine δ(t) by the following equation.

$$\gamma(t) := \gamma(t-1) + \mathrm{tr}[W(t)^T W(t)]$$

$$\delta(t) = \frac{1}{\gamma(t)}, \quad \gamma(0) = 1 \qquad (28)$$

Ⅳ. EXPERIMENTAL DEMONSTRATION

Experiment 1

- Estimate degradation parameters from artificially-degraded image.
- Testing stability of degradation parameters estimated by proposed method.
- Used image:
  - The normalized image without degradation
  - The degraded image (degradation parameters presented in TABLE I)

TABLE I

EXAMPLE OF TRUE DEGRADATION PARAMETERS

0.0042	0.0142	0.0260	0.0142	0.0042
0.0142	0.0616	0.0935	0.0616	0.0143
0.0260	0.0935	0.1442	0.0935	0.0260
0.0142	0.0616	0.0935	0.0616	0.0142
0.0042	0.0142	0.0260	0.0142	0.0042

Experiment 1

- To compare proposed algorithm and the other method, we tested the Least Squares method (LSM).
- We minimized Lagrange function L(θ) instead of J(θ).

$$L(\boldsymbol{\theta}) := \|S(t) - W(t)^T \boldsymbol{\theta}(t)\|^2$$
$$+ \varepsilon(\|\theta\|^2 - 1) + \sum_{j=-L}^{L} \sum_{i=-L}^{L} \varepsilon_{i,j} d(i, j)$$

TABLE II

ESTIMATED PARAMETERS BY PROPOSED METHOD

0.0074	0.0310	0.0461	0.0296	0.0063
0.0126	0.0501	0.0743	0.0501	0.0118
0.0408	0.0796	0.1114	0.0791	0.0394
0.0138	0.0500	0.0746	0.0500	0.0115
0.0095	0.0329	0.0466	0.0329	0.0074

These values are more diverged than true parameters.

TABLE III

ESTIMATED PARAMETERS BY LSM

0.0080	0.0338	0.0054	0.0044	0.0134
0.0084	0.0665	0.0944	0.0710	0.0103
0.0162	0.0816	0.1250	0.0990	0.0108
0.0254	0.0666	0.0826	0.0706	0.0211
0.0122	0.0352	0.0142	0.0145	0.0084

These values are biased and spreading asymmetrically.
These factors bring instability in classification of numbers.

Experiment 2

- Use the pseudo-degraded images as test data.
- Verify difference among proposed classification method and the other classification methods.

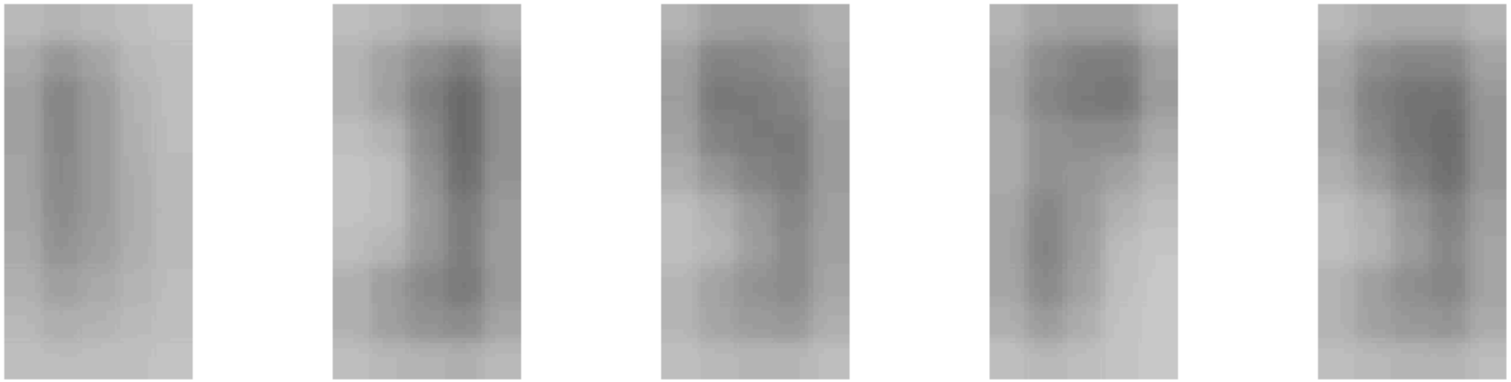


Fig. 4. Low resolution images which we scaled down.

Experiment 2

- Use 2,000 pseudo-degraded number pictures as training data
- 200 downscaled number pictures as test data for each resolution.
- Use performance of:
  - Support Vector Machine (SVM)
  - Linear Discriminant Analysis (LDA) method
  - classification with human eyes.

TABLE IV

RESULT OF CLASSIFICATION

	GP	SVM	LDA	Eyes
$w = 3.5$	92.5	95.5	66.0	95.0
$w = 3.0$	92.0	95.0	70.5	68.0
$w = 2.5$	90.5	90.0	76.0	64.0
$w = 2.0$	89.5	85.5	81.0	32.0

- LDA method does good work in low resolution than high resolution.
- Recognition rates of SVM decrease rapidly according to scaling down.
- those of the proposed method are almost constantly 90 %, nevertheless we down scaled their size.

Ⅴ. CONCLUSION

- This study proposed a system which recognizes numbers in degraded images.
- The problem includes both classification of numbers and estimation of degradation process.
- We focus on classification with the Gaussian process and MA model identification.
- As an advantage of our proposed system, instead of using training data, we can use pseudo-degraded number with estimated parameters.

The experiments result

- The proposed algorithm of identification can estimate degradation parameters stably .
- Our classification method is efficient for recognition of low resolution numbers which are difficult to discriminate with eyes.

Future work

- Research new method to handle more degraded images caused various factors.