

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

EIC | **電子情報通信学会** **D** | **論文誌**

情報・システム

VOL. J103-D NO. 4
APRIL 2020

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。
なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

情報・システムソサイエティ

一般社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

領域画像マッチングのための目標領域特徴量による初期位置探索法

山下 幸彦^{†a)} 若原 徹^{††b)}

Initial Location Search Method using Target Region Features for Region-Based Image Matching

Yukihiko YAMASHITA^{†a)} and Toru WAKAHARA^{††b)}

あらまし 画像マッチングは、画像処理におけるもっとも重要な手法の一つであり、特徴画像マッチングと領域画像マッチングに大別される。領域画像マッチングは高精度なマッチング能力を有するが、whole-to-part マッチングに適用する場合、マルチスケールでスライド窓探索を行うと処理時間が膨大化するという問題がある。本論文では領域画像マッチングのための高精度かつ高速な初期位置探索の実現を目指す。具体的には、画像劣化や幾何学的変換に頑健となるようにスライディング離散フーリエ変換を工夫した目標領域特徴量に基づく初期位置探索法を提案する。特徴画像マッチングの代表的手法である ASIFT/ASURF と RANSAC の組合せによる初期位置探索と提案手法の比較実験を行い、提案手法の優位性を示す。

キーワード 画像マッチング、特徴画像マッチング、領域画像マッチング、初期位置探索

1. ま え が き

画像マッチングは、物体検出、パターン認識などに応用され、画像処理における最も重要な手法の一つである。画像マッチングを実用的なものにするためには、入力画像での照明条件の変化、物体位置の変化や回転などの幾何学的変換、更には物体の非剛体変形に対して、テンプレート画像のマッチングが頑健である必要がある。そのために、多くの画像マッチング手法が提案されている。それらは大きく分けて、画像から特徴点を抽出し、それら特徴点の整合性を評価する特徴画像マッチングと、画像領域の整合性を直接評価する領域画像マッチングの二つに大別される。特徴画像マッチングでは、Scale-Invariant Feature Transform (SIFT) [1], [2] や Speeded-Up Robust Features (SURF) [3], [4] に基づく手法が広く用いられている。特に、SIFT を Affine 変換に関して頑健にした ASIFT [5] と RANSAC [6]

の組合せは、実用的で頑健な特徴画像マッチングを可能にしている。一方、領域画像マッチングでは、ガウス平滑化や固有値展開によりパターン変動を吸収した飯島による複合類似度 [7] や、変形パラメータを陽に反復解法で求める Lucas-Kanade アルゴリズム [8] に始まり、現在でもその性能を高めるべく研究が続けられている [9]。若原らによる GAT/GPT (Global Affine Transformation/Global Projection Transformation) 相関法 [9]~[11] は、アフィン変換/2 次元射影変換パラメータを含む正規化相互相関を目的関数として、それを微分して得られる非線形方程式を線形近似して解き、それを繰り返す反復解法で最適変換パラメータを決定している。領域画像マッチングは高精度なマッチングが可能であり、認識タスクでも高い性能が報告されている。GPT 相関法に同一エッジ方向平均最小距離を用いた k NN による MNIST の手書き数字認識 [10] では認識率 99.7% が達成されており、これは k NN に基づく手法 [12] としては非常に高い値となっている。

しかし、領域画像マッチングの場合、マッチング性能で優位に立てても、処理時間が膨大化する問題があった。実際、反復解法を用いる場合はなおさら、入力画像中に埋め込まれた物体を検出するためには、テンプレート画像をマルチスケールのスライド窓探索

[†] 東京工業大学, 東京都

Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552 Japan

^{††} 法政大学, 小金井市

Hosei University, 3-7-2 Kajino-cho, Koganei-shi, 184-8584 Japan

a) E-mail: yamasita@ide.titech.ac.jp

b) E-mail: wakahara@hosei.ac.jp

DOI:10.14923/transinfj.2019JDP7056

でマッチングする処理が必要となり、処理時間が増える。特に、マルチスケールでのスライド窓探索の時間を短縮するには、効率的な初期位置探索が有力である。初期位置探索には特徴画像マッチングを用いるのが自然である。ただし、特徴画像マッチングが有効に働くような画像マッチングのタスクに対して、特徴画像マッチングで得られた結果を取って領域画像マッチングと組み合わせることには余り意味がない。特徴画像マッチングで対応が難しいようなタスクを解決できるような領域画像マッチングにおける初期位置探索が課題である。その意味で、特徴画像マッチング自身の性能向上が必要となる。特に、特徴点のスケールが目標領域のスケールにより小さい場合に画像劣化に対する頑健性が不足となる。近年、ボケやブレに頑健な特徴点抽出法 [13] や特徴量 [14] も提案されているが、様々な変形や劣化に対する性能検証が十分になされているとはいえない。また、一般物体認識の先駆的研究として、dense SIFT [15] を用いたピラミッド型記述により ROI (Region of Interest) マッチングを実現するアプローチがある。更にその発展形として dense ASIFT [16] や SIFT Flow [17] が提案されているが、これらは画像検索や Optical Flow 推定を目的としており、本研究が目指す領域画像マッチングのための初期位置探索とは目的が異なる。

本論文では、SIFT 特徴量の考えに基づいた目標領域特徴量による、領域画像マッチングのための初期位置探索法を新たに提案する。目標領域特徴量に基づく手法は、スケールが大きく様々な劣化に対して頑健であるという利点があるが、初期位置探索法としての研究例はない。提案手法ではスライディング離散フーリエ変換を用いたガウス平滑化 [18]～[21] を利用した高速計算法 [22] を活用する。これにより、計算量のオーダーが平滑化量に依存せず、画素数に比例したものになる。特に、初期位置探索に設定するパラメータ数を最小化し、更に条件分岐がほとんどないループの単純構成によるアルゴリズムのスリム化を追求し、なおかつ計算量が対象画像の内容にほとんど依存しないという汎用性も実現している。計算機実験では、提案手法に対して、ASIFT/ASURF と RANSAC を組み合わせた手法による初期位置探索の性能を比較した。比較内容は、1) ガウス平滑化、移動平均、白色加法性雑音の重畳による入力画像の劣化に対する頑健性、2) 特徴点数が少ないテンプレート画像に対する頑健性、3) 計算時間、である。実験結果より、初期位置探索を目的

とする提案手法の有効性を示す。

2. 有限区間三角関数の畳み込み和によるガウス平滑化法

本章では、有限区間三角関数の畳み込み和によるガウス平滑化法 [19] において、微分画像に対する誤差、及び GPU での実装を視野に入れた単精度計算における誤差、についての詳細な定量的評価を行う。こうした評価はこれまでなされたことはなく、これにより本論文で提案する目標領域特徴量抽出への適用可能性を確認する。

2.1 ガウス平滑化の近似実現

1 次元離散点上の関数で記述する。ガウス関数は標準偏差の数倍で急速に 0 に近づくため、有限区間 $[-K, K]$ で考え、 $G[n] = e^{-(\pi n/K)^2/(2\sigma^2)}$ による平滑化を考える。 σ は、区間両端を $\pi, -\pi$ と考えたときの標準偏差で、画像上での標準偏差は $\sigma K/\pi$ である。関数 $f[n]$ に対して、ガウス平滑化やその 1 階、2 階微分の近似を求めるためには、次の値が求まれば良い。

$$\sum_{k=-K}^K G[k]f[n-k] \quad (1)$$

$$\sum_{k=-K}^K kG[k]f[n-k] \quad (2)$$

$$\sum_{k=-K}^K k^2G[k]f[n-k] \quad (3)$$

簡単のため、 $\beta = \pi/K$ とおき、次の近似を考える。

$$G[k] \simeq \sum_{p=0}^P a_p \cos(\beta pk)$$

$$kG[k] \simeq \sum_{p=1}^P b_p \sin(\beta pk)$$

$$k^2G[k] \simeq \sum_{p=0}^P d_p \cos(\beta pk)$$

P は計算量を減らすため、DFT を打ち切る次数である。ここでは、 P として 2 から 4 程度を想定している。

有限区間三角関数の畳み込み和を次式で定義すれば、

$$c_p[n] = \sum_{k=-K}^K f[n-k] \cos(\beta pk)$$

$$s_p[n] = \sum_{k=-K}^K f[n-k] \sin(\beta pk)$$

式 (1)-(3) の近似は、次式で表すことができる。

$$\sum_{k=-K}^K G[k]f[n-k] \simeq \sum_{p=0}^P a_p c_p[n]$$

$$\sum_{k=-K}^K kG[k]f[n-k] \simeq \sum_{p=1}^P b_p s_p[n]$$

$$\sum_{k=-K}^K k^2 G[k]f[n-k] \simeq \sum_{p=0}^P d_p c_p[n]$$

図 1 に、標準画像 Lenna を、有限区間三角関数の畳み込み和によりガウス平滑化した結果を示す。

2.2 有限区間三角関数の畳み込み和の高速計算
簡単のために、複素指数関数を使って表す。Kober [18] は、カーネル積分

$$u[n] = \sum_{k=1}^n f[k] e^{i\beta p k} \quad (4)$$

を使って、有限区間三角関数の畳み込み和 $c_p[n]$ と $s_p[n]$ を、次式で求めた。

$$\begin{aligned} c_p[n] - i s_p[n] \\ = e^{-i\beta p n} (u[n+K] - u[n-K-1]) \end{aligned} \quad (5)$$

この $u[n]$ は、次の漸化式で計算することができる。

$$u[n] = u[n-1] + f[n] e^{i\beta p n} \quad (6)$$

この式では、三角関数を 1 周期分記憶して、信号にかけていく必要がある。少し変形して、IIR フィルタの形にする。

$$v[n] = e^{-i\beta p} v[n-1] + f[n] \quad (7)$$

とすれば、 $c_p[n]$ と $s_p[n]$ は、次式で求まる。

$$\begin{aligned} c_p[n] - i s_p[n] \\ = (-1)^p (v[n+K] - v[n-K] + x[n-K]) \end{aligned} \quad (8)$$

また、積分値でなく、直接、区間和を計算すること

もできる。こうする方が数値計算精度は向上するが、 K , p を整数倍した場合への、積分値の再利用はできなくなる。

杉本ら [21] は 2 次のフィルタを使うことを提案している。

$$\begin{aligned} v[n+1] &= 2 \cos(\beta p) v[n] + v[n-1] \\ &\quad + f[n+1] - e^{i\beta p} f[n] \end{aligned} \quad (9)$$

この式では、実数部と虚数部が分離しているため、 $c_p[n]$ と $s_p[n]$ を別々に計算することが可能である。したがって、 $c_p[n]$ だけが欲しい場合などに計算量を削減することができる。しかしながら、2 階の微分方程式を解く形になっており、単精度浮動小数点演算で計算したときに、計算精度に問題が生じる場合が考えられる。

2.3 打ち切り誤差

区間長や次数を P で打ち切るため、ガウス関数の近似に誤差が生じる。画像上での標準偏差は $\sigma K/\pi$ を一定にしても、 K と σ に自由度が生じるが、打ち切り誤差は σ に大きく依存し、 K にはあまり依存しない。

図 2 は、 $P=2$ と $P=4$ のときの、 $G[n]$, $nG[n]$, LoG の近似誤差を、平均 2 乗誤差ともの関数の 2 乗平均の比の平方根で示したものである。

$P=2$ の場合、 $\sigma=1.1$ とすれば、誤差はガウス関数で 1.1%、微分ガウス関数で 5.3%、LoG で 8.2% 程度となる。 $P=4$ の場合、 $\sigma=0.85$ とすれば、それぞれで、0.049%、0.32%、0.68% 程度となる。

2.4 単精度浮動小数点演算誤差

倍精度浮動小数点演算を使えば、有限区間三角関数の畳み込み和の計算は十分な精度で計算できるが、今後重要になる安価な GPU は単精度浮動小数点演算器しか積んでいないため、単精度浮動小数点演算による精度評価を行う。標準画像 Lenna と Barbara をそれぞれ、1 次元信号として考え、連結した 524288 (= $512 \times 512 \times 2$) 次元のベクトルを入力とした。図 3 に、

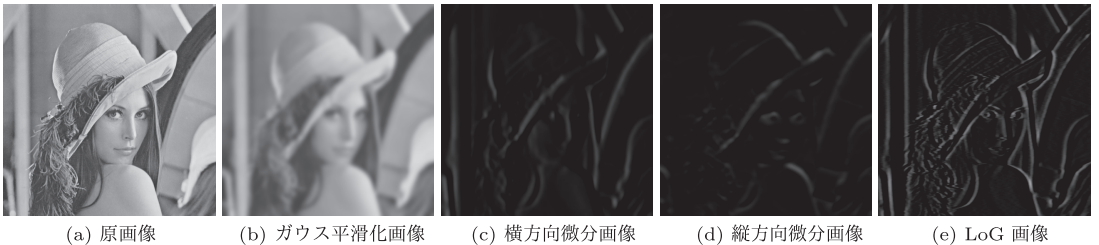


図 1 三角関数のスライディング畳み込み和によるガウス平滑化法

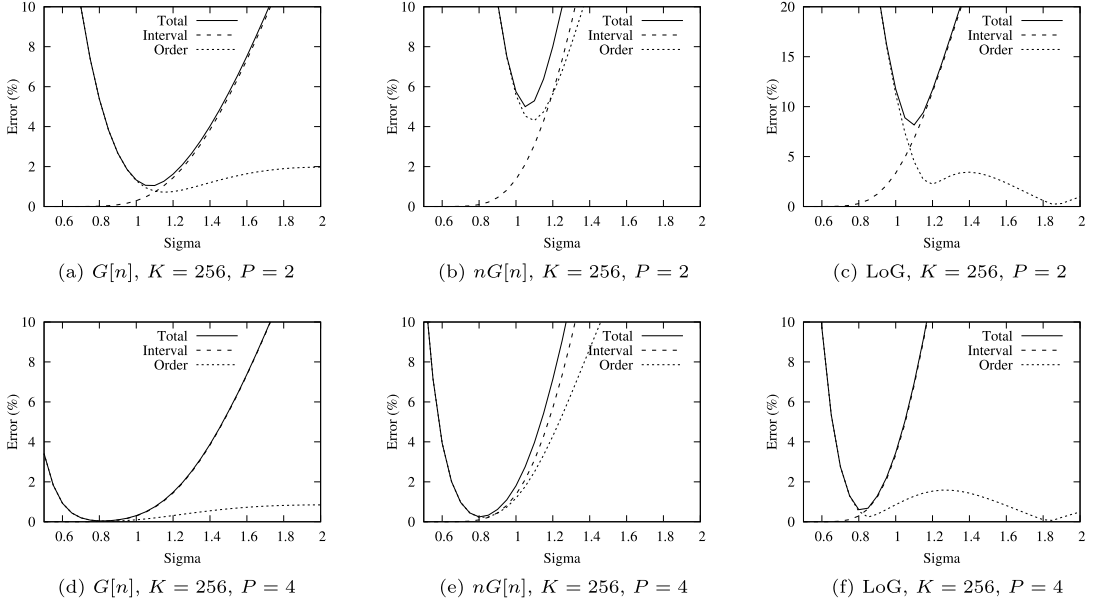


図 2 打ち切り誤差 (Interval: 区間打ち切り, Order: 次数打ち切り.)

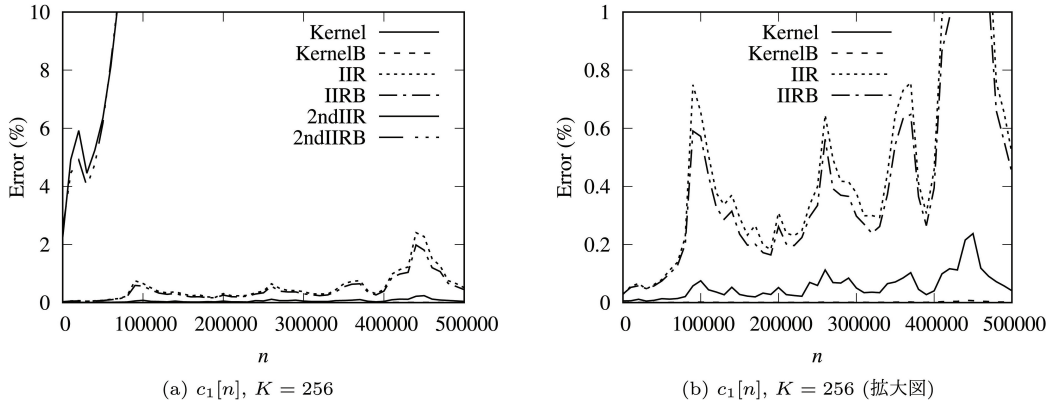


図 3 単精度計算誤差 (Kernel: カーネル積分 (式 (6)), IIR: 1 次 IIR (式 (7)), 2ndIIR: 2 次 IIR (式 (9)). B を付加したものは, 区間積分で計算.)

$K = 256$ のときの, $c_1[n]$ の区間長 10,000 での最大誤差と, 同区間での対象信号の 2 乗平均の平方根との比を示す.

単精度浮動小数点の計算では, 2 次 IIR では, 初めの 1 万点で 2% 程度の誤差があり, 十分とは言えない.

3. 提案手法

本章では, まず基本となるフーリエ記述エッジ方向ヒストグラム [23] について説明するが, 提案手法では GPU での実装を視野に入れ, 三角関数やその逆関数, 更に条件分岐を行わない簡便な計算方法を新たに導入

する. また, 特に, 回転共変性に注目する. 続いて, 提案手法の要となる目標領域特徴量について説明し, 具体的な計算法を示す. 最後に, 目標領域特徴量に基づく初期位置探索の手順を説明する.

3.1 フーリエ記述エッジ方向ヒストグラム

エッジ方向ヒストグラム $D(\theta)$ は, ある領域内の θ 方向のエッジの強さの (重み付き) 和である. これを, Q 次のフーリエ級数で記述する.

$$D(\theta) = \sum_{q=-Q}^Q d_q e^{iq\theta} \quad (10)$$

ここで、 $d_{-q} = \overline{d_q}$ が成立するため、係数 d_q の実数での自由度は $2Q + 1$ である。

入力画像 $I(x, y)$ をガウス平滑化し、横と縦方向に微分した画像を $I_X(x, y)$, $I_Y(x, y)$ とする。

$$\|I_1(x, y)\| = \sqrt{(I_X(x, y))^2 + (I_Y(x, y))^2} \quad (11)$$

とおけば、点 (x, y) におけるヒストグラムは、ディラックのデルタ関数 $\delta(\theta)$ を使って、 $\|I_1(x, y)\|\delta(\theta - \theta(x, y))$ である。このとき、 $I_X(x, y) = \|I_1(x, y)\| \cos \theta(x, y)$, $I_Y(x, y) = \|I_1(x, y)\| \sin \theta(x, y)$ が成立する。文献 [23] では、 $\arctan$, $\sin$, $\cos$ 関数と条件分岐を使って各点のフーリエ係数を求めている。本研究では、

$$\begin{aligned} d_q &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \|I_1(x, y)\| \delta(\theta - \theta(x, y)) e^{-iq\theta} d\theta \\ &= \frac{1}{2\pi} \|I_1(x, y)\| e^{-iq\theta(x, y)} \\ &\simeq \frac{\sqrt{\|I_1(x, y)\|^2 + \varepsilon}}{2\pi} \left(\frac{I_X(x, y) - iI_Y(x, y)}{\sqrt{\|I_1(x, y)\|^2 + \varepsilon}} \right)^q \end{aligned}$$

を使って求める。ここで、 ε は 0 による除算を防ぐための定数である（実験では、 $\varepsilon = 10^{-12}$ ）。この式を使えば、平方根の計算は必要であるが、量子化や三角関数の計算を行うことなく、フーリエ記述エッジ方向ヒストグラムを求めることができる。

また、エッジの基準方向を θ_0 回転したとき、回転したヒストグラムの係数 d'_q は、次式で得ることができる。

$$d'_q = d_q e^{-iq\theta_0} \quad (12)$$

ガウス平滑化を使ってエッジ方向ヒストグラムをフーリエ記述する利点は、回転に対する共変性である。例えば、ヒストグラムの最大位置を検出する場合、画像を 1 度回転させれば、検出最大位置もちょうど 1 度ずれることになる。その最大位置は次数による分解能の制限のため本当の最大位置とは限らないが、位置検出の基準方向のためには共変性で十分である。また、SIFT では特徴点抽出時に、基準方向を求めるために 36 方向に量子化するが、これを提案手法に直接適用するとヒストグラムを求めるためには平滑化を 36 回行う必要がある。フーリエ記述で $Q = 4$ とすれば、係数の数は 9 個であり、9 回の平滑化で済むことになり、計算量を削減することができる。この回転共変性については、4.1 の実験で確かめる。

3.2 目標領域特徴量

SIFT の特徴量の考え方に基づいた目標領域特徴量

に関して説明する。ただし、回転・スケール不変な目標領域特徴量の抽出について計算量の観点から汎用的に論じたのは本研究が初めてである。

入力画像 I の点 (x, y) において、スケール S の目標領域特徴量を記述する。 S は、ガウス関数の標準偏差で、その領域のおよその半径である。まず、 $P = 4$ とした有限区間三角関数の畳み込み和によるガウス平滑化法により、 I からスケール $S/8$ でガウス平滑化し、横と縦方向に微分した画像 I_X , I_Y を求める。正弦成分を使うことにより、ガウス平滑化と微分は同時に行うことができる。そして、画像の各点で、式 (12) を使って、フーリエ記述エッジ方向ヒストグラムを求める。その $2Q + 1$ の係数を、 $P = 2$ とした有限区間三角関数の畳み込み和によるガウス平滑化法により、スケールを S 及び $S/4$ として平滑化したものをそれぞれ求める。

提案手法の特徴として、画像からこのフーリエ記述エッジ方向ヒストグラムの計算まで、繰り返しのため以外に条件分岐を必要とせず、平滑化のための係数をテーブルにしておけば三角関数の計算も不要で、簡単なアルゴリズムで実装することができることである。

(x, y) におけるスケール S のフーリエ係数から、一周を L 等分した $\theta = 2\pi l/L$ ($l = 0, 1, \dots, L-1$) におけるヒストグラムの値を求め、それが極大となる方向を、ヒストグラムの値が大きい方から二つ求める。それぞれの角度がその対象領域の基準角度であり、特徴量はその角度で正規化して求める。両者のヒストグラムの値の比が 1 に近いときには、二つの角度に対する特徴量をそれぞれ求める。比が 1 から離れているときには、最大の角度に対してだけ特徴量を求める。

(x, y) での基準角度の一つを、 θ_0 とする。提案する目標領域特徴量は、 θ_0 で正規化したときの、 (x, y) の周りの M 点における、スケール $S/4$ でガウス平滑化したフーリエ記述エッジ方向ヒストグラムを並べたものである。

図 4 は、特徴量記述に使う原点を中心とした周辺の M 点の正規化座標 (x_m, y_m) ($m = 1, 2, \dots, M$) を図示したものである。実際にヒストグラムを特徴量とする点の画像上の座標 (x'_m, y'_m) ($m = 1, 2, \dots, M$) は、次式で与えられる。

$$x'_m = S(x_m \cos \theta_0 - y_m \sin \theta_0) + x \quad (13)$$

$$y'_m = S(x_m \sin \theta_0 + y_m \cos \theta_0) + y \quad (14)$$

スケール $S/4$ のフーリエ記述エッジ方向ヒストグラム

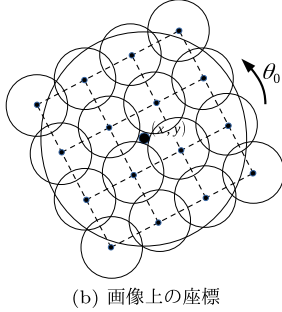
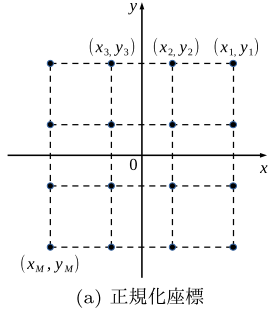


図 4 ヒストグラムを特徴量とする点の位置

は全ての点で計算済みであるため、その値を抽出し、角度正規化のためにヒストグラムを式 (12) で回転させる。これらの係数を並べ、 $M(2Q+1)$ 次元のベクトルが得られる。このベクトルをノルムを 1 に正規化したものが、 (x, y) での最終的な特徴量となる。

特徴量を求める点 (x, y) は、 S に比例 (比例係数は 0.2 程度) した値で画像中を格子点状に走査する。

画像から特徴量を求めるアルゴリズムを Algorithm 1 にまとめる。

3.3 目標領域特徴量に基づく初期位置探索法

提案する目標領域特徴量に基づく初期位置探索法の基本的方法は以下のとおりである。テンプレート画像の抽出したい物体領域の中心を、特徴量を抽出する点 (x, y) とし、そのおよその半径をスケール S として、特徴量抽出を行い、その特徴量を記憶する。その特徴量をテンプレート特徴量と呼ぶ。入力画像を複数スケールと格子点走査で特徴量を求め、その中から、テンプレート特徴量との内積が大きい特徴量を出力した座標を、検出位置とする。また、両者の基準角の差は、物体領域の回転角と考えることできる。後段処理のために、1 点ではなく複数の候補点の抽出を行いたい場合は、内積が大きい順番に複数個の特徴量を抽出することもできる。

Algorithm 1 目標領域特徴量計算法

Require: 入力画像 I , 初期スケール S_0 , スケール更新比 r_S , 最大スケール S_{Max} , スケール S において特徴量を求める点の集合 T_S , 特徴量となるヒストグラムを利用する点の特徴量を求める点からの正規化座標 (x_m, y_m) ($m = 1, 2, \dots, M$)

$S \leftarrow S_0$

while $S \leq S_{\text{Max}}$ **do**

I をスケール $S/8$ でガウス平滑化し、横と縦方向に微分した画像 I_X, I_Y を求める。

画像の各点でフーリエ記述エッジ方向ヒストグラムの係数を求める。

$2Q+1$ 個の係数を、スケール S 及び $S/4$ でガウス平滑化する。

for $(x, y) \in T_S$ **do**

スケール S で平滑化した係数を使って、基準方向を (複数) 求める。

for $\theta_0 \in (\text{得られた基準方向の集合})$ **do**

基準方向 θ_0 に対して、式 (13), (14) を使い、ヒストグラムを利用する点の座標を求める。

それらの点におけるスケール $S/4$ のフーリエ記述エッジ方向ヒストグラムを、式 (12) で回転させ、並べて $M(2Q+1)$ 次元のベクトルを得る。

得られたベクトルのノルムを 1 に正規化して、 (x, y) における特徴量の一つとして記憶する。

end for

end for

$S \leftarrow r_S S$

end while

しかしながら、画像の変形量が大きいときには、特徴ベクトルが異なってしまう、位置検出精度が下がることが考えられる。そこで、ASIFT の考え方を使得、画像を Affine 変換し、特徴量を抽出して位置検出させることを考える。入力画像を変形させると、全体の特徴抽出の計算時間が、変形しない場合の計算時間に変形パターン数をかけた値になってしまう。そこで、ここでは入力テンプレートだけ変形させることにする。そして、対象物体に対する特徴量を複数個用意し、入力画像の特徴量との内積を計算し、最大となるものを選び出す。内積の計算時間が増えることになるが、全体の計算時間が大きく変わることはない。

また、スケールに関しても、入力画像のスケールの種類におよそ比例して特徴抽出の計算時間が増加する。そこで、スケールの小さな変化に関してはテンプレートの方で対応することにする。例えば、実験では、入力画像に対するスケール変化は 2 倍 ($r_S = 2$) とし、テンプレートでは、 S_0 と $\sqrt{2}S_0$ の 2 種類のスケールを用いる。このようにすると、厳密には初期位置探索のための目標領域の大きさが多少異なることになるが、スケールを $\sqrt{2}$ ずつ変化させた初期位置探索が実現できる。

4. 計算機実験

4.1 フーリエ記述エッジ方向ヒストグラムの回転共変性

3.1 で説明した、フーリエ記述エッジ方向ヒストグラムの回転共変性を、最大位置検出の実験で示す。画像には Lenna を使い、線形補間による座標変換で画像を 1 度づつ回転させ、中心座標 (256, 256)、スケール 60 画素で $Q = 4$ のフーリエ記述エッジ方向ヒストグラムの最大位置を、1/16 度の精度で検出した。その結果を図 5 に示す。Lenna の原画像において、1/16 度の精度で検出した $Q = 4$ のフーリエ記述エッジ方向ヒストグラムの最大位置は 31.0625 度である。したがって、画像を回転させたときに検出される最大位置が、31.0625 と回転角を加算したものになれば、回転共変性が示されたことになる。図 5 から、フーリエ記述エッジ方向ヒストグラムによる最大位置検出に回転共変性があることが確認できる。また、検出した最大点の角度と画像回転角の差は、 31.0625 ± 0.0625 に収まっている。

4.2 初期位置探索

提案手法の性能を調べるために、画像を探索し目標とする領域のおよその位置を検出する実験を行った。また、その結果を ASIFT/ASURF+RANSAC と比較した。ASURF は、SIFT に対する ASIFT と同様に、Affine 変換した画像を利用することによって、SURF によるマッチングを Affine 変換に関して頑健にしたものである。実験に用いる画像データセットは、Graffiti (800×640 画素)、Boat (850×680)、Bark (765×512 画素)、UBC (800×640) [24] である。Graffiti は、グレイスケール画像に変換して用いた。それぞれ、6 種類の画像 (img1～img6) からなる。目標領域は、Graffiti では img1 の人物の領域、Boat では img1 のポート

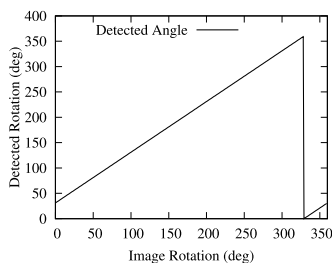


図 5 フーリエ記述エッジ方向ヒストグラムによる最大位置の回転共変性

の領域、Bark では左上の突出した領域、UBC では中心の窓の領域とするために、検出目標座標及び初期スケールを、それぞれ、(430, 310) 及び 150, (425, 340) 及び 200, (350, 255) 及び 100, (410, 240) 及び 100 とした。そして、テンプレート画像に、検出目標位置を中心として無変換も含めて 9 種類の Affine 変換を施し、初期スケールとその $\sqrt{2}$ 倍のスケールを用いて特徴量を求めた。入力画像には、img2～img6 を用いた、4 種類のデータベースの入力画像に対するスケールの組は、それぞれ、{80, 160}, {75, 150, 300}, {25, 50, 100}, {50, 100, 200} で、同じデータベースの画像に対しては同じスケールの組を用いた。特徴量を求める中心点をスケールの 0.2 倍の幅で移動させ、特徴量を求めた。ASIFT/ASURF+RANSAC のテンプレート画像は、img1 から切り出した。それぞれのテンプレート画像を図 6～9 の (a) に示す。入力画像は、提案手法と同じである。ASIFT/ASURF+RANSAC は、OpenCV (Version 3.4.1) の ASIFT/ASURF と RANSAC をライブラリーとして使う python の公開されているプログラム [25] を、パラメータを含めてほぼそのまま利用したが、計算時間短縮のために、テンプレートに対してだけアフィン変換を行い、入力画像に対しては変換を行わなかった。

図 6～9 の (b)～(f) に、提案手法によって特徴量の相関が最も大きくなった、中心点とスケールを図示する。図中の線分は、エッジ方向ヒストグラムが最大になる基準方向を示している。この方向が物体に対して一定していない理由は、複数種類あるテンプレートの基準方向が物体に対して異なるためである。

検出位置の評価には、検出位置をデータベースに記載されているホモグラフィー行列で逆変換し、テンプレート画像上での検出位置とその目標位置との誤差を用いた。この誤差が、走査ステップ幅に相当するテンプレート画像の最大スケールの 0.2 倍以内であるとき、正しい検出とした。提案手法と ASIFT+RANSAC では、この実験の全ての画像に対して、位置を正しく検出できた。ASURF+RANSAC では、Bark の最も難しい画像 img6 に対しては検出位置が画像範囲をはみ出し、正しく位置検出ができなかった。

表 1 に、計算時間とその標準偏差を示す。使用した計算機の CPU は、Intel(R) Core(TM) i5-6400 CPU @ 2.70GHz である。表中の値は、提案手法、ASIFT/ASURF+RANSAC とともに、テンプレートの作成時間は除いたものである。この結果から、

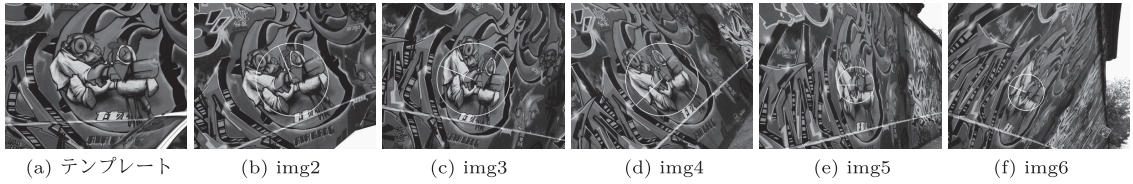


図 6 Graffiti の初期探索結果 ((b)–(f) の円の中心が検出位置, 半径が検出したスケールを表す.)

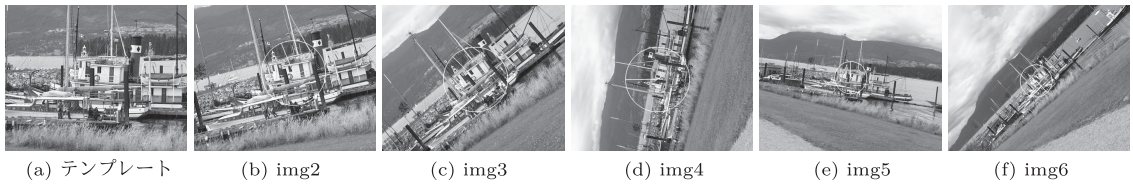


図 7 Boat の初期探索結果 ((b)–(f) の円の中心が検出位置, 半径が検出したスケールを表す.)

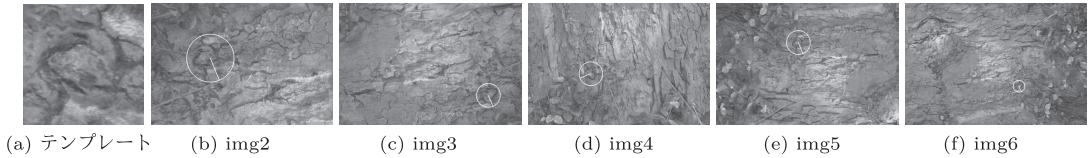


図 8 Bark の初期探索結果 ((b)–(f) の円の中心が検出位置, 半径が検出したスケールを表す.)

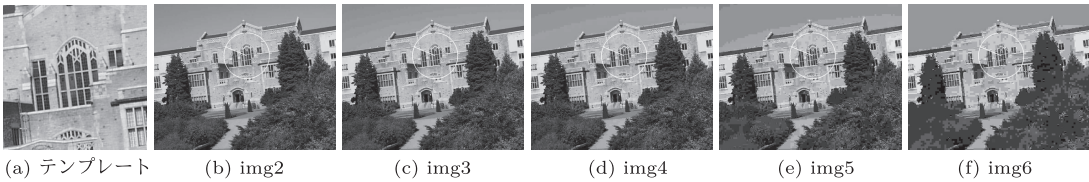


図 9 UBC の初期探索結果 ((b)–(f) の円の中心が検出位置, 半径が検出したスケールを表す.)

表 1 計算時間 (second)

データセット名 \ 初期位置検出手法	ASIFT+RANSAC	ASURF+RANSAC	提案手法
Graffiti	1.47±0.20	0.60±0.04	0.611±0.007
Boat	11.9±3.4	1.49±0.51	0.853±0.008
Bark	0.686±0.034	0.131±0.009	0.834±0.012
UBC	1.09±0.27	0.277±0.007	0.871±0.010

ASURF+RANSAC は多くの画像で提案手法よりも高速であるが, Boat の場合には, 平均で提案手法の 1.5 倍以上の時間がかかっており, 必ず高速ということはできない. ASIFT+RANSAC に対しては, Bark で多少劣っている以外は提案手法の方が高速である. 特に, Boat に関しては提案手法よりも計算時間が平均で 10 倍以上となった. そして, 提案手法では画素数とスケールだけで計算量/時間がほとんど決まり, 画

像内容に対してほとんど依存しないため, データセット内での計算時間の標準偏差が 1.5% 以下となった.

4.3 劣化した入力画像に対する初期位置探索

入力画像に対して, 標準偏差 6 画素のガウス平滑化, 角度 $\pi/6$ rad の長さ 25 画素の移動平均, 標準偏差 80 の加法的白色ガウス雑音の 3 種類の劣化を施した入力画像に対して初期位置探索を行った. 表 2 は, 劣化がない場合を含めて, その結果をまとめたものである.

表 2 劣化した入力画像に対する初期位置探索

劣化		なし					ガウス平滑化					移動平均					加法的白色ガウス雑音				
初期位置検出手法	DN \ INO	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
ASIFT +RANSAC	Graffiti	○	○	○	○	○	○	○	×	×	△	○	×	×	○	×	○	○	○	○	○
	Boat	○	○	○	○	○	×	×	×	×	×	×	○	△	×	△	○	○	○	○	×
	Bark	○	○	○	○	○	×	×	×	×	×	○	×	×	×	×	△	×	×	×	×
	UBC	○	○	○	○	○	×	×	△	○	×	×	×	×	×	×	○	○	○	○	○
ASURF +RANSAC	Graffiti	○	○	○	○	○	○	△	×	×	×	○	×	×	△	×	○	○	○	○	×
	Boat	○	○	○	○	×	×	×	×	×	×	○	×	×	△	×	○	○	○	○	×
	Bark	○	○	○	○	○	×	△	×	×	×	×	△	×	×	×	×	×	×	×	×
	UBC	○	○	○	○	○	○	○	○	○	○	○	△	△	△	△	○	○	○	○	○
提案手法	Graffiti	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	△	○	○	○
	Boat	○	○	○	○	○	○	○	○	○	○	○	○	△	○	○	○	○	○	○	○
	Bark	○	○	○	○	○	○	○	○	×	○	○	○	○	○	×	○	○	○	○	×
	UBC	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

(DN: データセット名, INO: 画像番号. ○: $e \leq 0.2s$, △: $0.2s < e \leq s$, ×: それ以外.
ここで, e は検出誤差, s はテンプレート画像での最大スケール.)

表 3 文字を埋め込んだ入力画像に対する初期位置検出率 (%)

初期位置検出手法	精度 \ 検出対象数字	0	1	2	3	4	5	6	7	8	9	全平均
SIFT +RANSAC	○	16	0	16	14	6	0	4	2	8	0	6.6
	○ + △	22	0	24	16	8	8	6	4	10	0	9.8
SURF +RANSAC	○	0	0	0	0	0	0	0	0	0	0	0.0
	○ + △	0	0	0	0	0	0	0	0	0	0	0.0
提案手法	○	92	50	90	86	74	76	86	90	78	70	79.2
	○ + △	100	84	94	92	78	82	88	92	82	90	88.2

(○: $e \leq 0.2s$, ○ + △: $e \leq s$. ここで, e は検出誤差, s はテンプレート画像での最大スケール.)

○ (正しい検出), △ (ほぼ正しい検出), × (誤検出/検出不可) は, それぞれ, 位置検出誤差がテンプレート画像上でその最大スケールの 0.2 倍以内, 0.2 倍より大きく最大スケール以内, 最大スケールより大きい検出されないことを表している. 表 2 から, 3 種類の劣化に対して提案手法が ASIFT/ASURF+RANSAC よりも頑健であることがわかる. 特徴点に基づく方法が劣化に対する頑健性が劣る理由は, 特徴点のスケールが目標領域のスケールより小さく, 平滑化する領域が狭く特徴点の位置や特徴量の変動が大きくなるためと考えられる.

4.4 特徴点数が少ないテンプレート画像に対する初期位置探索

特徴点数が少ないテンプレート画像に対する提案手法の優位性を示すため, 手書き数字の初期位置探索を行った. ここでの Bark に手書き数字を埋め込むという設定は, 岩石や砂に覆われた自然情景の中に特定の対象が紛れているような, 例えば岩肌にかかれた相対的に小さな手書き文字図形など, 人間でも検出が難しい状況での初期位置探索の実行を想定している. MNIST の各カテゴリーの訓練画像 (28×28 画素) から最初の 10 個づつを取り出し, 時計回りに 0.2π rad 回転, 4 倍に拡大し, Bark の img2~img6 に埋め込んだ計 500



図 10 手書き数字 (0) を埋め込んだ入力画像の例

枚の画像を入力画像とした. テンプレート画像は埋め込んだ数字だけの画像とした. 画像を変形しないため, 各手法ともテンプレート画像を変形せずに初期位置探索を行った. 提案手法で用いたスケールはテンプレート画像で $8, 8\sqrt{2}$, 入力画像で 27, 54 とした. テンプレート画像の 8 に対応するスケールは入力画像で 32 となるが, その値からは $1/2^{1/4}$ 倍程度ずらした. 画像の例を図 10 に, 実験結果を表 3 に示す. ○ 及び ○ + △ の行は, それぞれ, 目標位置をスケールの 0.2 倍以内及びスケール以内の誤差で検出できた割合である. 表 3 から, 次のことがわかる. SURF+RANSAC では全く検出できなかった. SIFT+RANSAC の場合, スケール以内の検出誤差でも平均 9.8% の検出率となった. 提案手法では, 検出誤差がスケールの 0.2 倍以内

で平均 79.2%, スケール以内で平均 88.2% となり, 特徴点の少ないテンプレート画像に対して, 提案手法が有効であることがわかる。

5. む す び

マッチングにおける領域画像マッチングの手法は, 特徴画像マッチングの手法と比較して, より高精度なマッチング性能の可能性をもつものの処理時間が膨大化するという課題がある。特に, whole-to-part の物体探索に用いる場合, マルチスケールでのスライド窓探索の処理負担が大きい。この課題を解決するため, 本論文では領域画像マッチングのための目標領域特徴量を用いた高速な初期位置探索の手法を提案した。特徴画像マッチングの手法である ASIFT/ASURF と RANSAC の組合せによる初期位置探索との比較実験を行った結果, 抽出される特徴点数が少ない場合や画像劣化が大きい場合に提案手法の優位性が明らかとなった。また, 初期位置探索に要する時間についても, 提案手法は, 特徴画像マッチング手法とほぼ同等であり, 画像内容にほとんど依存しなかった。今後の課題は, 領域画像マッチングとの組合せであり, 例えば, GAT/GPT 関連法の初期位置探索としての実装がある。

謝辞 本研究は JSPS 科研費 17K00250, 17H01760 の助成を受けたものである。

文 献

- [1] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol.60, no.2, pp.91–110, Nov. 2004.
- [2] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.506–513, CVPR'04, IEEE Comput. Soci., Washington, DC, USA, 2004. <http://dl.acm.org/citation.cfm?id=1896300.1896374>
- [3] H. Bay, T. Tuytelaars, and L.V. Gool, "Surf: Speeded up robust features," *Computer Vision -ECCV 2006*, pp.404–417, 2006.
- [4] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol.110, no.3, pp.346–359, June 2008.
- [5] J.-M. Morel and G. Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol.2, no.2, pp.438–469, April 2009.
- [6] M.A. Fischler and R.C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol.24, no.6, pp.381–395, June 1981.
- [7] 飯島泰蔵, バターン 認識理論, 森北出版, 東京, 日本, 1989.
- [8] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol.56, no.3, pp.221–255, Feb. 2004.
- [9] T. Wakahara, Y. Kimura, and A. Tomono, "Affine-invariant recognition of gray-scale characters using global affine transformation correlation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, no.4, pp.384–395, April 2001.
- [10] Y. Yamashita and T. Wakahara, "Affine-transformation and 2D-projection invariant k -NN classification of handwritten characters via a new matching measure," *Pattern Recognition*, vol.52, pp.459–470, April 2016.
- [11] S. Zhang, T. Wakahara, and Y. Yamashita, "Image matching using GPT correlation associated with simplified HOG patterns," *Proceedings of 2017 Seventh International Conference on Image Processing Theory, Tools and Applications*, vol.1, pp.1–6, Nov.–Dec. 2017.
- [12] Y. LeCun, C. Cortes, and C.J. Burges, "The MNIST DATABASE of handwritten digits". <http://yann.lecun.com/exdb/mnist/>
- [13] Q. Tong and T. Aoki, "A blur-invariant interest point detector based on momentsymmetry for motion and gaussian blurred image matching," *International J. Signal Processing*, vol.2, pp.96–106, 2017.
- [14] Q. Tong and T. Aoki, "A blur-invariant local feature for motion blurred image matching," *Proceedings of SPIE 10420, Ninth International Conference on Digital Image Processing*, pp.104201E–1–6, May 2017.
- [15] A. Bosch, A. Zisserman, and X. Muñoz, "Image classification using random forests and ferns," *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision*, pp.1–8, Oct. 2007.
- [16] P. Podbreznik and B. Potocnik, *Dense feature matching method based on ASIFT, LAP LAMBERT Academic Publishing*, Jan. 2014.
- [17] C. Liu, J. Yuen, and A. Torralba, "SIFT Flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.33, pp.978–994, Jan. 2011.
- [18] V. Kober, "Fast algorithms for the computation of sliding discrete sinusoidal transforms," *IEEE Trans. Signal Process.*, vol.52, no.6, pp.1704–1710, June 2004.
- [19] E. Elboher and M. Werman, "Cosine integral images for fast spatial and range filtering," *Proceedings of 2011 18th IEEE International Conference on Image Processing*, pp.89–92, Sept. 2011.
- [20] K. Sugimoto and S.-I. Kamata, "Compressive bilat-

- eral filtering,” IEEE Trans. Image Process., vol.24, no.11, pp.3357–3369, Nov. 2015.
- [21] 杉本憲治郎, 京地清介, 鎌田清一郎, “離散コサイン変換に基づく定数時間ガウシアンフィルタの包括的性能解析,” 信学技報, vol.117, no.48, IE2017-4, pp.19–24, May 2017.
- [22] 山下幸彦, 若原 徹, “領域画像マッチングのための初期探索法,” 信学技報, PRMU2018-53, Sept. 2018.
- [23] U.A. Ahmad, K. Kidiyo, and R. Joseph, “Texture features based on local fourier histogram: self-compensation against rotation,” IEEE Trans. Image Process., vol.17, no.3, pp.(030503–1)–(030503–3), July–Sept. 2000.
- [24] Visual Geometry Group, University of Oxford, “Affine covariant features”.
<http://www.robots.ox.ac.uk/~vgg/research/affine/>
- [25] A. Alekhin, “Affine invariant feature-based image matching sample”. <https://github.com/opencv/opencv/blob/master/samples/python/asift.py>
 (2019 年 6 月 26 日受付, 10 月 8 日再受付,
 12 月 13 日早期公開)



山下 幸彦 (正員)

昭和 58 東工大・工・電子電気工学科, 昭和 60 同大・理工・情報工学修士修了. 同年日本原子力研究所入社. 昭和 63 年 (株) アイザック入社, 現在, 東工大・環境社会・融合理工准教授. 画像処理の研究に従事. 博士 (工学).



若原 徹 (正員)

昭 50 東大・工・物理工学卒. 昭 52 同大大学院修士課程了. 同年日本電信電話公社 (現 NTT) 入社. 平 13 から法政大学情報科学研究科教授, 現在に至る. 文字・図形認識, 画像マッチングの研究に従事. 工博.