

色のクラスタリングを用いた指差し方向認識

Finger Direction Recognition Using Color Clustering

浅野秀胤†, 永易武‡, 織茂達也†, 寺林賢司‡, 太田睦†, 梅田和昇‡

Hidetsugu ASANO†, Takeshi NAGAYASU‡, Tatsuya ORIMO†,

Kenji TERABAYASHI‡, Mutsumi OOTA†, Kazunori UMEDA‡

†: パイオニア株式会社, {hidetsugu_asano,tatsuya_orimo,mutsumi_oota}@post.pioneer.co.jp

‡: 中央大学, nagayasu@sensor.mech.chuo-u.ac.jp, {terabayashi,umeda}@mech.chuo-u.ac.jp

概要: カメラを用いたリアルタイム・ジェスチャ・インタフェース・システムとして, 指差し方向の認識手法を提案する. 入力画像の HS 空間の 2 次元ヒストグラムを作成し, GMM により色空間のクラスタリングを行い, それを元に画像をセグメンテーションし, 領域毎に手らしさを評価する. 最も手らしい領域において, 領域の形状などから指領域を推定し, 指の方向を算出する. さらに複数カメラでの結果を組み合わせることにより, 三次元空間上での指差し方向を認識する. この手法を用いることで高精度かつ背景等にロバストなジェスチャインタフェースの実現が可能である. このシステムは, あらゆる種類の機器操作に有効である.

1. はじめに

我々は, 人間にストレスを与えない自然なヒューマン・マシン・インタフェースとして, 部屋の中に数台のカメラを設置して操作者の手振り検出し, 家電を操作するシステムを提案している[1-3]. リモコンを使わない類似のシステムはこれまでも多く提案されているが[4,5], ユーザはカメラの前に正対した上で, ジェスチャを行う必要があり, 使いやすさに問題があった. 我々の方法は, 操作する場所を拘束しないという点に特徴がある. さらに我々は, 指差すだけの動作で操作する機器を特定する手法も提案をしている[6]. これにあたっての手領域の抽出は, 特定の色を前提としたために照明変動や肌色の個人差に対して十分にロバストではなかった. さらに, 指差し方向を手領域全体の主軸として推定していたために正確な方向を取得できなかった.

本研究では, 部屋を俯瞰するカメラで撮影された画像から, 色情報などの前提知識を用いずに指領域を抽出し, 複数カメラを用いて指差し方向を認識する手法を提案する. また, 提案手法をリアルタイムシステムとして実装し, 効果の検証を行う.

2. アルゴリズム

本研究では, 指差ししている手が写っている画像



図 1 認識対象とする指差し画像の例

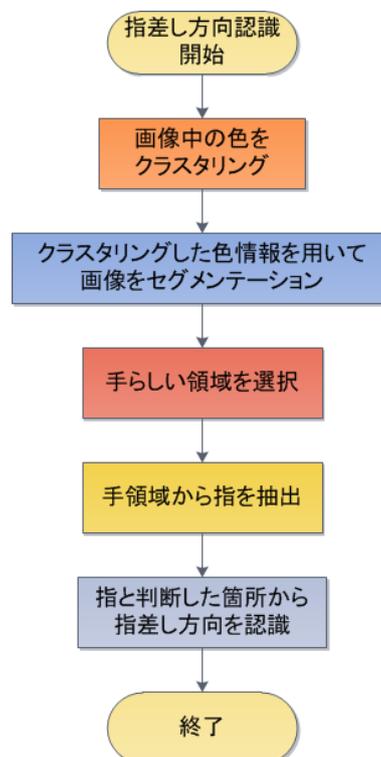


図 2 指差し方向認識のフローチャート

を認識対象とする(図 1) . 我々のシステムでは, 手振りを検出したカメラはそこに向けてズームを行うので[1-3], 目標とする手は画像内にある程度の大きさで写っている. ここから色情報などの前提知識を用いずに指差し方向を認識するアルゴリズムとして, 以下の手法を提案する(図 2). まず, 画面中の色をクラスタリングし, その結果を用いて画像のセグメンテーションを行い, いくつかの候補領域を取り出す. それぞれの領域に対して「手らしさ」の評価を行って手領域を特定し, 次いでその中での指の領域を特定し, 指の方向を求め, 複数カメラでの結果を組み合わせることで, 3次元空間上での指の方向を推定する. 以下に各処理について詳述する.

2.1 色のクラスタリング

手領域はひとつの色, あるいは多くとも 2~3 種類の色の組み合わせから構成されていると仮定する. 画像中の色情報のクラスタリングを行えば, 手領域はクラスタリングされた色の幾つかで特定可能であるし, 背景や照明環境の変動などに対し比較的頑健な領域抽出が実現できる. 本論文では HS 空間での 2次元ヒストグラムを対象とし, Gaussian Mixture Model(GMM)によるクラスタリングを行う. また, GMM のクラス数を 10 とした(図 3).

2.2 セグメンテーション

前項で作成したすべての色クラスについて, グラフカットによる画像のセグメンテーションを行う. エネルギー関数は Rother ら[7]が示すものとした.

本論文では単色を仮定して手領域を推定している

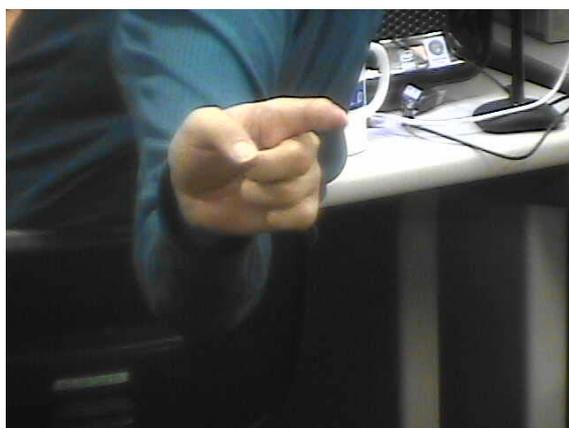


図 4 クラスタリング結果を用いて入力画像にグラフカットを適用し, クラス毎に画像をセグメンテーションした結果から最大の領域を抽出した例. この例ではクラス数 10 で GMM を適用し, 1 つのクラスが縮退した.

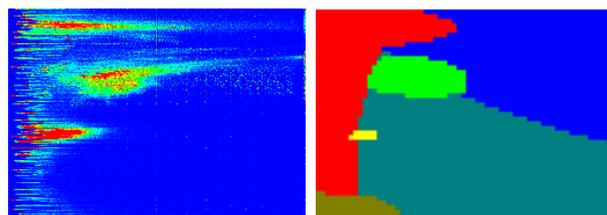


図 3 (左)HS 空間の2次元ヒストグラムの例. 青は低頻度, 赤は高頻度を表す. (右)GMM でクラスタリングした結果.

が, 複数クラスの組み合わせによるセグメンテーションも行うことで, 手袋をしている場合などにも対応できると考えられる. セグメンテーション結果のうち領域が大きい 2 つに対して穴埋めしたものを手領域の候補とする. 色クラス毎にセグメンテーションを行い, 最大領域を抽出した例を図 4 に示す .

2.3 手らしさの評価

前述の領域に対し, 手らしさの評価を行い, 手領域を選択する. 手らしさの評価は領域の以下の特徴を用いる.

- 領域の面積
ズームカメラを用いている我々のシステムでは手領域はある程度の面積で写っていることが期待できる.
- 領域重心と輪郭画素の距離の標準偏差
抽出した領域が手である場合, 輪郭が過剰に複雑な形状をしている可能性は低い. そこで, 領域の形状が滑らかな形をしているほど評価値が高くなるようにする.

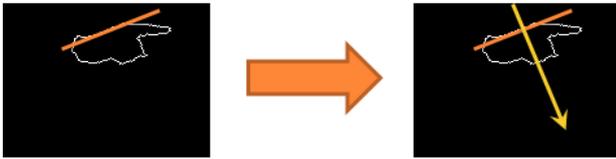


図 5 輪郭の法線ベクトルと、領域の深さ算出の例



図 6 輪郭の法線方向に存在する領域の画素数を輪郭方向に並べたグラフ. グラフ左方の指領域では D が小さい.

- 領域が画面端に接する割合

我々のシステムでは、手領域は画面中央に寄って写ることが期待できており、画面端に接する可能性は低く、接していてもその範囲は限られる。

これらの条件から、領域の面積を N 、重心と輪郭の距離の標準偏差を δ_r 、領域が画面端に接する割合を O_r として、以下の評価式を用いた。

$$E = \begin{cases} N \left(\frac{1}{\delta_r} + \lambda O_r^2 \right) & \text{if } N < M \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

ここで、 λ は δ_r と O_r のバランスを取る項で、 M は手領域が取りうる最大の面積である。本論文では $\lambda = 0.1$ 、 $M = \text{width} \times \text{height} / 2$ とした。

2.4 手領域からの指領域の推定

抽出した手領域から指領域を特定する。指は手の細い部分であるため、手領域の輪郭画素に注目し、以下の処理で指領域を抽出する。

- 手領域の各輪郭画素について、輪郭方向前後 3 画素を加えて最小二乗法で直線当てはめを行い、輪郭画素の接線とし、同時に法線を求める。この法線が通過する手領域中の画素数 D を算出する (図 5)。
- 前項の処理を全ての輪郭画素に適用し、 D の

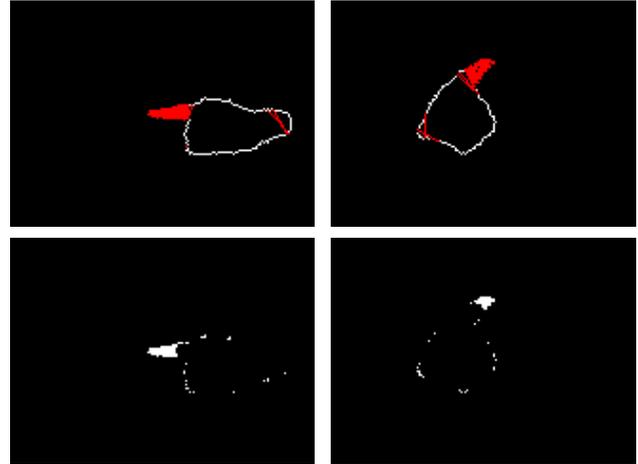


図 7 手領域から指領域を推定した例. 上段は本手法を、下段はモルフォロジー処理(Top-Hat 変換)を用いた。

平均を μ_D と、分散 δ_D を求める(図 6)

- 輪郭画素のうち、次式を満たす画素を指の輪郭の可能性があるとす。

$$D < \mu_D - a \times \delta_D \quad (2)$$

ここで、 a は指の細さを評価する係数である。本論文では $a = 0.75$ とした。式(2)の条件を満たす輪郭画素から法線方向に手領域の塗りつぶしを行い、塗りつぶされた箇所のうち最大のものを指領域とする。

以上のようにして手領域から指領域を推定する。本手法により指を抽出した結果と、モルフォロジー処理を用いた場合[8]との比較を図 7 に示す。本手法により指の領域をより正確に取得できている。指がはっきり写っている場合はどちらの処理でも結果にそれほど差はないが、指があまりはっきり写っていない場合、モルフォロジー処理では指の一部しか抽出できず、ここから指の方向を取得することは困難である。また、本手法でも指以外の相対的に箇所が指領域の候補として抽出されるが、最大領域を指として扱うことでこれを排除できる。

また、手領域を誤抽出している場合と、指がはっきり写っていない場合への対応として、式(3)~(5)の全ての条件を満たす場合のみを指領域とする。

$$L_D \leq n_L \quad (3)$$

$$\mu_D \geq n_\mu \quad (4)$$

$$\delta_D \geq n_\delta \quad (5)$$

L_D は輪郭の周長、 n_L 、 n_μ 、 n_δ はそれぞれ L_D 、 μ_D 、 δ_D の閾値である。本論文では $n_L = 300$ 、

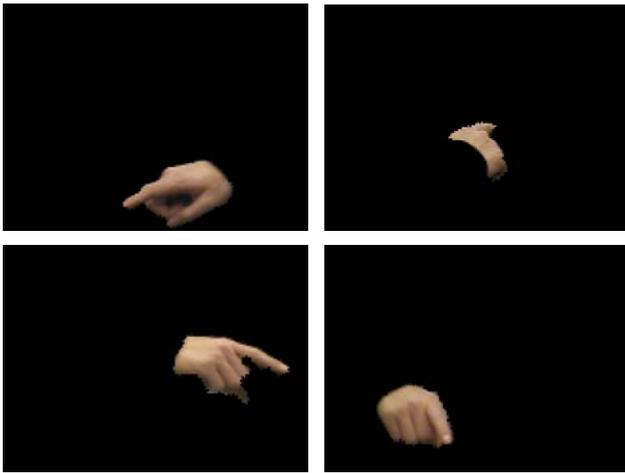


図 8 手領域抽出結果

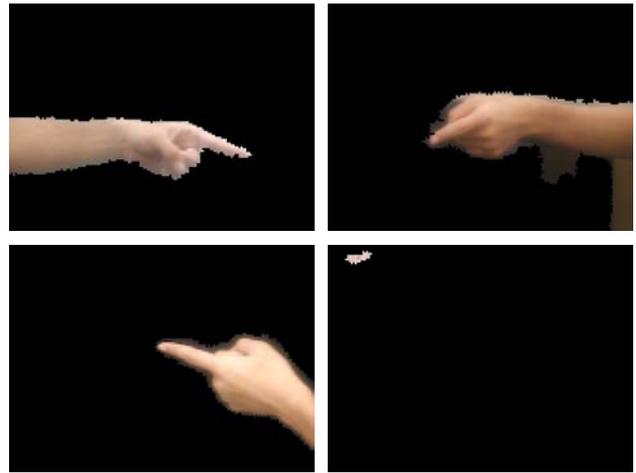


図 10 半袖着用時の手領域抽出結果

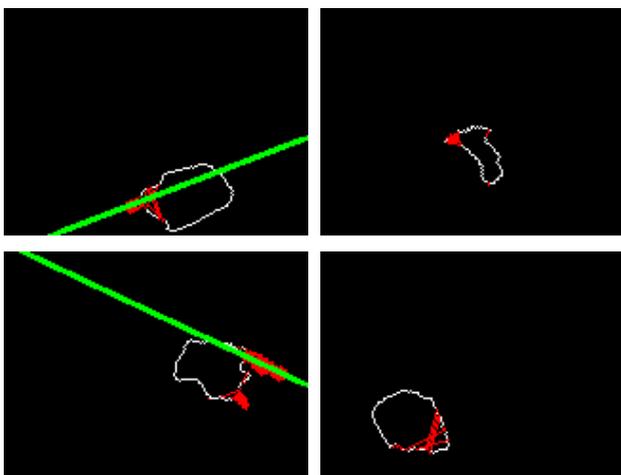


図 9 指方向算出結果

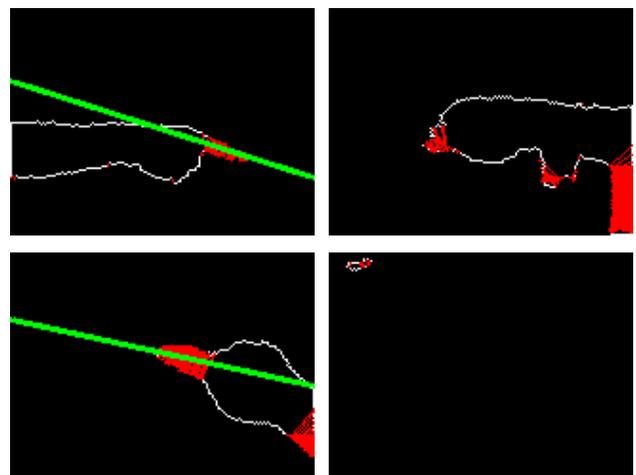


図 11 半袖着用時の指差し方向算出結果

$n_{\mu} = 14$, $n_{\delta} = 10$ としている.

2.5 指方向の算出

指領域に直線当てはめを行い、画像中での指の方向を算出する. この結果を複数カメラで組み合わせることにより、三次元空間上での指の向きを得ることができる[8]. 本手法では画像ごとに指の方向を求めているため、誤検出や直線当てはめの誤差などにより結果が安定しない. このため、本論文では算出した画像中の指の方向に 13 データでのメディアンフィルタを適用し、結果の安定化を行っている.

3. 実験

以上の手法を実装し、性能検証を行った. カメラには Axis 233D を使い、部屋の四隅に設置した. 部屋の広さは $6.9 \times 7.8\text{m}$ で、[2]の手法を用いて手振りを検出し、検出位置を中心にズームを行い、その後指差し方向認識を行なった.

色のクラスタリングから手らしさの評価までの処理については、処理量が多いが、映像が急激に変化することは稀であるため、それほど頻繁に結果を更新する必要がない. 一方で、手領域抽出から指方向の算出の処理量は少なく、また全てのフレームを処理することが望ましい. そこで、これらの処理を分離し、実装した. 前者については 100ms 毎に処理を行い、クラス情報を更新している. GMM とグラフカットについては GPU 上で処理を行い、その他については CPU で処理を行なった. 実験に用いた PC の構成は Intel®Core™i7 975・メモリ 6GB と、GeForce GTX 580 を 2 枚で、処理時間は、色のクラスタリングから手らしさの評価の処理が約 50ms、指方向の算出がカメラ一つにつき約 5ms 程度であった. 入力画像は 160×120 に縮小して使い、三次元空間上での指差し方向を 30fps で算出した.

四つのカメラで同時に取得した画像から手領域を抽出した結果を図 8 に示す. ここから、本手法により

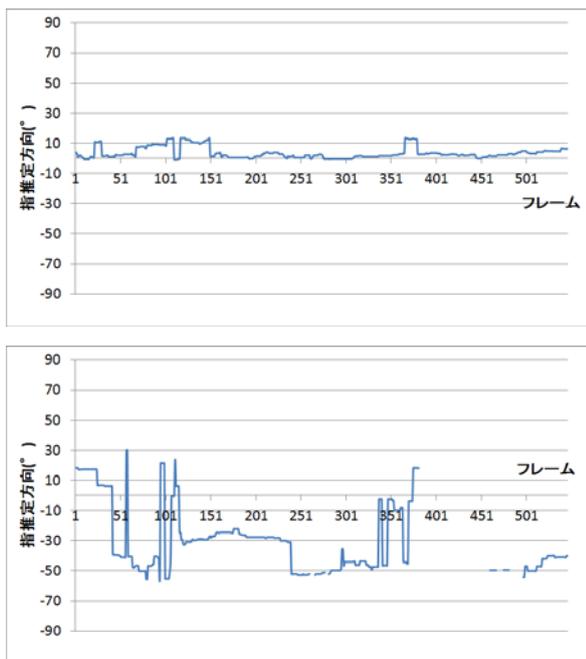


図 12 指方向推定結果(上:三次元空間上で X 軸に対し 0° 方向, 下: X 軸に対し -45° 方向の指差し)

手領域が正確に抽出できていることが分かる。図 9 に指領域推定結果と算出した指の方向を示す。指がカメラの方向を向いているなど、指がはっきり見えていない際には指なしとして判断し、指がはっきり見えている場合については正しく指の方向が算出できている。半袖着用時の検出例を図 10, 11 に示す。この場合でも指の方向が正しく算出できている。このように、本手法は操作者の服装を問わない認識処理が実現できている。

三次元空間上での指差し方向認識結果を図 12 に示す。結果は三次元方向として算出しているが、簡単のため水平方向のみ記載した。手らしさの評価を誤った場合や、手領域の抽出に失敗した場合など、様々な要因により指方向の算出結果が安定しない場合があるが、おおよそ指の方向が正しく推定できている。この結果を時間的に蓄積することにより指差し方向を正確に認識することが可能である。本手法の検出誤差を図 13 に示す。三次元空間上の X 軸と平行な向きに指差しを行い、Y 軸方向と Z 軸方向の誤差を示している。この例では、指差し位置から X 軸方向に 1m の位置に平面を置くと、概ね Y 軸方向に 40cm 程度、Z 軸方向に 20cm 程度の範囲に収まっている。

4. まとめ

本論文では、色のクラスタリングと画像のセグメンテ

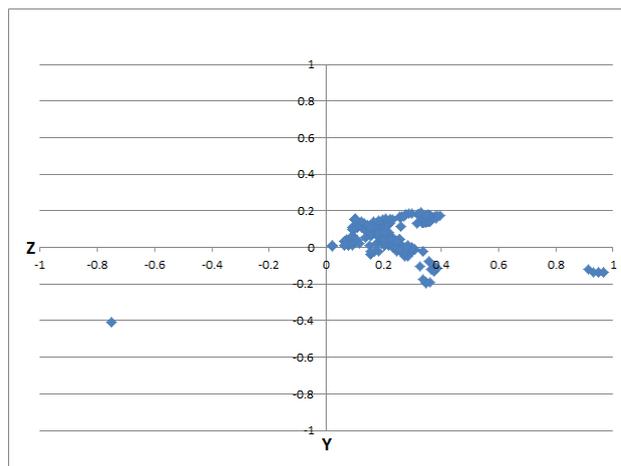


図 13 三次元空間上の X 軸と平行な方向への指差しの誤差。指差し位置から 1m の位置に射影した。

ーションにより、色情報などの事前知識を用いない指差し方向の認識手法を提案した。また、提案手法をリアルタイムシステムとして実装し、その有用性を確認した。これにより、手の色などの個人差や、カメラ・背景などに依らない検出が可能となった。

一方で本手法の制限として、指がカメラの方向を向いている場合に認識不可であることと、オクルージョンに弱いという点が挙げられる。これらについては他のカメラの情報を用い、機械学習などを用いて推定することで解決できると考えている。

参考文献

- [1] 入江耕太, 梅田和昇: 濃淡値の時系列変化を利用した画像からの手振り検出, 日本ロボット学会誌, vol.21, no.8, pp.923-931, 2003.
- [2] 浅野秀胤, 織茂達也, 高橋真人, 寺林賢司, 太田睦, 梅田和昇: フーリエ変換を用いた小さな手振りの検出, ビジョン技術の実利用ワークショップ (ViEW2010), pp.264-267, 2010.
- [3] 浅野秀胤, 織茂達也, 永易武, 寺林賢司, 太田睦, 梅田和昇: 小さな手振り検出を用いた家電操作システムの構築, 2011 年映像情報メディア学会年次大会予稿集, 9-9, 2011.
- [4] S. M. Dominguez, T. Keaton, and A.H. Sayed: A Robust Finger Tracking Method for Multimodal Wearable Computer Interfacing, IEEE Trans. on Multimedia, Vol.8, No.5, pp. 956-972, 2006.
- [5] X. Liu, K. Fujimura.: Hand gesture recognition using depth data, IEEE International Conference on Automatic Face and Gesture Recognition, pp.529-534, 2004.
- [6] Kota Irie, Naohiro Wakamura, Kazunori Umeda: Construction of an Intelligent Room Based on Gesture

Recognition -Operation of Electric Appliances with Hand Gestures-, Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp.193-198, 2004.9.

- [7] C. Rother, V. Kolmogorov, A. Blake: "GrabCut" Interactive Foreground Extraction Using Iterated Graph Cuts, ACM Trans. Graphics (SIGGRAPH '04), vol.23, no.3, pp.309-314, 2004.
- [8] 若村直弘, 鈴木健一郎, 入江耕太, 梅田和昇: インテリアジェントルールの構築 -直感的なジェスチャを用いた家電製品の操作-, 画像の認識・理解シンポジウム (MIRU2005), IS3-108, pp.1074-1081, 2005.7.

浅野秀胤:2000年横浜国立大学工学部電子情報工学科卒業,2002年同大学院電子情報工学専攻修士課程修了.2002年よりパイオニア(株)研究開発部に所属し,映像処理関連の研究に従事.

永易武:2010年中央大学理工学部精密機械工学科卒業.現在同大学大学院精密工学専攻修士課程前期課程在学中.画像処理を用いたヒューマンインタフェースの研究に従事.

織茂達也:1984年明治大学工学部電子通信工学科卒業,同年日本電気ホームエレクトロニクス(株)入社.開発研究所にてHDTV関連の研究開発に従事.1996年以降,PDP用映像信号処理ASIC開発に従事.現在,パイオニア(株)研究開発部に所属し,映像処理関連の研究に従事.

寺林賢司:2004年北海道大学大学院システム情報工学専攻修士課程修了.2008年東京大学大学院工学系研究科精密機械工学専攻博士課程修了,博士(工学).同年,中央大学理工学部精密機械工学科助教,現在に至る.ヒューマンインタフェース,ロボットビジョン等の研究に従事.精密工学会,日本ロボット学会,日本機械学会,IEEE等の会員.

太田睦:1981年京都大学工学部数理工学科卒,1983年同修士課程卒,NEC C&C研究所等で画像信号処理,主として圧縮符号化の研究を行う.2000年東京大学より博士(工学)の学位取得.現パイオニア株式会社研究開発部.電子情報通信学会,映像情報メディア学会,IEEE会員.

梅田和昇:1989年東京大学工学部精密機械工学科卒,1994年同博士課程修了.同年中央大学理工学部精密機械工学科専任講師,2006年同教授,現在に至る.ロボットビジョン,画像処理の研究に従事.博士(工学).日本ロボット学会,精密工学会,日本機械学会,電子情報通信学会,IEEE等の会員.