

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 †,
Hidetsugu Asano †,
寺林 賢司 ‡,
Kenji Terabayashi ‡,

永易 武 ‡,
Takeshi Nagayasu ‡,
太田 睦 †,
Mutsumi Ohta †,

織茂 達也 †,
Tatsuya Orimo †,
梅田 和昇 ‡,
Kazunori Umeda ‡

†: パイオニア株式会社

‡: 中央大学

要旨 複数カメラの映像から指差し方向を認識する手法を提案する。従来の手法で問題となっていた操作位置の制限や肌色の限定などを外し、色のクラスタリングと形状解析によって、安定的な指差し方向認識が実現できた。本手法は、人が何も持たずに身振り・手振りだけで機器を操作するジェスチャインタフェースに有効である。本提案はVIEW2011で発表済である[1]。

Summary We propose a method to recognize finger direction using multiple cameras. Input images are segmented by color clustering, and shapes of the segments is analyzed. With these techniques, some restrictions in the previous method, such as limitations of operation area and predefined skin color, are removed. The proposed method realizes a stable gesture interface that controls equipments without a hand-held device. Main part of the method in this paper has already been proposed in ViEW2011[1].

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を検出し、家電を操作するシステムを提案している[2-4]。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[5,6], 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。また我々は、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[7]。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

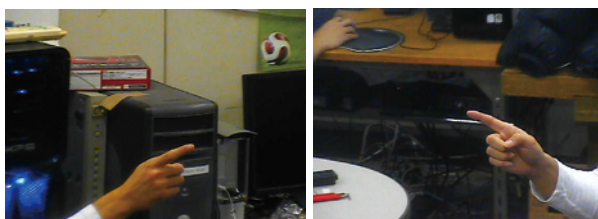


図1 認識対象とする指差し画像の例

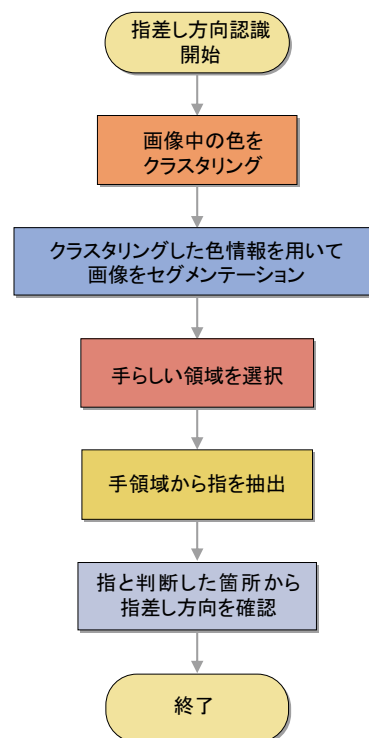


図2 指差し方向認識のフローチャート

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、肌色情報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ングを行うので[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。ここから肌色情報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、画面中の色を分類する。この分類をクラスタリングと言う。この分類した色情報を元に、画像を複数の領域に分割する。これをセグメンテーションと言う。この結果として、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算出する。以下に各処理について詳述する。

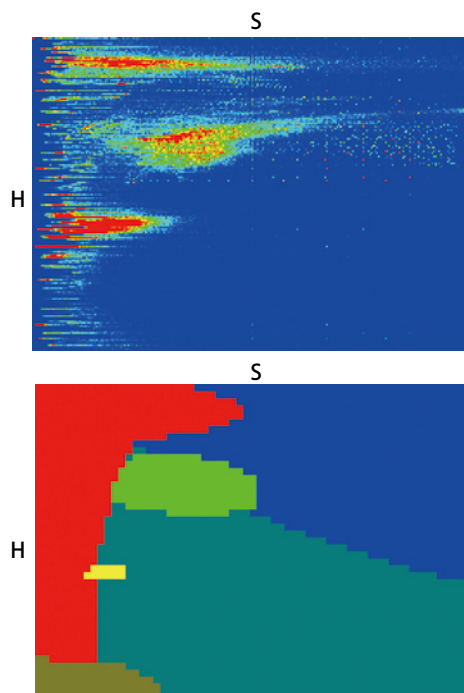


図3 (上)HS空間の2次元ヒストグラムの例。
青は低頻度、赤は高頻度を表す。
(下)GMMでクラスタリングした結果。

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。

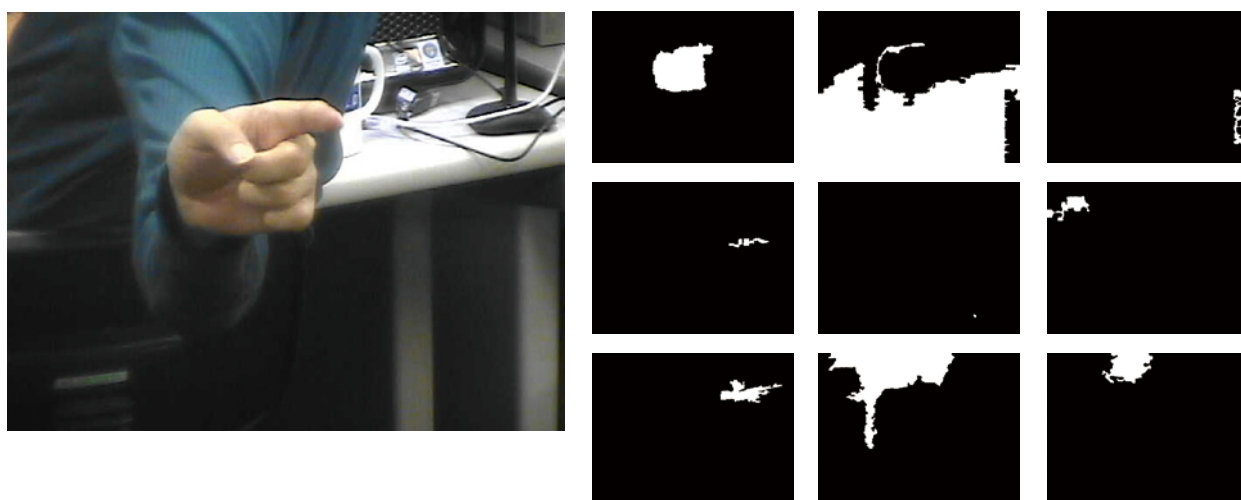


図4 クラスタリング結果を用いて入力画像にグラフカットを適用し、クラス毎に画像をセグメンテーションした結果から最大の領域を抽出した例。この例ではクラス数10でGMMを適用し、1つのクラスが縮退した。



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。

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。GMMは、データが正規分布の混合であると見なし、繰り返し処理により正規分布のパラメータを推測して、そのパラメータでデータの分布を表現する手法である。本論文では、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[8]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。

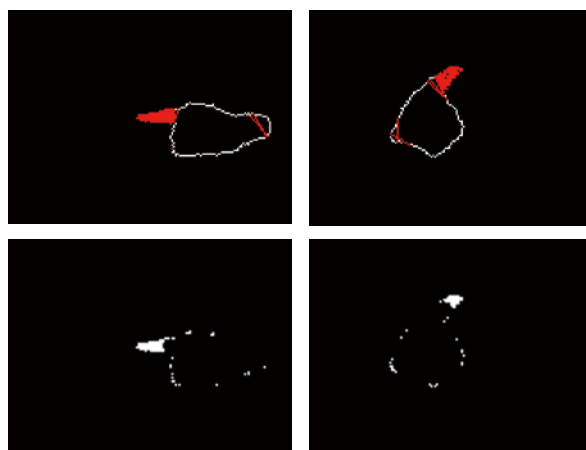


図7 手領域から指領域を推定した例。上段は本手法を、下段はモルフォロジー処理(Top-Hat変換)を用いた。

- 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。
- 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。

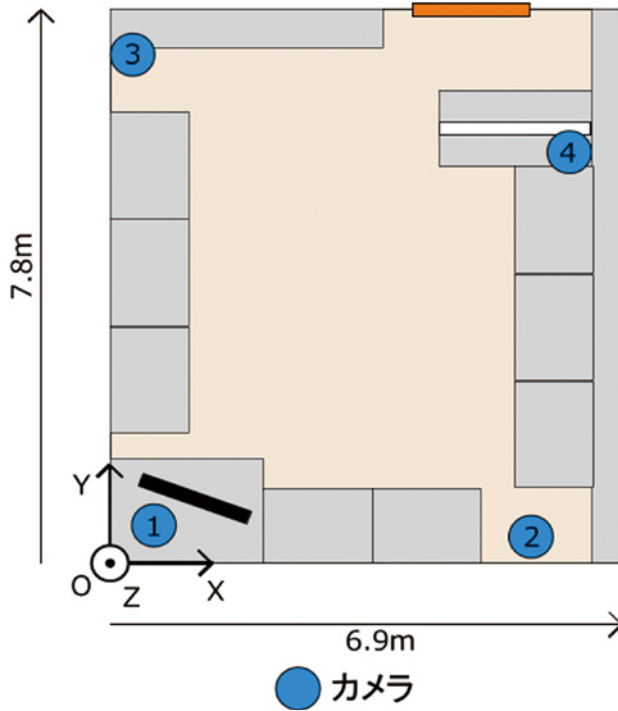


図8 実験を行った部屋の見取り図

- 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る。

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

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

表1 カメラ設置位置[m]

	X	Y	Z
カメラ1	0.59	0.53	2.11
カメラ2	5.67	0.37	2.18
カメラ3	0.13	7.15	2.47
カメラ4	6.22	5.78	2.39

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 $n_L = 300$ 、 $n_\mu = 14$ 、 $n_\delta = 10$ 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を用い、部屋の四隅上部に設置した。部屋の広さは6.9×7.8mで、機器の設置状況は図8、表1に示す通りである。この中で、[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はない。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

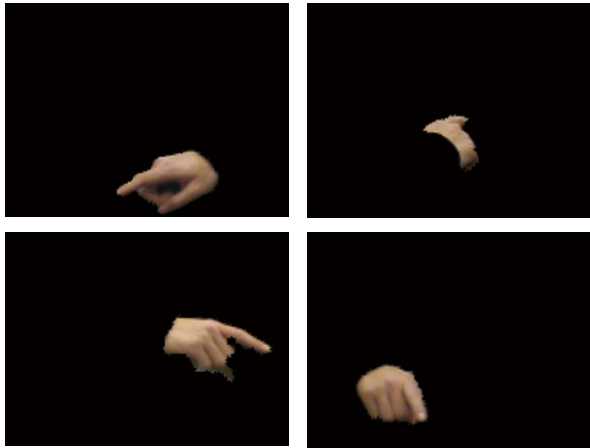


図9 四つのカメラ入力から、本手法により手領域を抽出した結果

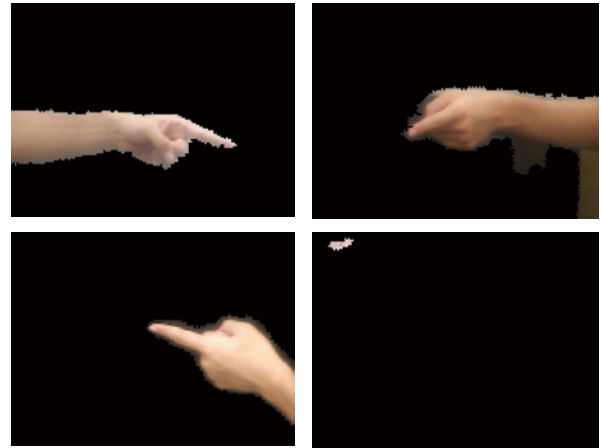


図11 半袖着用時の手領域抽出結果

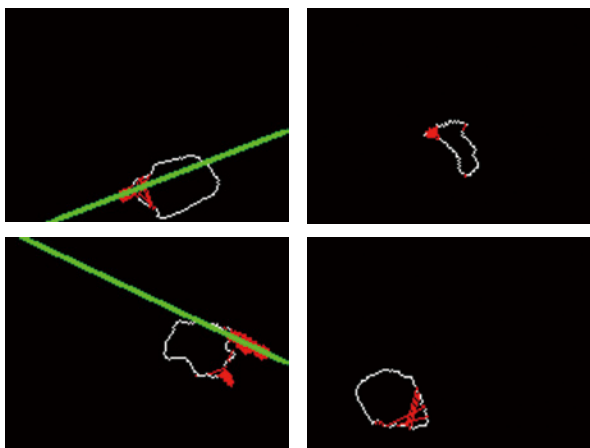


図10 指方向算出結果。赤い領域は指の候補領域を示し、緑の線は推定した指の方向である。

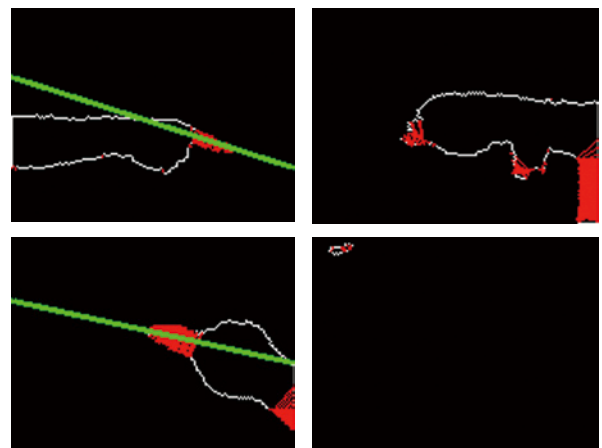


図12 半袖着用時の指差し方向算出結果

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、NVIDIA® 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640×480で、160×120に縮小して処理を行い、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9に示す。ここから、本手法により手領域が正確に抽出できていることが分かる。図10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11、12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13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14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。

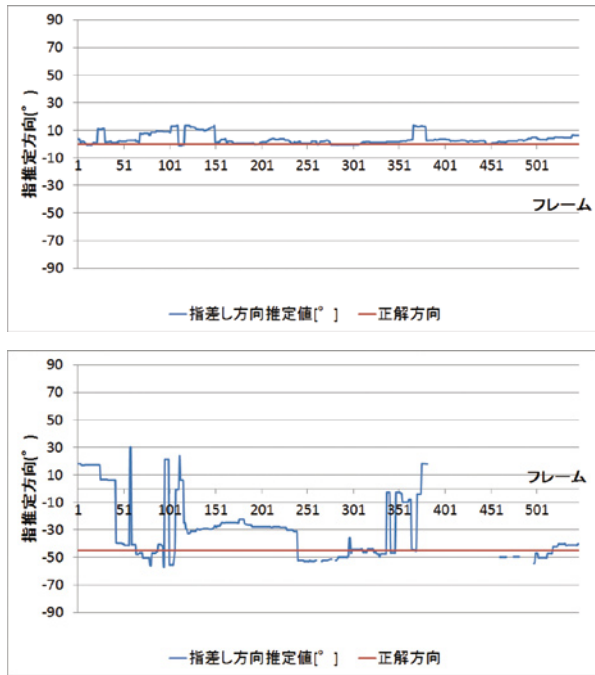


図 13 指方向推定結果(上：三次元空間上でX軸に対し0° 方向, 下：X軸に対し-45° 方向の指差し)

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。

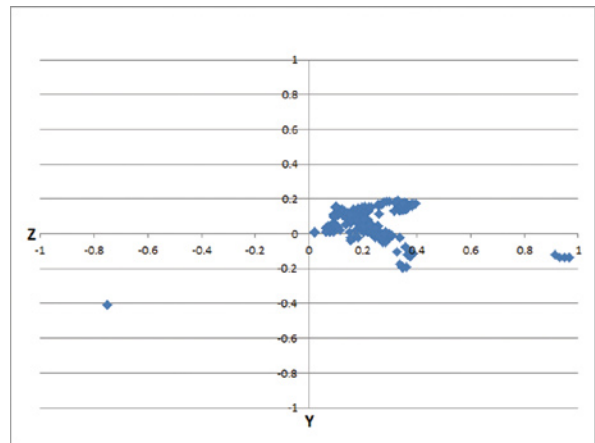


図 14 三次元空間上のX軸と平行な方向への指差しの誤差。指差し位置から1mの位置に射影した。

参考文献

- (1) 浅野秀胤, 永易武, 織茂達也, 寺林賢司, 太田睦, 梅田和昇: 色のクラスタリングを用いた指差し方向認識, ビジョン技術の実利用ワークショップ (ViEW2011), pp.85-90, 2011.
- (2) 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.
- (3) 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.
- (4) 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.
- (5) 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.
- (6) X. Liu, K. Fujimura,.: Hand gesture recognition using depth data, IEEE International Conference on Automatic Face and Gesture Recognition, pp.529-534, 2004.
- (7) 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.
- (8) 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.
- (9) 若村直弘, 鈴木健一郎, 入江耕太, 梅田和昇: インテリジェントルームの構築 -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.

筆者紹介

浅野 秀胤 (あさの ひでつぐ)

研究開発部 第4研究部 所属。映像処理関連の研究に従事。

永易 武 (ながやす たけし)

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 (おりも たつや)

研究開発部 第4研究部 所属。PDP用映像信号処理ASIC開発を経て、映像処理関連の研究に従事。

寺林 賢司 (てらばやし けんじ)

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 (おおた むつみ)

研究開発部 第4研究部 部長。画像信号処理, 主として圧縮符号化の研究を経て, 現在 映像処理関連の研究開発責任者。工学博士。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 (うめだ かずのり)

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。