

# 手話認識のための距離情報を用いた 隠蔽を含んだ顔・手領域抽出

森 昭太<sup>†1,a)</sup> 松尾 直志<sup>†1,b)</sup> 白井 良明<sup>†1,c)</sup> 島田 伸敬<sup>†1,d)</sup>

**概要:** 画像から手話を認識するためには、顔・手領域を抽出する必要がある、特に手形状の抽出は非常に重要である。しかし手話によっては手同士、顔と手で隠蔽がおこる場合がある。その場合、肌色やエッジ検出のみでは顔と手を分離することが困難であった。そこで深度センサ (Kinect for Xbox360) を用いることが考えられる。しかし手話によっては顔と手が密着する場合には、顔・手を分離することが困難である。本手法では、1 フレーム前の手領域を現在フレームの手領域があると推測される領域に対して顔の3次元テンプレートマッチングを行うことによって、手領域候補を抽出する。その領域に手以外の領域が含まれていないか距離のヒストグラムから判別分析法により自動で判断し手領域を求める。手同士の隠蔽時も同様の処理を行う。その結果、隠蔽が起こる手話に対して隠蔽中の顔・手領域を正確に分離し、手話認識に用いる領域特徴を抽出する。以上をの手法の有効性を実験によって確かめた。

**キーワード:** 手話認識, 手形状抽出, Kinect,

## 1. はじめに

近年、高齢化に伴う難聴者が年々増加しており、難聴者のコミュニケーションの1つである手話が注目されている。手話を扱える人は多くなく、難聴者とのコミュニケーションで広く使われる手話を理解できない場合が多々ある。手話を理解できない人は手話通訳者を雇うが、手話通訳者は年々増えているが全国約3000人(平成24年2月)しかおらず、手話通訳者が不足している。そのため手話通訳システムの開発が求められている。

手話を認識するためには、手形状の抽出が必要である。従来の研究では、肌色抽出法で顔と手領域を抽出する [1] [2]。しかし肌色に近い服装や、撮影環境に影響されるなどの条件から、肌領域の抽出ができないことがある。また、隠蔽中の手形状の変化には対応できないという欠点がある。この問題を解決するために、色情報の他に深度センサーから距離情報を取得し統合することにより、正確な手形状抽出の研究がなされている。武用ら [4] はステレオカメラ、佐藤ら [5] は ToF カメラ、竹内ら [6]、渡辺ら [7] は CCD

カメラと近赤外光照射装置、西村ら [3] らは Kinect を使用し手形状を抽出した。しかしこれらの手法は顔と手の間の距離差を単純に評価しているため、顔と手が重なる手話で顔から手までの距離の差が小さい場合には、顔領域と手領域を分離して抽出することが困難であった。

本稿では、Kinect [8] から得られた人体骨格情報 (図 1)、距離情報 (図 2) を用いて手形状抽出を行う。Kinect とは 2010 年に Microsoft から販売された Xbox360 向けのゲームデバイスである。Kinect は可視光センサーと深度センサーを搭載しており、2つのセンサーから得られた情報を元に手話の肌領域抽出を行う。

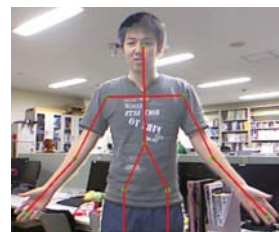


図 1 骨格情報



図 2 深度画像

## 2. 提案手法の概要

提案する手形状抽出法のフローチャートを図 3 で示す。まず始めに手話を Kinect で撮影し、OpenNI ライブラリを

<sup>†1</sup> 現在, 立命館大学  
Presently with Ritsumeikan University 〒 525-8577 滋賀県草津市野路東 1 丁目 1-1  
a) mori@i.ci.ritsumei.ac.jp  
b) matsuo@i.ci.ritsumei.ac.jp  
c) shirai@i.ci.ritsumei.ac.jp  
d) shimada@i.ci.ritsumei.ac.jp

用いて人体骨格情報を取得する。話者の腕がどの方向に伸びているか調べ、腕が前方に伸びているかどうかを判定することで手領域を抽出する手法を切り替える。手話開始時は顔・手領域は重なっていないと仮定し、人体骨格情報から顔・手候補領域を抽出する。手話によっては顔や手が重なる場合があるので、重なっている領域があるか判定する。隠蔽領域がなければ、各候補領域をそれぞれ手・顔領域として確定し抽出を終了する。隠蔽領域があると判定される場合は、どの領域で隠蔽が起こっているか調べる。手同士で隠蔽が起こっている場合、1フレーム前の手領域と現フレームの手領域候補範囲に対して3次元テンプレートマッチングを行う。顔と手の候補領域が重なっている場合、隠蔽直前の顔領域と現フレームの顔候補領域の範囲で3次元テンプレートマッチングを行う。その結果から顔領域を求め、顔以外の候補領域を手領域として抽出する。



図3 提案手法のフローチャート

### 3. 手話動作開始位置の顔・手領域抽出

手話動作を撮影する時に、話者には手話の開始位置からスタートしてもらう。手話の開始位置は、顔・右手・左手で隠蔽が起こっていない位置とする(図4, 図5)。



図4 腕が前に伸びている



図5 腕が下に伸びている

### 3.1 骨格情報の取得

本研究では、オープンソースライブラリであるOpenNI[9]を使用する。OpenNIは骨格追跡を行うことができ、各関節の3次元座標 $(\alpha, \beta, \gamma)$ を取得することができる。 $(\alpha, \beta, \gamma)$ を画像平面 $(x, y)$ に投影して奥行き情報 $z = \gamma$ とともに保持する。顔 $(x_F, y_F, z_F)$ ・右手 $(x_{HR}, y_{HR}, z_{HR})$ ・左手 $(x_{HL}, y_{HL}, z_{HL})$ ・右肘 $(x_{ER}, y_{ER}, z_{ER})$ ・左肘 $(x_{EL}, y_{EL}, z_{EL})$ がそれぞれ取得できる。これらを手がかりに顔・手候補領域抽出を行う。

### 3.2 前腕の方向ベクトルによって異なる手領域抽出

手話によっては、手領域によって肘が隠れており骨格情報が取得できない場合がある。そこで肘座標が正しく取得できているか調べるために、左肘座標 $(\alpha_{EL}, \beta_{EL}, \gamma_{EL})$ 、左手座標 $(\alpha_{HL}, \beta_{HL}, \gamma_{HL})$ から式(1)を用いて前腕の方向ベクトル成分 $(\Delta\alpha_L, \Delta\beta_L, \Delta\gamma_L)$ を求める。

$$\begin{pmatrix} \Delta\alpha_L = \alpha_{EL} - \alpha_{HL} \\ \Delta\beta_L = \beta_{EL} - \beta_{HL} \\ \Delta\gamma_L = \gamma_{EL} - \gamma_{HL} \end{pmatrix} \quad (1)$$

$|\Delta\gamma_L| > |\Delta\alpha_L|$  かつ  $|\Delta\gamma_L| > |\Delta\beta_L|$  を満たすとき前腕が前に伸びていると判定し、それ以外は腕が横に伸びている前腕領域がカメラから見えていると判定する。その判定結果に基づいて以下に述べる手法により顔と手領域を抽出する。

### 3.3 骨格情報に基づく顔・手領域抽出

前腕が前に伸びていると判定された場合、まず左手中心座標 $(x_{HL}, y_{HL}, z_{HL})$ の画素を仮の手領域とし、半径 $r$ (ここでは画像に映り込む手の大きさの前提から50ピクセルと仮定)の範囲内で仮の手領域に隣接する深度値が僅差の画素を仮の手領域にマージすることを繰り返す(以下この処理を肌領域復元法と記述する)ことで、仮の手領域を抽出する。このままでは中心から遠い指先の部分が抽出できないことがあるため、抽出された仮の手領域の重心座標 $(x_{HL_g}, y_{HL_g})$ を求め、重心から半径50ピクセルの範囲内で再度肌領域復元法を行う。この処理を重心が移動しなくなるまで繰り返すが、手話によっては重心が元の左手座標 $(x_{HL}, y_{HL})$ から離れてしまう。そこで肌領域復元法を実行する回数を最大で3回までとする。この処理で抽出した領域を左手候補領域とする。同様の処理を右手 $(x_{HR}, y_{HR}, z_{HR})$ に対しても行う。

しかしこの処理は、図4のように腕が話者から見て前方に伸びている時は有効であるが、図5のように腕が横方向に伸びている場合、手の平と腕領域が類似の深度値になるため、重心が腕領域に移動してしまう可能性がある。そこで前腕が横に伸びていると判定された場合は、画面に対して前腕領域(肘から手の領域)が平行なので、手首位置を

検出する. 左肘 ( $x_{EL}, y_{EL}$ ) から左手 ( $x_{HL}, y_{HL}$ ) に対して垂直方向に前腕領域の幅を調べ, 前腕領域幅が最小になる位置を手首位置とする (図 6). 重心が手首を越えて前腕領域まで移動しないように手首位置に境界線を設定した上で肌領域復元法を適用する. 図 7 は腕が横に伸びている場合に, 左手座標 ( $x_{HL}, y_{HL}$ ) から手候補領域を抽出した結果である.

また顔座標 ( $x_F, y_F$ ) から肌領域抽出法を適用し抽出した領域を顔候補領域とする. 図 5 の顔・手領域抽出結果が図 8 である.

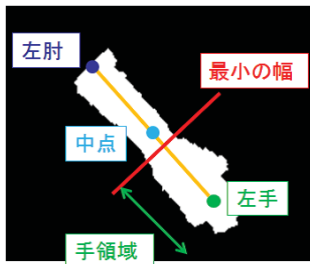


図 6 手首位置検出



図 7 手領域



図 8 顔・手領域抽出

## 4. 隠蔽が起こった場合の顔・手領域抽出

### 4.1 隠蔽の有無の判定

前節に述べた手法により顔・右手・左手の各候補領域が抽出される. もしいずれかの領域間で隠蔽が起こっている場合は当該候補領域どうしが接したり重なったりする. ここでは両手の隠蔽と顔と片手の隠蔽のみを取り扱うものとし, どの候補領域どうしが重なったのかを検知することで, どの領域間で隠蔽が起こっているのかを判定する.

手同士が重なり隠蔽が起こった場合と, 顔と手が重なって隠蔽が起こった場合でそれぞれ専用の顔・手領域抽出法を適用する.

### 4.2 手同士が重なった場合

手話は両手を使って表現するものが多いので必然的に手同士が重なる手話が多々ある. 手同士が重なり 1 つになっ

た領域を隠蔽手候補領域とする.

手同士が重なると判定された 1 フレーム前の手領域の深度値の値  $T(i, j)$  をテンプレート画像とし, 手同士が重なっている現在フレームの手候補領域の深度値  $I(i, j)$  を探索画像とする. テンプレート画像の横幅を  $M$  画素, 高さを  $N$  画素とし, 3 次元テンプレートマッチングを行う (式 (2)).

$$SAD(x, y) = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} |I(x+i, y+j) - T(i, j)| \quad (2)$$

$SAD$  が最小になる座標位置  $(x, y)$  に図 11 のようにテンプレートを重ねて現在フレームとの深度の差を画素ごとに計算する. この処理を左右両方の手についてそれぞれ行う



図 9 左手候補範囲



図 10 前フレームの左手領域

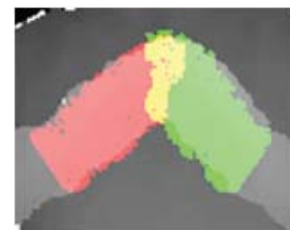


図 11 3次元テンプレートマッチング結果

テンプレート画像から深度の変化が小さい画素は現在フレームも手領域であると推測されるため, 仮の手候補領域とする. 深度の変化が大きい画素のうち, 正の方向に変化している画素は別の手が隠蔽している領域, 負の方向に変化している画素は体幹や背景の領域を誤って含んだものと推測されるので除去する. 仮の手候補領域の重心位置を基に肌領域復元法を行う.

手話認識は手形状の特徴抽出を行うので, 図 12 のように左手によって右手の一部が隠蔽している場合でも, 隠蔽している右手の手形状を推測して抽出する必要がある. そこで, 距離変化が正方向に変化している隠蔽領域 (図 11 の黄色領域) の面積が大きい方を隠蔽された手と判定し, 別の手によって隠蔽された手領域とする. その領域の深度情報は, 前フレームで取得したテンプレート画像の深度情報を追加して補完する. これにより図 13 のように隠れて見えない領域を推測して手形状を抽出することができる.

補完された手領域内の深度情報を, 次のフレームで利用

する深度テンプレート画像として登録する。この処理により、隠蔽中に手形状の変形が起こった場合にも対応が可能となる



図 12 右手候補範囲



図 13 前フレームの右手領域追加

### 4.3 顔と手が重なった場合

#### 4.3.1 3次元テンプレートマッチングによる手領域抽出

重なっている顔と手領域が充分離れていれば、深度情報で容易に顔と手領域を分離することが可能であるが、顔と手領域が近い場合は分離することが困難である。そこで4.2で述べた3次元テンプレートマッチングを行う。顔と手が重なり1つになった領域を顔候補領域とする。顔と手が重なる直前のフレームでの顔領域の深度値  $T(i, j)$  をテンプレート画像とし、現在フレームの顔候補領域の深度値  $I(i, j)$  を探索画像とする。テンプレート画像の横幅を  $M$  画素、高さを  $N$  画素とし、3次元テンプレートマッチングを行う。式(2)の  $SAD$  が最小になる座標位置  $(x, y)$  にテンプレートを重ねて隠蔽直前の顔領域との深度の変化を調べる。深度の変化が小さいところ画素は現在フレームも顔領域であると推測されるため、仮の顔候補領域とする。深度の変化が大きい画素のうち、正方向に変化している画素は手によって隠蔽された画素、負方向に変化している画素は体幹や背景の領域を誤って含んだものと推測されるので除去する。図14に深度の変化をグラデーションで表示する。隠蔽前の顔領域より前にある領域ほどを赤く、奥にある領域ほどを青く表示する。隠蔽前の顔領域よりも前にある領域で一番面積の大きい領域を手候補領域として抽出する。手領域候補に対し肌領域復元法を適用することによって顔を隠蔽している部分手候補領域（隠蔽手候補領域）を抽出する。手領域のうち顔と重なっていない部分は3.3で求めた手領域抽出法で求めることが出来る。そこでこの2つの手領域候補を接合することにより、全体の手候補領域が抽出できる。

しかし話者の顔が前後に動いてしまい隠蔽前と現在で顔の深度が違う場合では、たとえば図15のように手領域以外も隠蔽前の顔領域よりも前に出ていると処理されてしまうことがある。

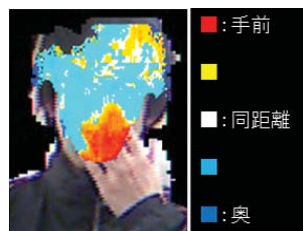


図 14 顔の距離変化 1

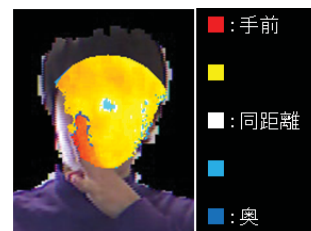


図 15 顔の距離変化 2

そこで話者が前後に動くことを考慮するために式(2)で求めた  $SAD$ 、顔テンプレートの面積  $S$  を用いて、顔の前後の移動量  $E$  を式(3)を計算する。

$$E = \frac{SAD}{S} \quad (3)$$

顔の前後移動を補償するために、 $E$  だけ顔テンプレートを前後に移動し(図16)テンプレート画像と現在フレームの深度の差を画素ごとに評価し直す(実際には先に計算した深度差を一樣にオフセットするだけでよい)。この処理により大部分の顔領域は除去できるが、手と鼻の深度が近い場合、額や鼻領域も抽出されることがある(図17)。

そこで、深度が最も手前に観測されている画素を含む領域を切り出し、その領域に対して深度差のヒストグラム(図18)を求め、判別分析法によって決定した閾値を用いて顔と手を分離する。

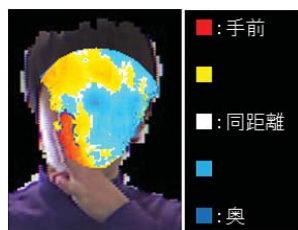


図 16 顔テンプレート移動後



図 17 顔よりも前にある領域

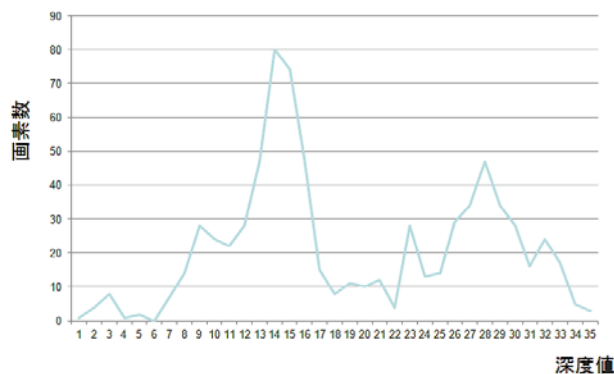


図 18 深度差ヒストグラム

この例では判別分析により深度差の閾値が20となり、手領域を分離した結果が図22である。

この手法は、図 16 のように手領域以外が抽出されている場合に有効であるが、図 20 のように手領域のみで余計な領域が含まれていない場合に判別分析法に適用すると、必ずなんらかの閾値が得られるため図 20 のように除去すべきではない手領域が除かれてしまう。そこで、クラス間分散  $\sigma_b^2$  とクラス内分散  $\sigma_w^2$  から分離度  $\lambda$

$$\lambda = \frac{\sigma_b^2}{\sigma_w^2} \quad (4)$$

を計算し、( $\lambda$  が閾値以上の場合、手領域以外が含まれているので有効と判断し、判別分析法の処理を反映させ、そうでない場合は手領域のみが含まれているのでそれ以上分離しない。今回の例では分離度の閾値を 3 と設定した。図 18 の例に対しては  $\lambda=4.5393$ 、図 21 の例では  $\lambda=1.2304$  になった。以上の一連の処理を行い最終的な手領域を抽出した結果が図 23 の緑線内の領域である。



図 19 顔の距離変化 3

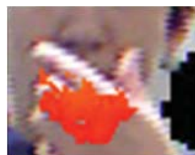


図 20 領域除去の失敗例

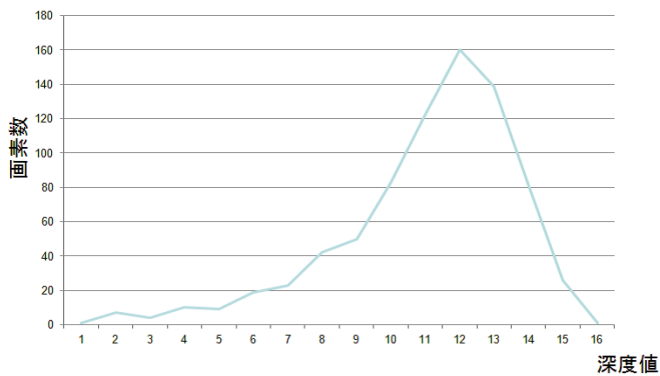


図 21 距離の差ヒストグラム

#### 4.3.2 3次元テンプレートマッチングによる顔領域抽出

4.3.1 の処理で顔と手が重なっている領域から手領域を抽出することができる。顔と手が重なっている領域から手領域を除いた領域が顔領域となる。顔座標  $(x_F, y_F)$  から手領域を除いた領域に対し肌領域抽出法を行う。その抽出した領域を顔領域とする。顔領域を抽出した結果が図 23 の赤線内の領域である。



図 22 閾値で分けた手領域

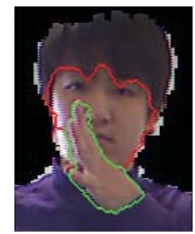


図 23 隠蔽中の顔・手領域抽出

## 5. 実験結果

実際に隠蔽を含む手話単語を 20 種類撮影し、本手法を用いて顔・手領域を抽出して、従来法「引用」と比べて顔・手領域抽出が向上したか比較した。図 24、図 25 は共に、顔と手の隠蔽が発生し、顔と手領域の深度の差が比較的小さい手話単語である。従来法による顔・手領域抽出結果(図 24)では手領域のうち抽出できていない部分があり、また背景が誤って含まれている。一方、本手法による抽出結果(図 25)では手領域ほぼ全て抽出できていることがわかる。一部、中指および親指の指先が検出できていない。



図 24 色情報から得た顔と手領域



図 25 距離情報から得た顔と手領域

図 26 は隠蔽中に手形状が変化する手話単語「年齢」に対して従来法によって顔・手領域を抽出した結果例である。従来法では正しく領域抽出できていないが、深度情報を使用した本手法では図 27 のように顔・手領域がより正確に抽出できていることがわかる。さらに本手法ではこの例のように隠蔽中に指が変形して形が変わる場合にも領域抽出ができていくことがわかる。

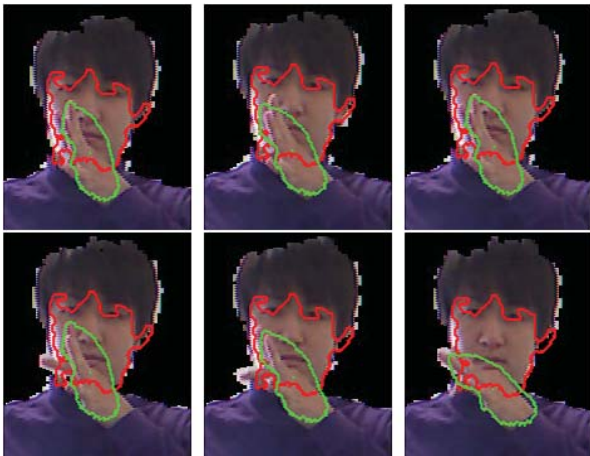


図 26 隠蔽中に手形状変化する手話 (従来法)

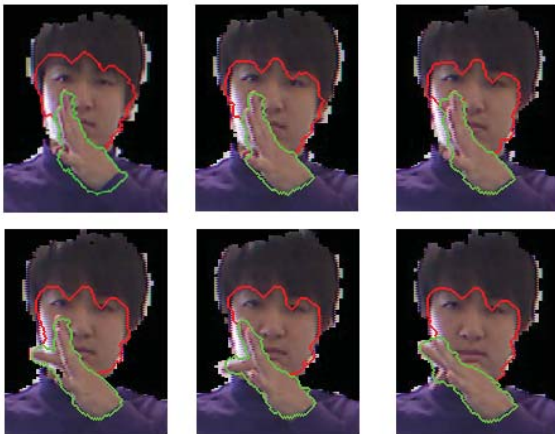


図 27 隠蔽中に手形状変化する手話 (提案法)

## 6. おわりに

本論文では、Kinect の OpenNI から得られた人体骨格情報に基づき手話の顔・手領域の抽出を行う手法を提案した。手領域によって肘が隠れて骨格情報が取得できない場合には、骨格情報から取得した手領域から深度の近い領域をマージして手領域を抽出した。腕が横に伸びており肘が見えている場合には、手と前腕領域で深度が近いため、手首位置を検出して境界を設定した上で肌領域復元法を適用した。

手同士で隠蔽が起こったと判断された場合、隠蔽前のフレームにおける手領域と現在フレームの手候補領域で深度画像のテンプレートマッチングを行い、同様に顔と手領域で隠蔽か?起こったと判断された場合は、隠蔽直前の顔領域と現在フレームの顔候補領域に対して、顔が前後に移動することを考慮しつつ深度画像テンプレートマッチングを行うことで、顔・手領域を抽出することができた。

今後の課題は、手と顔の隠蔽時、顔が回転を含んだ動きを考慮して手と顔を分離する手法の検討があげられる。ま

た本手法で抽出した両手の形状特徴を入力として手話単語の時系列モデルを学習し認識を行うことである。

## 参考文献

- [1] 松橋 聡, 藤本 研司, 中村 納, 南 敏”顔領域抽出に有効な修正 HSV 表色系の提案”, テレビジョン学会誌 49(6), pp.787-797, 1995-06-20.
- [2] 相本 幸治, 山田 寛, 松尾 直志, 白井 良明”手話認識のための動き特徴に基づく学習データの自動合成”, 画像の認識・理解シンポジウム (MIRU2011) 論文集, pp.710-717, 2011/
- [3] 西村 洋介, 今村 大輔, 堀内 靖雄, 川本 一彦, 篠崎 隆宏, 黒岩 眞吾” Kinect とパーティクルフィルタを用いた HMM 手話認識手法の検討”, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 111(430), pp.161-166, 2012-02-02.
- [4] 武用 麻弥, 芦川 裕太, 岡本 教佳”手話認識のための奥行き情報を用いた手領域抽出”, 映像情報メディア学会技術報告 32(35), pp.79-80, 2008-08-21.
- [5] 佐藤 新, 篠田 浩一, 古井 貞熙”ToF カメラによる 3D 手話認識”, 画像の認識・理解シンポジウム (MIRU2010), pp.1861-1868, 2010.
- [6] 竹内 麻梨子, 加藤 邦人, 山本 和彦”近赤外肌検出法を用いた手振りインタフェースシステムの構築”, 映像情報メディア学会技術報告 34(34), pp.35-38, 2010-08-30.
- [7] 渡辺 泰斗, 山本 和彦, 加藤 邦人, 小島 真一, 中西 悟”LH-002 近赤外光を用いた肌領域検出について有効性の考察”, 情報科学技術レターズ 6, pp.211-214, 2007-08-22.
- [8] Xbox 公式サイト. Kinect. <http://www.xbox.com/ja-JP/kinect>
- [9] OpenNI. <http://www.openni.org/>