

複数人の利用者の手の動きに追従する ユーザインタフェースの研究

山崎雄太† 大塚友章† 高笠綾華† 水谷晃三† 荒井正之†

帝京大学理工学部ヒューマン情報システム学科†

1. はじめに

センサを設置して利用者の手の動きを認識し、その動きに追従するようにプロジェクタからユーザインタフェース (UI) を投影する研究が行われている。例えば奥村らの Lumipen ではハイスピードカメラを用いることによって手の動きを捉え、プロジェクタからの投影軸をミラーによって制御することで手の動きに追従しながら UI を投影する[1]。手の動きに瞬時に追従する動作が実現されているが、複数人の手の動きに追従する方法については議論されていない。

複数人の手にそれぞれ異なる UI を投影するための容易な実現方法として、プロジェクタを固定し、プロジェクタ投影範囲の一部分を使って UI を投影する方法が考えられる。例えば Wilson らの LightSpace[2]では深度センサと複数のプロジェクタを使い、手のひらに追従する形でオブジェクトを投影する仕組みを実現した。しかしながら、簡易的なメニュー機能は実現されているものの、手に投影された UI をジェスチャにより操作する手法については検討されていない。

筆者らは、深度センサを用いて手の位置とジェスチャの認識方法を検討し、簡易的な UI を手の動きに合わせて投影するシステムを試作した[3,4]。しかしこのシステムでは、原理的に複数人で利用できないという問題点がある。そこで本研究では、UI の投影場所を決定する際に手の認識技術を用いて解決する方法を検討する。

2. システム概要

図 1 に検討システムの概念図を示す。上方にプロジェクタとセンサを設置する。プロジェクタの投影範囲を利用者の移動可能範囲として利用する。UI はプロジェクタ出力の一部を使用して手のひらに投影し、センサを用いて投影範囲内の手のひらの位置を認識する。

本研究では RGB カメラや深度センサを内蔵している Kinect (V2) センサ 1 台と液晶プロジェクタ 1 台を用いる。プロジェクタは床から 2.1m の高さに、Kinect は床から 2.0m の高さに設置す

A study of a followable user interface to a hand of multiple user.
†Yuta Yamazaki, Yusyo Otsuka, Ayaka Takakasa, Kozo Mizutani, Masayuki Arai : Department of Human Information Systems, Faculty of Science and Engineering, Teikyo University.

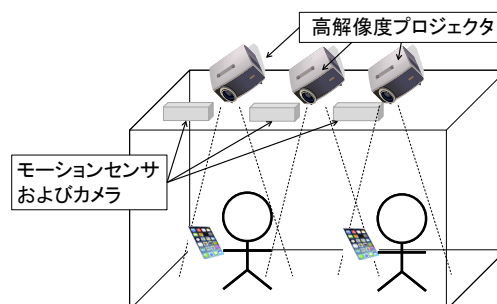


図 1 検討システムの概念図
(複数プロジェクタとセンサの配置例)



図 2 画像認識に用いる深度画像の例
(左: RGB 画像, 右: 深度情報の可視化画像)

る。手の認識範囲は床から 0.9m から 1.3m の高さとする。

3. 画像分類器の検討と試作プログラムの設計

本研究では、複数人での同時利用を可能にするために利用者の手を認識する。手の認識には OpenCV[5]の `opencv_traincascade` と呼ばれる画像分類器を使用することにした。opencv_traincascade により手の画像の特徴を学習させ、その結果を用いた分類器により手を認識する。認識した位置を用いて UI の投影を行う。

3.1. 学習画像の検討

検討中のシステムではセンサの位置と手の認識範囲を固定しているため、深度情報を可視化した画像を用いて手の認識を行うことにした。具体的には、図 2 に例示するように深度センサから深度情報を取得し、手の認識範囲にグレースケールを割り当てる。それ以外には白を割り当てた画像を可視化画像 (以下深度画像と呼ぶ) とする。

3.2. 学習特徴量の検討

opencv_traincascade で分類器を作成する際に、Haar-like, HOG (Histograms of Oriented Gradients), LBP (Local Binary Pattern) の 3 つの特徴量が選択できる。深度画像による手の認識に適した特

微量を検討するため比較実験を行う。

3.3. UI 投影プログラムの設計

分類器による認識結果を利用した、手へリアルタイムに UI を投影するプログラムを試作する。センサから深度情報を取得して深度画像に変換する。その画像に対して分類器を用いて手を認識し、認識した領域の座標を用いてプロジェクタから投影する UI の位置を決定する。

分類器による認識はセンサから深度情報が送られてくるフレームごとに行う。そのため UI を連続して手の位置に投影するには各フレームで認識される手の位置の関係を考慮する必要がある。本試作プログラムでは、手と認識された領域が前後フレームで一定範囲以上重なっている場合に同一の手の位置と判定することにする。また、一方の手の位置がほぼ停止状態であるところへ、他方の手の位置が衝突してきたときジェスチャの発生と認識するようにする。こうすることで手をスライドして画面を切り替えるような動作の実現を試みる。

4. 学習特徴量の比較実験の結果と試作プログラムの実装

3.2 で述べた比較実験のため、同じ条件下で取得した深度画像から正例画像（手が写っている画像）900 枚と負例画像（胴体部分のみで手が写っていない画像などを含む）1300 枚を作成し、3 つの特徴量ごとに学習させた。学習結果を用いて、別途用意した 180 枚の画像中にある 310 個の手の認識結果を比較した。実験結果を表 1 に示す。

図 3 には試作プログラムの実行例を示す。図 3（左）の画面上部が深度画像、画面下部が画像を反転させ分類器で手を認識した結果となっている。認識結果は赤色の矩形で表示される。センサの認識範囲とプロジェクタの投影範囲の重なる範囲を画面下部で指定でき、それにより実際に投影される際の領域全体の調整や UI のサイズを変化させることができる。図 3（右）は実際の投影結果の例である。本例では UI として色の違う四角形のみを出力している。また、手をスライドさせるようにして重ねた際に投影物体を変化させるジェスチャ認識の動作も確認した。

5. 考察

学習特徴量の比較実験では Haar-like 及び LBP の検出率が HOG に比べ高い。その一方で、HOG は他 2 つの特徴量より誤検出は少なく未検出が多い。これらの結果について検証が必要である。

試作プログラムでは、認識する際に手の向きを考慮していないため、UI を投影する際の向きの決定方法について検討が必要である。これに

表 1 学習特徴量比較

	検出	検出率	誤検出	未検出
Haar-like	306	98.7%	20	4
HOG	246	79.4%	2	64
LBP	305	98.4%	92	5

※誤検出とは誤った領域を検出したもの、未検出とは検出すべき領域を検出できなかった数を示す。

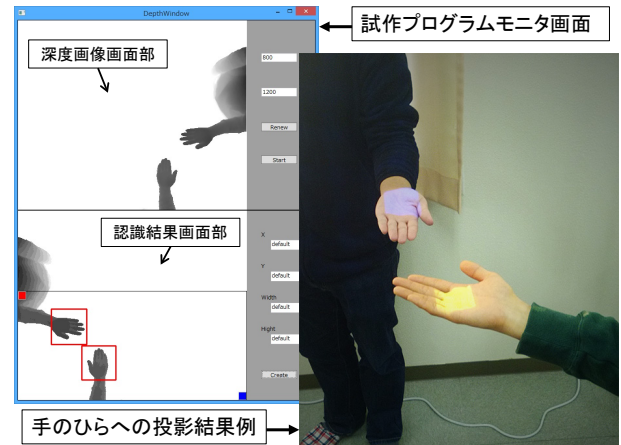


図 3 試作システムの動作結果例

ついては、認識結果の矩形の辺ごとの画素の輝度値を用い、その平均値の低い方向を手首として判定して向きを決定する方法が考えられる。

6. おわりに

本研究では、上方に設置したセンサを利用し、特定の範囲内の利用者の手の動きに追従するユーザインタフェースの実現方法を検討し、物体を投影し、操作できるプログラムを試作した。

参考文献

- [1] 奥村 光平, 奥 寛雅, 石川 正俊: 高速光軸制御を用いた動的物体への投影型拡張現実感, 東京大学 大学院情報理工学系研究科, Vol.67, No.7, J204-J211, 2013.
- [2] Andrew D. Wilson, Hrvoje Benko : Combining Multiple Depth Cameras and Projectors for Interactions On, Above, and Between Surfaces, Proceedings of the 23rd annual ACM symposium on User interface software and technology, pp.273-282, 2010.
- [3] 山口拓也, 伊藤匠, 水谷晃三, 荒井正之 : 利用者の手の動きに追従するユーザインタフェースの研究, 2015 年 IEICE 総合大会, ISS-P-3, 2015.
- [4] Takuya Yamaguchi, Kozo Mizutani and Masayuki Arai: A Study of Followable User Interface to Hand Behavior, International Journal of Knowledge Engineering, Vol.1, No.3, pp.240-243, 2015.
- [5] OpenCV, <http://opencv.org/>