プロアクティブ対話型観光案内システムのための顔向き・頭部動作認識

Recognition of Face Direction and Head Gesture for Proactive Sightseeing Guide System

香山 健太郎[*1] Kentaro KAYAMA
小林 亮博[*1] Akihiro KOBAYASHI
Lee Dongwook[*2]
角 薫[*1] Kaoru SUMI
加藤 丈和[*1] Takekazu KATO
門林 理恵子[*1] Rieko KADOBAYASHI
山崎 達也[*1] Tatsuya YAMAZAKI

[*1]独立行政法人 情報通信研究機構
National Institute of Information and Communications Technology (NICT)
[*2]Information and Communications University

We have been proposed proactive dialog system which actively presents acceptable information to users with a good sense of timing. To realize such dialog system it is indispensable to recognize non-verbal information of users and to make full use of them. Therefore we have developed following algorithms: to detect parts of face by using active appearance model (AAM), to estimate face direction based on the coordinates of the parts and three dimensional face model, and to recognize head gesture based on trajectories of the parts of face. We also developed the information display system which consists of 50 inches plasma display and three cameras whose poses are controllable. We implemented dialog scenario of sightseeing guide on the system and evaluated its efficiency on the demonstration exhibition session on the international conference.

1. はじめに

近年，認識技術の進歩を背景として，広告媒体として必要に応じて必要な情報を適宜切り替えて表示できるようなデジタルサイネージ技術が盛んに研究されており，様々なシステムが提案されている．

その中で使われる認識技術としては，オムロン・東芝・NECなどで開発されている，年齢・性別等の判別も可能な顔認識技術がある [Lao 09]．また，北陽電機は，レーザー測距センサを利用して，周辺を移動する人間の動線を取得できるようなシステムを開発している．しかし，情報の提示方法としては，年齢・性別等に応じた情報をシステム側が選択して一方的に提示するだけのものがほとんどである．ユーザが望ましい情報を引き出すためのシステムは用意されていないか，あるいはタッチパネル等の明示的な入力装置が必要となる．

一方，観光地の主要駅周辺などでは，対話型観光案内システムが設置されていることが多い．そのような対話システムでは，ユーザの意思や希望を伝えるため，キーボードやタッチパネル，ボタン等が用意されている．しかし，このようなシステムを使ってみようとする人はほとんどおらず，また，使ってみようとした人でも，対話のテンポの悪さに途中で投げ出してしまうことも多い．

このような問題に対して，我々は，人間と機械とのプロアクティブな対話を可能にする新しいインタラクティブ情報ディスプレイシステムを提案している [小林 09]．プロアクティブな対話システムとは，システム側からも積極的に気の利いた情報を気の利いたタイミングで提示するものである [河原 08]．そして，そのような対話の実現のためには，ユーザが対話中に見せる視線や顔向きの変化，頭部動作などの非言語情報を検出することが重要であると我々は考え，このようなディスプレイにおける観光案内において，人間がどのような非言語情報を表出するか，およびそれをどの程度画像認識により判別できるか，ということについての研究を行っている．

図 1: 情報ディスプレイシステムの概観

本稿では，まず，2. 節で試作した情報ディスプレイについて説明する．そして，非言語情報認識アルゴリズムとして，3. 節では顔向き検出を，4. 節では頭部動作認識を行うアルゴリズムについて述べる．

さらに，我々はこのアルゴリズムを用いてプロアクティブ対話型観光案内システムのプロトタイプを構築し，デモ展示を行った．その際に得られた情報を分析し，頭部動作認識の性能評価を行ったので，それについて 5. 節で述べる．

連絡先: 香山 健太郎 kayama@nict.go.jp

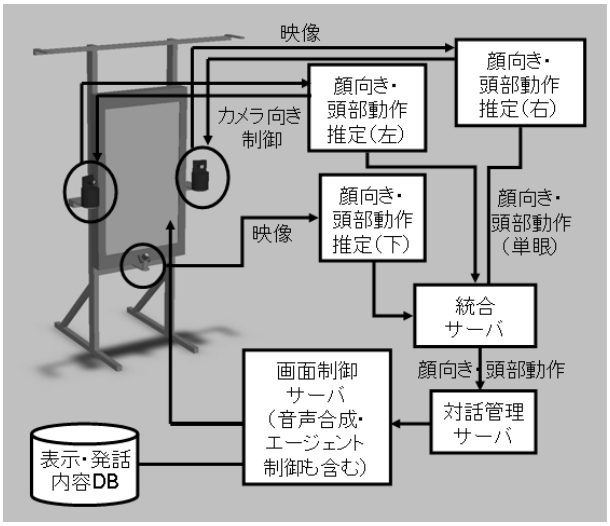

図 2: 情報ディスプレイシステムのソフトウェア構成

2. 情報ディスプレイ

これまでに，我々は，プロアクティブな対話システムのプロトタイプとして，50 インチプラズマディスプレイを縦置きにし，その周囲にユーザ認識用カメラを設置した「情報ディスプレイ」システムを構築し，ユーザの視線方向及び顔向きを用いた情報提示システムを提案してきた [水口 07].

今回構築したシステムの様子を図 1 に示す.

我々は，このシステムを用いた対話型観光案内の実現を目指している．そのためのソフトウェア構成を図 2 に示す.

3. 顔向きの検出

本システムでは，次のような処理を行って顔向きの検出を行っている (図 3).

1. Haar-like 特徴量を用いた顔領域検出
2. Active Appearance Model (AAM) を用いた顔パーツの検出・追跡 [Kobayashi 08, 小林 08]
3. 最急降下法による顔パーツの 3 次元顔モデルへのあてはめ (顔向き推定)
4. 楕円マッチングによる眼球モデルへのあてはめ (視線推定) [佐竹 08]

各処理の詳細は次の通りである.

顔領域検出 本システムでは，1 秒間に 15 フレームの 800 × 600pixel の画像が入力される．直前のフレームで画像中に顔が存在しなかった，あるいは次で述べる顔パーツの追跡に失敗した場合，画像に Haar-like 特徴量を用いた顔領域検出アルゴリズムを適用する.

顔パーツの検出・追跡 前フレームで顔パーツが検出されていた場合はその座標値を，新たに顔が検出された場合は規定の座標値を初期値とした上で，Active Appearance Model (AAM) を用いて顔の特徴点 45 点を抽出する．AAM は，各特徴点の画像上の座標を並べたものを主成分分析により圧縮したベクトル，および特徴点を頂点とする三角領域内の画素の輝度値を並べたものを主成分分析により圧縮したベクトルを組み合わせ，再度，主成分分析することで作成される.

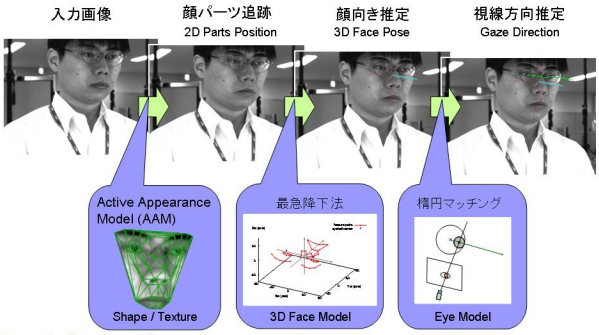

図 3: 顔向き検出

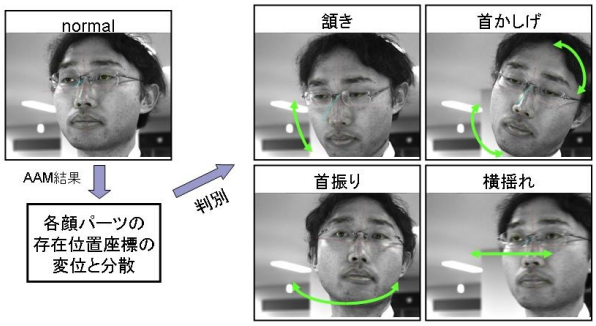

図 4: 頭部動作認識

顔向き推定 あらかじめ作成してある 3 次元顔形状モデルにおける各特徴点の 3 次元座標と，上記で得られた各特徴点の画像上の座標から，最急降下法を用いて 6 自由度 (回転 3 自由度・並進 3 自由度) の顔向きパラメータを求める.

視線推定 目の領域に対して二値化を行った上で楕円あてはめを行い，虹彩領域候補を検出する．そして，顔向き推定の際に得られた眼球中心の 3 次元座標と，虹彩領域候補の画像上座標および顔向き推定結果から計算した虹彩中心の 3 次元座標を結ぶ直線を視線方向とする.

4. 頭部動作認識

さらに，前節で検出した顔パーツの特徴点 45 点の座標およびその位置関係の時間的な変化から，頭部動作を認識する.

認識する頭部動作は次の 4 つである (図 4).

- 頷き
- 首かしげ
- 首振り
- 横揺れ

i 番目の顔パーツの時刻 t における画像上の座標値を $(x_i(t), y_i(t))$ とする．また，各種判定のための基準値として，

初期状態の左右の目の外側の端点間の距離を用いることとし，それを D_{base} とする．なお，初期状態には，顔が最初に認識されたフレームの状態を用いる．

認識には，次の値を用いる．

- 直前フレームからの各特徴点の x 方向の移動量 $dx_i(t) = x_i(t) - x_i(t-1)$，および y 方向の移動量 $dy_i(t) = y_i(t) - y_i(t-1)$ の標準偏差 $\sigma_{xall}(t), \sigma_{yall}(t)$
- 主要特徴点の 10 フレーム間の x, y 座標値の標準偏差 $\sigma_{xi}(t), \sigma_{yi}(t)$

まず，そのフレームで動きがあったかどうかを判断する．これは，

$$\sigma_{xall}(t) > 0.04 \times D_{base} \text{または} \sigma_{yall}(t) > 0.04 \times D_{base}$$

であれば，何らかの動作があったと判定する．揺れなどの平行移動の場合は $\sigma_{xall}(t), \sigma_{yall}(t)$ の値は小さくなるため，これではじいている．

また，現在の顔が傾いているかどうかを調べる．これには次のような特徴点を結ぶ線分を用いる．

- 左右の目の外側の端を結ぶ線分
- 口の両端を結ぶ線分
- 左目の外側の端と口の左端を結ぶ線分
- 右目の外側の端と口の右端を結ぶ線分

これらの線分の初期状態からの傾きを調べ，4 つとも閾値以上であれば，現在の顔が傾いているとみなす．

次に，顔パーツの主要特徴点についての最近 10 フレームの動きを調べる．主要特徴点としては，左右の目および口の両端点の合計 6 点を用いる．そして，これらの特徴点について，10 フレーム間の x, y 座標値の標準偏差 $\sigma_{xi}(t), \sigma_{yi}(t)$ が

$$\sigma_{xi}(t) > 0.08 \times D_{base}, \sigma_{yi}(t) > 0.05 \times D_{base}$$

となれば，それぞれ x 方向 (横)・y 方向 (縦) に動きがあったとみなす．そして，動きがあったと見なされる特徴点がそれぞれ 4 つ以上あれば，顔全体としてその方向に動いているとみなす．

以上により，現在のフレームの動きを決定する．その優先順位は

1. 横方向に動きがあり，かつ平行移動ではない
2. 顔が傾いている
3. 縦方向に動きがある

としている．これは，頷きはほとんど横方向の動きを伴わないが，首振りは縦方向の動きを伴うことが多いためである．

そして，これらの動きを数フレームまとめて頭部動作を判別する．

- 上記 1. が 3 フレーム続けば首振り
- 上記 2. が 5 フレーム続けば首かしげ
- 上記 3. が 1 フレーム検出されてかつその後 3 フレーム首振り・首かしげがなければ頷き

これは，それぞれの動作の持続時間の違いに基づいている．

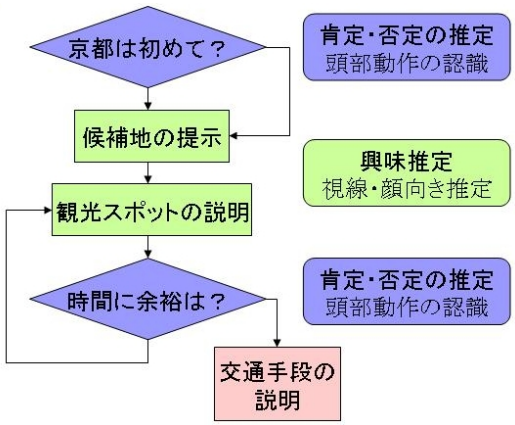

図 5: 対話シナリオ

5. 認識性能評価

5.1 対話シナリオ

本稿で述べた情報ディスプレイシステム，および認識アルゴリズムの性能を評価するため，これらを用いた対話シナリオを設計し実装した．

今回はこのシステムが京都駅に置かれていると仮定し，利用者を，特に目的地は決まっていないが少し時間が空いたので観光をしようとしている人と設定した．そして，そのような利用者に対しシステム主導で観光スポットを提示するというタスクを実装した．

今回の対話シナリオでは，ボタン・タッチパネル・マイク・キーボード等の明示的な入力機器を用いず，利用者の非言語情報の認識のみを用いて対話を進めることとした．そのシナリオを図 5 に，画面表示を図 6 に示す．詳細は次の通りである．

1. システムは利用者に京都が初めてかどうかを問いかけ，それに対し頷きか首振りで答えるよう求める．
2. その答えに応じて 4 つの観光スポットの名前と写真を提示する (図 6 左上)．システムは利用者がどこを見ているかを推定し，それに基づき興味があると思われるスポットを選択する．
3. 選択したスポットの説明および写真の提示を行う．ここでも，利用者がある写真を見るとその写真が拡大表示される (図 6 右上)．
4. システムは利用者にもう 1ヶ所行く余裕があるかどうかを問いかけ，それに対し頷きか首振りで答えるよう求める (図 6 左下)．
5. もし余裕があるならば，選択したスポットの近くの場所についても 3. と同様に説明を行う．
6. 最後に，説明したスポットへの交通手段を示して終了する (図 6 右下)．

5.2 国際会議におけるデモ展示

2008 年 12 月 15 日,16 日に大阪国際会議場にて行われた 2nd International Symposium on Universal Communication (ISUC) のデモセッションにおいて本システムの展示を行い (図 7)，約 70 名がこの観光スポット提案デモを体験した．

図 6: 対話画面

その結果，おおむね好意的な反応が得られたものの，勝手に対話が進んでいくように感じられる，ユーザからの割り込みも入れられるようにするべきという指摘が複数得られた．

また，対話中に約 300 回頭部動作を認識する機会があり，システムによる認識と人間の目視による認識との一致率は約 70%であった．

認識に失敗した理由としては，まず，頭部動作の際に画像がぶれてしまうことが多く，顔パーツ検出自体がうまく行われていないことがあげられる．また，頷きの速度や大きさは個人によってかなり異なっており，頷きではなく少し動いただけと思われる動作を頷きと認識してしまったり，小さくかつ動きの速い頷きを見落としてしまったりということが多かった．

6. まとめ

本稿では，対話中に表れる顔向き・頭部動作の認識を行うアルゴリズムについて述べた．また，それらを組み込んだ，大型ディスプレイによるプロアクティブ対話型観光案内システムを構築し，国際会議のデモセッションにおいて展示を行い認識性能を評価したので，その結果について述べた．

今後は，顔向き・頭部動作の認識性能の向上を図るとともに，このような対話中に表れる非言語情報の解析を行い，ユーザにとってより満足度の高いプロアクティブな対話を実現するための対話戦略についても検討して改良を加えていく予定である．

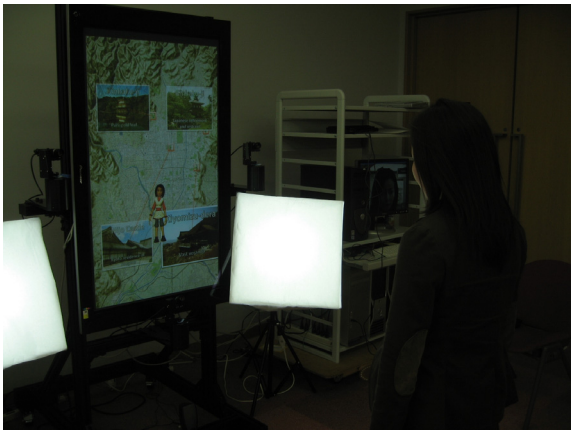

図 7: 対話型観光案内システムのデモ展示

謝辞

本研究にあたってご指導・ご助言いただいた京都大学松山研究室の松山隆司教授，川嶋宏彰講師，平山高嗣特任助手に感謝します．

参考文献

[河原 08] 河原達也, 川嶋宏彰, 平山高嗣, 松山隆司: 対話を通じてユーザの意図・興味を探り情報検索・提示する情報コンシェルジェ, 情報処理, Vol. 49, No.8, pp. 912–918 (2008).

[Kobayashi 08] A. Kobayashi, J. Satake, T. Hirayama, H. Kawashima, T. Matsuyama: Person-Independent Face Tracking Based on Dynamic AAM Selection, IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG) (2008).

[小林 08] 小林亮博, 佐竹純二, 平山高嗣, 川嶋宏彰, 松山隆司: AAM の動的選択に基づく不特定人物の顔追跡, 情処学 CVIM 研報, No. 161, pp.35–40 (2008).

[小林 09] 小林亮博, 香山健太郎, Lee Dongwook, 角薫, 加藤丈和, 門林理恵子: 顔向き・頭部動作推定を用いたプロアクティブ情報提示システムの提案, 信学会全国大会 (2009).

[Lao 09] Lao ShiHong, 山口修: 顔画像処理技術の動向 (前編), 情報処理, Vol. 50, No. 4, pp. 319–326 (2009).

[水口 07] 水口充, 浅野哲, 佐竹純二, 小林亮博, 平山高嗣, 川嶋宏彰, 小嶋秀樹, 松山隆司: Mind Probing: システムの積極的な働きかけによる視線パタンからの興味推定, 情処学ヒューマンコンピュータインタラクション研報, No. 125, pp.1–8 (2007).

[佐竹 08] 佐竹純二, 小林亮博, 平山高嗣, 川嶋宏彰, 松山隆司: 高解像度撮影における実時間視線推定の高精度化, 信学技報 PRMU2007, Vol. 107, No. 491, pp.137–142 (2008).