マルチモーダル対話における頭部ジェスチャ認識

田中太喜 † 嶋田和孝 ‡ 遠藤 勉 ‡

† 九州工業大学大学院情報工学研究科情報科学専攻
‡ 九州工業大学情報工学部知能情報工学科

概要

本論文では，固定したカメラで取り込んだ動画像から，頭部ジェスチャの認識を行うことを目指す．提案する手法は，まず，動画像中から色情報に基づかない手法により，顔・目・鼻・口の領域を抽出する．そして，一定フレーム間において，抽出した領域の位置・オプティカルフローなどの素性を基に，帰納推論システム「C4.5」を用いて頭部ジェスチャを認識する決定木を生成し，頭部ジェスチャの認識を行う．実験の結果より，頭部ジェスチャ認識を高精度で行えることを確認した．また，C4.5 で生成した頭部ジェスチャの決定木が，人手で作成したシンプルな頭部ジェスチャの決定木よりも有効であることを確認した．

Head Gesture Recognition in Multimodal Conversation

Taiki TANAKA†, Tsutomu ENDO‡, and Kazutaka SHIMADA‡

†Graduate Scool of Computer Science and Systems Engineering, Kyushu Institute of Technology
‡Department of Artificial Intelligence, Kyushu Institute of Technology

Abstract

In this paper we aim to recognize head gestures from moving images with a USB camera. First our method extracts the area of the face, eyes, nose and mouth without color information. For the process we use Haar-like features. We also employ integral images to detect features in real time. Next it generates a decision tree by using C4.5 for the head gesture recognition. We use the locations and optical flows of each part, e.g., face and eyes, for features of C4.5. In the experiment we obtained high accuracy.

1　はじめに

近年，ロボットや，デパートなどのインフォメーションを行うシステムを構築する上での，マンマシンインタフェース技術の重要性が高まってきている．これらの技術においては，人間が自然な行動で意図した内容を，機械が認識できることが重要となっている．日常，人間同士のコミュニケーションは，音声言語と身振り手振り，表情などの情報を組み合わせたマルチモーダルなコミュニケーションを行っており，一方のみでは，コミュニケーションが成立しにくい．なぜなら，音声発話には文法からの逸脱・省略や指示語の頻繁な出現などという話し言葉特有の問題点があり，ジェスチャには，それが身体の空間的・時間的な連続動作であるため，動作のどの部分にどのような意味があるかを決定しにくいという問題点がある．よって，音声情報や映像情報といった複数のメディアをモダリティとして用いることで，音声発話とジェスチャそれぞれが持つ問題点をうまく相互補完することができれば，人間と機械との自然なコミュニケーションが可能となる．

ここで，我々はモダリティの一つとして，頭部ジェスチャに注目する．頭部ジェスチャは，単独で情報を伝達するだけではなく，音声情報を補完する重要な情報伝達手段であり，音声情報のみでは困難な，意図や心境なども汲み取ることができる．これまでに頭部ジェスチャ認識に関する様々な研究が行われている [1][2]．これらの頭部ジェスチャ認識システムの多くは，あるジェスチャのモデル化を行うために，顔や目の位置や，動きなどの情報を素性として利用している [3]．そのため，ジェスチャ認識を行う前処理として，顔や目の特定領域の抽出が重要であり，ジェスチャ認識と同様に研究が行われている [4]．これら領域の抽出に最も利用されている手法は，カラー画像の色情報に基づいて肌色領域を特定し，抽出する方法 [5][6] である．しかし，色情報は個人差があり，さらに照明の変化に影響を受け易く，ロバスト性に問題が生じてしまう．また，目領域を抽出する方法として，近赤外線を利用して目の虹彩を抽出し，目領域の抽出を行う方法 [7] があるが，専用の高価なカメラが必要となり，容易に使用することが難しい．

そこで，本研究では，顔の構成要素領域の濃淡画像を矩形特徴化することで，色情報に基づかない領域抽出を行う．そして、抽出された顔や目領域の位置やオプティカルフローなどに基づき，一定フレーム間での頭部ジェスチャ認識を行う．本稿は，顔の構成要素領域の特徴に基づいた頭部ジェスチャ認識手法について報告するものであり，まず，領域抽出手法に用いる関連研究の手法について述べ，それを用いた，頭部ジェスチャ認識のための領域抽出手法や，頭部ジェスチャ認識のための素性の項目について述べる．そして，頭部ジェスチャ認識実験・結果について考察する．

2　関連研究

ここでは，ジェスチャ認識を行うための顔領域の抽出に Rainer ら [8] が提案した「Haar like 特徴に基づいた特徴抽出」と，処理を高速化するために，Paul ら [9] が提案した「Integral Image」について述べる．

2.1　Haar like 特徴

Haar like 特徴とは，画像における特徴量として，任意の画像矩形領域の濃淡に基づいた特徴である．各画素の明度値をそのまま用いるのではなく，近接する 2 つの矩形領域の明度差を求めることで得られる特徴である (図 1)．

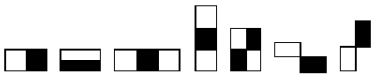


図 1: 基本的な Haar like 特徴のセット

図 1 で求めた Haar like 特徴を組み合わせ，領域抽出モデルとなる Haar like 特徴に拡張を行う (図 2)．図 3 に目領域の Haar like 特徴例を示す．

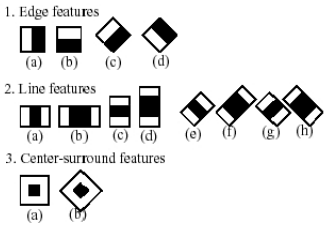


図 2: 拡張した Haar like 特徴

図 3: Haar like 特徴による目領域抽出例

2.2　Integral Image

Haar like 特徴を高速に計算し，リアルタイムな処理を行うためには，Paul ら [9] が画像の中間表現として用いた「Integral Image」を利用する．

画像左上が $(x, y) = (0, 0)$ とした場合，Integral Image の画素値は，元画像の，(x, y) 座標位置を含む左上の全画素の総和である．Integral Image の計算は，走査線上の累積和などを用いることでワンパスで計算することができ，処理対象となる矩形領域の大きさに依存せず，一定の時間で Haar like 特徴を抽出することができ，リアルタイムに領域抽出を行うことが可能となる．

3　頭部ジェスチャ認識手法

2 節で述べた手法を用いて，USB カメラから取り込んだ動画像から連続して特定の領域を抽出し，領域の情報などを基に頭部ジェスチャ認識を行う手法について述べる．

3.1　認識する頭部ジェスチャの種類

本研究では，頭部ジェスチャにおいて,「うなずき」(図 4(a)),「首振り」(図 4(b)),「首傾げ」(図 4(c)) の 3 種類の認識を行う．それぞれのジェスチャ特徴に基づき，決定木を生成し，頭部ジェスチャの認識を行う．

(a) 「うなずき」

(b) 「首振り」

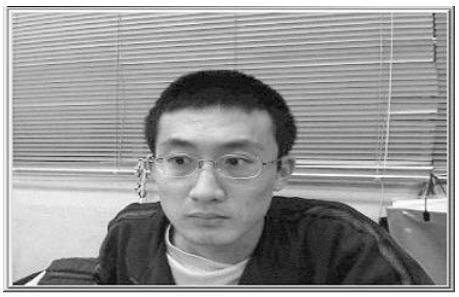
(c) 「首かしげ」

図 4: 認識する頭部ジェスチャの種類

3.2　領域抽出処理

顔の構成要素である，目・鼻・口は極端に形が変化することはない．この特徴を利用し，2.1 節で述べた Haar like 特徴を用いて顔・目・鼻・口の領域を抽出する．

まず，キャプチャしたフレーム画像を，グレイスケール画像に変換して，任意の矩形サイズを膨張しながら画面上を走査し，顔，目，鼻，口の順に各構成要素の Haar like 特徴との合致数が，3 箇所以上の領域を各領域として抽出する．ただし，Haar like 特徴による領域抽出では，複数の領域を抽出してしまうので，領域の位置の特徴に基づき処理を行う．例えば,「口領域は顔領域内の低い箇所に位置する」といった特徴である．また，処理の高速化のために，目・鼻・口の順で領域を抽出する上で，例えば「目の上に鼻や口は位置しない」というヒューリスティックに基づき，走査領域の絞込み (図 5(b) 図 5(c) 図 5(d)) を行う．また，同じ構成要素において，既に候補領域があり，新たに抽出した候補領域の中心が，他の候補領域内にある場合，候補として除去する (図 6)．また，顔を抽出できなかった場合は，他の構成要素の領域の誤抽出を防ぐため，そのフレーム画像での目，鼻，口の領域抽出は行わない．



(a) 顔走査範囲

(b) 目走査範囲　(c) 鼻走査範囲　(d) 口走査範囲

図 5: 各領域の走査範囲例

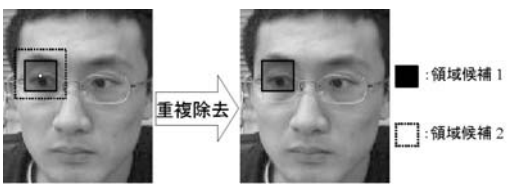


図 6: 重複除去例

抽出処理手順を以下に示す.

処理 1 全画面中 (図 5(a)) から顔の Haar like 特徴の合致数を調べ，顔領域を抽出する．顔の Haar like 特徴は，顔の向きにより以下の 4 種類あり，それぞれ合致数を調べ，領域抽出を行う．ただし，抽出する顔の向きは以下の順に優先順位があり，ある向きにおいて顔領域を抽出した場合，それ以降の向きの顔領域の抽出は行わない．また，複数の顔領域を抽出した場合，候補領域が最大の領域を顔領域とする．

1. 正面
2. 左側面
3. 右側面
4. 30 度傾いている

処理 2 顔領域を抽出した場合，顔領域内 (図 5(b)) から，目の Haar like 特徴の合致数を調べ，目領域を抽出する．ただし，3 箇所以上の目領域を抽出した場合，抽出領域の中心位置が高い領域の上位 2 領域を目領域とする．また，2 箇所目領域を抽出した場合には，条件式 1，2 に基づき左右の目を判定する．ただし，正面の顔に対して，1 箇所しか抽出されない場合には，条件式 3 に基づいて，それが左右の目のいずれかであるか判定する．

$$right\_eye = \begin{cases} eye\_area_i \\ (if\ eye\_area\_x_i > eye\_area\_x_j) \\ eye\_area_j \\ (if\ eye\_area\_x_i < eye\_area\_x_j) \end{cases} \quad (1)$$

$$left\_eye = \begin{cases} eye\_area_i \\ (if\ eye\_area\_x_i < eye\_area\_x_j) \\ eye\_area_j \\ (if\ eye\_area\_x_i > eye\_area\_x_j) \end{cases} \quad (2)$$

$$\begin{array}{l} right\_eye = eye\_area_i\ (if\ eye\_area\_x_i > face\_x) \\ left\_eye = eye\_area_i\ (if\ eye\_area\_x_i < face\_x) \end{array} \quad (3)$$

$right\_eye$: 右目　　$left\_eye$: 左目
eye_area : 抽出した目領域
(i, j は領域箇所, x は横幅の中心位置)
$face\_x$: 抽出した顔の横幅の中心位置

処理 3 処理 2 の終了後，鼻の Haar like 特徴の合致数を調べ，鼻領域を抽出する．ただし，処理 2 において，目領域を抽出した場合は，目領域の中心位置よりも低い顔領域 (図 5(c)) を走査し，目領域を抽出しなかった場合は，処理 2 と同じ領域を走査する．また，2 箇所以上の鼻領域を抽出した場合，抽出領域の中心位置が高い領域を鼻領域とする．

処理 4 処理 3 の終了後，口の Haar like 特徴の合致数を調べ，口領域を抽出する．ただし，処理 3 において，鼻領域を抽出した場合は，鼻領域の中心位置よりも低い顔領域 (図 5(c)) を走査し，鼻領域を抽出しなかった場合は，処理 3 と同じ領域を走査する．また，2 箇所以上の口領域を抽出した場合，抽出領域の中心位置が低い領域を口領域とする．

3.3 頭部ジェスチャ認識に用いる素性抽出

頭部を上下に振る「うなずく」，左右に振る「首振り」という連続した動きである頭部ジェスチャを認識するためには，時系列を考慮した素性を抽出することが有効である．そこで，今回は 20 フレーム間において素性を抽出する．素性値として各領域の位置に加えて，各領域のオプティカルフローを抽出する．そして，頭部ジェスチャを認識する素性として用いることで，連続した動きの頭部ジェスチャを認識する．

3.3.1 オプティカルフロー計算方法

フレーム間の領域の移動方向，大きさはフレーム間の処理時間に基づいて計算を行う．グレイスケール画像に対して，以下の手順により処理を行う．

処理 1　フレーム間差分画像を式 4 を用いて取得.

$$\begin{array}{l} fdi(x, y) = \\ |\ front\_image(x, y) - present\_image(x, y)\ | \\ fdi(x, y) : \text{フレーム間差分画像} \\ front\_image(x, y) : \text{前フレーム画像} \\ present\_image(x, y) : \text{現フレーム画像} \end{array} \quad (4)$$

処理 2　処理 1 で求めた画像に対して，式 5 を用いてモーション履歴画像を取得.

$$\begin{array}{l} mhi(x, y) = capture\_time\ (if\ fdi(x, y) > thresh) \\ mhi(x, y) : \text{モーション履歴画像} \\ capture\_time : \text{キャプチャ時間}, thresh : \text{閾値 } 30 \end{array} \quad (5)$$

処理 3　処理 2 で求めた画像に対して，式 6 を用いて画素毎のオプティカルフローを取得

$$\begin{array}{l} ori(x, y) = arctan(Dy(x, y)/Dx(x, y)) \\ ori(x, y) : \text{座標 } (x, y) \text{ におけるオプティカルフロー} \\ Dy(x, y), Dx(x, y) : mhi(x, y) \text{ の } y \text{ 軸}, \\ x \text{ 軸方向における微分係数} \end{array} \quad (6)$$

処理 4　3.2 節で求めた各領域内において，処理 3 で求めた値のヒストグラムを生成し，最大となるヒストグラム値の方向を，各領域におけるオプティカルフローの方向 ang とし，最大値をオプティカルフローの大きさとする．ただし，ヒストグラムの最大値が 0 の場合，ang は「方向なし」とする．

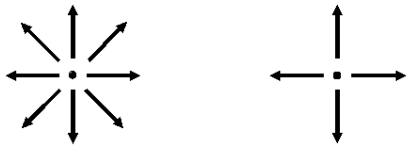
処理 5　処理 4 で求めた ang を式 7 を用いて 8 方向に，もしくは式 8 を用いて 4 方向に変換する．オプティカルフローの簡易化により，領域の移動方向を特徴化することができる．ただし，ang が「方向なし」の場合は計算は行わない．

$$8angle = \begin{cases} 0\text{ 度} & (0\text{ 度} \leqq ang < 45\text{ 度}) \\ 45\text{ 度} & (45\text{ 度} \leqq ang < 90\text{ 度}) \\ 90\text{ 度} & (90\text{ 度} \leqq ang < 135\text{ 度}) \\ 135\text{ 度} & (135\text{ 度} \leqq ang < 180\text{ 度}) \\ 180\text{ 度} & (180\text{ 度} \leqq ang < 225\text{ 度}) \\ 225\text{ 度} & (225\text{ 度} \leqq ang < 270\text{ 度}) \\ 270\text{ 度} & (270\text{ 度} \leqq ang < 315\text{ 度}) \\ 315\text{ 度} & (315\text{ 度} \leqq ang < 360\text{ 度}) \end{cases} \quad (7)$$

$$4angle = \begin{cases} 0\text{ 度} & (0\text{ 度} \leqq ang < 90\text{ 度}) \\ 90\text{ 度} & (90\text{ 度} \leqq ang < 180\text{ 度}) \\ 180\text{ 度} & (180\text{ 度} \leqq ang < 270\text{ 度}) \\ 270\text{ 度} & (270\text{ 度} \leqq ang < 360\text{ 度}) \end{cases} \quad (8)$$

$8angle$: 8 方向のいずれかに変換したオプティカルフローの方向
$4angle$: 4 方向のいずれかに変換したオプティカルフローの方向
ang : 変換前のオプティカルフローの方向

式 7 で求めた 8 方向と「方向なし」の合計 9 方向は図 7(a) で示される方向になり，式 8 で求めた 4 方向と「方向なし」の合計 5 方向は図 7(b) に示される方向になる．



(a) 9 方向　　(b) 5 方向

図 7: オプティカルフロー

3.3.2　素性抽出項目

3.2 節の処理後，3.3.1 節で述べた処理に基づき各領域のオプティカルフローを計算し，各フレームにおける素性として，各フレーム毎に 18 素性 (表 1 は $8angle$ の場合，表 2 は $4angle$ の場合)，計 360 素性と，20 フレーム間における 26 素性 (表 3)，合計 386 素性を抽出する．ここで表 1 の素性は，オプティカルフローの方向が図 7(a) の値となり，表 2 の素性は，オプティカルフローの方向が図 7(b) の値となる．また，ジェスチャ認識に用いる素性は,「表 1，表 3」の組み合わせ，もしくは「表 2，表 3」とする．

表 1: 各フレームにおける素性 ($8angle$ を使用)

素性項目	素性値
顔の向き (1 素性，5 項目)	正面，左側面， 右側面，傾き，非抽出
顔・目・鼻・口の オプティカルフローの方向 (5 素性，9 項目)	0 度，45 度，90 度， 135 度，180 度，225 度， 270 度，315 度，方向なし
顔・目・鼻・口の オプティカルフローの大きさ (5 素性)	自然数，非抽出
両目の位置のずれ (1 素性)	傾きあり，傾きなし， 片目のみ，非抽出
顔の中心に対する 両目の縦と横の位置 (4 素性)	整数，非抽出
顔の中心に対する 口の横の位置 (1 素性)	整数，非抽出
抽出した領域の種類 (1 素性，12 項目)	顔・両目・鼻・口， 顔・両目・鼻， 顔・両目・口， 顔・両目， 顔・片目・鼻・口， 顔・片目・鼻， 顔・片目・口， 顔・鼻・口， 顔・鼻， 顔・口， 顔， 非抽出

表 2: 各フレームにおける素性 ($4angle$ を使用)

素性項目	素性値
顔の向き (1 素性，5 項目)	正面，左側面， 右側面，傾き，非抽出
顔・目・鼻・口の オプティカルフローの方向 (5 素性，5 項目)	0 度，90 度， 180 度，270 度， 方向なし
顔・目・鼻・口の オプティカルフローの大きさ (5 素性)	自然数，非抽出
両目の位置のずれ (1 素性)	傾きあり，傾きなし， 片目のみ，非抽出
顔の中心に対する 両目の縦と横の位置 (4 素性)	整数，非抽出
顔の中心に対する 口の横の位置 (1 素性)	整数，非抽出
抽出した領域の種類 (1 素性，12 項目)	顔・両目・鼻・口， 顔・両目・鼻， 顔・両目・口， 顔・両目， 顔・片目・鼻・口， 顔・片目・鼻， 顔・片目・口， 顔・鼻・口， 顔・鼻， 顔・口， 顔， 非抽出

表 3: 20 フレーム間での素性

素性項目	項目値
顔の向き (正面，左側面，右側面，傾き，非抽出) の抽出合計値 (5 素性)	自然数
顔・目・鼻・口の オプティカルフロー各方向 (0 度，90 度，180 度，270 度，方向なし) の抽出合計値 (25 素性)	自然数
横位置に対する，口と顔の中心のずれの合計値 (1 素性)	自然数

素性「顔の向き」は，顔領域を抽出した Haar like 特徴の種類 (3.2 節の処理 1 の正面，左側面，右側面，傾き) に対応している．また，素性「両目の位置のずれ」の値は，以下の条件に基づき決定する．

$$\text{[条件] 両目の位置のずれ} = \begin{cases} \text{傾きあり} \begin{pmatrix} \text{両目を抽出し，} \\ \text{中心の高さの差が} \\ \text{5 ピクセル以上} \end{pmatrix} \\ \text{傾きなし} \begin{pmatrix} \text{両目を抽出し，} \\ \text{中心の高さの差が} \\ \text{5 ピクセル以内} \end{pmatrix} \\ \text{片目のみ (片目のみ抽出)} \\ \text{非抽出 (目を抽出しなかった)} \end{cases}$$

3.4 頭部ジェスチャの決定木生成

一定のフレーム間における抽出した素性を基に，Ross Quinlan[10] らが開発した，帰納推論システム「C4.5」を用い，頭部ジェスチャの決定木を生成する．

4 頭部ジェスチャ認識実験

この節では，3 節で述べた手法を用いて，頭部ジェスチャ認識実験を行い，その結果と考察を述べる．

4.1 実験方法

3.1 節で述べた頭部ジェスチャ「うなずき」・「首振り」・「首かしげ」に加えて，3 つの頭部ジェスチャ以外のジェスチャを「ジェスチャなし」とし，各ジェスチャの動画像を事前に 600 動画像保存しておき，学習データとして，各ジェスチャ500 動画像，合計 2000 動画像を C4.5 を用いて決定木を生成し，テストデータとして，各ジェスチャ 100 動画像，合計 400 動画像を作成した決定木によりジェスチャ認識を行った．素性として，オプティカルフローが 5 方向のもの (表 2)，9 方向のもの (表 1) を比較した．

4.2 実験結果

オプティカルフローが 5 方向の素性結果を表 4 に，9 方向の素性結果を表 5 に示す．

表 4: 5 方向の素性 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	81	0	0	19
首振り	2	93	0	5
首かしげ	0	0	88	12
ジェスチャなし	4	3	1	92

表 5: 9 方向の素性 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	73	1	2	24
首振り	0	99	0	1
首かしげ	0	0	87	13
ジェスチャなし	3	3	0	94

表 6: 方向別の平均認識率 (単位:%)

5 方向	9 方向
88.2	88.5

表 6 から分かるように，オプティカルフローの方向数の違いで精度の変化は生じなかった．その理由として，今回の頭部ジェスチャが，頭部を上下，もしくは左右に移動させるものに限定していたためだと考えられる．また，誤認識の多くが「ジェスチャなし」と認識されてしまうが，これは，曖昧なジェスチャや，顔や目などの領域を抽出できない場合,「ジェスチャなし」に認識されてしまうためだと考えられる．また,「うなずき」の認識率が低いが，これは,「うなずき」のジェスチャの際，3.2 節の処理 1 において,「正面の顔」で顔領域を抽出できなかった場合に,「傾いた顔」で，顔領域を抽出してしまうと,「うなずき」のパターンのノイズとなり，認識精度に影響を及ぼしたと考えられる．図 8 に例を示す．

(a) キャプチャ画像　　(b) 誤抽出画像

図 8: 誤抽出例

そこで，追加実験として，領域抽出条件と素

性の再検討を行い，再び頭部ジェスチャ認識を行う．

4.3 追加実験

4.1 節と同様の学習データ，テストデータを用いて実験を行い，頭部ジェスチャ認識を行った．4.2 節での結果を参考にして，領域抽出・素性を変更する．変更点は次の通りである．

- 3.2 節の処理 1 に対して「傾いた顔」の Haar like 特徴 (処理 1．(4)) を使用しない
- 決定木生成の際に「傾いた顔」に関する素性を削除

また，顔や目，鼻，口の領域抽出条件が，認識精度に影響を及ぼすのかを調べるために，3.2 節の各領域抽出処理の領域抽出条件である「Haar like 特徴の合致数」を 3 箇所以上の場合と 1 箇所以上の場合でそれぞれ抽出を行い，それに基づいて，素性値を求め決定木を作成し，認識精度を比較する．ただし，4.2 節より，オプティカルフローが 9 方向の方が高精度だったため，実験の簡易化のために，オプティカルフローが 9 方向の場合のみを検証した．

4.3.1 実験結果

領域抽出の合致条件が 1 以上，3 以上の認識結果を，それぞれ表 7，表 8 に示す．

表 7: 「傾き顔」素性なし，合致条件 1 以上 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	74	2	0	24
首振り	6	87	0	7
首かしげ	0	0	93	7
ジェスチャなし	6	7	1	86

表 8: 「傾き顔」素性なし，合致条件 3 以上 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	90	1	0	9
首振り	0	96	0	4
首かしげ	0	0	99	1
ジェスチャなし	3	4	5	88

表 9: 合致条件別の平均認識率 (単位:%)

傾き顔削除		傾き顔利用
1 箇所以上	3 箇所以上	3 箇所以上
85.0	93.2	88.5

表 9 から，追加実験の方が認識精度が高いことが分かる．さらに，表 5 と表 8 を比べると，追加実験の方が,「首かしげ」の認識精度が高い．これらのことから，3.2 節の処理 1 で，本来「首かしげ」の認識に有効であると考えた「傾き顔」を用いることが，全体としてジェスチャ特徴のノイズを生み出す結果となってしまい，有効な素性として利用することができないことが分かった．

また，表 7 と表 8 や，表 9 のそれぞれを比較すると，合致条件が 3 箇所以上の場合の方が認識精度が高い．この要因として，合致条件が 3 箇所以上の場合，Haar like 特徴領域が明確に存在する場合のみ領域抽出を行う．これにより，背景を顔などの領域として誤抽出することを防ぐことができる[1]．その結果，信頼性の低い顔・目・鼻・口それぞれの領域は,「非抽出」となり，領域抽出が難しいような場合に，一環した結果を得ることができるため，結果として，決定木を作成する素性値のノイズが減少する．そのため，ロバストなジェスチャ認識を行うことができると考えられる．

今回の実験では，オフラインな動画像を利用し，実験を行った．このため，フレームを省くことなく処理を行うことができた．しかし，提案手法をリアルタイム処理に適応すると，計算機のスペックによっては，ジェスチャ認識処理のための処理時間がかかりすぎ，オプティカルフローなどを計算する場合にフレーム飛ばしが原因で正しい値が抽出されない可能性がある．リアルタイム処理を実現するためには，ジェスチャ認識処理の計算コスト削減を行う必要がある．例えば，現在は常に 20 フレーム間の素性値を利用しているが，利用フレーム数の削減や，有効な素性を用いることでの素性などの再検討を行うことで，処理軽減を行うことができる思われる．そして，リアルタイム処理を実現し，発話とジェスチャが相互作用し，機械とのスムーズなコミュニケーションへの利用を行っていく．

4.4 シンプルなルールとの比較

補足実験として，人手による極めてシンプルな認識モデルを作成し，頭部ジェスチャの認識を行う．まず，頭部ジェスチャの判別には見た目の動きを基に，表 10 の素性を利用し，人間の判断で以下の認識条件を作成した．

[1] すなわち，合致条件が厳しければ，図 8(b) のような誤抽出が生じにくい．

表 10: シンプルなジェスチャに用いる素性

項目素性	項目値
20 フレーム間の顔の向き (正面，左側面，右側面，非抽出 ※「傾き」を削除) の抽出合計値 (4 素性)	自然数
20 フレーム間の顔のオプティカルフロー各方向 (0 度，90 度，180，270 度，方向なし) の抽出合計値 (5 素性)	自然数
20 フレーム間の横位置に対する 口と顔の中心のずれの合計値 (1 素性)	自然数

うなずき条件

正面の顔の向きの合計値 ≥ 15

0 度方向の合計値, 180 度方向の合計 ≤ 3

90 度方向の合計値, 270 度の方向の合計 ≥ 3

首振り条件

0 度方向の合計値, 180 度方向の合計 ≥ 3

90 度方向の合計値, 270 度の方向の合計 ≤ 3

首かしげ条件

横位置に対する，口と顔の中心のずれの合計 ≥ 100

そして，4.1 節と同様のテストデータを用い，頭部ジェスチャ認識実験を行った．ただし，特徴の合致数を 1 箇所以上と 3 箇所以上の場合に分け，実験の簡易化のために，オプティカルフローは 9 方向のみで検証した．

4.4.1 実験結果

領域抽出の合致条件が 1 以上の認識結果を表 11 に示し，領域抽出の合致条件が 3 以上の認識結果を表 12 に示す．

表 11: 合致条件 1 以上で領域抽出 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	56	1	0	43
首振り	0	56	0	44
首かしげ	0	0	72	28
ジェスチャなし	13	18	1	68

表 12: 合致条件 3 以上で領域抽出 (単位:%)

テストデータ / 認識結果	うなずき	首振り	首かしげ	ジェスチャなし
うなずき	27	0	0	73
首振り	0	78	0	22
首かしげ	0	0	5	95
ジェスチャなし	9	20	0	71

表 13: 合致条件別の平均認識率 (単位:%)

1 箇所以上	3 箇所以上
63.0	45.3

実験結果から，合致条件が 3 箇所の方では，シンプルな素性の場合では認識率は大きく低下してしまったことが分かる．この原因として，顔領域などの抽出に厳しい制約を用いると，決定木生成の際に必要な素性が十分に得られないことが挙げられる．その結果,「ジェスチャなし」に認識する頻度が高くなってしまい，認識率が低下しまうと考えられる．また，合致条件が緩やかな場合であっても，C4.5 を用いた決定木による頭部ジェスチャ認識の精度には及ばなかった．この結果から，明示的な特徴だけでは頭部ジェスチャ認識を行うことは困難であり，C4.5 を用いて生成した決定木が，単純な素性によるジェスチャ認識よりも有効であることが確認できた．

5 おわりに

本稿では，固定したカメラから取り込んだ動画像から，色情報を使用せずに頭部ジェスチャを認識する手法を提案した．今後は，リアルタイム処理の実現を目指す．

参考文献

[1] James W. Davis and Serge Vaks, “A Perceptual User Interface for Recognizing Head Gesture Acknowledgements”, In IEEE PUI, Orlando, FL, 2001.

[2] P. Ravindra De SILVA, Minetada OSANO, Ashu MARASINGHE and Ajith P. MADURAPPERUMA, “A Computational Model for Recognizing Emotion with Intensity for Machine Vision Applications”, IEICE TRANS. INF. &SYST., VOL. E89-D, NO. 7 JULY 2006.

[3] 中島慶, 江尻康, 藤江真也, 小川哲司, 松坂要佐, 小林哲則, “対話ロボットの動作に頑健な頭部ジェスチャ認識”, 電子情報通信学会論文誌 D Vol. J89-D No. 7 pp. 1514-1522.

[4] Shinjiro KAWATO, Nobuji TETSUTANI, and Kenichi HOSAKA, “Scale-Adaptive Face Detection and Tracking in Real Time with SSR Filters and Support Vector Machine”, IEICE Trans. Inf. & Syst. VOL. E88-D NO. 12 pp. 2857-2863, DECEMBER 2005.

[5] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, “Face detection in color images”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-706, May 2002.

[6] K. Schwerdt and J. L. Crowley, “Robust face tracking using color”, in Proc. of 4th International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000, pp. 90-95.

[7] Zhiwei Zhu, Kikuo Fujimura, Qiang Ji, “Real-Time Eye Detection and Tracking Under Various Light Conditions”

[8] Rainer Lienhart and Jochen Maydt, “An Extended Set of Haar-like Features for Rapid Object Detection”, IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep.

[9] Paul Viola and Michael J. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE CVPR, 2001.

[10] J. Ross Quinlan, C4.5：Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.

[11] Intel, OpenCV, http://www.intel.com/technology/computing/opencv/