

## 距離動画像を用いた頭部ジェスチャの認識

梅田 和昇<sup>\*1</sup> 鈴木 範史<sup>\*2</sup>

## Recognition of Head Gestures Using a Sequence of Range Images

Kazunori Umeda<sup>\*1</sup> and Norihito Suzuki<sup>\*2</sup>

Gesture recognition using images is recently studied as a technology of human friendly man-machine interface. In this paper, recognition of 3D gesture by using range images is discussed. Two methods for recognizing head gestures from a sequence of range images are proposed. One uses differential range images for identifying 'Yes' motion, and the other detects and tracks nose position for identifying 'Yes' and 'No' motion. Experiments to recognize real person's gestures using a range sensor system are performed to show the effectiveness of the presented methods.

**Key Words:** Gesture Recognition, Range Image, Motion Recognition, Man-Machine Interface

## 1. 序 論

今日、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、人間に対する自然さが重要な要素の一つである。人間は言語、表情、ジェスチャなどのチャンネルを用いてコミュニケーションを行っており、これらを用いたインタフェース技術を構築すれば、人間に対して自然なものになると期待される[1]。自然なインタフェースを目指した技術の一つとして、ジェスチャの認識により人間から機械へ情報伝達する手法が近年多く研究されている[2][3]。これらの研究ではジェスチャを観測するためのセンサ情報として、濃淡画像が用いられている。しかしながら、濃淡画像は三次元の位置情報が二次元に射影された情報しか持たない。そこで、三次元のジェスチャを認識するために、三次元情報を直接持つ距離動画像の利用が有効であると考えられる。かつては距離動画像は入力に時間を要し、実時間性を要する目的には適さなかった。しかしながら最近、距離動画像を実時間で入力可能なセンサが種々開発されてきており[4]~[8]、距離動画像が実時間性を要する目的に利用できるようになってきた。ジェスチャ認識のためのセンサ情報として、距離動画像は以下の2点の長所を持つと考えられる。

- 三次元情報を持つため、三次元的なジェスチャの認識に適している

- 背景と対象領域との間に距離の不連続性が存在するため、背景からの対象領域(人物)の抽出が容易かつロバストである

以上より、本論文では、距離動画像を用いたジェスチャ認識手法を提案することを目的とする。具体的な認識対象として、頭部のジェスチャを取り上げ、特にうなずき動作、首振り動作を対象として、それらの識別手法を構築する。まず距離動画像を用いたジェスチャ認識手法の概要を示した後、頭部のジェスチャの具体的な認識手法を2種提案し、実験でこれらの手法の有効性、ひいては距離動画像を用いたジェスチャ認識の有効性を検証する。

## 2. 距離動画像を用いたジェスチャ認識手法の概要

何らかの認識手法の構築を考える場合、汎用的な手法を構築する場合と、特定の認識対象に固有の手法を構築する場合とが考えられる。これまでの濃淡画像を用いたジェスチャ認識の研究では、汎用的な手法の構築を目指したものが多く[2][3](例えば[2]ではHMMを、[3]では連続DPを利用)。本論文では、うなずき動作、首振り動作の認識のみに特化した手法を構築するという立場をとる。

画像の時系列からのジェスチャ認識は、大きく、特徴抽出、抽出された特徴の時系列を用いた認識、の2段階の過程で行われると考えられる。抽出される特徴に着目すると、ジェスチャ認識手法は次の2通りに分類されると考えられる。

- (1) 各画像から静的な特徴を抽出する手法
- (2) 連続する画像から動的な特徴を抽出する手法

両方法論のいずれが適しているかは、認識対象、計算コストなどに依存すると考えられる。以下、頭部のジェスチャ認識に

原稿受付 1998年5月8日

<sup>\*1</sup>中央大学理工学部

<sup>\*2</sup>中央大学大学院理工学研究科(現NTT)

<sup>\*1</sup>Faculty of Science and Engineering, Chuo University

<sup>\*2</sup>Graduate School of Science and Engineering, Chuo University

において、両者を検討する。

(1)に属する手法: これまでの濃淡画像を用いた手法はこちらに属しているものが多い[2][3]。距離画像を用いる場合、各画像で顔の認識を行って位置・姿勢を計測し、位置・姿勢を特徴として用いるという方法がまず考えられる。顔は曲面であり、これまでに提案されている距離画像を用いた曲面形状の認識手法が適用しうると考えられる[9][10]。これらの研究では、形状を、局所的な曲率やエッジ情報を用いて領域分割し、その結果に基づいてモデルマッチングを行って位置・姿勢の計測を行っている。このように各画像で顔の認識を行う場合、安定に顔の位置・姿勢の追跡が実現できると予想される反面、計算時間を要するため、実時間でのジェスチャ認識を行うのは困難である。そこで本論文では、距離画像中の最も近い点を鼻とみなしてその追跡を行うという、単純な手法を提案し、検証することとする。

(2)に属する手法: 距離画像から得られる動的な特徴として、顔はほぼ剛体であると仮定できるので、まず剛体の三次元運動パラメータ[11][12]が考えられる。この三次元運動パラメータの時系列を利用すれば、汎用的な運動パターン認識が可能であると予想される。しかしながら、顔の形状が比較的平面に近いので、三次元運動パラメータの6成分すべてをロバストに求めるのは困難である[12]。そこで、本論文では、うなずき、首振り動作さえ認識できればよいという立場から、簡単に計測可能な差分距離画像を利用して、距離画像の誤差にロバストな認識手法を構築することとする。

以上をまとめ、本論文では、頭部のジェスチャの認識手法として、上記のそれぞれのカテゴリに属する手法として、差分距離画像を動的な特徴として利用する(2)に属する手法を3で提案し、鼻の各画像での位置を特徴として利用する(1)に属する手法を4で提案することとする。

### 3. 差分距離画像を利用したうなずき動作の識別

#### 3.1 手法構築の指針

頭部の動きによるジェスチャとして、

- うなずき動作 ('Yes' motion)
- 首振り動作 ('No' motion)
- 首をかしげる動作 ('Question' motion)

などが考えられる。本節ではうなずき動作を他の動作から識別する手法を提案する。頭部のジェスチャでは、位置の変動が頭部の大きさと比較して相対的に小さく、2枚の連続する画像で重なりが大きいと考えられる。そこで、差分距離画像が有効に利用できると考えられる。差分距離画像は、2枚の連続する距離画像の距離値の差を要素とする画像として定義される。以下では、2枚の画像が重なっている領域のみを処理対象とする。画像が重ならない領域では、距離値の差が重なる領域より大きくなると考えられる。そこで、画像が重なっている領域のみでの差分距離画像は、距離値の差が適当なしきい値より大きな箇所を除くことで求められる。このようにして求められる差分距離画像の各要素の差分値を  $d_i$ 、座標値を  $(u_i, v_i)$  ( $i = 1 \dots N$ ) とおく。  $N$  は差分距離画像の要素数である。

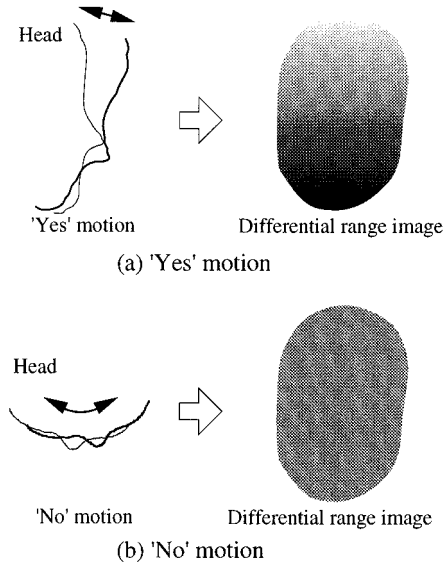


Fig. 1 Differential range image in 'Yes'/'No' motions (Intensity represents differential value)

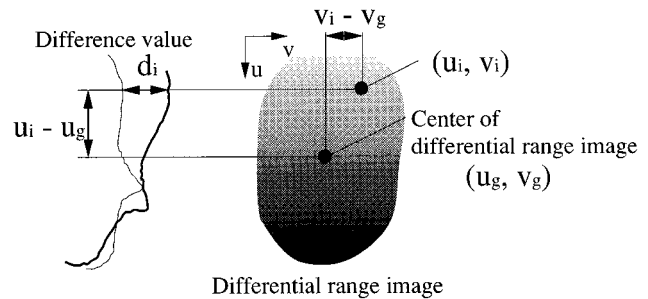


Fig. 2 Differential range image in 'Yes' motion

差分距離画像中の差分値は、うなずき動作において、縦方向には連続的に大きく変化し、横方向には変化しないと考えられる (Fig. 1 (a) 参照)。この性質を利用して識別手法を提案する。なお、首振り動作においては、回転中心が頭部の回転中心を通過しており、差分距離画像の横方向に差分値はそれほど変化しない (Fig. 1 (b) 参照)。

#### 3.2 識別手法の提案

上記の指針より、以下の式(1)(2)をうなずき動作の識別のための基本式としてまず提案する。

$$I_v = \frac{1}{N} \sum_{i=1}^N (u_i - u_g) d_i \quad (1)$$

$$I_h = \frac{1}{N} \sum_{i=1}^N (v_i - v_g) d_i \quad (2)$$

ただし、 $(u_g, v_g)$  は差分距離画像の中心、すなわち、

$$u_g = \frac{1}{N} \sum_{i=1}^N u_i, \quad v_g = \frac{1}{N} \sum_{i=1}^N v_i$$

とする。これらのパラメータを Fig. 2 に示す。

式(10)は、それぞれ差分距離画像の中心からの縦方向、横方向の変位で重みづけされた差分値の、差分距離画像全体の平均値である。うなずき動作では、式(1)が大、式(2)が小になると考えられる。なお、式(10)は頭部の並進運動によっては値を生じない。

式(10)の値は、動作の反復に応じて振動する。振動成分を抽出するにはいくつか手法が考えられるが、ここでは単純に、連続する適当枚数の差分距離画像で式(10)の絶対値を積算し平均する。そしてこれらの値を、識別のための評価値として用いることとする。すなわち、評価値として、

$$J_v(t) = \sum_{i=t-n+1}^t \frac{|I_v(i)|}{n} \quad (3)$$

$$J_h(t) = \sum_{i=t-n+1}^t \frac{|I_h(i)|}{n} \quad (4)$$

を用いる。ここで、 $n$  は積算する差分距離画像の枚数である。あるいは、式(4)の  $J_h(t)$  を、式(3)の  $J_v(t)$  で規格化して(すなわち  $J_h(t)/J_v(t)$ )を用いることも可能である。これらの評価値を用いれば、以下の基準でうなずき動作を識別できる。

$$\text{if}(J_v \text{ 大 and } J_h \text{ 小}) \text{ then うなずき動作} \quad (5)$$

この識別は、 $n$  枚の差分距離画像(すなわち  $(n+1)$  枚の距離画像)が計測されて以後は各時刻  $t$  において行われる。また、実際に認識システムを構築する場合には複数の時刻での識別結果を組み合わせることも考えられる。

以上、差分距離画像を用いたうなずき動作の識別手法を提案した。

### 3.3 差分距離画像の積算枚数などに関する考察

以上で提案した手法における、差分距離画像の積算枚数  $n$  などに関して考察する。式(10)の  $I_v, I_h$  は、Fig.3に模式的に示すように、うなずき動作の反復に応じて正弦波的に振動すると考えられる。また、式(30)の  $J_v, J_h$  は、それぞれ  $I_v, I_h$  のある区間の面積を積分して求めた量(を積算枚数  $n$  で割った量)に相当する。このとき、積分する区間が短いと、振動する  $I_v, I_h$  のどの位相で計測するかにより、 $J_v, J_h$  が変動してしまうことになる。積分区間の取り方による値の変動を最小値と最大値との比  $r$  で求めると、 $x$  を積分区間の幅(正弦波の周期で規格化)、 $[x]$  を  $x$  を越えない最大の整数として、

$$r = \begin{cases} \frac{2[x]+1-\cos \pi(x-[x])}{2[x]+\sin \pi(x-[x])} & (x-[x] < \frac{1}{2}) \\ \frac{2[x]+2-\sin \pi(x-[x])}{2[x]+1-\cos \pi(x-[x])} & (\frac{1}{2} \leq x-[x]) \end{cases}$$

となる。これを図示すると Fig.4 のようになる。

Fig.4より、計測のタイミングによる  $J_v, J_h$  の値の変動が大きくならないようにするには、最低でも積算時間をうなずき動作の周期の  $1/2$  以上、可能ならば1周期以上にすることが望ましいことが分かる(また、実際の応用を考えた場合、うなずき動作の周期は一定ではなく変動する。変動分を考慮して、積算時間には余裕を持たすことが必要である)。また、積算時間が長くなることは、うなずき以外の動作を分離する点からも

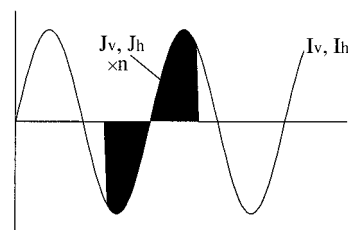


Fig.3 Relation of  $I_v, I_h$  and  $J_v, J_h$

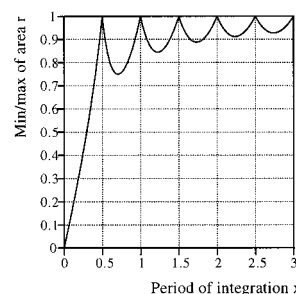


Fig.4 Min/max of area according to period of integration

望ましいと考えられる。ただし、時間が長くなりすぎると、識別を行うのに要する時間がそれに依りて長くなってしまふ。

積算時間は、差分距離画像の積算枚数  $n$  と距離動画像計測のサンプリングタイムとの積で与えられる。サンプリングタイムは、一般的に計測システムの能力で上限が設定される。サンプリング定理より、最低でもうなずき動作1周期あたり2回以上の計測が必要であると考えられる。サンプリングタイムはなるべく短い方がよいと考えられるが、短すぎると差分距離画像の差分値が小さくなってしまふ。

以上、差分距離画像の積算枚数  $n$  (とサンプリングタイムとの積で与えられる積算時間)とサンプリングタイムに関しての定性的な考察を行った。

## 4. 鼻の抽出・追跡によるうなずき・首振り動作の識別

前章で差分距離画像という動的な特徴を利用した識別手法を提案したのに対し、本節では静的な特徴を利用した手法を提案する。静的な特徴として鼻の位置を用いる。これは、

- 鼻が顔の中心にあり、顔を代表する特徴として適切であると考えられる
- 距離画像からの抽出が容易であると考えられる

という理由による。

### (1) 鼻の位置の抽出・追跡

鼻は、顔のほぼ中心にあり、周辺で距離が最も近い領域であると仮定できる。そこで、以下の手順で時系列距離画像からの鼻の位置の抽出・追跡を行う。

1 枚目の画像：顔の距離画像の中心付近で最も距離に近い点を探索し、鼻とみなす。

2 枚目以降の画像：前の画像の鼻の抽出位置の近傍で距離が最も近い点を探索し、鼻とみなす。

同様の追跡は、近年よく用いられているトラッキングビジョンなどを利用することによって、濃淡画像でも可能であると考え

えられる[13]．本手法と、トラッキングビジョンなどにおいて用いられている相関演算による追跡とを比較すると、濃淡画像の利用よりも本手法のように距離値を利用する方が、照明などの環境の変化に対してはロバストであると考えられる．ただし、本手法では、距離画像中の1点のみの抽出を行うため、距離画像の距離値の、通常の正規分布に従わない大きなノイズには弱い．これは利用する距離画像センサの特性にも依存し、使用するセンサの誤差が正規分布によく従うならば、問題にはならない．この問題に本質的に対処するには、2で述べたような位置・姿勢の計測を行うか、あるいは距離画像においても近傍領域での相関演算を行えばよいと考えられる．

## (2) 鼻の位置の時系列からのうなずき・首振り動作の識別

距離画像の時系列から得られる鼻の位置の時系列より、うなずき動作・首振り動作を識別する．うなずき動作では、鼻の位置は画像中で上下に、首振り動作では左右に振動する（Fig. 5 参照）．鼻の位置の振動成分を求めるにもいくつか手法が考えられるが、ここでは3と同様の単純な手法を用いる．すなわち、

$$a_v(t) = \sum_{i=t-n+1}^t \frac{|u_{nose}(i) - u_{nose\_ave}(t)|}{n} \quad (6)$$

$$a_h(t) = \sum_{i=t-n+1}^t \frac{|v_{nose}(i) - v_{nose\_ave}(t)|}{n} \quad (7)$$

により縦方向の振動成分  $a_v$ 、横方向の振動成分  $a_h$  を計測する．ただし、 $n$  は積算する距離画像の枚数、 $u_{nose}(i)$ 、 $v_{nose}(i)$  は画像  $i$  での鼻の縦・横の位置、 $u_{nose\_ave}(t)$ 、 $v_{nose\_ave}(t)$  は画像  $t$  から  $n$  枚前までの画像での鼻の縦・横の位置の平均、すなわち

$$u_{nose\_ave}(t) = \frac{1}{n} \sum_{i=t-n+1}^t u_{nose}(i),$$

$$v_{nose\_ave}(t) = \frac{1}{n} \sum_{i=t-n+1}^t v_{nose}(i)$$

である．

こうして求められた  $a_v$ 、 $a_h$  を用いて、以下のように識別すればよい．

$$\text{if}(a_v \text{ 大 and } a_h \text{ 小}) \text{ then うなずき動作} \quad (8)$$

$$\text{if}(a_v \text{ 小 and } a_h \text{ 大}) \text{ then 首振り動作} \quad (9)$$

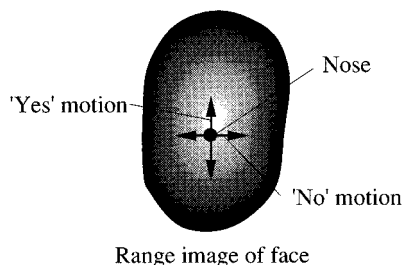


Fig. 5 Vibration of nose in 'Yes'/'No' motion (Intensity represents distance value)

この識別は、3.2 節と同様に  $n$  枚の距離画像（3.2 節では差分距離画像）が計測されて以後は各時刻  $t$  において行われる．また、距離画像の積算枚数  $n$  およびサンプリングタイムに関しては 3.3 節での議論がまったく同様に当てはまる．

以上、鼻の位置を用いたうなずき動作・首振り動作の識別手法を提案した．

## 5. 認識実験

提案した二つのジェスチャ認識手法の有効性を検証するための実験を行う．5.1 節で実験システムの構成を示し、5.2、5.3 節でそれぞれ差分距離画像を用いた手法、鼻の追跡を用いた手法の実験結果を示す．5.4 節では、3.3 節で考察した差分距離画像の積算枚数およびサンプリングタイムが与える影響を評価する実験を行う．また、5.5 節では 2 手法の比較実験を行う．

### 5.1 実験システムの構成

実験システムの構成を Fig. 6 に示す．距離画像センサ、パーソナルコンピュータ（DX4 100 [MHz]）が主な構成要素である．距離画像センサには文献 [5] で示されているものを改良したセンサを用いた．本センサは、Fig. 6 に示すように、多点のスポット光を投影し、各スポット像までの 8 ビットの距離値を三角測量によりビデオレートで計測する．計測点数は 1,056 点で、測定レンジは約 760 [mm]～850 [mm] である．8 ビットの距離値は、三角測量の視差に相当する量を表している [5]．実際の距離値への対応は、0 のときが約 760 [mm]、255 のときが約 850 [mm] となり、この間で 8 ビットの距離値と実際の距離値とはほぼ線形に対応づけられる．本実験では処理の高速化のため実際の距離値への変換は行わず 8 ビットの距離値をそのまま用いた．

Fig. 7 に示すように被験者にセンサの前に座ってもらい、頭部のジェスチャを観察した．本センサではハロゲンランプによる可視光のスポット光を投影して距離を計測するため、外乱光に弱い．通常の室内程度の明るさでの計測は可能であるが、以下の実験は半暗室で行い（5.4 節は蛍光灯照明下）、外乱光の影響を除去している．

なお、多点のスポット光の投影によるセンサを用いた本実験システムは、あくまで提案手法の有効性を検証するための一例であり、他の能動手法やステレオ、Shape from focus などを用いて得られた距離画像に対しても本論文の手法はまったく同様に適用可能である．

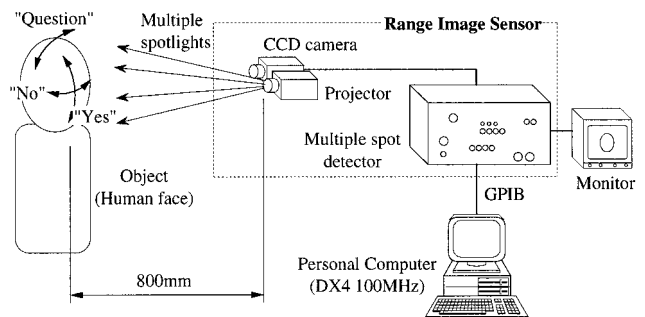


Fig. 6 Experimental system



Fig. 7 Experimental environment—projector, CCD camera and an object person

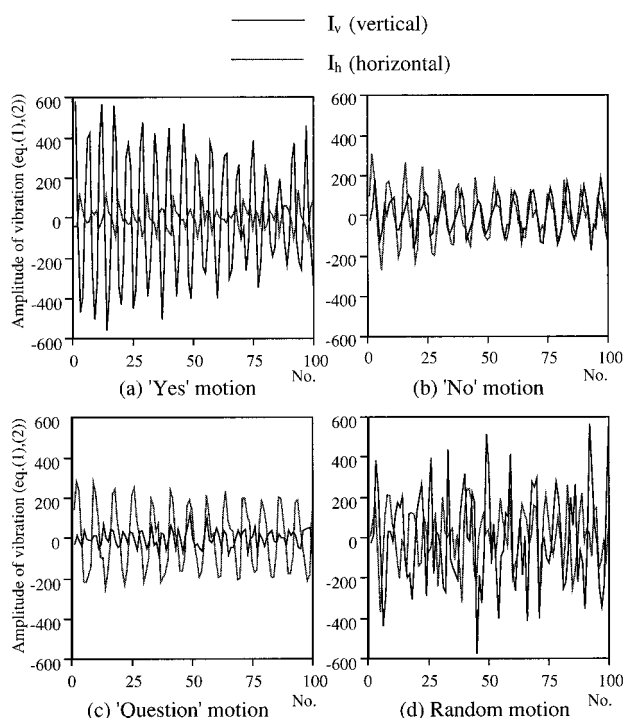


Fig. 8 Weighted averages of differential values for 'Yes', 'No', 'Question' and random motions

## 5.2 差分距離画像を利用したうなずき動作の識別実験

うなずき動作, 首振り動作, 首をかしげる動作, 頭部をランダムに動かす動作の4種の動作に対して, 3の手法を適用した. 差分距離画像の積算枚数(式(3)の $n$ )はここでは適当に10とした. 動作中の顔に対し, 100枚の距離画像を時系列で入力し, オンラインで処理を行った. うなずき動作, 首振り動作, 首をかしげる動作で周期が1.2~1.6秒, サンプルタイムは0.3秒以下であった.

Fig. 8に差分距離画像における差分値の重みつき平均(式(1)の $2$ )を示す. 本図より, うなずき動作では縦方向の平均の振幅が他の動作の場合より大きく, また横方向の平均の振幅はそれほど他の動作と差はないが, 相対的に小さいことが示されている.

縦・横それぞれの評価関数(式(3)の $4$ )をFig. 9に示す.

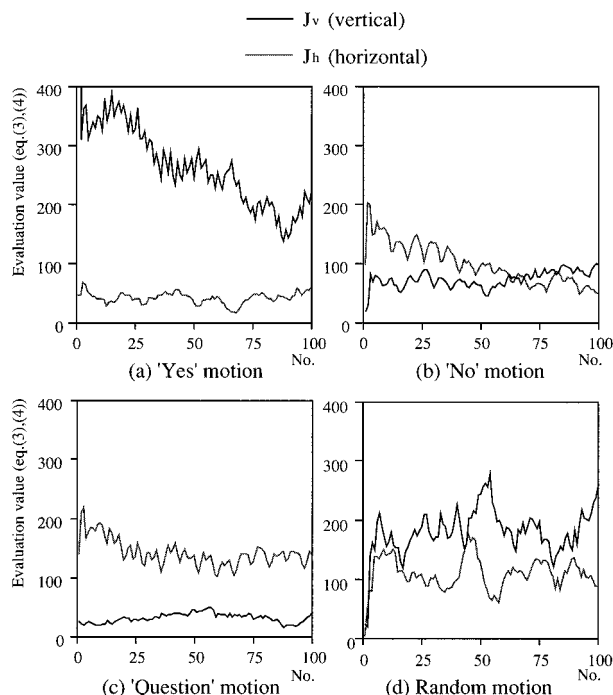


Fig. 9 Evaluation functions for 'Yes', 'No', 'Question' and random motions

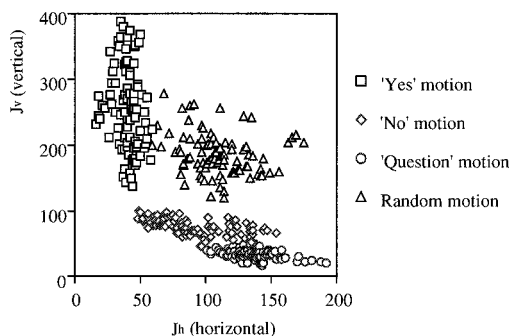


Fig. 10 Distribution of vertical-horizontal evaluation functions of differential range images

本図より, うなずき動作では縦方向の評価値が高い値で, また横方向の評価値が低い値で推移している. さらにFig. 10に縦方向, 横方向の評価値を縦軸・横軸にとった図を示す. ランダムな動作での値が一部接近しているものの, ほぼうなずき動作を分離可能である.

以上の結果より, 3章で示した判断基準でうなずき動作が他の動作より識別可能であることが示されている. 一方, うなずき動作以外の識別は, これらの値からでは困難である.

## 5.3 鼻の抽出・追跡によるうなずき・首振り動作の識別実験

うなずき動作, 首振り動作の2種の動作に対して, 4の手法を適用した. 距離画像の積算枚数(式(6)の $n$ )は前項の実験と同じく適当に10とした. 動作中の顔に対し, 30枚の距離画像を時系列で入力し, 鼻の抽出・追跡をオンラインで, 評価値の算出のみをオフラインで行った.

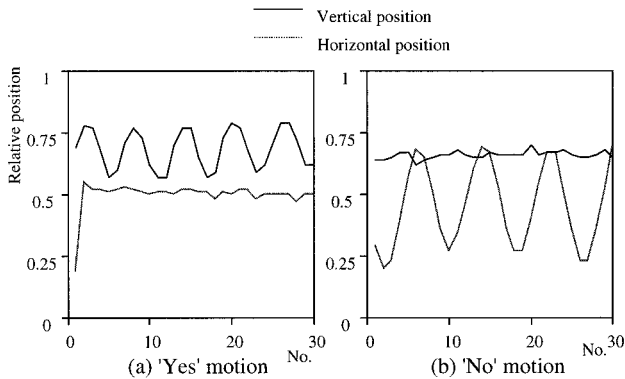


Fig. 11 Tracking of nose position for 'Yes' and 'No' motions

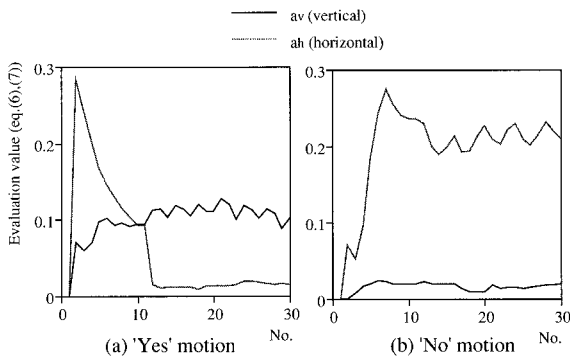


Fig. 12 Evaluation functions for 'Yes' and 'No' motions

Fig. 11 より, うなずき動作においては縦方向の振幅が大きく, また横方向の振幅が小さく計測され, 一方首振り動作においては逆に横方向の振幅が大きく, また縦方向の振幅が小さく計測されていることが分かる. さらに, Fig. 12 では, 評価値が, うなずき動作における初期の値以外, 式(8)(9)に示す基準に従っている. Fig. 13 に評価値の分布を示す. 本図での両動作の空間的な分布は特徴的であり, これらの他の動作からの分離は容易であると考えられる.

以上の結果より, 4 章で示した判断基準でうなずき動作, 首振り動作が識別可能であることが示されている. また, うなずき動作と首振り動作とでは, うなずき動作の方が振幅が大きくなることも Fig. 11, 12 に示されている.

#### 5.4 積算枚数およびサンプリングタイムの影響の評価実験

差分距離画像を用いたうなずき動作の識別において, 積算枚数およびサンプリングタイムが識別能力に与える影響を評価する実験を行った. 本実験では距離画像の入力のみをオンラインで行い, その後の識別処理に関してはすべてオフラインで行った. 距離画像入力のサンプリングタイムはビデオレートの 5 倍で 5/30 秒であった. なお, 実験で用いた距離画像センサ自体はビデオレートでの計測を行う能力があるが, GPIB インタフェースなどでのオーバーヘッドでこれだけの時間を要している. また, オフラインで行った識別処理で 1 サンプリングの処理に要する時間は 30 [msec] 以下と高速であった.

識別を行うためのしきい値は実験により適当に定めた. 式(5)において  $J_v > 50$  かつ  $J_h/J_v < 0.2$  を満たしたときにう

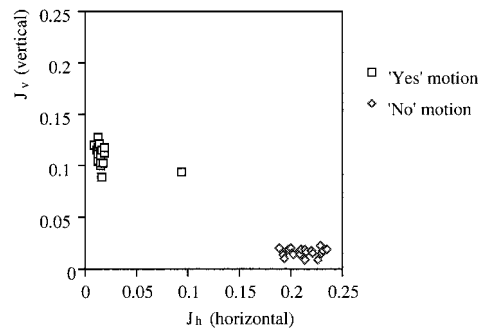


Fig. 13 Distribution of vertical-horizontal evaluation functions of nose tracking

なずき動作と識別した. ここでは, うなずき動作を正しくうなずき動作であると識別する割合, およびうなずき動作以外をうなずき動作と誤って識別する割合を計測し, 識別能力を評価した. Fig. 10 に示されているように, うなずき動作以外で最もうなずき動作と誤って識別しやすいのはランダム動作である. そこで, うなずき動作以外としてランダム動作を対象とした.

#### (1) 差分距離画像の積算枚数の影響

差分距離画像の積算枚数(式(3)(4)の  $n$ )を 1~20 枚に変化させたときのうなずき動作の識別率を検証した. うなずき動作, ランダム動作を行っている人の顔の距離画像を 5/30 秒間隔で 200 枚あまり, 計 5 回ずつ計測した. その結果, うなずき動作, ランダム動作それぞれに対して合計 985 回の識別結果を得た. また, うなずき動作の周期は 1.6~1.8 秒で, 1 周期あたりの差分距離画像の計測数は約 10 枚であった.

うなずき動作を正しくうなずき動作と識別した割合およびランダム動作をうなずき動作と誤って識別した割合を Fig. 14 に示す. Fig. 14 より, 差分距離画像の積算枚数が 10 あたりから(すなわち 1 周期分の積算時間あたりから)識別能力が高くなり, 積算枚数が増加するに従って識別能力が向上し続けることが示されている. しかしながら, むやみに枚数を増加させれば, それだけ入力時間が長くなってしまふ. 例えば  $n = 20$  では積算時間は 3.3 秒以上であり, 動作を行う人物は最低それだけの時間, うなずき動作を続ける必要がある. 現実の問題では, 識別能力と入力時間とのトレードオフを考える必要がある.

#### (2) サンプリングタイムの影響

サンプリングタイムを 1~10 倍に変化させたときのうなずき動作の識別率を検証した. 差分距離画像の積算枚数  $n$  は 10 とした. 上記と同じ距離動画像から  $k$  枚ごとに画像を選択することで  $k$  倍のサンプリングタイムを擬似的に実現した. そのため識別結果を得た数はサンプリングタイムにより異なっており, 1 倍から 10 倍まで順に, 1035, 495, 315, 225, 170, 135, 105, 90, 75, 60 回であった. うなずき動作を正しくうなずき動作と識別した割合およびランダム動作をうなずき動作と誤って識別した割合を Fig. 15 に示す. Fig. 15 を見る限りでは, 8 倍のサンプリングタイムまで高い識別能力を示している. しかしながら, サンプリングタイムが 5 倍以上では, 1 倍での一周あたりの距離画像の計測数が約 10 枚であるので一周に 2 枚以下の計測となってサンプリング定理を満たさなくなる. Fig. 16 に 1, 3, 5, 6

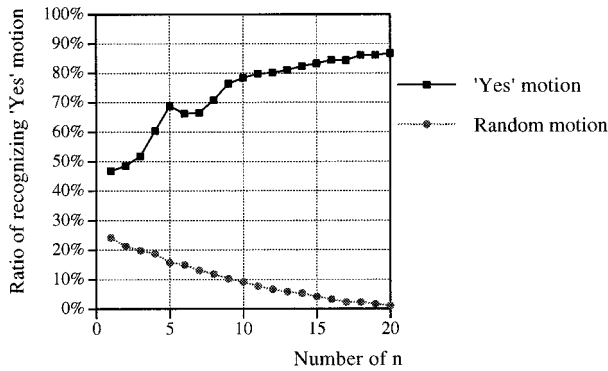


Fig. 14 Change of ratio of recognizing 'Yes' motion according to number of differential images

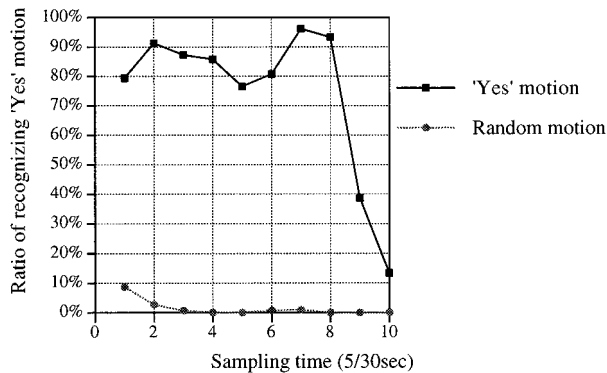


Fig. 15 Change of ratio of recognizing 'Yes' motion according to sampling time

倍のサンプリングタイムでの式(1)の縦方向の重みつき平均の値を示す。四つの図で横方向の枚数はサンプリングタイムに応じて調整してある。Fig. 16より1.3倍で22個計測されているピークが、5.6倍ではすべては計測されていないことが分かる。この結果から、サンプリング定理を満たす範囲内でのサンプリングタイムを用いることが望ましいと考えられる。実際、Fig. 15では9.10倍のサンプリングタイムで識別率が大きく落ちている。これはサンプリングタイムがほぼうなずき動作の周期と一致し、差分が小さくなってしまったためと考えられる。また、1倍より2倍以上の方がランダム動作を誤ってうなずき動作と識別する割合が小さくなっているが、これは積算時間が長くなり、ランダム動作中で部分的にうなずきの動作が入ってしまうのを誤ってうなずき動作と識別することが少なくなったためと思われる。

### 5.5 両手法の識別率比較

3, 4章で提案した両手法によるうなずき・首振り動作の識別率の評価実験を行った。識別のためのしきい値は実験により適当に定めた。式(5)において  $J_v > 80$  かつ  $J_h/J_v < 0.4$  を満たしたときにうなずき動作と識別した。また、式(8)において  $a_v > 0.4$  かつ  $a_h < 0.4$  を満たしたときにうなずき動作、式(9)において  $a_v < 0.4$  かつ  $a_h > 0.5$  を満たしたときに首振り動作と識別した。五人の被験者に対し、うなずき動作、首振り動作を行ってもらい、100枚の距離画像を入力して識別し

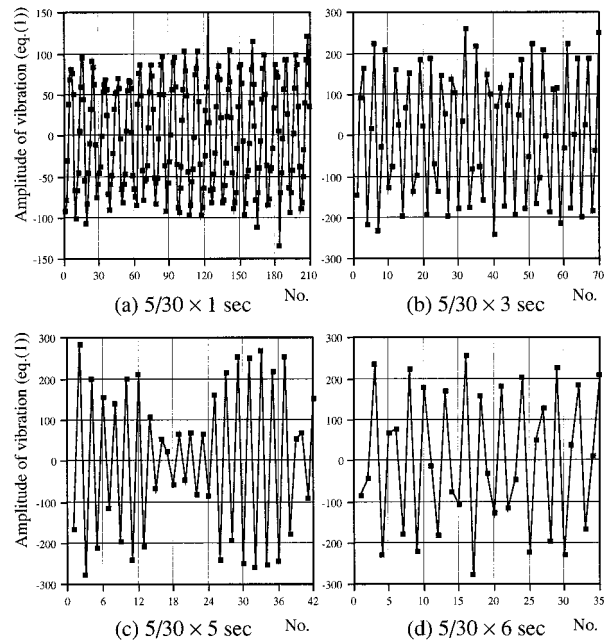


Fig. 16 Weighted averages of differential values in eq.(1) for several sampling times

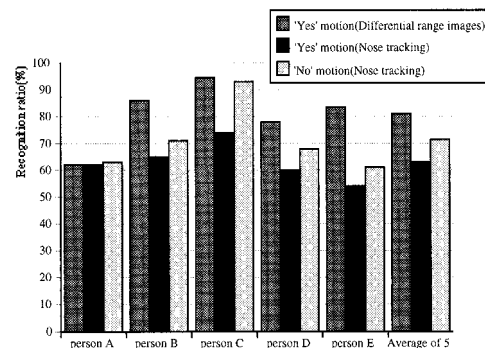


Fig. 17 Recognition rate of 'Yes' motion and 'No' motion

た結果を Fig. 17 に示す。ここで示す識別率は、うなずき、首振りそれぞれの動作を正しく識別した割合である。

本図より、以下の結果が示されている。

- それぞれの手法で比較的高い識別率を実現している。なお、これらの実験での識別率は各サンプリング時においてのものであるため、実際に認識システムを構築する場合には複数の時刻での識別結果を組み合わせることで実質的な識別率を向上させることが可能である。
- うなずき動作の識別では、差分距離画像を用いた方が識別率が高い。これは、4で示したように、本論文での鼻の抽出が距離画像中の1点のみの抽出により行われているのに対し、差分距離画像による識別では、多くの点での値を積分するため、ノイズの影響が平均化されキャンセルされるためと考えられる。例えば Fig. 11(a) のうなずき動作における横方向の振動の最初のデータが距離値の誤差の影響を受けて不正確になっているのもこのためである。

- 鼻追跡手法では、鼻の振幅が大きい首振り動作の方が、うなずき動作より識別率が良い。

以上の結果（特に後の2項）を考慮すると、本センサシステムを用いた頭部ジェスチャ認識システムとしては、うなずき動作の識別を行うには差分距離画像を利用し、鼻の抽出・追跡による識別は首振り動作の識別にのみ利用するのが妥当であると考えられる。実際、このように2手法を組み合わせて同様に被験者5名に対して100枚の距離画像を入力して実験を行ったところ、平均でうなずき動作77.6%、首振り動作82.6%であった。2手法の併用による処理時間のオーバーヘッドがあるにもかかわらず、鼻追跡手法のみでうなずき動作・首振り動作の認識を行う場合（Fig. 17 参照）に比べて高い識別率を得ることができた。

## 6. 結 論

機械へより自然に情報を伝達するための方法として、距離画像を用いたジェスチャ認識手法を取り上げ、頭部のジェスチャの認識手法として、以下の二つの手法を提案し、距離画像センサを用いた基礎的な実験により、これらの手法の有効性を示した。

- 差分距離画像を利用したうなずき動作の識別
- 鼻の抽出・追跡によるうなずき・首振り動作の識別

これら二つは、識別のための特徴として、複数画像から抽出される動的なものをういたか、1枚の画像から得られる静的なものをういたか、で位置づけが異なるものである。

本論文は、ジェスチャ認識における距離画像の有効性の例示、ならびに具体的な認識手法の提案という二つの側面を持つものであり、それらが実験結果により示されたといえよう。

今後は理論・実験の両面から、識別におけるしきい値の決定、手法の適用範囲・ロバスト性の検証などを行っていくことが課題である。さらに、モデルベース認識により得られる位置・姿勢、三次元運動パラメータなど、他の特徴の時系列を用いた、汎用性の高い認識手法の実現、ならびに本論文での手法との比較を行うことも今後の重要な課題である。

また、本論文で対象としたうなずき・首振り動作は、4で簡

単に触れたように、濃淡画像を用いて認識を行うことも可能であると考えられる。距離画像を用いる場合と濃淡画像を用いる場合との認識特性の比較を行うこと、さらには距離画像と濃淡画像とを組み合わせよりロバストな認識システムを構築することも今後の重要な課題である。

謝 辞 実験装置に関してご助力下さったバイオニア（株）の武藤守男氏、実験などでサポートをしてくれた中央大学大学院の田中信也氏に深謝致します。

## 参 考 文 献

- [1] 黒川隆夫：ノンバーバルインタフェース。オーム社，1994。
- [2] J. Yamato, J. Ohya and K. Ishii: "Recognizing human action in time-sequential images using hidden markov model," Proc. CVPR'92, pp.379-385, 1992.
- [3] 高橋勝彦，関進，小島浩，岡隆一："動画像による人物動作のリアルタイム・スポッティング認識"，MIRU'94, pp.215-222, 1994.
- [4] M. Hebert and E. Krotkov: "3-D measurements from imaging laser radars: how good are they?," IEEE/RSJ Int. Workshop on Intelligent Robots and Systems '91, pp.359-364, 1991.
- [5] K. Sorimachi: "Active range pattern sensor," J. Robotics & Mechatronics, vol.1, pp.269-273, 1989.
- [6] K. Sato, A. Yokoyama and S. Inokuchi: "Silicon range finder - a realtime range finding VLSI," Proc. Custom Integrated Circuits Conf., pp.339-342, 1994.
- [7] T. Kanade, A. Gruss and L.R. Carley: "A VLSI sensor based rangefinding system," Robotics Research Fifth International Symposium, pp.49-56, 1990.
- [8] T. Kanade, H. Kano, S. Kimura, A. Yoshida and K. Oda: "Development of a video-rate stereo machine," Proc. IEEE Int. Conf. on IROS'95, vol.3, pp.95-100, 1995.
- [9] 脇迫仁，鎌田清一郎，河口英二："距離画像処理による3次元物体検出システム"，信学論，vol.J77-D-II, no.7, pp.1282-1292, 1994.
- [10] 藤田武洋，佐藤宏介，井口征士："局所曲面形状解析に基づくピンピッキングのためのビジョンシステム"，信学論，vol.J73-D-II, no.1, pp.46-53, 1990.
- [11] 山本正信他："距離動画像を用いた非剛体運動パラメータの直接的推定法"，情処学論，vol.32, no.9, pp.1129-1141, 1991.
- [12] T. Arai and K. Umeda: "Measurement of 3D Motion Parameters from Range Images," IEEE/RSJ Int. Workshop on Intelligent Robots and Systems '91, pp.383-388, 1991.
- [13] H. Inoue et al: "Robot vision system with a correlation chip for real-time tracking, optical flow and depth map generation," Proc. 1992. IEEE Int. Conf RA, pp.1621-1626, 1992.



梅田和昇 (Kazunori Umeda)

1967年2月6日生。1989年東京大学工学部精密機械工学科卒業。1994年同大学大学院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専任講師，1998年より同助教授。ロボットビジョン，センサフュージョン，マンマシンインタフェースの研究に従事。電子情報通信学会，情報処理学会，精密工学会，日本時計学会，IEEEの会員。(日本ロボット学会正会員)



鈴木範史 (Norihito Suzuki)

1972年1月17日生。1995年中央大学理工学部精密機械工学科卒業。1997年同大学大学院精密工学専攻修士課程修了。現在，NTTに所属。