

人間の動きのセンシング - DigitEye3D™ とジェスチャセンサ -

平井 重行^{*1} 三輪 祥太郎^{*1} 金森 務^{*1}

片寄 晴弘^{*2} 坂口 貴司^{*3} 井口 征士^{*4}

Sensing Devices of Human Motions-DigitEye3D™ and Gesture Sensor-

Shigeyuki Hirai,^{*1} Shotaro Miwa,^{*1} Tsutomu Kanamori,^{*1}

Haruhiro Katayose,^{*2} Takashi Sakaguchi^{*3} and Seiji Inokuchi^{*4}

Abstract - This paper describes our sensing devices of human motions. We introduce two original sensing devices, one is the "DigitEye3D™", and the other is the "Gesture Sensor". The DigitEye3D™ is a motion capturing device, which is much cheaper than other same types of capturing devices, and can process 3-D data of human motions on the spot with no wires. The Gesture Sensor is a very small, light weight gesture-recognition device composed of accelerometers and gyroscopes. We describe mechanism of these two devices, and show two applications using them.

Keywords: Motion Capture, Gesture Recognition, Sensor

1. はじめに

近年、ヒューマンインタフェース、手話認識、バーチャルリアリティ等の分野において、人間の動作の計測および認識を行う研究が盛んに行われている。しかし、画像処理による方法^[1]は非接触式であるという利点はあるが、画像抽出およびトラッキングの処理が複雑になるため、リアルタイム計測が困難であり、またシステムが高価になる等の問題点をもつ。一方、磁気センサを用いて位置および姿勢を検出する方法は、周囲の磁界環境に影響を受けると同時に、計測範囲が限られる等の問題がある^[2]。

そこで我々は、身体の計測点に複数の赤外線球状発光体を装着し、各光源の発光をフレーム単位で能動的に切替ることによって複数発光点の位置を同定する手法を用いることで、非接触でありながらリアルタイムでの計測が可能な装置 "DigitEye3D™" を開発した。(DigitEye3D™ は (財) イメージ情報科学研究所の登録商標です。)

また、人間の動作計測の中でも、特にジェスチャは対人コミュニケーションにおいて重要な役割を果たしており、マンマシンおよびマンマンインタフェースの有効な手段と考えられる。ジェスチャには大きく分けて、シンボル情報(ジェスチャの語義を表す情報)とバリュウ情報(「とても」「やや」など語義内容の程度を表す情報)とがある。ジェスチャ認識の第一義はシンボル情報認識であり、よってシンボル情報を抽出しやすい位置センサを用いるのが一般的とされている。しかし、ヒューマンコミュニケーションシステムにジェスチャ認識を利用する場合、

シンボル情報だけでなくバリュウ情報の認識がより重要となる。

そこで我々は、バリュウ情報を抽出しやすいと言われる加速度、速度を一次情報として得られる加速度センサおよびジャイロセンサを用いたジェスチャ認識システム (Gesture Watcher) を開発した。

本論文では、以上で述べた2つの身体動作の計測デバイス、DigitEye3D™システムとジェスチャ認識システムに関して、その基本測定原理、処理メカニズム、及びその適用例について報告し、最後に今後の計画について述べる。

2. DigitEye3D™

市販のモーションキャプチャシステムとしては、モーションアナリシス社等数社が製品化しているが、いずれも非常に高価で、簡単に使用できるものではない。安価な実現が期待できる簡易な方式では二台のPSDカメラによるものや、三本のCCDラインセンサによる研究がある^{[3][4][5]}。これらはいずれもカメラやセンサの出力をADコンバータを持つパーソナルコンピュータで処理をする構成になっておりシステム簡略化の余地を残している。我々は一台のCCDモノクロカメラでリアルタイムに複数の光源の三次元位置を認識する、三次元簡易モーションキャプチャとしてDigitEye3D™の開発に取り組んできており^{[6][7][8]}、昨年には、KRI((株)関西新技術研究所)の協力により15万円という低価格で製品化に成功した。低価格なモーションキャプチャシステムを実現するために、DigitEye3D™システムは、以下の特徴を持つ。

(1) カメラ上に写った点光源の面積を用いて光源からカメラ迄の距離を算出する。

(2) 点灯させる点光源をフィールド単位で切り替えていくことによって、各光源の判別を行う。

以下、DigitEye3D™の概要と技術仕様について述べ、具体的なアプリケーション例を示す。

2.1 一般的な画像計測について

一般的にこうした画像処理による位置認識は映像信号

*1: (財) イメージ情報科学研究所

*2: 和歌山大学

*3: 三菱電機(株)

*4: 大阪大学

*1: Laboratories of Image Information Science and Technology

*2: Wakayama University

*3: Mitsubishi Electric Corp.

*4: Osaka University

をフレームバッファメモリ内に展開し、そのピクセル情報から求める方法がとられることが多い。この方法では認識に大量のデータを高速で処理するハードウェアを必要とする。本システムはフレームバッファメモリを持たず、汎用のシングルチップコンピュータで制御された標準ロジックで構成され、さらにこうした問題を回避するために以下の原理に基づいて動作している。

(1) 同期信号を分離した映像信号を直接二値化し、走査線上での光源の始点と終点の位置情報を得る。次にこれらの垂直同期信号と水平同期信号に対する時間計測を行い、光源の輪郭情報を取得し、この重心を求めて二次元情報とする。

(2) 複数の赤外線光源を垂直同期信号に従って能動的に制御し、フレーム毎に順次一個ずつ点灯させる。これにより1フレームで観測される光源数は常に1個になり対応付け問題を回避している。

(3) 全方位にわたって発光強度が一定な球状発光体を光源にすることで映像上での面積から奥行き情報を取得する。

以下、DigitEye3D™ について実装面を中心に述べる。

2.2 DigitEye3D™ のH/W構成

図1にシステムの構成図、図2にDigitEye3D™の写真を示す。システムは赤外線透過フィルタを装着したCCDモノクロカメラ、映像信号から位置認識を行うDigitEye3D™本体、複数の赤外線球状発光体から構成されている。CCDモノクロカメラは通常赤外領域で高感度な特性を持つ。システムではレンズ系にも赤外線光源だけを選択的に透過するIRフィルタを装着し、室内等の一般的な環境での使用を可能としている。光源は球状発光体で内部に数個の赤外線発光ダイオードを持ち、発光強度は支持棒を除く全方位にわたって一定に保たれている。

2.3 面積に基づいた距離の推定

物体は、近くに有るもの程大きく見えるという単純な事実を利用して、奥行き情報を計測する。まず、レンズ系を調節することで、カメラ中で観測される物体の大きさを、目的にあった精度で切り出す。切り出しの際には、計測点に使用する光源に近赤外LEDを用い、可視光カット



図2 DigitEye3D™
Fig.2 DigitEye3D™

フィルタをカメラに装着し、透過画像を閾値処理する。

ここで、光源は、ある程度大きさを持ったものが望ましいので、2.2で述べたように、球状の発光体で内部に数個の赤外線発光ダイオードを持ったものを使用する。また、ぼかしフィルタの利用、あるいは意識的に焦点をずらすことにより、距離に応じた総体的な面積に加えて、光量そのものを面積として計測することが可能となる。

2.4 点光源の検出

次に、カメラへの投影面と平行な2次元平面における位置を測定する手法について述べる。まず最初に、カメラ画像中の背景画像はゼラチンフィルタにより除去され、その結果、光源はフィールド毎に一つの円として認識される(図3)。この円の重心を求めることにより、2次元平面上での位置を計測する。光源の走査線上での位置情報を取り出すために、適当なレベルでコンパレータにより二値化し、円の輪郭位置として水平同期信号からの時刻T1、T2(24MHzでカウントアップ)を得る。光源を横切るすべての走査線からのT1、T2についての相加平均から重心を求め、水平位置としている。垂直位置は走査線が光源を横切り始めた線と終わった線の垂直同期信号からの位置の平均で求めている。また、奥行き情報は、各T1、T2

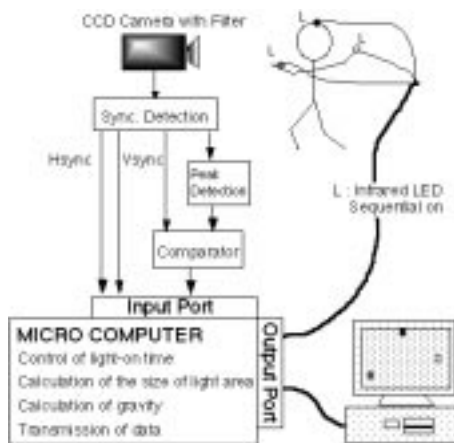


図1 システム構成
Fig.1 System Configuration

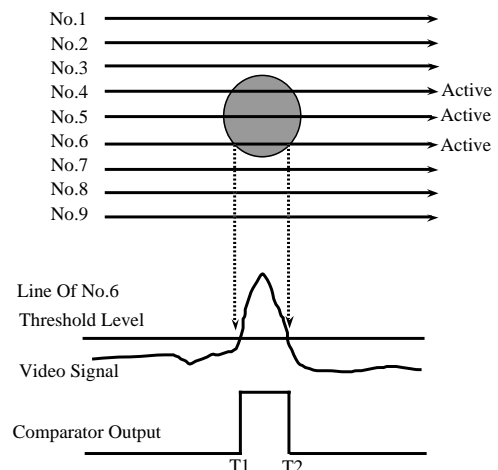


図3 位置認識の原理
Fig.3 Mechanism of Location Recognition

の差分を加算することで求めた面積から得る。

2.5 点光源の時間制御

本システムでの複数点のポイント識別の手法は、点光源の発光の時分割制御に基づいている。本手法を同時性という観点から見た場合には、正確な位置計測ができないが、想定した使用条件からはさほど重要な問題にはならない。赤外線光源の点滅制御のタイミングは、通常のカメラを用いた場合に考えられる時分割の最大スペックとして、フィールド単位で行っており、時間分解能は、ポイント数 \times 16.7msで計算される。以下、位置情報の送出迄の動作を簡単に説明する(図4)。システムは、映像信号の垂直同期に基づいて、そのフィールド内で光源の点灯制御を行う。それと同時に、映像信号上での(実際には、1フィールド前の)発光点の検出処理を行う。

映像信号の同期信号を用いることで、点滅制御回路が簡略化されると共に、映像信号から直接発光エリアの計算をするために、フレームバッファが不要である。このことは、高速カメラへの対応に有利な条件である。

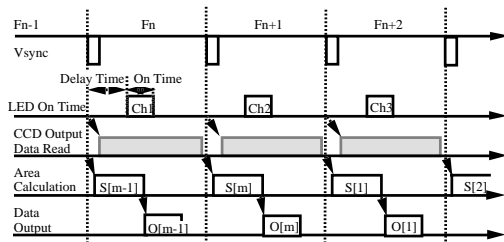


図4 点光源の制御

Fig.4 Switching of lights

2.6 アプリケーション - 動作による音楽演奏システム -

DigitEye3D™は低価格でパーソナルユースを狙ったモーションキャプチャセンサである。ここでは、このセンサの応用例として、家庭で個人が楽しめるような動作による音楽演奏システム "PlaytheDE" を紹介する。

PlaytheDEは、特別な知識や経験を必要とせず、幅広い世代に渡って感覚的に音楽を表現できる演奏システムである。図5に画面例を示す。発光体は計4つ使用し、この内2点の情報が音のON/OFF制御、残りの2点の情報が音色の制御を担う。まず音のON/OFFのスイッチの役割を果たす2つの発光点はユーザの手のひらにつけられており、手の開閉に合わせて発光点が見え隠れすることで、音のON/OFF制御を行う。次に、残りの音色を制御する2点であるが、認識された3次元情報のうち、水平方向が音の高さ、垂直方向が音の大きさ、奥行き方向が音の出現頻度にマッピングされている。この結果、ユーザが手のひらで音のON/OFFを行いながら、様々な音色で発音するという構成をとっている。また、音の高さに関しては、単に位置情報をマッピングするのではなく、ブルースや四七抜き音階などへの変換を行っている。BGMを流しながら、ソロを演奏することも可能である。

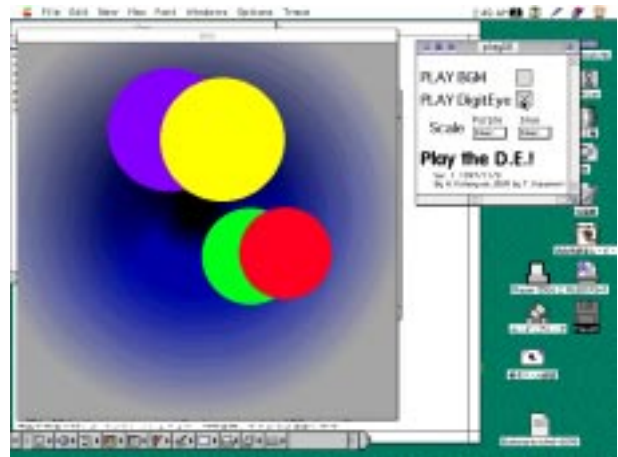


図5 PlaytheDE

Fig.5 PlaytheDE

3. ジェスチャセンサ

1. はじめに、人間のコミュニケーションにおけるジェスチャの重要性、中でもバリュー情報の重要性について述べた。

本章では、加速度センサとジャイロセンサとを組み合わせたジェスチャセンサを用いニューラルネットワークで統合化するジェスチャ認識手法について述べ、最後にこのシステムを用いて、実際に12種類のジェスチャ(手話動作)認識実験を行い、シンボル情報認識およびバリュー情報認識における本手法の有効性を検討した。

3.1 加速度センサとジャイロセンサを用いた統合型センシング

人間の基本動作は関節による屈曲動作である^[9]。よって動作の計測・認識にはジャイロセンサなどの回転系センサが適切である。しかし、ジャイロセンサのみを用いたジェスチャ認識^[10]は、絶対的な姿勢情報の取得が行えない、そのうえ出力信号のDCドリフト成分によって積分誤差を生じるという欠点をもつためシンボル情報認識に適していない。一方、半導体加速度センサはDC成分を高精度に測ることができ、重力加速度の検出により、静止および低速動作時の姿勢推定を精度良く行うことができる。

そこで、加速度センサとジャイロセンサとを組み合わせたジェスチャセンサを用い、ニューラルネットワークで動作認識を行うというジェスチャ認識システム (Gesture Watcher)^[11]を開発した。

3.2 加速度センサとジャイロセンサを用いたジェスチャ認識システム

3.2.1 システム構成

本システムのシステム構成を図6に示す。システムは、加速度センサとジャイロセンサとから構成される3軸ジェスチャセンサとセンシングデータからジェスチャの特徴量を抽出しジェスチャ認識を行うPC上のニューラルネットワークのプログラムとからなる。図7に実物の開発した3軸ジェスチャセンサとその装着方法を示す。センサは、ユーザの手の甲にバンドで止められ、ジェスチャを行うとその間の3軸の加速度および角速度データが検出

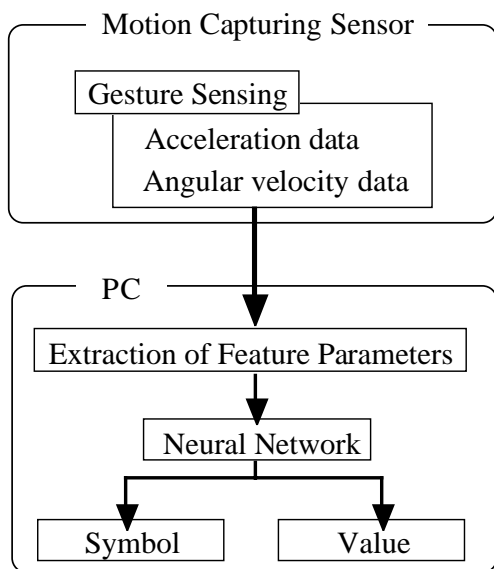


図6 システム構成
Fig.6 System Configuration

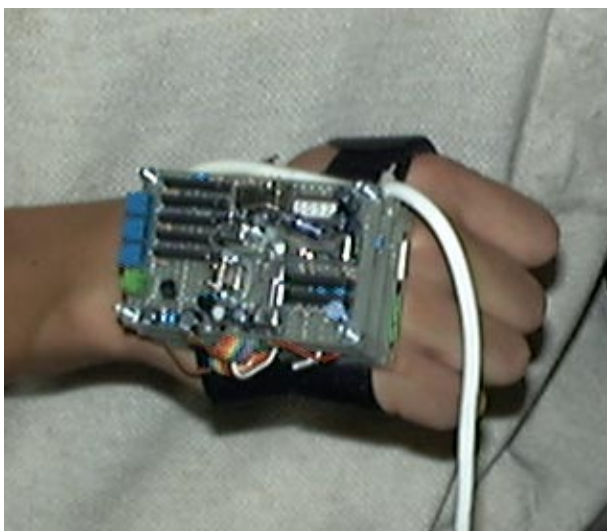


図7 ジェスチャセンサ
Fig.7 Gesture Sensor

される。検出されたデータから特徴量を抽出し、あらかじめ教示データによって学習させてあるニューラルネットワークを用いて、そのジェスチャのシンボル (Symbol) 情報やバリュー (Value) 情報を出力させる。

3.2.2 ニューラルネットワークを用いたジェスチャ認識部

本節では、認識のポイントとなるニューラルネットワークを用いた認識部に関して、ジェスチャ動作の定義、ジェスチャの特徴パラメータ、そしてネットワークの構造に関して説明する。

(1) ジェスチャ動作の定義

まず、ジェスチャ動作は以下のように基準化した。動作の開始点は静止判定を利用して自動認識させる。まず最初に原点で静止し、つぎにジェスチャ開始点で静止し、その後ジェスチャを行い、動作が終わったら再び静止する。一般的にジェスチャ動作は連続表現する場合でも、1

ジェスチャ毎の区切りをはっきりさせるためにほぼ静止する動作を間にはさむ。また手話を知らない人あるいはコンピュータに意図・感情を伝達することを目的として区切りの静止を明示的にはさむことは、ジェスチャ表現にさほど影響を与えない。そこで、ジェスチャの開始点および終了点検出は静止判定によって行った。具体的にはある一定時間(約0.5秒間)すべてのジャイロセンサの出力値がしきい値を下回ったら静止と判断した。

(2) ジェスチャ特徴パラメータ

本システムではリアルタイム認識を前提としているので、入力ノード数を減らすため、時系列データを直接用いるのではなく、下記の特徴量を使用した。表1に本システムにおいて検討したジェスチャ特徴量およびパラメータ数を示す。

表1 特徴パラメータ及びその数

Table.1 Feature Parameters and their numbers

	Accelerometer	Gyroscope
Mean	3	3
P-P	3	3
Start	3	3
Completion	3	3
Freq.	0	1
Func.[P]	2	2

Mean: mean during a gesture

P-P: peak amplitude during a gesture

Start: value at a gesture start

Completion: value at a gesture completion

Freq.: gesture frequency

Func.[P]: evaluation functions for Value information

特徴量候補にはジェスチャ動作中の動的情報を表すものとして両センサの各軸毎の平均値および1周期動作間のピーク振幅(記号P)、ジェスチャ周波数平均値(記号F)、計算式(式1)、(式2)で求めたバリュー情報評価関数およびその自乗値のそれぞれ平均値(記号P)を選択した。

$$P_a(t) = \sqrt{(a_x(t)^2 + a_y(t)^2 + a_z(t)^2) / 3} \quad (式1)$$

$$P_g(t) = \sqrt{(g_x(t)^2 + g_y(t)^2 + g_z(t)^2) / 3} \quad (式2)$$

また、ジェスチャ前後の静的情報としてジェスチャ開始・終了点情報を選択した。

周期検出には角速度信号のゼロクロス点を利用し、ゼロクロス点前後の角速度サンプル値の大きさをもとに加重配分して求めた。また、両端の周期データは除外した。

非周期的動作の場合、ジェスチャ間の最大振幅をピーク振幅とし、周波数は0とした。

ジェスチャ開始・終了点情報とはジェスチャ開始・終了時点での位置・姿勢情報のことであり、姿勢(重力加速度)情報として加速度データを使用した。また、位置情報として基準点(システム起動時)からのジャイロ出力(角速度)の積分値を用いた。

バリュウ情報評価関数Pとは3次元空間内での加速度、角速度の大きさを表現したものであり、自乗することによりエネルギー要素を取り入れた。式中 $a_x, a_y, a_z, g_x, g_y, g_z$ はそれぞれ加速度センサ、ジャイロセンサのサンプリング時間毎の3軸方向の出力である。

(3) ニューラルネットの構成

本システムで使用したニューラルネットの構成を図8に示す。教示データから3.2.3(2)のジェスチャの特徴量を抽出し、バックプロパゲーションによって中間層のノード数が16の3層のニューラルネットを構築し、これを用いてジェスチャのシンボル情報や、バリュウ情報を出力する。

3.3 ジェスチャ認識実験

12種類の明示的なジェスチャ動作を対象とした認識実験を行い、シンボル情報認識およびバリュウ情報認識における本システムの有効性を検討した。

3.3.1 実験概要

認識対象ジェスチャは日本手話(JSLs)とし、バリュウ情報を多く含んでいる形容詞などを表す手話の中から、主に上肢運動によってその意味/程度を表現しているものを10種類選択した。選択した単語は「寒い」「暑い」「嬉しい」「涼しい」「悲しい」「難しい」「美味しい」「明るい」「強い」「話す」であり、並進/回転動作、繰り返し/非繰り返し動作を含んでいる。なお、実験は単独のジェスチャ

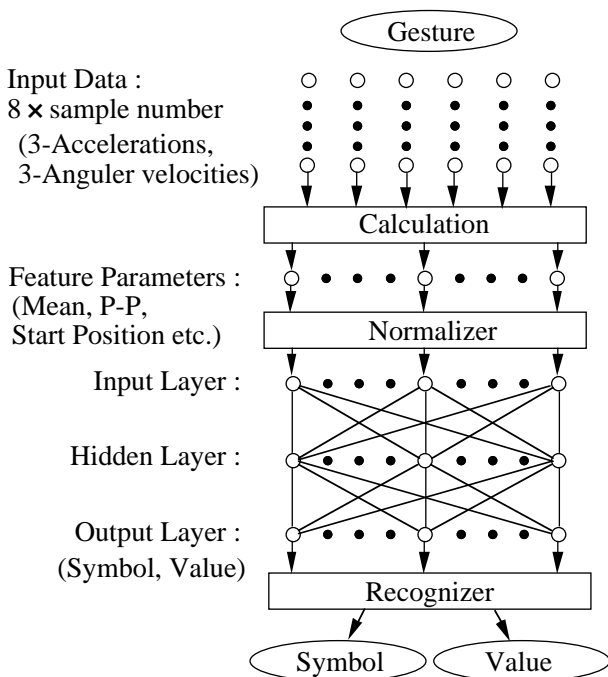


図8 ニューラルネットワークの構成
Fig.8 Neural Network Configuration

を認識対象とした。

3.3.2 バリュウ(程度)情報認識

バリュウ情報は0.0~1.0までの値で表現し、被験者の心理スケールを基準にして、程度の小さい方から順に0.1, 0.3, 0.5, 0.7, 0.9の5段階を収集した。出力ノードはバリュウノード1つとし、出力値(0.0~1.0)をバリュウ情報認識結果とした。なお、認識は特定者に限定した。テストデータに対するバリュウ情報認識のずれを表2に示す。

3.3.3 シンボル(意味)情報認識

出力ノード数にはジェスチャ数である10を割り当て、教示時に対応ノードのみを1としその他を0とした。認識時には最大値を示すノードに対応するジェスチャを認識結果とした。ここでは特定者認識(A) 不特定者認識(B) グループBの被験者のジェスチャを教示データに加えた場合の認識(C)の3種類の認識実験を行った。不特定者には手話熟練者1名を含む5名を選定した。

表3に認識結果を表す。特定者ジェスチャ認識の場合、単独センサでもほぼ100%認識することができた。これに対して不特定者ジェスチャ認識の場合、個人差の影響で認識率は低下するが、単独センサを用いた場合に比べて、両センサを併用した場合の認識率の低下は小さい。また不特定者のジェスチャも教示データに加えた場合、学習に幅が広がるため認識率は上がるが、特に両センサを併用した場合の認識率は100%となった。

表2 バリュウ情報認識のずれ

Table 2 Recognition Error of Quantative Information

Accelerometers	Gyroscopes	Both
0.146	0.155	0.063

表3 シンボル情報の認識結果

Table 3 Recognition Rate of Symbol Information

	unit : %		
	A	B	C
Accelerometers	99.6	70.0	92.0
Gyroscopes	100.0	72.5	94.0
Both	98.8	85.0	100.0

A : Particular (teaching)
B : Non-Particular (no teaching)
C : Member B, teaching

4. おわりに

人間の動作を簡単かつ安価なシステムでリアルタイムに計測する手法を2例、提案した。両手法共、センシングユニットは非常に小型軽量であり、人間動作計測に向いている。

DigitEye3D™の採用する光源の発光を能動的に切り替える方法は、高価なフレームメモリを使用することなく、また膨大で不完全な画像処理演算を行うことなく、複数発光点の位置を同定することができる。また、加速度セン

サとジャイロセンサを用いた統合型のジェスチャセンサは、人間動作の角速度、加速度が高精度に得られることから人間のアナログ的情報の計測・認識に向いていること、オクルージョンの影響を受けないことなどの特長をもつ。

今後の計画として、DigitEye3D™の場合、各発光体の時間軸方向での分割発光を採用した現在の手法では、計測点の数が増えた場合に時間方向の解像度が低下してしまうという問題がある。そこで、今後は、時間軸方向の解像度をあげられるような改良を行っていく予定である。これにより、軽量、小型、安価、そしてリアルタイム処理が可能であるというメリットを持つDigitEye3D™を、さらに時間分解能の高い認識も可能なモーションキャプチャセンサへと改良していく。

ジェスチャ認識システムでは、加速度センサとジャイロセンサとを組み込んだジェスチャセンサを用い、ニューラルネットワークで統合化する認識手法を提案した。また、このジェスチャセンサを用いて、12種類のジェスチャ(手話動作)の認識実験を行い、シンボル情報に加えて、バリュウ情報の取得が可能であること、シンボル情報認識においてもバリュウ情報認識においても両センサの併用が有効であること、1つのニューラルネットワークでシンボル情報およびバリュウ情報の同時認識が可能であることを確認した。本方式では、腕の交差による隠れや磁界環境の影響を受けずにリアルタイムかつ安定に動作計測することが可能である。また、骨格モデルに基づかない認識手法であることから、不特定者ジェスチャ認識に強みを発揮するだけでなく、センサの取付位置および取付姿勢のずれにあまり影響されないという利点がある。一方で、動作というよりも力みによる身体表現を行うジェスチャに関しては、認識率が低いという欠点がある。

このジャイロセンサと加速度センサを組み合わせたジェスチャセンサは、現在、EMG(皮膚表面筋電)を測定することで、力みを認識できるジェスチャ認識システムへと発展している。

5. 参考文献

- [1] 高田, 他: 肘と手の関係を用いた3次元動作解析法; 第1回画像センシングシンポジウム講演論文集, pp.5-10(1995).
- [2] 城山, 他: 小数センサデータからの上半身動作の再構成に関する一検討; 信学会総合大会, pp.262(1995).
- [3] 畑田, 久米, 石川: 視覚情報提示の研究; 平成7年度重点領域研究, 研究成果報告書, 人工現実感の研究, pp.33-34(1996).
- [4] 白井, 久米, 津田, 畑田: 光学的3次元位置検出法を用いたリアルタイム人間動作入力デバイス; 平成7年度重点領域研究, 研究成果報告書, 人工現実感の研究, pp.209-214(1996).
- [5] G.Bianchi, F.Gazzani and V.Macellari: The COSTEL system for human motion measurement and analysis; SPIE, Vol.1356, Image-Based Motion Measurement, pp.38-50(1990).
- [6] 片寄, 金森: 三次元位置認識方法、および装置; 特許出願番号 特願平 8-47155号(1996).
- [7] 片寄, 金森, 井口: 簡易モーションキャプチャセンサ DigitEye3D- 原理と基本特性 - 信学論, Vol.J80-D-II, No.10, pp.2889-2892(1998).
- [8] 金森, 片寄, 井口, 戸島, 西山: モーションキャプチャ

「DigitEye3D」の実装; 信学論, Vol.J81-D-II, No.5, pp.804-809(1998).

- [9] 田所, 相川, 市道: 上肢運動の動作解析に関する研究; 人間工学, Vol.26, No.1, pp.41-47(1990).
- [10] 坂口, 金森, 片寄, 井口: ジャイロセンサを用いた人間の動作計測; 第50回情処全, 4, pp.329-330(1995).
- [11] 坂口, 金森, 片寄, 佐藤, 井口: 加速度センサとジャイロセンサを用いたジェスチャ認識; SICE論文誌, Vol.33, No.12, pp.1171-1177(1997)