2004－CVIM－144　(27)
2004／5／7

マルチフレームレート制御に基づく身振りのリアルタイム画像認識

桐島 俊之†　佐藤 宏介††　千原 國宏†††
†奈良工業高等専門学校電気工学科
††大阪大学大学院基礎工学研究科
†††奈良先端科学技術大学院大学情報科学研究科

概要：人物の自然な振る舞いをリアルタイムで認識する技術は，映像メディアや知能ロボットなどとの直接的インタラクションのみならず，セキュリティや介護支援といった生活環境領域においてもその重要性が急速に高まってきている．従来の身振り認識システムの多くは，単一の処理フレームレートで身振り動作を観測し，認識処理を行っている．しかしながら，生活環境領域における多種多様な運動特性を持つ動作すべてを単一処理フレームレートで認識できる保証はない．本論文では，異なる運動特性を持つ身振りを，異なる処理フレームレートで同時観測して認識する，マルチフレームレート制御によるリアルタイム身振り認識手法を提案する．さらに，評価実験により提案手法の有効性を検証する．

Real-Time Gesture Recognition by Multiple Frame Rate Control

Toshiyuki KIRISHIMA†, Kosuke SATO††, and Kunihiro CHIHARA†††
†Department of Electrical Engineering, Nara National College of Technology
††Graduate School of Engineering Science, Osaka University
†††Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: Recognizing human's natural behaviours in real-time is indispensable for not only direct interaction with visual media and intelligent robotic systems but also in security and nursing support systems to be used in our daily activity domains. Most of the foregoing gesture recognition systems observe gestures under single processing-frame-rate condition. But there is no guarantee that all gestures can be successfully recognized, for the gestures are so rich in their variability, e.g. variation and speed. In this paper, a Multiple Frame Rate Control(MFRC) approach is presented, and its recognition performance is demonstrated through the evaluation experiments.

1 まえがき

我々の日常生活にコンピュータが急速に浸透し，コンピュータを媒介としたコミュニケーションが常態化するにつれて，コンピュータに人間の活動状況や意図を読み取らせることの意義は益々大きくなっている．現状ではセキュリティ分野，近い将来としては介護支援分野およびヒューマンロボットインタラクション分野などにおいて，そのニーズは一層高まると考えられる．こうした状況の下，身振り認識・理解へのニーズは，従来の直接対話型から，認識システムが生活環境などに組み込まれて機能する環境型へと拡大・多様化してきている．これら要求に応えるには，複雑な環境条件下において非接触かつ実時間で人物の自然な振る舞いを把握する画像認識手法が不可欠である．

実環境でリアルタイム動作する身振り認識システ

ムには，観測および認識に必要な処理時間を含めた上で，所定の処理フレームレートでの動作が要求される．処理フレームレートを一定に保つことは，すなわち，一定時間間隔で動作を観測することであり，認識精度の向上が期待できる．しかしながら、従来の身振り認識システムの多くは、処理フレームレート安定化の問題を考慮しておらず，コンピュータ性能に依存した不安定な処理フレームレートでの認識処理を行っている．この問題に対処するために，筆者らは，多注視点選択制御法を提案し，任意の処理フレームレートでの認識処理を実現している．

しかしながら，日常生活環境における身振り動作には，スポーツ動作のような高速なものからストレッチ動作のような低速なものまで幅広く存在している．これら動作すべてを単一の処理フレームレートで認識できる保証はない．そこで，本論文では，異なる運動特性を有する身振りを異なる処理フレームレートで同時に観測および認識した場合の認識率への影響について検証する．

2 関連研究

非接触・非装着での身振り認識のために，画像処理技術に基づく認識手法がこれまでに数多く提案されている．それらの手法は，身体モデルに基づく手法と図形パターンに基づく手法に大別できる．前者の例としては，円筒モデルなどの幾何プリミティブを入力画像にフィッティングさせる方法や，複数のマーカーを体に取り付けてそれらの位置関係から姿勢などを推定する方法がある．後者の例としては，入力画像をパターン情報とみなし，それが属するカテゴリーを推定する図形パターン認識技術に基づく手法がある．その代表的手法としては，DP照合法に基づく手法，ファジー連想記憶に基づく手法，HMM（Hidden Markov Model）に基づく手法，図形の固有空間内の軌跡の類似性に基づく手法，ニューラルネットワークに基づく手法，更にZernike Momentなどの図形モーメント特徴に基づく手法などがある．

一方，認識プロセスを複数動作させ協調させるマルチエージェント型の身振り認識システムは，研究の途上にある．協調分散視覚研究[(1)(2)]では，複数カメラからの入力画像を複数のコンピュータにより協調処理させることで人物の移動経路などを推定している．また，いわゆる多視点型の身振り認識システムでは，複数カメラからの画像を単一あるいは複数の認識プロセスにより処理している．

上述の従来システムにおける共通点の一つとして，実際の処理フレームレートが実装環境に依存するということが挙げられる．身振り画像を処理する場合，単位時間当たりのデータ量が膨大となるため，実時間認識するには入力画像列から有意な空間的特徴量を選択的に認識する手法が必要となる．また，標準パターンの増加に伴う処理フレームレートの低下や，アプリケーションシステムとの接続に伴う処理フレームレートの低下，更にＯＳ環境下の他プロセスの影響による処理フレームレートの不安定化などの問題に対処する必要がある．

もう一つの共通点は，上述の従来システムが単一の処理フレームレートで認識処理を行っているということである．すでに述べたように，日常生活環境における身振り動作には，スポーツ動作のような高速なものからストレッチ動作のような低速なものまで幅広く存在している．これら動作すべてを単一の処理フレームレートで認識できる保証はない．協調分散視覚研究の考え方を時間軸に適用することで，多様かつ複雑な動作の認識がよりロバストに行えるようになる可能性がある．この場合，複数の認識プロセスを異なる処理フレームレート下で動作させることになる．しかしながら，複数プロセスかつ異なる処理フレームレートによる観測により，認識性能が具体的にどのような影響を受けるかについては十分な検討はなされていないのが現状である．

そこで，本論文では，多注視点身振り認識法[(3)]と多注視点選択制御法[(4)]を活用した複数プロセスかつ異なる処理フレームレートでの身振り認識方式，すなわち，マルチフレームレート制御による身振りの実時間画像認識手法を提案し，認識性能の処理フレームレート依存性評価を行う．

3 マルチフレームレート制御

マルチフレームレート制御に基づく身振り認識では，複数の認識プロセスを並行動作させ，異なる処理フレームレート下で身振りを観測・認識させる．このような認識システムを実現する鍵は，複数の認識プロセスを，異なるパラメータ条件下で並行動作させることにあり，スケーラビリティが高く自律性を備えた認識手法が不可欠である．本論文では，多注視点身振り認識法に基づくプロトコル学習，多注視点選択制御法による処理フレームレート制御機能を活用し，マルチフレームレート制御に基づく身振り認識を実現する．具体的には，これら手法に基づ

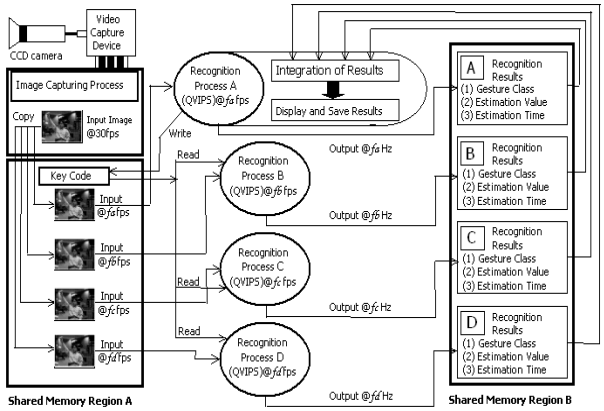

図 1: マルチフレームレート制御による身振り認識の枠組

く認識プロセスを複数個並行動作させ，各プロセスから出力される認識結果を統合する．以降，マルチフレームレート制御の枠組みを提案した後，多注視点身振り認識法および多注視点選択制御法の概要を説明する．

3.1 マルチフレームレート制御の枠組み

図 1 にマルチフレームレート制御による身振り認識の枠組みを示す．図 1中には，４個の認識プロセスが描かれているが，提案手法においては，認識対象とする身振り動作数，および，認識プロセス数自体には理論的な制約はない．なお，本論文で提案するシステムでは，NTSC 準拠の画像入力装置を使用するため，実装時の処理フレームレートは最大 30[fps] に制限される．

CCD カメラから出力される映像信号は，画像キャプチャプロセスによりビデオキャプチャ装置を介して，コンピュータに入力される．画像キャプチャプロセスは，共有メモリ領域 A 上に設けてある各認識プロセスに対応する画像データ領域に入力画像をコピーする．これは，複数の認識プロセスを並行動作させる際，同一の共有メモリ領域を同時にアクセスすることで生じる待ち時間を回避するためのものである．

認識システムへの動作指示は，認識プロセス A のコントロールパネル上から行う．認識プロセス A 上でのキー入力データは，共有メモリ領域 A 内キーコード領域に格納され，他の認識プロセス上で共有され，同一の操作を実行する．これにより，複数の認識プロセスがあたかも１つの認識プロセスのように動作することになる．

続いて，共有メモリ領域 A に格納された画像データを各認識プロセスがそれぞれの処理フレームレートで読み出し，認識処理を行う．各認識プロセスは，それぞれの処理フレームレートで認識結果（最有力クラス名と評価値）を出力する．この際，それぞれの認識結果は，共有メモリ領域 B 上に設けてある各認識プロセスに対応するデータ領域に格納される．入力された動作の判定は，処理フレームレートが最も低い認識プロセスＡ上で行う．この際，認識プロセスＡは各認識プロセスからの評価値を共有メモリ領域 B から読み出し，次式により統合評価値 E_{final} を算出し，統合評価値が最大となる身振りクラスに入力された動作が属するものと判定する．

$$E_{final} = W_A E_A + W_B E_B + W_C E_C + W_D E_D \quad (1)$$

ここで，W_A，W_B，W_C，W_D は各認識プロセスにおけるクラス重みであり，E_A，E_B，E_C，E_D は各認識プロセスが出力する動作終了時の累積評価値である．

3.2 多注視点身振り認識法

多注視点身振り認識法における基本的アイディアは，同一種類の身振りとして定義される動作から視覚的な共通項を学習させ，人物身振りをより柔軟に認識させることにある．こうした要求に応えるために，多注視点身振り認識法では，同一種類として与えられる身振り画像列から，注視すべき視覚的特徴を見出し，それらに対してより大きな重みを与える．多注視点身振り認識法は，特徴量に基づく照合処理，活性化マップによる特徴統合処理，身振りプロトコルに基づく認識処理の３段階により構成される階層型身振り認識機構である．

図 2 を用いて処理の概要について説明する．時々刻々と入力される身振り画像は，複数種類の特徴抽出フィルタにかけられ，その後，各種注視点に対応する特徴量が抽出される．抽出された特徴量は，学習時に身振り標準パターンとして登録される．一方，認識時には身振り標準パターンとの照合処理が行われ，照合結果は活性化マップとして出力される．各注視点の重み付けのために，活性化マップを利用した身振りプロトコルの学習（以降，プロトコル学習

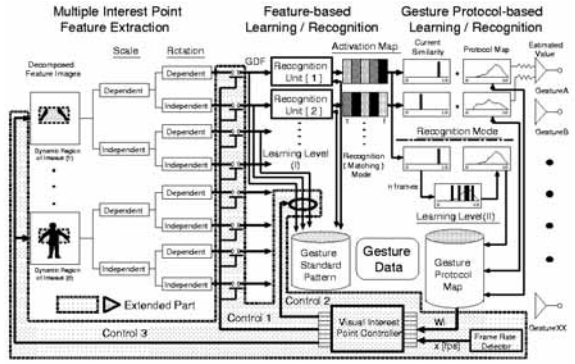

図 2: 各認識プロセスにおける処理の枠組

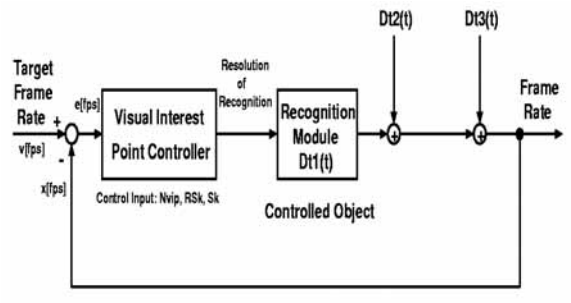

図 3: フィードバック制御のブロック図

と呼ぶ）が行われる．プロトコル学習では，ある身振りを認識する際に時間領域で安定している注視点に，相対的に大きな重みが割り当てられる．更に，各注視点に対応するゆう度分布（以降，プロトコルマップと呼ぶ）を生成・登録し，以降，このプロトコルマップに基づいた認識処理が行われる．

3.3 多注視点選択制御法

多注視点身振り認識法では，認識対象となる身振りの種類およびその動作時間が増大すればする程，処理フレームレートが低下し，認識システム自体のリアルタイム性が損なわれてしまう．認識対象となる身振りの種類およびその動作時間が増大しても，任意かつ一定の処理フレームレートを維持させるために，本研究では多注視点選択制御法を提案し，多注視点身振り認識法に組み込んでいる．多注視点選択制御法では，図 2 の斜線部に示す有効注視点の選択制御（Control 1），パターン照合間隔の選択制御（Control 2），パターン走査間隔の選択制御（Control 3）を実行する．選択制御を操作量 3 変数・制御量 1 変数のフィードバック制御問題として捉え，図 3 に示す処理負荷の大きさを動的に選択するフィードバック制御系を構成する． 具体的には，制御量を処理フレームレート x[frames/s]（[frames/s] は以降 [fps] と表記）（目標フレームレート v[fps]）とし，操作量をパターン走査間隔 S_k（特徴画像を走査する際の刻み間隔を指す），パターン照合間隔 RS_k（入力画像から得る特徴パターンと既に登録してある特徴パターンとを照合する際のステップ間隔を指す），有効注視点数 N_{vip}（認識処理で考慮される注視点の数を指す）の 3 変数としている．想定される制御対象は，認識モジュール自体の負荷 $D_{t1}(t)$，応用システムとのプロセス間通信に伴うネットワーク負荷 $D_{t2}(t)$，さらに応用システム自体の負荷 $D_{t3}(t)$ の 3 種類である．本論文では認識モジュールの負荷 $D_{t1}(t)$ を操作することにより，システム全体での処理フレームレートを目標値へと安定化させる。なお，マルチユーザ・マルチプロセス環境での動作を想定しているので，$D_{t3}(t)$ には予測不可能な他ユーザなどのプロセスによる影響も含めている．フィードバック制御では，微小操作量を適用した際の処理時間変化を逐次検出し，以降の操作量を決定している．この際，制御偏差 $e(= v - x)$ が最小となるよう負帰還をかけている．

4 評価実験

4.1 実験環境および実験条件

提案手法に基づく身振り認識システムをパーソナルコンピュータ（Pentium 4(2.26GHz,Single CPU), Memory 768MB，OS:Linux）上にC 言語で実装し評価実験を行った．CCD カメラからの画像は画像キャプチャボード（WinFast TV）を通してパーソナルコンピュータに解像度横 320[dot] 縦 240[dot]，16 ビットカラー (R:5[bit]，G:6[bit]，B:5[bit]）で取り込まれる．入力画像は，解像度横 80[dot] 縦 60[dot] の 8 ビット濃淡画像に変換された後，共有メモリ領域に格納され，各認識プロセスに利用される．

上記以外のハードウェア，および，実験時の特殊な照明や背景は使用していない．なお，本実験では，認識段階での解像度を低く設定（横 80[dot] 縦 60[dot] の二値画像）しているため，手指や顔表情などの細部は考慮されず，主に体全体の大雑把な動きが対象となる．

表 1: 評価実験対象動作（I）

A群	
動作記号	動作カテゴリ（名称）
GA-A	ストレッチ（深呼吸）
GA-B	ストレッチ（水平腕伸ばし）
GA-C	ストレッチ（体前屈）
GA-D	スポーツ（砲丸投げ）
GA-E	スポーツ（棒スイング）
GA-F	スポーツ（拳パンチ）

B群	
動作記号	動作カテゴリ（名称）
GB-A	ストレッチ（体 (両) 側面伸ばし）
GB-B	ストレッチ（体 (左) 側面伸ばし）
GB-C	ストレッチ（体 (斜め方向) ねじり）
GB-D	スポーツ（バレーレシーブ (低)）
GB-E	スポーツ（バレーアタック）
GB-F	スポーツ（テニスフォアハンド）

C群	
動作記号	動作カテゴリ（名称）
GC-A	ストレッチ（体 (右) 側面伸ばし）
GC-B	スポーツ（バレーレシーブ（高））
GC-C	スポーツ（腕振りダッシュ）
GC-D	スポーツ（テニスバックハンド）
GC-E	ストレッチ（体 (水平) 横ねじり）
GC-F	ストレッチ（バンザイ）

表 2: 各動作群の設定フレームレート（I）

動作群A	
認識プロセス名	処理フレームレート [fps]
PA-A	10
PA-B	18
PA-C	25
PA-D	30

動作群B	
認識プロセス名	処理フレームレート [fps]
PB-A	15
PB-B	20
PB-C	25
PB-D	30

動作群C	
認識プロセス名	処理フレームレート [fps]
PC-A	12
PC-B	16
PC-C	25
PC-D	30

4.2 マルチフレームレート制御による認識実験

本節では，ストレッチ動作９種類とスポーツ動作９種類の合計１８種類の動作について行った評価実験について述べる．各動作の軌跡画像を図 9 に示す．本実験は，次に示す手順で実施した．

（１） １８種類の動作を表 1 に示す３グループ（A群・B群・C群）に分けて，以下の手順に従って実験を行った．

【手順１】 各動作群に対応した４個の認識プロセスを立ち上げ，各動作群６種類の対象動作をプロトコル学習（標準動作１回，類似動作１回）させる．

【手順２】 各動作群に対応して，表 2 および表 4 に示す目標フレームレートを設定し，多注視点選択制御を開始する．

【手順３】 認識対象動作を入力する．

【手順４】 共有メモリ領域 B に格納された各認識プロセスの評価値を認識プロセスA上で統合し，最終の認識結果を表示し，ファイルに保存する．

【手順５】 各動作についてそれぞれ２０個ずつ，合計１２０個のテストサンプルについて上記の手順３と手順４を繰り返し適用する．

（２） (1)により得られた実験結果（図 4，図 5，図 6）において，認識率が処理フレームレートに依存しないと判断した表 3 内のD群６動作について，(1) と同様の手順で実験を行った．得られた実験結果を図 7に示す．

（３） (1)により得られた実験結果（図 4，図 5，図 6）において，認識率が処理フレームレートに依存すると判断した表 3 内のE群６動作について，(1) と同様の手順で実験を行った．なお，対

表 3: 評価実験対象動作（II）

D群	
動作記号	動作カテゴリ（名称）
GA-A	ストレッチ（深呼吸）
GA-D	スポーツ（砲丸投げ）
GA-E	スポーツ（棒スイング）
GB-A	ストレッチ（体 (両) 側面伸ばし）
GB-B	ストレッチ（体 (左) 側面伸ばし）
GB-D	スポーツ（バレーレシーブ (低)）

E群	
動作記号	動作カテゴリ（名称）
GA-C	ストレッチ（体前屈）
GB-F	スポーツ（テニスフォアハンド）
GC-F	ストレッチ（バンザイ）
GB-E	スポーツ（バレーアタック）
GC-B	スポーツ（バレーレシーブ（高））
GC-D	スポーツ（テニスバックハンド）

象とした６動作の内，前半の３動作については「処理フレームレートが向上すると認識率も向上する動作」，後半の３動作については逆に「処理フレームレートが向上すると認識率が低下する動作」を選出した．得られた実験結果を図 8 に示す．

ここでの認識率は，認識処理時の有効画像枚数に占める（正答枚数の割合）を,（誤答枚数の割合）の総和と（正答枚数の割合）を加えたもので割ることにより算出した.（正答枚数の割合）とは，正答枚数を処理画像枚数で割った数値である．一方,（誤答枚数の割合）とは誤答枚数を処理画像枚数で割った数値であり，正答クラス以外のクラスそれぞれについて算出する．これにより，誤認識率も考慮した上での認識率が算出される．

5 考察・検討

前節で示した実験結果について，平均認識率の最大値と最小値の差が 5(%) に収まるものを「処理フレームレートに依存しない傾向を持つ動作」，それ以外のものを「処理フレームレートに依存する傾向を持つ動作」として分類すると以下のようになる．

動作群Aについては，認識率が処理フレームレートに依存しない傾向を持つ動作（GA-A／GA-D／GA-E）と，依存する傾向を持つ動作（GA-B／GA-C／GA-F）が見出せる．統合評価値に基づく認識率（98.1(%)）が，認識プロセス PA-B（18[fps]）の平均認識率（98.6(%)）を除く単一プロセスのみによる認識率を上回る結果が得られた．

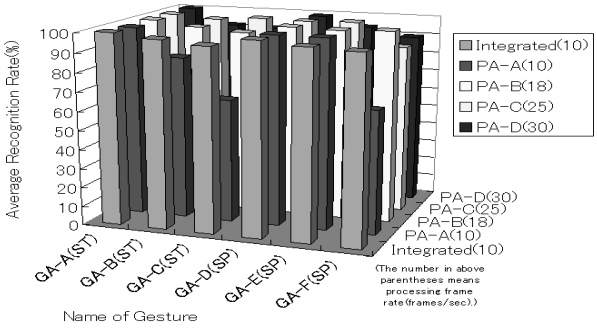

図 4: 各動作における平均認識率（動作群Ａ）

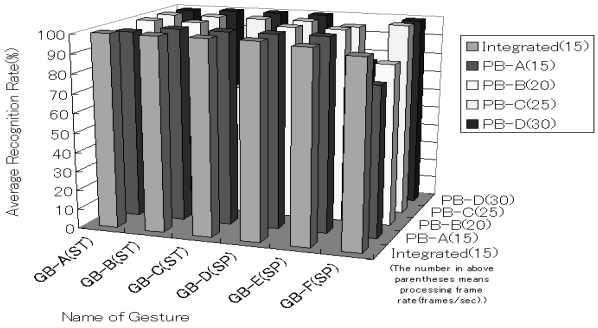

図 5: 各動作における平均認識率（動作群Ｂ）

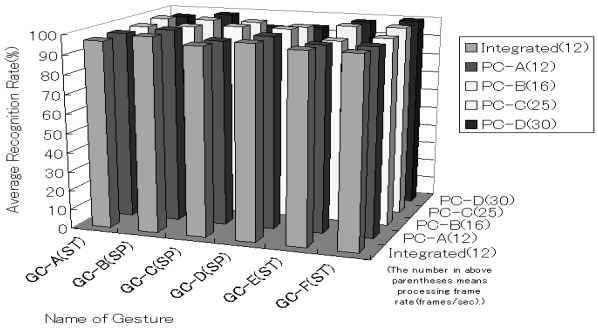

図 6: 各動作における平均認識率（動作群Ｃ）

動作群Ｂについては，認識率が処理フレームレートに依存しない傾向を持つ動作（GB-A／GB-B／GB-D）と，依存する傾向を持つ動作（GB-C／GB-E／GB-F）が見出せる．統合評価値に基づく認識率（99.0(%)）が，認識プロセス PB-C（25[fps]）の平均認識率（99.7(%)）を除く単一プロセスのみに

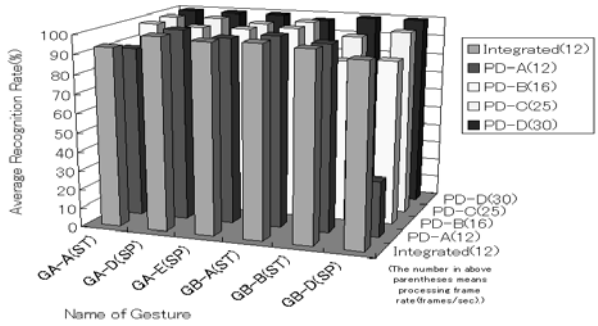

図 7: 各動作における平均認識率（動作群Ｄ）

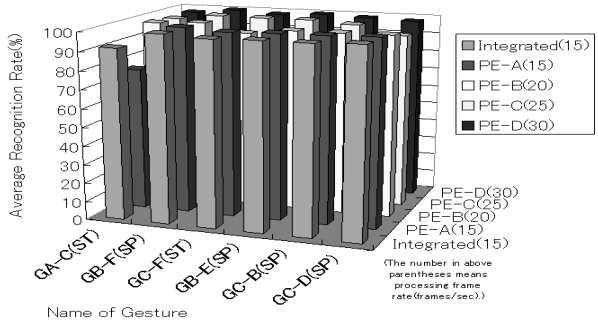

図 8: 各動作における平均認識率（動作群Ｅ）

よる認識率を上回る結果が得られた．

動作群Ｃについては，認識率が処理フレームレートに依存しない傾向を持つ動作（GC-A ／ GC-B）と，依存する傾向を持つ動作（GC-C ／ GC-D ／ GC-E ／ GC-F）が見出せる．統合評価値に基づく認識率（97.9(%)）が，認識プロセス PC-C（25[fps]）の平均認識率（98.6(%)）を除く単一プロセスのみによる認識率を上回る結果が得られた．

動作群Ｄについては，認識率が処理フレームレートに依存しない傾向を持つ動作（GA-D ／ GA-E ／ GB-A）と，依存する傾向を持つ動作（GA-A ／ GB-B ／ GB-D）が見出せる．統合評価値に基づく認識率（97.3(%)）が，認識プロセス PD-D（30[fps]）の平均認識率（99.4(%)）を除く単一プロセスのみによる認識率と同等あるいは上回る結果が得られた．

動作群Ｅについては，認識率が処理フレームレートに依存しない傾向を持つ動作（GB-F ／ GB-E ／ GC-B）と，依存する傾向を持つ動作（GA-C ／ GC-F ／ GC-D）が見出せる．統合評価値に基づく認識率（98.1(%)）が，認識プロセス PE-D（30[fps]）の平均認識率（98.9(%)）を除く単一プロセスのみによる認識率と同等あるいは上回る結果が得られた．

表 4: 各動作群の設定フレームレート（II）

動作群Ｄ	
認識プロセス名	処理フレームレート [fps]
PD-A	12
PD-B	16
PD-C	25
PD-D	30

動作群Ｅ	
認識プロセス名	処理フレームレート [fps]
PE-A	15
PE-B	20
PE-C	25
PE-D	30

以上をまとめると，認識率が処理フレームレートに依存する傾向を持つ動作が 56(%)（１８個中１０個）存在したことになる．さらに，Ｄ群とＥ群の約 58(%) の動作（１２動作中５動作（GA-A，GB-B，GB-D，GB-F，GB-E）を除く）については，別の組合せで実験したにも関わらず，処理フレームレート依存性の分類上の変化はなかった．この結果は，各動作および動作の組合わせに最適な処理フレームレートが存在する可能性を示唆している．また，すべての動作群において，統合評価値に基づく認識率が常時２位以内に入っている．これは，マルチフレームレート制御に基づく身振り認識により，特定の処理フレームレートでの観測に偏ることで生じるリスクを回避できていることを示している．

6 まとめ

本論文では，マルチフレームレート制御による身振りの実時間認識手法を提案した．提案手法の特長は，認識プロセスを並行動作させることにより，異なる処理フレームレートで対象動作を同時認識できる点にある．評価実験では，スポーツ動作およびストレッチ動作における典型的な動作合計１８種類を，５グループ各６動作に分けた際の平均認識率を求めた．その結果，各動作および動作の組合わせに最適な処理フレームレートが存在する可能性が見出された．

認識対象となる身振り動作の組合せを事前に特定できない場合，最適な処理フレームレートを事前に予測することは困難であり，本論文で提案したマル

図 9: 各動作群における身振りの軌跡画像

チフレームレート制御による身振り認識が特に有効である．ただし，本実験ではビデオレートを超える領域での評価を行っていないため，これら微小時間領域での評価実験を行い，提案手法の有効性をさらに検証することが今後の課題である．

参考文献

(1) Takashi Matsuyama：“Cooperative Distributed Vision : Dynamic Integration of Visual Perception, Camera Action, and Network Communication”, Proc. Fourth Int'l Workshop on Cooperative Distributed Vision, pp.1-25, Mar., 2001

(2) Hiroaki Kawashima, Takashi Matsuyama：“Integrated Event Recognition from Multiple Sources”, Proc. First Int'l Workshop on Man-Machine Symbiotic Systems, pp.243-258, Nov., 2002

(3) 桐島俊之，佐藤宏介，千原國宏：“プロトコル学習による身振りの実時間画像認識”，信学論（D–II）,Vol.J81-D-II,No.5, pp.785-794, May, 1998

(4) 桐島俊之，佐藤宏介，千原國宏：“多注視点の選択制御による身振りの実時間画像認識”，信学論（D–II）, Vol.J84-D-II, No.11, pp.2398-2407, Nov., 2001