

# 多フレームレート制御による身振りの実時間画像認識

桐島 俊之 佐藤 宏介\*1 千原 國宏\*2

Real-time Gesture Recognition by Multiple Frame Rate Control

Toshiyuki KIRISHIMA, Kosuke SATO and Kunihiro CHIHARA

In real-time gesture recognition systems, there is potentially a great need to observe and to recognize human actions at constant time intervals, for the improvements in recognition accuracy can be expected. Gestures are so rich in its variability, e.g. variation and speed. Some are easily understood at lower sampling rate, while others require higher sampling rate to be understood. For the accurate and robust recognition of gestures, the system needs to concurrently observe human gestures at differing processing frame rates and to integrate the recognition results at constant time intervals. In this paper, Multiple Frame Rate Control (MFRC) method is proposed and its effectiveness 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、コンピュータとの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、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。これらの要求に応えるには、非接触かつ実時間で人物の状態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では、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、従来の身振り認識手法の多くは、任意処理フレームレートでの安定化問題を考慮しておらず、コンピュータの処理性能に依存した不安定なサンプリング速度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<sup>1)</sup>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、従来手法<sup>1),2)</sup>は、単一のプロセスでの実装および動作を想定していたため、複数の処理フレームレートでの認識処理には対応していない。身振りには、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、本研究では、複数種類の身振りを複数の処理フレームレートで観測し、認識結果を得ることで、認識率の向上が得られるのかどうかを検証することとする。本論文では、多注視点身振り認識法<sup>2)</sup>と多注視点選択制御法を活用して、複数プロセスかつ複数フレームレートでの身振り認識方式、すなわち、多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する。

## 2 関連研究

従来手法の多くは、単一プロセスでの動作を想定した身振り認識手法である。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<sup>3)</sup>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<sup>4),6)</sup>では、複数カメラからの入力画像を複数PC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<sup>5),6),7)</sup>。

\*1大阪大学大学院 基礎工学研究科

\*2奈良先端科学技術大学院大学 情報科学研究科

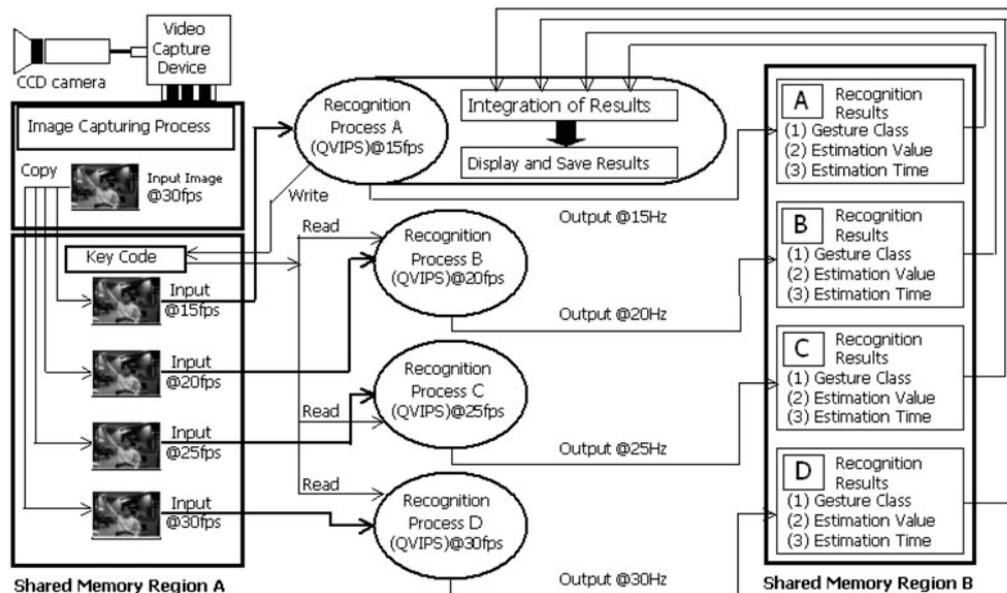


図1 多フレームレート制御による身振り認識

多くの従来システムに共通することは、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点であり、複数プロセス、かつ、複数の処理フレームレートによる観測により、人物身振りの認識性能・ロバスト性がどのような影響を受けるかについては、十分な検討がなされていないのが現状である。特に、人物動作の検出やセグメンテーションの際、時間差分操作を行っている手法については、処理フレームレートおよび対象動作により認識性能がどのような影響を受けるかについて評価する必要がある。以降、多フレームレート制御手法を提案し、評価実験により提案手法の有効性を実証する。

### 3 多フレームレート制御

#### 3.1 提案手法の概要

本稿で提案する多フレームレート制御に基づく身振り認識とは、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・認識させる手法である。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。筆者らは、これまでに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提案している。これら手法に基づ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ことにより、より高いロバスト性を備えた認識システムの実現が期待できる。

#### 3.1.1 多フレームレート制御の枠組み

図1に多フレームレート制御による身振り認識の枠組みを示す。本稿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、処理フレームレートは最大30[fps]に制限される。認識プロセス数自体には制限はないが、4通り程度の処理フレームレートで身振りを観測すれば、多フレームレート制御の有効性は評価できると考え、表1に示す4種類の認識プロセスを使用することにする。

表1 認識プロセスの設定

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

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。

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

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**

多フレームレート制御による身振り認識の核心となる認識手法として、本研究で提案している多注視点身振り認識法(QVIPS)を適用する。

多注視点身振り認識法における基本的発想は、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。

**3.3 多注視点選択制御法**

画像パターン学習に基づく身振り認識手法においては、認識対象身振りの種類・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も著しく低下する問題が生じている。認識対象身振りの種類・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ことを目的として、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むことで上述の問題に対応し

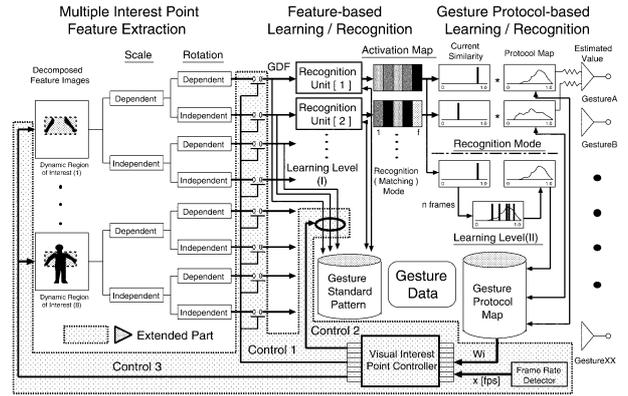


図2 多注視点身振り認識法と多注視点選択制御法による身振り認識の枠組み

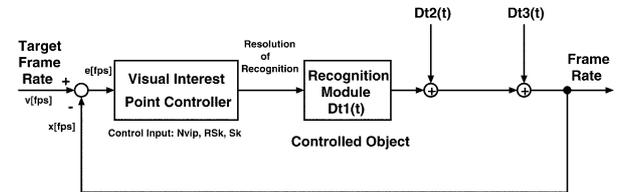


図3 フィードバック制御のブロック図

ている。具体的には、図2に示すように、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) の3種類の選択制御系を多注視点身振り認識法に組み込んでいる。

**3.3.1 多注視点選択制御の枠組み**

多注視点選択制御法では、図3に示すような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ことにより、システム全体の応答の安定化を図っている。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  $D_{i1}(t)$ 、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  $D_{i2}(t)$ 、更に応用システム自体の負荷  $D_{i3}(t)$  の3種類であるが、本論文では認識モジュールの負荷  $D_{i1}(t)$  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ため、 $D_{i3}(t)$ 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。各種パラメータの設定は、初期有効注視点数 $N_{vip}=32$ 、初期パターン照合間隔 $RS_k=1$ 、初期パターン走査間隔 $S_k=1$ である。なお、本実験はすべてマルチユーザーモード、X-Window上で行った。

### 4.2 多フレームレート制御による認識実験

本節では、図4および表2に示す3種類のストレッチ動作A~Cと3種類のスポーツ動作D~Fについて提案システムにより認識実験を行った結果を示す。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。評価実験は以降の手順に従って行う。

**【手順1】** 認識プロセスA~Dを立ち上げ、認識対象動作A~Fについてプロトコル学習 (標準動作1回、類似動作1回) させる。

**【手順2】** 処理フレームレートを認識プロセスAについては15[fps]、認識プロセスBについては20[fps]、認識プロセスCについては25[fps]、認識プロセスDについては30[fps]に設定し、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】** 認識対象動作A~F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。

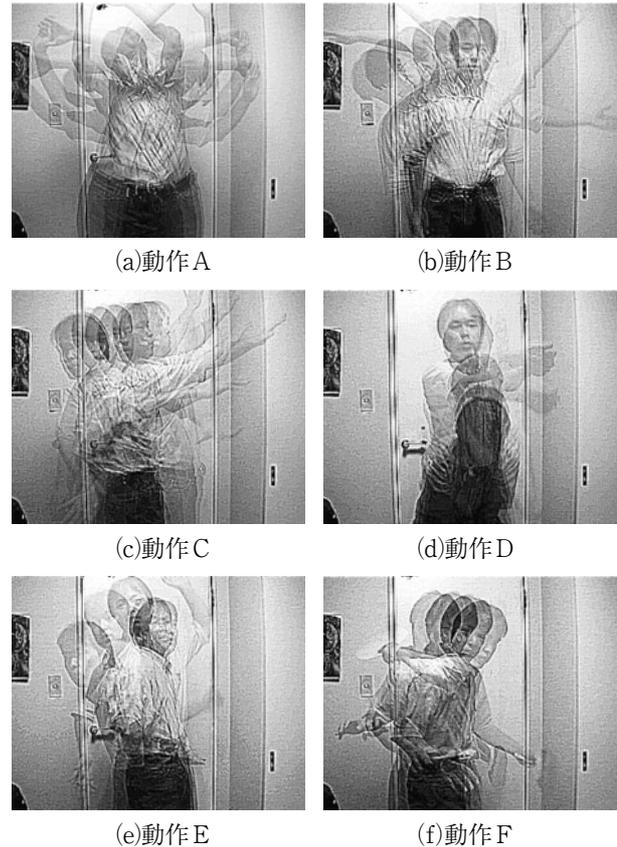


図4 認識対象動作の軌跡画像

表2 評価実験で対象となる動作

動作記号	動作名 (内容)
A	ストレッチ1 (体両側面伸ばし)
B	ストレッチ2 (体側面伸ばし)
C	ストレッチ3 (体横ねじり)
D	スポーツ1 (バレレーシーブ)
E	スポーツ2 (バレレーアタック)
F	スポーツ3 (テニスフォアハンド)

以上の実験手順により得られた実験結果を表3に示す。また、各動作における平均認識率を図5に示し、各認識プロセスにおける平均認識率を図6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。なお、図・表に示した認識率は、上記認識率を各動作20個のテストサンプルについて算出し、それを平均したものである。

表3 各動作の各認識プロセスにおける認識率 (%)

各動作の認識率	統合認識率	プロセスA	プロセスB	プロセスC	プロセスD	平均認識率
動作A	100.0	97.5	100.0	100.0	99.5	99.3
動作B	100.0	100.0	100.0	100.0	100.0	100.0
動作C	100.0	100.0	67.1	100.0	100.0	91.8
動作D	100.0	100.0	100.0	100.0	100.0	100.0
動作E	98.5	100.0	100.0	98.1	75.8	93.5
動作F	95.6	77.8	83.9	100.0	99.2	90.2
平均認識率	99.0	95.9	91.8	99.7	95.8	95.8

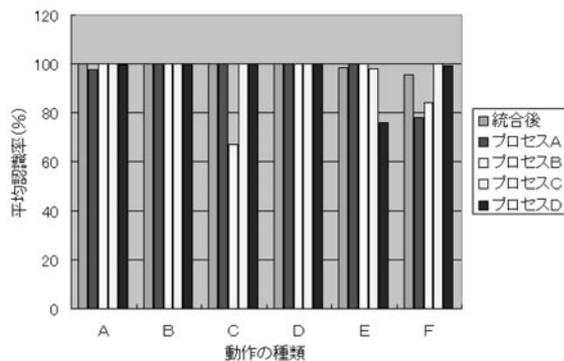


図5 各動作における平均認識率

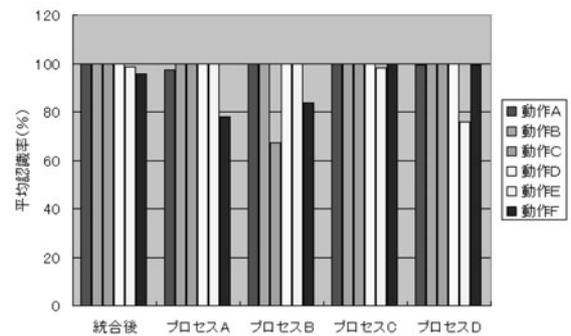


図6 各認識プロセスにおける平均認識率

## 5 考察・検討

図5より、動作A・B・Dの認識率については処理フレームレートに依存しない傾向が見出せる一方で、動作C・E・Fについては依存する傾向が見出せる。図6より、今回の実験では認識プロセスCの処理フレームレートである25[fps]付近で平均認識率が最大(99.7%)となった。また、統合認識率が最低でも95.6%(動作F)となり、今回選定した動作については良好な認識率が得られた。

統合認識率の全体平均は99.0%であり、認識プロセスCの平均認識率(99.7%)とほぼ並んでいる。なお、認識プロセスC以外については、統合認識率が各プロセスの平均認識率を上回っている。これは、複数の処理フレームレートにおける認識結果を統合することで、最大認識率に準じた認識率が得られたことを示している。

動作Eについては、処理フレームレートが低くなるにつれて平均認識率が向上している反面、動作Fについては、処理フレームレートが低くなるにつれて平均認識率が低下する現象が見られる。また、動作Cについては、認識プロセスBにおいて認識率が67.1%と極端に低下している。これらの結果は、認識対象動作によって最適な処理フレームレートが異なることを示唆していると考えられる。さらに、認識プロセスCにおいては、認識プ

ロセスDよりも処理フレームレートが低いにも関わらず平均認識率が最大となったことは、単純に高い処理フレームレートにすれば低速な動作から高速な動作まで幅広く認識できることにはならないことを示唆している。

本論文の目的は、低速な動作から高速な動作までを単一の処理フレームレートで認識することが果たして最善なのかどうかを明らかにすることにあった。本評価実験では、単一プロセス・単一処理フレームレートでは得られない平均認識率が、複数プロセスの認識結果の統合により得られており、このことは、単一プロセス・単一処理フレームレートによる身振り認識が必ずしも最善の選択ではないことを示唆している。認識対象動作を事前に特定できない条件下では、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、本論文で提案した多フレームレート制御が有効と思われる。

## 6 まとめ

本論文では、多注視点身振り認識法と多注視点選択制御法に基づき、複数の処理フレームレートでの認識処理を実現する、多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とする身振り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について、4通りの処理フレームレート下での認識率を求めた。その結果、異なる処理フレームレート下での認識結果を統合処理することにより、単一の処理フレームレート下での結果を上回る認識率が得られることが分かった。これは、多フレームレート制御により、認識対象動作群に関する時間領域での情報量が増加したため、認識率が向上したものと考えられる。また、認識対象動作の性質に応じてある程度適切な処理フレームレートが存在することが分かった。対象動作が特定できない場合、事前に最適な処理フレームレートを予測することは困難であり、本論文で提案した多フレームレート制御による身振り認識が有効であると考えられる。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効果を確認することが今後の課題である。

なお、本実験では、単一PC上での評価実験にとどめたが、マルチプロセッサベースのPCや、ネットワーク化されたPC群上に提案手法を拡張し実装することによりその有効性をさらに引き出すことも今後の課題の一つである。

## 参考文献

- (1) 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, Nov., 2001, pp.2398-2407
- (2) 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, May, 1998, pp.785-794
- (3) 梅澤猛, 佐藤一郎, 安西祐一郎: “モバイルエージェントを用いたセンサネットワーク向けフレームワーク”, 情処論, Vol.44, No.3, Mar., 2003, pp.779-788
- (4) Takashi Matsuyama: “Cooperative Distributed Vision: Dynamic Integration of Visual Perception, Camera Action, and Network Communication”, Proc. Fourth Int'l Workshop on Cooperative Distributed Vision, Mar., 2001, pp.1-25
- (5) Hiroaki Kawashima, Takashi Matsuyama: “Integrated Event Recognition from Multiple Sources”, Proc. First Int'l Workshop on Man-Machine Symbiotic Systems, Nov., 2002, pp.243-258
- (6) Rin-ichiro Taniguchi, Daisaku Arita, Satoshi Yonemoto: “Real-time Multi-view Image Analysis on PC-cluster and Its Application”, Proc. Fourth Int'l Workshop on Cooperative Distributed Vision, Mar., 2001, pp.431-453
- (7) 齋藤真希子, 佐藤洋一, 小池秀樹: “Perceptual Glove: 多視点画像に基づく手形状・姿勢の実時間入力とその応用”, 情処論, Vol.43, No.1, Jan., 2002, pp.185-194