

CHLAC 特徴とグリッドコンピューティングを併用した実時間動作認識

白井達也, 斎藤秀雄, 吉本晴洋, 鴨志田良和, 白木孝義, 石黒勝彦, 深野亮, 大武美保子, 佐藤知正, 田浦健次郎 (東京大学), 大津展之 (東京大学, 産業技術総合研究所)

I. はじめに

我々は、立体高次局所自己相関(以下, CHLAC)特徴を用いた実時間動作認識システムを構築した。CHLAC 特徴は、動作認識において計算コストの低さと高い識別能力を両立するが、特徴数の増加に伴い実時間処理が困難になるという問題があった。今回、我々はグリッド計算機群を用いた並列化を行い、上記の課題を克服した。我々のシステムを、格闘動作という素早い全身動作に対して適用したところ、カメラ1台で実時間処理と高い認識性能の両立に成功した。

II. システムの概要

我々が構築した動作認識システムは、カメラからの動画像入力を実時間処理し、シーン中の人物の動作を認識する。

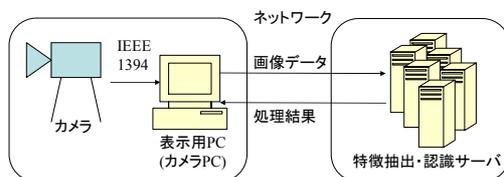


図1: システム構成

システムは図1に示される構成である。まず動画撮影用のカメラで撮影された画像は表示用PCで前処理され、グリッド計算機群へ送信される。グリッド計算機群上では画像フレーム毎に並列にCHLAC特徴抽出と認識処理を行い、認識結果を表示用PCに返す。認識結果は、格闘ゲームに見立てて、ジャブ、ストレート、ローキック、ミドルキックの4動作を行うとそれに合わせて攻撃したような演出の画像を表示する(図2)。特徴抽出・

認識処理を行うグリッド計算機群には、表示用PCとは別ネットワークにあるクラスタ内の計算機群を用いる。人物の動作を実時間的に画面表示に反映させるため、1秒あたり30フレームの画像認識を行い、画像フレーム毎の処理開始から終了までの時間は100ms以内で実現する。



図2: 格闘動作の認識結果への演出の付加

III. システムの実装

CHLAC 特徴を用いた動作認識

CHLAC 特徴[1]は、高次局所自己相関特徴[2]を時間軸方向へ拡張したものである。CHLAC 特徴 \mathbf{x} は以下の式で定義される:

$$\mathbf{x} = (x_1, x_2, \dots, x_N)$$
$$x_i = \int_A f(\mathbf{r}) f(\mathbf{r} + \mathbf{a}_1^i) \dots f(\mathbf{r} + \mathbf{a}_D^i) d\mathbf{r}$$

ここで、 A は Z 枚のフレームから構成される3次元のピクセル集合であり、 \mathbf{r} は A 中の1ピクセル(位置ベクトル)、 \mathbf{a} は \mathbf{r} からの変位ベクトル、 $f(\mathbf{r})$ は位置 \mathbf{r} の輝度値である。CHLAC 特徴は時空間中の近傍点の相関を抽出する特徴量であることから、人間の動作のように時空間中を連続的に変化する動きの性質を捉える上で好ましい特徴となっている。また、積分特徴であることから、対象が時空間方向に平行移動してもまったく同じ特徴量が得られる点もCHLAC特徴の利点である。

このCHLAC特徴を用いて、我々は高精度の動作認識を実現するアルゴリズムを開発した。図3にその流れを示す。まず、動きのある部分を検出するため、カメラから取得した画像から、隣接フレーム間の差分画像を計算する。次に、ノイズ除去と特徴抽出処理の計算コスト削減のため、差分画像に対して二値化を行う。認識プログラムは、

A CHLAC Method-Based Real-time Motion Recognition System for Grid Computers

Tatsuya Shirai, Hideo Saito, Haruhiro Yoshimoto, Yoshikazu Kamoshida, Takayoshi Shiraki, Katsuhiko Ishiguro, Ryo Fukano, Mihoko Otake, Tomomasa Sato and Kenjiro Taura (Univ. of Tokyo), Nobuyuki Otsu (Univ. of Tokyo, AIST)

本研究は、「100時間ワークショップ@情報理工COE」の支援を受けた。

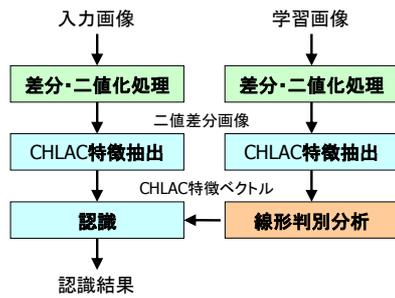


図 3：認識アルゴリズム

この二値差分画像から CHLAC 特徴を抽出する。

そして、多変量解析手法の一種である線形判別分析によって、各動作を識別する上で最適な判別特徴を計算する。学習時には、判別特徴空間に学習サンプルを射影し、動作の種類毎に中心を求める。認識時には、カメラ画像を判別特徴空間に射影し、最も中心点の距離が近い動作を求める。

また、認識率をさらに向上させるために、様々なパラメタ（画像の解像度、変位ベクトルの変位幅）で CHLAC 特徴を抽出する。パラメタ毎に認識を行い、これらの認識結果の多数決を全体の認識結果とする。

並列化の方法

パラメタ毎の認識処理のうち、特徴抽出が主に計算力を必要とする処理である。特徴抽出は隣接する Z フレームの二値差分画像に対して行うが、この処理はフレーム毎に並列化できる。 f 番目の画像を受け取ったプロセスは、1 フレーム毎の特徴抽出には隣接する 3 フレームの二値差分画像が必要になるため、 $f-1, f-2$ 番目の二値差分画像を持つプロセスと通信して特徴抽出を行う(図 4)。そして、隣接する Z フレーム分の特徴ベクトルを集めて和を計算し、認識を行う。通信にはクラスタ内では MPICH[3]を、表示用 PC とクラスタ間はファイヤーウォールのため片側からしか接続できない場合に対応するため、ソケット通信を用いる。

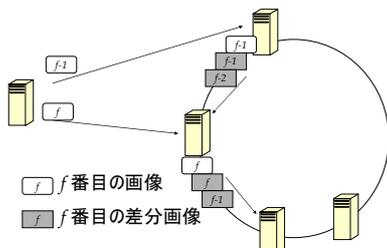


図 4：クラスタ内のデータの流れ

IV. 実験結果

実験には、表示用 PC, クラスタ計算機ともに、東京大学本郷キャンパスに設置した Xeon

2.8GHz (dual), 2GB メモリの計算機を合計 22 台用いた(表示用 PC に 1 台, パラメタ毎に 5 台, パラメタ数 4(表 1), 1 台は表示用 PC との通信用)。クラスタ内の通信帯域幅は 1Gbps である。CHLAC 特徴抽出に用いるフレーム数(Z)は 25 とした。99,911 フレームの画像データについて、18-fold cross validation を実行したところ、認識率は 79.9%になった。また、2,677 フレームの画像データについて、画像フレーム毎の処理開始から終了までの時間を計測したところ、表示用 PC をクラスタ計算機と別のネットワーク(東京大学柏キャンパス, 通信帯域幅は最大 100Mbps)に設置した場合でも最悪で 64ms であり、100ms 以内に完了することを確認した(図 5)。

V. まとめと今後の課題

我々が構築した実時間動作認識システムは、人間の格闘動作という、全身を用いた素早い動作を対象とした、珍しい試みである。実験はクラスタ計算機でのみ行われたが、今後は、今回達成された実時間性を保持しつつ、広域分散環境におけるファイヤーウォール・通信遅延・帯域幅など多様な制約の下で動作するような、さらに高精度の認識アルゴリズムを開発することが求められる。

参考文献

- [1] TKobayashi and N.Otsu, "Action and Simultaneous Multiple-Person Identification Using Cubic Higher Order Local Auto-Correlation", Int. Conf. on Pattern Recognition, Vol. 4, pp.741-744, 2004.
- [2] N.Otsu and TKurita, "A new scheme for practical, flexible and intelligent vision systems", Proc. IAPR Workshop on Computer Vision, pp.431-435, 1988.
- [3] <http://www-unix.mcs.anl.gov/mpi/mpich>

解像度(WxH)	変位ベクトルの変位幅(x,y)
180 x 120	1,1
120 x 80	3,3
120 x 80	1,3
120 x 80	1,1

表 1：認識で使したパラメタ

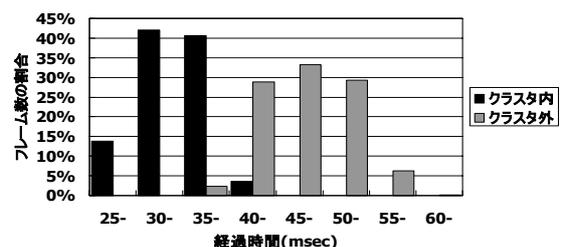


図 5：処理時間の分布