

Dual CNN オートエンコーダによる 体操競技の技認識

阪田雅人^{1,a)} 本田崇¹ 矢吹彰彦¹ 梶井昇一¹ 佐々木和雄²

概要：富士通は国際体操連盟と共同で、体操選手の演技における三次元骨格座標や実施された技を認識する機能を備えた体操採点支援システムの開発を行っている。後者の技認識においては、従来はルールベースにより判定ロジックや閾値を手手で定義する方式を採用していたが、平均台演技では認識正答率が60%台に留まるという課題があった。また、平均台などの種目によっては技と技の間の「つなぎ動作」から技と技の組み合わせ加点を決定する必要がある。我々は Dual CNN オートエンコーダを用いて、採点に関わる技の判定と、技と技の組み合わせの成否判定を並行して実施する手法を開発し、技認識 96.4%、組み合わせ加点認識 95.2%の正答率を達成した。

Element Recognition in Gymnastics Using Dual CNN Autoencoder

MASATO SAKATA^{1,a)} TAKASHI HONDA¹ AKIHIKO YABUKI¹
SHOICHI MASUI¹ KAZUO SASAKI²

1. はじめに

富士通は、国際体操連盟・日本体操協会との連携により、体操競技における正確かつ公平な採点の実現を目指して、採点支援技術の開発に取り組んでいる[1]。体操採点支援システムの模式図を図1に示す。反射マーカを使用せずにアスリートの3次元の動きを数値化するため、3D レーザセンサーを使用して人体表面の凹凸情報を表す深度画像を取得する。この深度画像から学習型骨格認識と幾何モデルフィッティングにより主要関節の三次元座標を高精度に求め、得られた結果から肘や膝、背骨等のまがり角度を正確に決定する。さらに、得られた角度の時系列変化を入力と

し、技のデータベース(辞書)を参照することで技認識を行う。我々は、3D レーザセンサーによる深度画像生成とその後の骨格認識を合わせて3D センシングと呼んでいる。

3D センシング・技認識技術によって得られた三次元骨格座標と技認識結果から採点支援アプリを駆動する。図1に示すように、採点支援アプリには、体操選手の演技におけるフレームごとの関節角度を詳細に確認できるマルチアングルビューと、実施された技ごとに技認識により得られる技名等を示す技認識ビューで構成されている。本システムは、2019年シュツットガルト世界選手権から正式採用されており、現在までに、つり輪、あん馬、跳馬(男子・女子)、平均台5種目の対応技術を確認している。

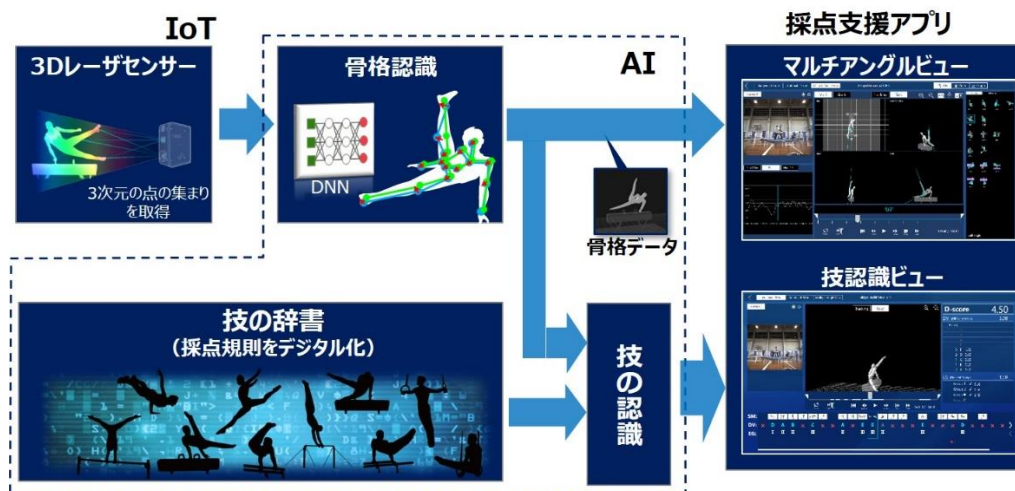


図1 3D センシング・技認識技術による体操採点支援の概要

¹ (株)富士通研究所
Fujitsu Laboratories Ltd., <http://fujitsu.com/jp/labs/>
a) sakata.masato@fujitsu.com
² 富士通(株)

Fujitsu Limited., <http://fujitsu.com/jp/>

本論文では、技認識技術に関し、審判間の判定精度確保が最も難しい平均台演技の技認識に関し、Dual CNN オートエンコーダの採用により、採点に関わる技の判定と、技と技の組み合わせの成否判定を同時に実施することで、正答率の大幅改善を達成したので、詳細を解説する。

2. 体操採点と技認識技術

2.1 体操採点の概要

体操競技は、陸上競技や水泳競技のように時間や距離を競う競技ではなく、演技（アスリートの動き）の複雑さを採点してその高さを競う採点競技である。男子 6 種目（ゆか、あん馬、つり輪、跳馬、平行棒、鉄棒）、女子 4 種目（ゆか、跳馬、平均台、段違い平行棒）のそれぞれの演技に対して、技の難度を示す D (Difficulty) スコア、演技の出来映えを反映する E (Execution) スコア、および、演技領域からの逸脱（線審が判定）や時間超過（計時審判が判定）などによる減点の合計によって採点される。

女子種目を例にとると、D スコアは、採点規則に記載された技の難度価値点（例えば、A 難度であれば 0.1、B 難度は 0.2 というように、難度のランクが 1 レベル上がれば得点も 0.1 ずつ増加する）に従って演技された 8 技の価値点の合計に加え、高難度の技の連続実施による組合せ加点と、実施された技の構成要求に関わる価値点の合計により求められる[2]。構成要求の価値点は、演技中に要求された項目を 1 つ達成するごとに 0.5 点加算され、合計 2.0 点まで加算される。このように、技認識の正確さは D スコアに大きな影響を与える。

E スコアは、10 点を満点として、美しさ、実施、技術、

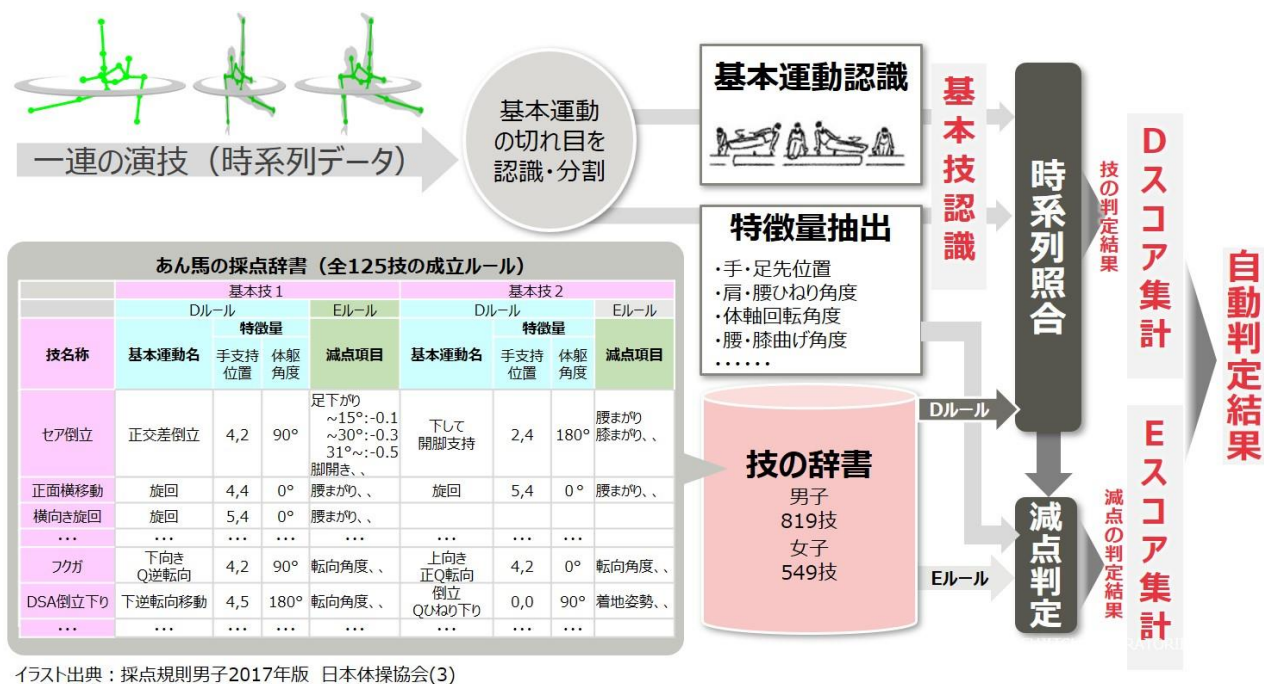
構成上の欠点による減点を合計したものとなる。減点の対象となるのは、着地における完璧な姿勢からの逸脱、演技における膝・腕・体等のまがりや脚の開き、静止技における完璧な姿勢からの逸脱などであり、逸脱の量やまがり角度により、減点の量が決まる。本論文が対象とするのは、D スコアに関わる技認識技術である。

国際審判資格を持った審判員は高い技能を備えているが、採点の基準を記載した採点規則が、人が審判することを前提とするために、後述する組み合わせ加点のような、厳密に定義できずに、あいまいさを持っていることや、採点規則に減点要素として記載されている関節の理想からのまがり角度を人が目視で正確に認識することが不可能であることから、採点は経験に頼らざるをえず、正確性・公平性の担保には本質的な困難が存在し、採点支援技術の実用化に高い期待が寄せられている。

2.2 技認識技術

採点支援向け技認識技術の概要を、あん馬を例にして図 2 を用いて説明する。新しい技が追加された際に最小限のデータ追加で対応することが要求されるため、まず一つ一つの演技を構成する基本運動を定義し、基本運動に関連した特徴量(図 2 の例では手の支持位置と体躯とあん馬のなす角度)に基づいて基本技を認識し、連続する基本技を技の辞書と時系列照合して採点の対象となる技を認識する構成となっている。図 2 の左下の例では、最初的基本運動として「正交差倒立」、続いて「下ろして開脚支持」が認識された結果、最初に演技された技が「セア倒立」とであると判定している。

基本運動の数は、男子 6 種目 819 技に対して 475、女子 4 種目 549 技に対して 318 と、技総数と比較して削減できて



イラスト出典：採点規則男子2017年版 日本体操協会(3)

図 2 採点支援向け技認識技術の概要

おり、基本運動認識による共通要素の抽出によって技認識精度の向上が期待できる方式となっている。一旦採点対象となる技として認識できれば、採点規則から技の難易度価値点や構成要求の価値点は一意に決まり、種目ごとに異なる組合せ加点を決定できれば、D スコアを自動的に生成できる。

技認識において最も重要な技術要素は、3D 骨格座標の時系列データから、基本運動間の切れ目を認識し、分割された時系列データに対して特徴量と基本運動を決定することにある。基本運動間の切れ目や特徴量、基本運動、特徴量を決定するアルゴリズムには、ルールベース(判定ロジックと閾値により決定)と機械学習(Deep Learning による学習・識別の実施)の2つの手法があるが、我々は種目によって使い分けている。次章では平均台における技認識の特徴と技術課題をまとめる。

3. 平均台技認識の特徴と技術課題

3.1 平均台技認識の特徴と課題

平均台技認識において、審判間の判定精度の確保を最も難しくしている要因は、組み合わせ加点の判定にある。組み合わせ加点は、一定の難度以上の技同士が連続して演技された場合に加点が与えられる。組み合わせの成否は、技と技の間の「つなぎ動作」を元に判定される。このつなぎ動作は、ジャンプ、ターンといったダンス系を含んだ技と技の間で発生し、止まり、余分なステップ、足が台に触れる、バランスを失う、余分な腕の振りのいずれも無い場合に組み合わせが承認されると採点規則には記載されている

が、目視による数値化ができないため審判毎の判断が一致しない場合が発生する。

図 2 に示す技認識術をルールベースで行うと一部認識が出来ない技があり、正答率が低くなっていた。また、平均台では演技の構成上技と技の間の「つなぎ動作」が半分以上を占めるので、技と「つなぎ動作」に対して、ルールベースの判定ロジックと閾値を決定することが容易ではなく、組み合わせ加点の判定精度が実用レベルに到達しなかった。国際体操連盟の技術委員会と協議しながら採点規則から判定ロジックや閾値を決定した後、実際に取得したデータで評価した結果を審判員の意見を元に判定ロジックや閾値を再調整するサイクルを繰り返す必要があるためである。このような課題から、平均台では機械学習による技認識システムに導入することとした。

3.2 関連技術と課題

機械学習を用いた技術としては、3D 骨格座標の時系列データを RNN に入力する運動認識手法が提案されている[3]。しかし前述の通り平均台では技と「つなぎ動作」を含む多様な動きを含むため、技とつなぎ動作の境界があいまいとなり、単純な時系列データの学習・認識では正答率の向上が期待できない。

4. 提案手法

本論文では、入力時系列データから Dual 構成のオートエンコーダ(AE)を用いて、技とつなぎ動作の境界にマージンを持たせて両動作領域を抽出することで、技判定と、技と技の組み合わせ成否判定を高精度に実施する手法を提案す

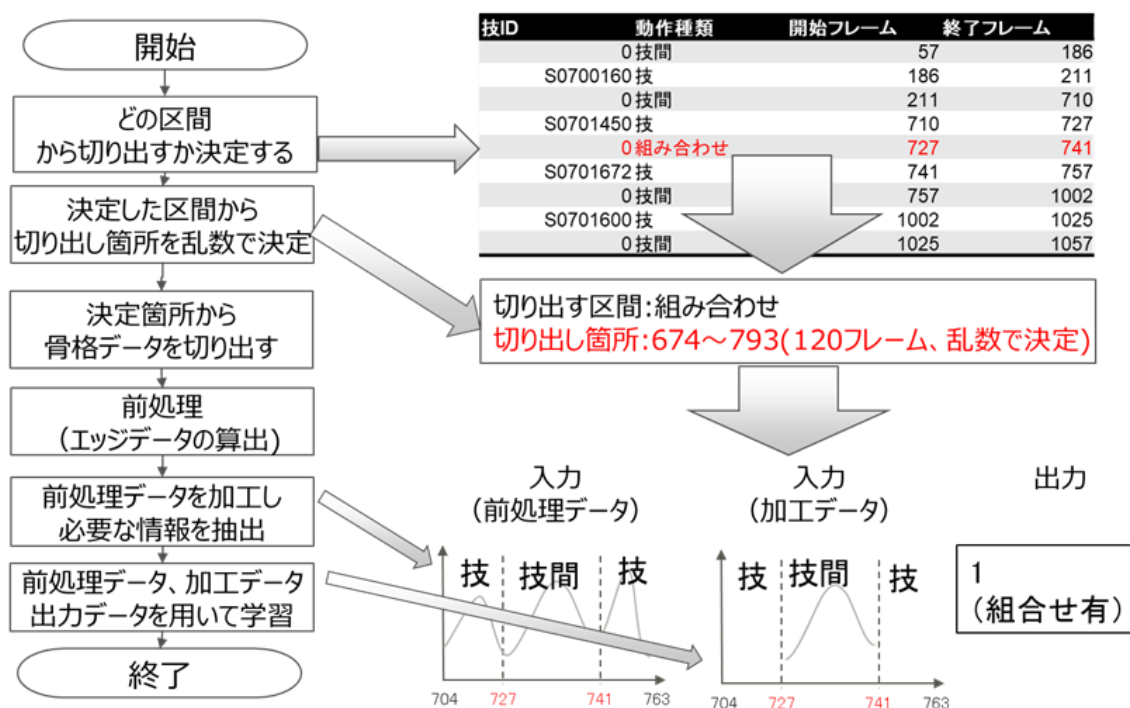


図 3 技認識の学習の処理フロー (組み合わせ)

る。本章では、学習に用いるデータや、Dual CNN オートエンコーダについて述べる。

4.1 学習に用いるデータ

採点支援システムでは、3D センシングにより、予め決定した器具中心からの相対座標で、選手の三次元関節座標を求める。座標の位置は、頭、首、腕、手、胴、腰、膝、脚の合計 18 関節を元のデータとして用いる。ただし座標値をそのまま学習しても、選手は器具上の様々な場所で演技を行うため、値が分散して学習が収束しない。そこで関節座標をそのまま用いるのではなく、関節同士を結んだ骨格の向きデータ（エッジデータ）を用いている。

学習用入力データの作成方法を図 3 に示す。まずは選手から取得した時系列の演技データに対して、認識する動作対象を一定時間分含んだデータの切り取りを行う。平均台では、1 つの技あたりの実施時間や、つなぎ動作の時間がおおよそ 4 秒以内に収まっているため、4 秒分のデータを基本単位として抽出する。抽出されたデータは、認識対象が中央付近になっているものの丁度中央ではなく、学習毎に確率密度関数に基づいた乱数を用いて、抽出位置を中央からずらすようにする(図 4)。認識対象が技や組み合わせの場合は正規分布を採用し、認識対象が抽出データの中央付近になる際に重点的に学習する。これは識別の際に、1 フレームずつずらして連続入力されたデータに対しても、正しく認識結果を出力するためである。反対に認識対象が技や組み合わせに無関係の「つなぎ動作」の場合は、ベータ分布を採用する。ベータ分布はベータ関数 $B(\alpha, \beta)$ を用いて、 α と β の 2 つのパラメータを指定して表される。ベータ分布の確率密度関数は以下の通りである。

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta}}{B(\alpha, \beta)}$$

α と β に対して、1 未満の値かつ同じ値を指定することで、確率密度は凹型の関数となる。この関数で、技の動作との境界値付近を重点的に学習する。このように乱数を用いた切り取ったデータに対して、さらにデータ拡張を行う。元のデータに対して頭や脚、腕、膝の位置に対する微小な変更、前後および左右の鏡像反転を行うことで、疑似的にデ

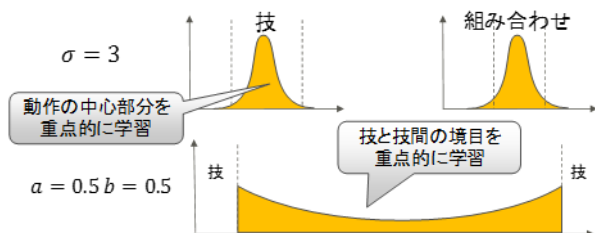


図 4 データ抽出に使用される確率分布
正規分布（上）とベータ分布（下）

ータを拡張している。入力データは、この拡張データから算出したエッジデータを用いる。

4.2 Dual CNN オートエンコーダ

提案手法を実現する Dual CNN オートエンコーダの構成と入力・内部時系列データを図 5 に示す。先述の 3D 骨格の時系列エッジデータを入力として、技動作抽出 AE とつなぎ動作 AE の 2 つのネットワークを通じて判定結果を出力する。技動作抽出 AE では、技動作領域のみを抽出した時系列データを生成し、後段の CNN により技判定結果を出力する。これに並行して、つなぎ動作抽出 AE では、つなぎ動作領域のみの時系列データを生成し、後段の CNN によるつながり判定を経て、前述した技判定結果と合わせて技同士の組み合わせ成否判定結果を出力する。学習時には、3D 骨格のエッジデータと、技とつなぎ動作の境界データに加え、審判から提供された技と組み合わせ両判定の正解値を使用する。本研究では Dual 構成のオートエンコーダを採用し、技とつなぎ動作の境界それぞれに対して、判定箇所が含まれるように、最適な境界を持たせて学習する。この結果として、境界のあいまいさを排除し、正答率の向上を目論んでいる。

5. 評価結果

平均台において、システムで取得した国際大会や選手の練習データ（技数 5,095、組み合わせ数 1444）を前章の手法で学習した。学習済みモデルを用いて、国際大会出場選手ののべ 243 人（技数 3,095、組み合わせ数 890）の平均台データに対して、技と組み合わせ成否判定を実施した。審判資格員が判定した結果と比較して、判定結果の一致している割合を正答率として評価した。その結果を表 1 に示す。

表 1 技認識と組み合わせ判定の正答率の比較

	技認識	組み合わせ判定
ルールベースによる手法	68.3%	67.9%
本手法	97.0%	96.3%

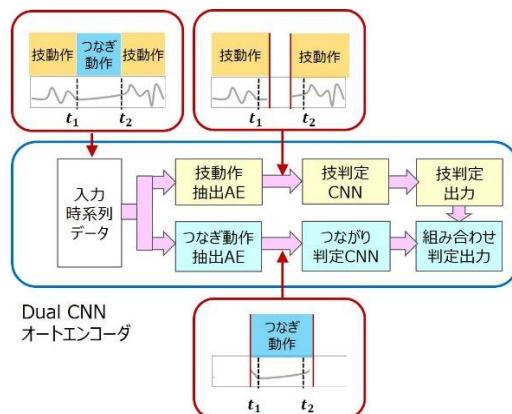


図 5 Dual CNN オートエンコーダと入力・内部データ

従来のルールベースにより基本運動から定義した技認識の手法では、技認識の正答率が 68.3%，組み合わせ判定の正答率が 67.9%であったが，本手法を適用した結果，技認識の正答率が 97.0%，組み合わせ判定の正答率が 96.3%まで向上した．技動作とつなぎ動作を両方抽出することで，組み合わせ判定に加え，技認識の精度も向上した．従来手法では，頻度の高い A 難度のターン（実施数 139）や B 難度のバク転（実施数 257）で未検出が発生していたため，正答率が低くなっていた．本手法を適用することで，ターンは取得数が 98 から 137，バク転は 200 から 254 となり，これらの技の未取得の割合が減少したことで正答率が向上した．従来未取得であった技の組み合わせも正しく取得することが出来，組み合わせの正答率も向上した．

6. おわりに

本論文では，Dual CNN オートエンコーダを用いた体操競技の技認識手法を提案した．本手法は，体操競技の他種目では，平均台と組み合わせ加点の形態に近い床運動に適用可能と考えており，今後は他の採点競技への適用可能性を検討していきたい．

参考文献

- [1] 榊井 ほか, 3D センシング・技認識による体操採点支援, 通信学誌, Vol. 103, No.1, pp.5-14 (2020)
- [2] 公式財団法人 日本体操協会, 採点規則 体操女子 2017 年度版
- [3] Y. Du et al., Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition, CVPR, pp.1110-1118 (2015)