

ものの把持に基づくコンテンツ検索支援

山本 景子¹² 南部 俊輔¹ 佐藤 宏介¹

1) 大阪大学大学院基礎工学研究科 2) 日本学術振興会 特別研究員

A Contents Retrieval Supporting Technique Based on Grasping Pressure

Keiko Yamamoto¹² Shunsuke Nambu¹ Kosuke Sato¹

1) Graduate School of Engineering Science, Osaka University

2) Japan Society for the Promotion of Science (JSPS)

概要

キャプチャデバイスの進歩により、長時間のコンテンツ収録が可能になったことから、限られた動作だけを記録するのではなく、動作の一部始終をキャプチャする機会が増えている。それに伴い、コンテンツ検索、すなわち長時間に亘るキャプチャデータの中から、ユーザが必要とするシーンを選び出すための技術が必要とされている。本論文ではこの問題に対して、人が中心となる動作風景を撮影した映像を対象とし、動作時に映像と共に対象者の把持圧の変化を記録し、把持圧情報から求められる動作状態を提示することで、内容検索を支援する手法を提案する。筆者らは把持圧から、把持状態、把持対象、把持して行う動作についての情報が得られると考える。そこで、把持圧計測装置を構築し、動作映像と同時に取得した把持圧を分析することで、対象者の状態を判別する手法を提案し、実験により判別精度の検証を行った。ユーザが動作毎に所望のシーンを効率的に閲覧する支援となるコンテンツ検索支援システムを構築し、提案手法の動作確認を行った。

Abstract

This paper presents a video content retrieval system of operators' performance areas for camera monitoring and life-logging on a plant operator, fine artist etc. As the system focuses on grasping actions and tool handling operations, it records not only audio-video data, but also grasping pressures by special finger pressure sensors. A sensing module consists of tiny film sensors and a bluetooth wireless data transmitter, in order not to interfere workers' free actions. Unlike other video retrieval system with voice or image recognition, tag text analysis, the proposed system searches the desired video scene based on grasping pressure data, so that it distinguishes whether the worker operates a tool physically while the worker's outlook shows holding. As its content browser places only interested video clips on a time line, which the worker handles the tool with a certain force, the user easily knows actual actions among a longer and monotonous raw video sequence. By applying a supervised learning process with known actions in NN-method, the system extracts the same actions from a video sequence when the user tells the representative action as a search key in the content browser.

1 はじめに

マルチメディア技術の発達により、各種動作の一部始終を素材として収録する機会が増えている。映像入力の場合では、記録デバイスの技術進歩により長時間撮影が容易となり、録画制御などを行うカメラマンを伴わずに撮影可能となったことも起因しているだろう。しかし、収録時に適切にカット点データを入れずに作成されたコンテンツは、閲覧・編集時に冗長なコンテンツ素材全編の中から目的のシーンを発見するための時間と労力がかかるという問題がある。

コンテンツ応用の分野においては動画やモーションキャプチャによる各種素材の組み合わせによる表現の需要が高い。例えば、スポーツや舞踊などにおける動きの分析 [1]、美術工芸品制作などにおける手技のデジタルアーカイブ、作品の組み立て映像マニュアルなどである。動画のメリットは、各状態の時間的変化が連続的に記録されている点にあり、そのためユーザは任意時間の状態を映像から確認することができる。しかし、動画は紙の資料や静止画像のように一覧できず、冗長なシーンの中からユーザが所望する内容を含む箇所を取り出す検索が、コンテンツの円滑な活用の鍵となる。

そこで本論文では、任意の映像ではなく人の動作がコンテンツの中心となる映像における検索を目的とし、「人と道具の関係」に着目し、シーンの状況認識手法として人の行動判別に焦点を当てる。具体的には、人が何らかの道具を用いて動作するシーンを抽出するために動作時の把持情報を映像と共にコンテンツ管理システムの中に組み込むことを試みる。人間が道具を用いて何かを行うときには、手で握る、触るという行為を必ず伴う [2]。そして、この手で握る、触る行為の際には、手と道具間に接触圧が生じる。この圧力を本論文では「把持圧」と呼ぶ。筆者らはこの把持圧には3種の意味情報が含まれていると考える。第一に、物を手にしたかという情報、第二には、手にしたものの性質に関する情報、そして第三には手にしたものを使用して行った動作に関する情報である。これら3種の情報は、コンテンツの中にある人の動作の様子を知る上で重要な要素そのものであるため、把持圧からこれらの情報を抽出することが可能であれば、把持圧の分析により人を含むコンテンツにおいて行動判別の手がかりとなる。

2 関連研究

コンテンツ内容検索として現在主流の手法は、画像群に対しクエリとしてテキスト情報をタグ付けする方法と、映像を分析しカット点抽出や人物抽出等を行う方法がある。前者において、自動的に映像データへのタグ付けを行うためには、シーンの状況を読み取り適切にテキスト情報を付与する技術が必要であり、状況認識が重要な技術要素となる。後者においては、画像に含まれる特徴量をもとに自動的に状況認識する手法 [3][4]、およびナレーションの音声認識やテロップの文字認識を組み合わせたもの [5] など、シーン内容認識を組み込んだ映像検索技術には様々あるが、本論

文では画像処理による認識技術は取り扱わないこととする。

状況認識のうち、映像以外のセンシング手法を用いて人の行動判別を行う研究も数多くなされており、その手法は光学的に遠隔計測する手法を用いるものとウェアラブルセンサを用いるものの2種類に大別できる。光学的手法の例としてはモーションキャプチャシステムがある。そのうち、環境設置のカメラで撮影するモーションキャプチャシステムでは、ユーザは身体にデバイスを装着する必要がなく、固定環境における人物の長時間の計測に向いているが、死角やオクルージョンの問題がある。一方、ウェアラブルセンサでは、センサの装着が身体への負担や行動を妨げる要因になりうるが、計測環境が限定されず、装着者のみのデータを精度よく取得できる。そこで、本論文では動作と把持圧の関係に着目し、ウェアラブルセンサで把持圧を測定することとする。

本システム同様、状況認識として様々な要素をトリガとして用い、取得データを編集・検索するシステムが盛んに研究されている。尾関ら [6] は、映像や音声から提示や呼びかけなどの注目喚起行動を取得することで、重要な行動をトリガとしプレゼンテーション映像の自動編集を行うシステムを提案している。また、現在タクシーなどで普及しているドライブレコーダもその例と言える。これは、車両事故の発生状況と原因を明確化するために、事故発生時前後の走行情報を記録するシステムである。走行時の映像を常時撮影しており、事故が発生した場合にのみ映像が保存される。事故発生のトリガは、加速度センサによって感知する衝撃・急ブレーキ・急ハンドルなどである。

ライフログの分野においては、身体に装着した加速度センサにより人の行動パターンや忙しさを認識する手法 [7] や、RFID を用いて物への接触状況や接触対象の情報を取得する手法を利用した日常記憶支援システム [8] などが提案されている。ライフログサービスの種類の増加や質の向上を目指し、脈拍数を例として、行動の種類・データの周期性を考慮したライフログの種類に依らない汎用的なラベル [9] も新たに提案されている。また、志村ら [10] は、ウェアラブル機器を用いてユーザ視点の体験映像を記録し、その映像に対してユーザ自身が体験時の感情を付与できる日記インタフェースを構築している。この研究では記録された膨大な体験記録の中からユーザの興味がある体験区間の抽出のために、興味があった体験の前後でユーザ自身にインデックスを付与させているが、ライフログの自動的な要約を目的に、脳波を用いてユーザの覚醒状態を推定することで、喜びや怒り、悲しみなどの強い感情を抱いた部分を抽出する手法 [11] も検討されている。

本論文では、人の動作を対象とするコンテンツ素材内の状況認識を行うために、人が道具と関わる際に必ず発生する「把持圧」をコンテンツ収録とともに記録し、把持状態を判別することで状況認識を行う手法、およびコンテンツの代表として映像を用いた映像検索支援システムについて述べる。

3 把持圧に基づく映像検索

動作と動作の合間の冗長なシーンを多く含む動作映像一般において、動作部分の効率的な検索を実現するために、把持状態の判別や把持対象の判別、動作の瞬間の検出は重要である。そこで、把持圧の時間変化から把持状態の情報、瞬時的な動作の発生情報を取得し、動作記録映像の検索を支援するシステムを構築する。

図1にシステムの概要を示す。本検索支援システムを使用するには、事前に、撮影時に把持圧データの同時取得が必要である。撮影対象者は手に把持圧計測装置(4章に記述)を装着し、カメラの前で動作を行う。続いて、取得した把持圧データの変動を基に、把持状態の判別、把持対象の判別、瞬時的な動作の検出を行う(5章に記述)。なお、把持状態とは把持しているか否かのことであり、瞬時的な動作とは例えばバットで球を打つなど、把持圧値が急激に上昇または下降した後戻るといった変動を示す一過性の運動のことである。また、把持対象の判別は再現可能な単純な把持動作に特定した上で取得される把持圧値の変動をもとに学習および判別する。システムはその判別および検出結果を基に、一連の映像内で対象者が物を把持した箇所、特定の把持対象を把持している箇所、道具を使用し動作している箇所をピックアップしユーザに提示することが可能となる。

映像検索を行う際、従来の動画再生ソフトではイベントの発見にコストを要するが、本システムでは物体を把持している箇所や道具を使用して動作している瞬間が提示されるため、ユーザは効率的に映像内のイベントを検索できる。さらに、映像データと把持圧データを事前に見比べ、ユーザが任意区間の把持圧データに対しテキスト情報をラベル(例:把持対象名、動作名等)として付加することで、映像データと把持圧データ、ラベルの対応関係の学習を行い、それを基に把持対象や動作をシステムに自動分類させることも可能である。それに基づき、ラベルをキーワードとするテキスト検索を用いた絞り込みによって、把持対象や動作毎の提示が可能となり、より容易にイベントの検索を行える。また、ラベル付けされた把持圧データ間を比較することで類似度を算出し、把持対象や動作の類似度順にソートして提示する機能も備える。これにより、類似した把持圧データに対応する映像データのみを次々に閲覧することが可能となる。

これらの機能を用いることで、例えば、テニスの練習を撮影した一連の映像においてスイング動作の確認を行いたい場合に、スイング動作の瞬間として抽出され提示されている画像群をクリックしていき、一連の練習風景の映像からスイングのシーンのみを効率よく確認することが可能である。また、器材組み立ての映像マニュアルでラベルに対するテキスト検索により使用工具の絞り込みを行うことや、類似度に基づくソーティング機能とテキスト検索を組み合わせ、特定の映像クリップ群を効率的に閲覧することも可能になると考えられる。

本システムの事前処理として、映像と同期した把持圧データを取得する必要がある。次章では、そのために構築した把持圧計測装置について述べる。

4 把持圧計測

本コンテンツ検索手法では、人の動作判別のために把持圧を用いるため、物を握った際に把持圧が大きく発生し、動作時に把持圧が顕著に変化する部分へセンサを配置することが望ましい。把持状態への移行は掌の形状を変化させることにより行うことから、人間が把持の際コントロールしているのは指にかける力であると仮定し、センサ配置は母指、示指、中指、薬指の指腹部とした。これは、日常生活に現れる把握を14種類の型に分類した鎌倉の研究[12]に基づく、手作業認識のための触覚センサ素子配置の決定手法[13]を参考とした。指腹部のどの位置にセンサを配置するかに関しては、この手法によると、指の第一関節から先の領域と第一関節と第二関節の間の領域が共に有効な範囲に含まれているため、どちらの領域に圧力センサを配置することが動作情報の取得として適しているかを予備実験により検証した。その結果、より顕著な把持圧変動を示す領域は握る対象の形状ごとに変わることがわかった。例えば太い円柱状物体を把持する場合には第一関節までの領域、細い円柱状物体を握る場合には第一関節と第二関節の間の領域であった。しかし、指先で摘むように持つ場合には第一関節から先の領域でしか物体と接触しないため、第一関節から第二関節の間の領域では把持圧が発生しない上、道具を把持し、力を込めて動作する場合には第一関節から先の領域のほうが把持圧の上昇が顕著になる傾向があった。そこで、本計測装置では第一関節から先の領域にセンサを配置することとした。

また、動作時の把持圧を測定するに当たり、実験装置の配線による負担をなくすため、無線でデータ通信を行う仕様とし、複数の圧力センサの着脱を簡便にするため、図2(上)に示すように、センサは手袋に組み込み一体型にした。

把持圧を取得するための圧力センサはニッタ(株)FlexiForce(A201-1)(図2(下))を使用した。このセンサはフィルムのように薄く、柔軟性に富んでいる。先端の感圧部に圧力が加わることによって両端の電極間の抵抗値が変化する。最大測定可能荷重値は4.4[N]である。データ通信部はBluetooth機能付き通信機として任天堂ゲーム機Wiiのコントローラ回路を用い、センサからの圧力データの出力は、同Wii拡張コントローラ「ヌンチャク」のジョイスティック部AD変換器を利用した。通信部を含めたサンプリングレートは5.0[Hz]である。なお、本計測装置により取得された把持圧データを以降では「計測把持圧」と呼ぶ。

5 判別手法

提案する映像検索支援システムでは、冗長なシーンを含む動作映像において、ユーザが閲覧したい箇所、つまり検索対象となる動作における重要な箇所を自動的に取り出す必要がある。本章では、把持圧データの変動から、動作における重要な箇所として、人が何を持っているかや動作をしている瞬間を判別する手法について述べる。

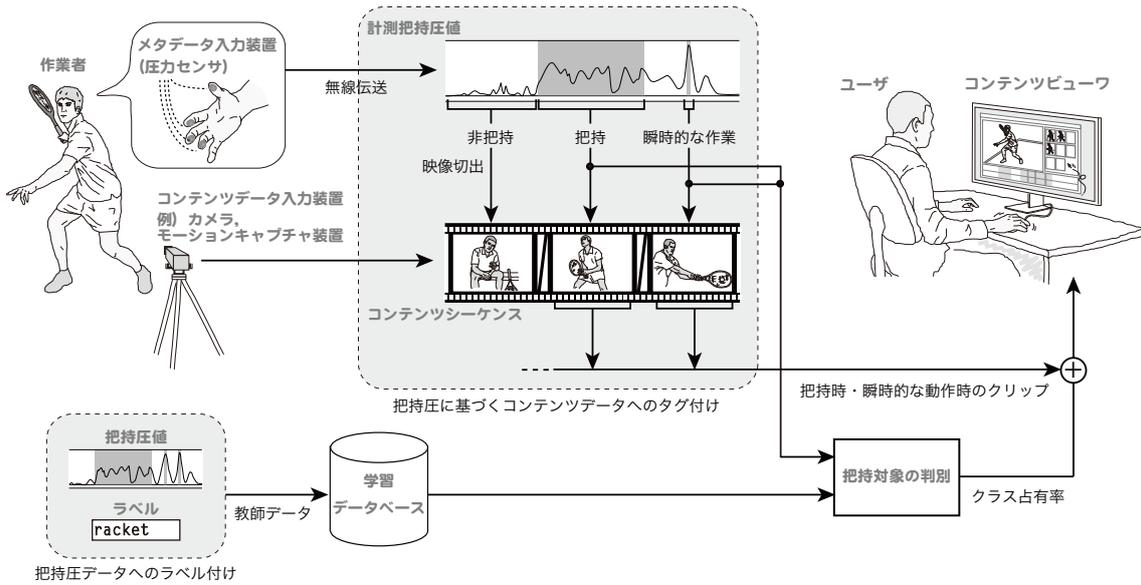


図 1: システムの概要

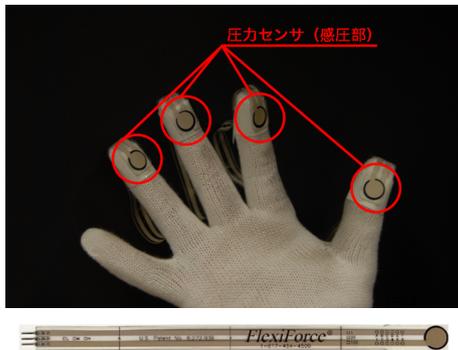


図 2: 把持圧測定装置 (上: 手指への装着の様子, 下: 使用した圧力センサ)

5.1 把持状態 (把持・非把持) の判別

物を把持する際には把持圧が発生する。そのため、取得した把持圧に対して閾値を設け、計測把持圧が閾値を超えていれば把持しているとみなすことにより把持・非把持の判別を行うことができる。物を把持している際でも、把持対象に接触する指としない指が存在するため、把持圧センサを配置した母指、示指、中指、薬指に対して各々閾値処理を行い、閾値を超えている計測把持圧が一箇所以上で存在していれば把持状態とみなす。閾値処理の閾値には予備実験により最大測定荷重値の 100 分の 1 である $0.044[N]$ を用いた。

5.2 把持対象の判別

物を把持する際に発生する把持圧の変化には、動作を特定のものにした場合は純粋に把持対象の性質 (形状や質量, 弾力性など) に関する情報が含まれると考えられる。そこで、動作を単純な把持動作に限定するという条件下で取得した把持圧データの学習および分類により把持対象の判別を行う。

把持対象の判別には事前にラベル付けされた訓練例を用いる手法として、k-NN (k 近傍法, $k=1$) 法を適用した。統計的判別に用いる把持圧データからの特徴量の抽出には、データを一定時間の移動窓に分割し、その移動窓毎に特徴量を求めるという手法 [14] を用いる。移動窓のシフト量はサンプリングレートと同じ 0.2 秒で行い、移動窓幅は 1.0 秒とする。移動窓内の平均, 標準偏差, エネルギー, エントロピー, 最大値, 最小値, 最大最小差分, 勾配の 8 個の特徴量を指 4 本分計 32 個, また 4 本の指の平均値の平均の 1 個および指 2 本間の相関係数計 6 個を加えた合計 39 個の特徴量を要素とし、各学習事例を 39 次元特徴空間の座標点で表し、距離は 39 次元空間のユークリッド距離を用いて判別を行った。なお、エネルギー, エントロピー, 相関係数の計算方法は付録 A.1 に示す。

5.3 瞬時的な動作の判別

計測把持圧のインパルス変化を検出することで瞬時的な動作を検出する。移動窓を用い、特徴量として移動窓内の最大最小差分と移動窓の両端の差分を抽出する。急激な上昇や下降が発生する場合には最大最小差分は大きな値を示す。また、元の強さに戻ることから変化の前後では移動窓の両端の差分は小さな値とな

る。そこで、把持圧の最大最小差分と、最大最小差分を移動窓の両端の把持圧の差分で割った値のそれぞれに閾値を設けることでインパルス変化の検出を行う。移動窓幅にはサンプリングレートの2倍にあたる0.4秒を使用する。

また、ノイズの影響を除去するために以下の二つの条件を設定し動作検出を行う。一つ目の条件は、最大最小差分の閾値として比較的大きめな値を設定し、いずれかの指の計測把持圧が閾値を超えた場合に瞬時的な動作として検出する。実装では事前の予備実験の結果、最大測定荷重値の30%にあたる1.23[N]を閾値として用いた。二つ目の条件は、瞬時的な動作では複数の指の計測把持圧が同時に変化する傾向があることから、最大最小差分の閾値を比較的小きめに設定する代わりに、二箇所以上の指の計測把持圧に同時にインパルス変化が起こった場合のみ瞬時的な動作として検出を行う。実装では予備実験の結果、最大測定荷重値の10%にあたる0.44[N]を閾値として用いた。これら二つの条件のいずれかを満たしたインパルス変化を瞬時的な動作として検出する。

6 実験

4章で構築した計測装置と5章で提案した判別手法を用い、把持状態の判別、把持対象の判別、瞬時的な動作の検出が可能かを実験により検証した。

6.1 把持状態（把持・非把持）の判別実験

6.1.1 実験目的

5.1節で述べた計測把持圧の閾値処理による把持状態の判別の確認を目的とする。

6.1.2 実験方法

把持対象を、A群：りんご、みかん、B群：金づち、ドライバ、テニスラケットとする。A群に対しては、空中に持ち上げ、一時停止し、置き直す動作を、B群には、持ち上げ、各々の用途で使用し、置き直す動作を行う。把持の仕方は特に指示せず被験者の自然な動作に任せしたが、結果的には全被験者とも、A群に関しては鎌倉ら[12]が定義した「包囲軽屈曲把握(CMF: Circular Mild Flexion Grip)」に分類される型、B群の金づち、テニスラケットに関しては「握力把握一標準型(PoS: Power Grip-Standard Type)」, B群のドライバは「三面把握一標準型(Tpd: Tripod Grip-Stanford Type)」に分類される型で把持した。このとき、把持圧変化を測定して記録し、同時に動作風景の映像撮影を行う。記録した計測把持圧値に対して閾値処理を行い、A, B群に対して把持・非把持の判別検証を行う。また、B群に対しては、使用中も把持として判別するかを検証する。そして、判別結果と

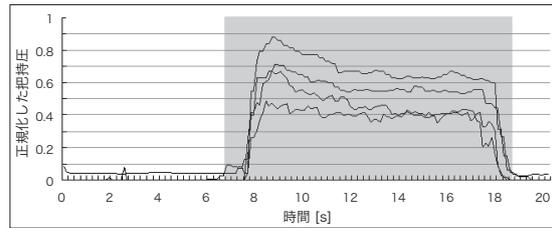


図 3: ある被験者のりんご把持時の計測把持圧の変化

映像を比較することで判別精度を検証する。20代被験者3名で、それぞれ利き手にセンサを装着させ計測を行った。

6.1.3 結果と考察

図3はある被験者がりんごを持ち上げた際の4指の把持圧変化を示すグラフである。グラフ横軸は撮影開始後からの経過時間[s]、縦軸は最大測定荷重値4.4[N]で正規化した計測把持圧値で、グラフ中の灰色領域は把持と判別された領域である。目視により把持と判別された領域での撮影映像を確認したところ、映像中で把持と確認できる全時間帯で本手法により把持と判別することができていた。りんごと同様に、みかん、金づちに関しては、把持・非把持を100[%]判別できた。ドライバ使用時に関しては、指先がドライバと接触しない瞬間では把持と分類されなかった。また、テニスラケット使用時に関しては、素振りの合間に見られる構えの際に小指以外の指先の力を解放する瞬間に、把持と分類されていないことがわかった。前者の問題に関しては、把持対象の学習と学習データベースを充実させることで解決できるものと考えられるが、後者の問題に関しては、把持圧計測装置のセンサの手指装着配置を工夫する必要がある。鎌倉らの把持の分類で、CMFとTpdに関しては小指と把持物体の接触がないが、テニスやゴルフなどのスポーツの際によく現れる把持の型Posでは小指が把持対象と接触することが多いとされている。より適切で汎用的なセンサ配置の検討に関しては今後の課題としたい。

6.2 把持対象の判別実験

6.2.1 実験目的

5.2節で述べた判別手法を用いた把持対象の判別可能性を検証する。

表 1: 把持対象の属性

特徴量	りんご	ボール	バナナ	缶	はっさく	ブロック	みかん
質量 [g]	332	188	148	540	374	50	62
直径 [mm]	90	70	30	70	130	64	60

表 2: 把持対象の判別結果の混同行列

真のクラス	りんご	ボール	バナナ	缶	はっさく	ブロック	みかん
りんご	41	0	9	1	5	0	0
ボール	0	67	8	0	0	0	0
バナナ	0	0	65	0	0	0	0
缶	0	0	0	65	0	0	0
はっさく	26	0	0	0	29	0	0
ブロック	0	12	35	0	1	15	14
みかん	2	15	12	0	3	10	20

6.2.2 実験方法

把持対象を、りんご、みかん、はっさく、バナナ、テニスボール、レゴブロック、コーラ 500ml 缶とする。これらの対象を利き手で持った際の把持圧変化を測定して記録する。把持対象の属性表を表 1 に示す。そして、把持圧情報から特徴量を抽出し k-NN 法で分類器を作成し、把持対象の判別を行う。20 代被験者 4 名に対して約 5 秒間物体を空中で把持した状態で静止させるという特定の動作を各 4 回試行させ計測した。その際、どのような握り方にするかは、特に指示せず被験者の自然な動作に任せた。

6.2.3 結果と考察

ある被験者が行った 4 回の試行のうちの 3 試行分の把持圧データを用いて分類器を作成し、残りのデータを判別した場合の判別率 (TP 率 [True Positive Rate] と FP 率 [False Positive Rate]) を図 4 に、各把持物体を真のクラスとしそれがどのクラスに判別されたかの回数を示す Confusion Matrix (混同行列) を表 2 に例として示す。把持対象の判別を行った結果、同一把持領域内の判別結果の多数決処理から 7 対象中 6 対象を判別できることを確認した。また、学習に使用するデータ量が少ないと著しく判別率が低下する傾向が見られたが、複数人のデータが学習に使用するデータに混じることによる判別率の大きな変動は見られないことが分かった。

また、把持の仕方は被験者の自然な動作に任せたが、結果的には全被験者とも、球状の形状をしているりんご、みかん、はっさく、テニスボールに関しては鎌倉ら [12] が定義した「包囲軽屈曲把握 (CMF: Circular Mild Flexion Grip)」に分類される型、柱状の形状をしているバナナ、レゴブロック、コーラ缶に関しては「並列軽屈曲把握 (PMF: Parallel Mild Flexion Grip)」に分類される型で把持した。

次に、対象毎の判別率のばらつきに関して考察する。図 4 で TP が低いのははっさく、ブロック、みかんは、

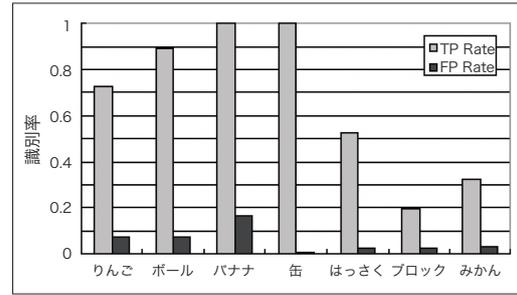


図 4: 把持対象ごとの判別率

TP が高いりんご、バナナ、ボールに比べ、特徴空間上で把持圧データのちらばりが大きいと言える。その要因として、後者三つに比べ前者三つは人の手にとって把持しにくい特徴を持つことが考えられる。はっさくは直径が約 130mm であり、包囲軽屈曲把握型で把持しても指先が確実に物体表面に接触する状態は保ちにくい。また、ブロックは表面が平面で構成されており、かつその質感がつるつるしていることから、センサを装着した指先が滑りやすい状況であったと考えられる。みかんに関しては形状と質量が比較的小さいことから把持が安定せず、同一の把持動作が再現しにくかったことが考えられる。

さらに、表 2 からブロックがバナナと誤判別されやすいことがわかるが、その理由に関しては、両者ともに並列軽屈曲把握型で把持されることから、把持圧データのちらばりが大きく不安定であるブロックではなく、ちらばりが小さく安定しているバナナに多く分類されたと考えられる。この際、バナナと缶はそれぞれ明確に分類されており、ブロックの把持パターンがどちらかという質量の近いバナナに類似するものだったと推測できる。

6.3 瞬時的な動作の判別実験

6.3.1 実験目的

5.3 節で述べた手法による計測把持圧のインパルス変化の抽出可能性、およびインパルス変化と瞬時的な動作の関係について検証する。

6.3.2 実験方法

20 代被験者 1 名に対して、ゴルフのスイング、テニスラケットのスイング、キャッチボール、金づち打ちの 4 種類の動作を行わせ、それぞれ把持圧変化を測定して記録する。5.3 節の手法によりインパルス変化を検出することで瞬時的な動作の判別を行う。

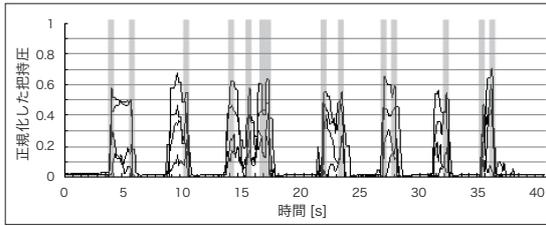


図 5: キャッチボール時の計測把持圧の変化

6.3.3 結果と考察

各動作から得られたデータに判別手法を適用した結果のうち、キャッチボールの例を図5に示す。グラフ横軸は撮影開始後からの経過時間 [s]、縦軸は最大測定荷重値 4.4[N] で正規化した計測把持圧値で、グラフ中の灰色に色付けされた棒状の領域がインパルス変化として検出された区間である。図より明らかなように、撮影映像全体から動作の瞬間を抽出することで、有効な区間以外を圧縮することができている。それぞれの動作区間において撮影映像を確認した結果、インパルスを検出することで瞬時的な動作として抽出されたものは、それぞれの道具をそれぞれの用途で使用しているときの動作（キャッチボール時には投球と捕球）と、道具を握り直している動作（キャッチボール時にはボールの握り直し）の二種類あることがわかった。

映像検索を行う際、意味のある動作として抽出されるべきなのは、道具を使用しているときであり、握り直し動作とは区別できたほうが検索支援システムとして有用だろう。しかし握り直し動作時には大きき分離は難しい。ただし、道具を使用している動作の方が握り直しより動作が大きい傾向があるため、オプティカルフローを用いた手腕の動きの解析などの動画処理を含めると分離できる可能性がある。

7 コンテンツビューワの実装

実装した映像検索支援システムのコンテンツビューワの外観を図6に示す。映像再生部（図中、左上部）では、ユーザが選択した映像を再生し表示する。閲覧したいシーンの選択は、クリップ化されたサムネイル画像または映像タイムライン上をクリックすることで行う。サムネイル表示部（図中、右上部）では、5.1節で述べた手法により物体を把持したと判別された直後、5.3節で述べた手法により瞬時的な動作を開始した直後のフレームのサムネイル画像を表示する。サムネイル画像をクリックすることで表示されている動作の瞬間の前後一定時間（0.5秒間）の映像が再生される。タイムライン表示部（図中、下部）では、撮影時間全体のうち任意時間の映像をクリックにより再生可能であり、把持を行っている時間帯と道具を使用し瞬時的な動作を行っている時間帯を色のついた帯で提示する。表示部では、単純な把持を行っている時間帯で

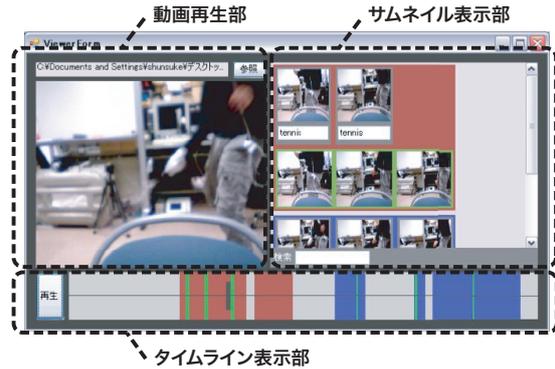


図 6: コンテンツビューワの外観

は5.2節で述べた把持対象の判別手法により、それ以外の時間帯では手動によるラベリングに基づく学習を行い、その結果である把持対象や動作内容の違いを色の違いによりユーザに提示する。

サムネイル表示部においては、動作映像内でクリップ毎に道具把持時の把持圧変化の類似度順にソートする機能も実装した。類似度の基準として、5.2節で述べたk-NN法を用いて分類したクラスの占有率を使用する。現在再生している映像のクラス要素が多く含まれている順に映像をソートし、サムネイル表示部に提示する。クラスの占有率は、把持と判別された一区間の時間に対する各クラスに分類された区間の時間の割合として算出する。

本検索支援システムを用いると、6.1節で述べたように、把持・非把持に関しては100[%]判別可能であることから、図3で例示した20秒間の動作記録映像の中から把持の様子を観察する場合、ビューワによりピックアップされた映像開始7秒後から18秒後までのクリップを閲覧すればよく、把持区間の頭出しもサムネイルをクリックすることにより瞬時に行える。これは、映像のみのコンテンツを閲覧する場合、20秒間全ての区間を連続再生する必要があることに比べ、冗長な情報が圧縮されており、効率のよい閲覧が可能となっていると言える。また、6.3節の図5で例示した40秒間のキャッチボールの映像では、8回の捕球のうち7回の抽出に成功していることから、サムネイル画像を順次クリックすることで捕球動作のみの閲覧がほぼ可能であり、映像全体を探索し、所望の瞬間を発見していく作業より効率的と言える。

さらに、本システムで実装したソート表示機能により、抽出されたサムネイル画像を把持圧変化の類似度順に並べ替え、閲覧していくことも可能である。稲葉らの研究[1]のようにスポーツの動作分析を目的とした場合などに、上達の順に映像を確認する作業などに特に有用だと考えられる。

8 おわりに

動作時に、映像と共に対象者の把持圧の変化を記録し、把持圧情報から推定される動作状態を有効にユーザに提示することで映像データの検索支援を行う手法を提案した。

コンテンツ映像撮影と同時に把持圧を計測することで、その把持圧に基づき、把持状態（把持・非把持）の判別、道具を用いた瞬時的な動作の検出が可能であること、さらに事前学習を行うことで把持対象を分類する処理法を示した。

また、把持状態を利用した検索支援システムの構築において、動作状態のタイムライン上への提示、動作時のサムネイル画像の表示、指定した動作区間のみの再生、クラス分類に基づくソート表示を実装し、動作映像の検索においてコンテンツを抽出し色分けする検索ビューワの実装を行った。

本論文で述べたコンテンツ映像検索法では、把持圧データは動作を判別する手がかりとして用いた。しかし、芸術活動における熟達した技能の伝承においては、把持圧情報を手指の細かな状態や詳細な力加減などのノンバーバルな情報と捉え、その技能を解明するための有用なデータとして活用できると考えられる。本システムを発展させることで、熟練技術を要する手作業における手元映像と把持圧を記録し、提示・閲覧する技能記録再生システムとして有用となる可能性がある。今後は、指別の把持圧変化を解析支援する手法に取り組み、それを組み込んだ技能伝承支援システムを構築したい。

A 特徴量の算出方法

5.2 節のエネルギー E 、エントロピー e 、相関係数 corr は以下のように定義する。ただし n は移動窓内のサンプル数、 $f(i, t)$ は指番号 i の時刻 t における把持圧である。また σ は移動窓内における標準偏差、 $\bar{f}(z)$ は時刻を t として窓区間にわたる $f(z, t)$ の平均を表す。

$$E = \frac{1}{n} \sum_{t=0}^n f^2(i, t) \quad (1)$$

$$e = - \sum_{t=0}^n p(i, t) \log p(i, t) \quad (2)$$

ここで、

$$p(i, t) = \frac{f^2(i, t)}{\sum_{s=0}^n f^2(i, s)} \quad (3)$$

とした。また、

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad (4)$$

ここで、

$$\text{cov}(x, y) = \frac{1}{n-2} \sum_{t=0}^n (f(x, t) - \bar{f}(x)) (f(y, t) - \bar{f}(y)) \quad (5)$$

とした。

参考文献

- [1] 稲葉洋, 瀧剛志, 宮崎慎也, 長谷川純一, 肥田満裕, 山本英弘, 北川薫: スポーツ動作分析の支援を目的とした人体センシング情報の可視化提示法, 芸術科学会論文誌, 2(3), pp. 94-100, 2003.
- [2] 山本景子, 金谷一朗, 佐藤宏介: 三次元形状デザインのための道具握り判別型インタフェース, 芸術科学会論文誌, Vol. 7, No. 3, pp. 102-112, 2008.
- [3] 南里卓也, 大津展之: 複数人動画からの異常動作検出, 映像処理学会研究報告, Vol. 2004, No. 91, pp. 179-186, 2004.
- [4] 社本裕司, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: 放送映像における準同一映像区間の出現パターンによる分類, 電子情報通信学会技術研究報告, Vol. 108, No. 484, pp. 165-170, 2009.
- [5] 桑野秀豪, 松尾義博, 川添雄彦: 映像・音声認識, 自然言語処理の適用によるメタデータ生成の作業コスト削減効果に関する研究, 映像情報メディア学会誌, Vol. 61, No. 6, pp. 842-852, 2007.
- [6] 尾関基行, 中村裕一, 大田友: 注目喚起行動に基づいた机上作業映像の編集, 電子情報通信学会論文誌 D, Vol. 88, 2005.
- [7] Kern, N. and Schiele, B., Junker, H., Lukowicz, P., Troster, G., Schmidt, A.: Context Annotation for a Live Life Recording, Journal of Personal and Ubiquitous Computing, Special Issue on Memory and Sharing of Experiences, 2006.
- [8] 河村竜幸, 福原知宏, 村田賢, 武田英明, 河野恭之, 木戸出正継: 対象物に「触れる」行為と記憶の遍在化による日常記憶支援, 電子情報通信学会論文誌 D, Vol. 88, 2005.
- [9] 村山卓弥, 河田博昭, 手塚博久, 山田智広, 武藤伸洋, 阿部匡伸: ライフログサービスのための脈拍を用いた平常・非平常ラベルの検討, 電子情報通信学会信学技報, Vol. 108, pp. 77-82, 2009.
- [10] 志村将吾, 平野靖, 梶田将司, 間瀬健二: 体験映像への感情付与インタフェース, インタラクシオン 2005 論文集, pp. 59-60, 2005.
- [11] 相澤清晴, 石島健一郎, 椎名誠: ウエアラブル映像の構造化と要約: 個人の主観を考慮した要約生成の試み, 電子情報通信学会論文誌 D-II Vol. J86-D-II, No. 6, pp. 807-815, 2003.
- [12] 鎌倉矩子: 手のかたち 手のうごき, 医歯薬出版株式会社, 1989.
- [13] 松尾一矢, 村上剛司, 長谷川勉: 手作業認識のための接触センサ配置の決定手法, TVRSJ, Vol. 13, No. 1, pp. 93-96, 2008.

- [14] 田淵勝宏, 納屋太, 大村廉, 野間春生, 小暮潔, 岸野文郎: 加速度センサを用いた日常行動識別におけるデータ収集条件の識別性能への影響評価, 電子情報通信学会技術研究報告, Vol. 106, No. PRUM2006-27, pp. 43-48, 2006.

山本 景子

2005年大阪大学基礎工学部システム科学科卒業。現在、同大学大学院博士後期課程在籍。日本学術振興会特別研究員 DC2, 2008年度芸術科学会論文賞受賞, ハプティックインタフェース, 投影型複合現実感, ヒューマンコンピュータインタラクションに関する研究に従事。芸術科学会, 人間工学会会員。

南部 俊輔

2007年大阪大学基礎工学部卒業。2009年同大学大学院博士前期課程修了。主として映像情報メディアに関する研究に従事。

佐藤 宏介

1983年大阪大学基礎工学部制御工学科卒業。1985年同大学大学院博士前期課程修了。1986年同大助手。1988~1990年カーネギーメロン大学ロボット工学研究所客員研究員。1994年奈良先端科学技術大学院大学情報科学研究科助教授。2003年大阪大学大学院基礎工学研究科教授。三次元画像計測, 映像情報メディア, 実世界インタフェース等の研究に従事。工学博士。電子情報通信学会, VR学会, SICE, IEEE ほか各会員。