

【特許請求の範囲】**【請求項 1】**

映像から人物オブジェクトを抽出し、動的な人物オブジェクトに対応する触覚メタデータを生成する触覚メタデータ生成装置であって、

入力された映像について、現フレーム画像を含む複数フレーム分の過去のフレーム画像を抽出する複数フレーム抽出手段と、

前記現フレーム画像を含む複数フレーム分のフレーム画像の各々について、骨格検出アルゴリズムに基づき、各人物オブジェクトの第 1 の骨格座標集合を生成する人物骨格抽出手段と、

前記現フレーム画像を含む複数フレーム分のフレーム画像の各々について、前記第 1 の骨格座標集合を基に探索範囲を可変設定し、各人物オブジェクトの骨格の位置及びサイズと、その周辺画像情報を抽出することにより人物オブジェクトを識別し、人物 ID を付与した第 2 の骨格座標集合を生成する人物識別手段と、

10

前記現フレーム画像を基準に、前記複数フレーム分のフレーム画像における前記第 2 の骨格座標集合を時系列に連結し、人物オブジェクト毎の骨格の軌跡を示す軌跡特徴量の集合として骨格軌跡集合を生成する軌跡特徴量生成手段と、

前記現フレーム画像を基準に、当該複数フレーム分のフレーム画像における骨格軌跡集合を基準とした各人物オブジェクト間の距離、及び人物オブジェクト毎の各関節のオペティカルフロー量から各人物オブジェクトの人物動きの状況変化量を算出するとともに、骨格検出の成否、人物骨格の重心の移動量、及び前記探索範囲を表す骨格外接矩形のアスペクト比の変化量を算出し、これらを要素とする特徴ベクトルを、動作状況を表す動作状況特徴量として生成する動作状況計測手段と、

20

前記骨格軌跡集合の軌跡特徴量と、当該骨格軌跡集合の軌跡特徴量に対応する動作状況特徴量とを基に、機械学習により、触覚提示デバイスを作動させる衝撃提示用の情報を検出する人物動作認識手段と、

前記現フレーム画像に対応して、前記人物動作認識手段から得られる当該触覚提示デバイスを作動させる衝撃提示用の情報を含む第 1 の触覚メタデータを生成し、フレーム単位で外部出力する第 1 のメタデータ生成手段と、

前記現フレーム画像に対応して、前記動作状況計測手段から得られる各人物オブジェクトの人物動きの状況変化量を示す動作状況提示用の情報を含む第 2 の触覚メタデータを生成し、フレーム単位で外部出力する第 2 のメタデータ生成手段と、
を備えることを特徴とする触覚メタデータ生成装置。

30

【請求項 2】

前記人物識別手段は、前記探索範囲として、最大で人物骨格の全体を囲む人物探索範囲に限定し、最小で人物骨格のうち所定領域を注目探索範囲として定めた絞り込みによる可変設定を行い、状態推定アルゴリズムで得られる人物の骨格の状態遷移推定値に基づいて、少なくとも前記注目探索範囲を含むように前記探索範囲を決定して、当該人物オブジェクトを識別する処理を行う手段を有することを特徴とする、請求項 1 に記載の触覚メタデータ生成装置。

【請求項 3】

40

前記現フレーム画像を含む複数フレーム分のフレーム画像の各々を用いて隣接フレーム間の差分画像を基に動オブジェクトを検出し、各差分画像から検出した動オブジェクトのうち全ての人物オブジェクトの前記骨格軌跡集合を用いて特定の動オブジェクトを選定し、各差分画像から得られる特定の動オブジェクトの座標位置、大きさ、移動方向を要素とし連結した動オブジェクト情報を生成する動オブジェクト検出手段を更に備え、

前記人物動作認識手段は、前記動オブジェクト情報を基に、全ての人物オブジェクトの前記骨格軌跡集合のうち前記触覚提示デバイスを作動させるための骨格軌跡集合を選定し、選定した骨格軌跡集合の軌跡特徴量と、その選定した骨格軌跡集合に対応する動作状況特徴量とを基に、機械学習により、人物オブジェクト毎の衝撃提示用のタイミング及び速さを示す情報を検出することを特徴とする、請求項 1 又は 2 に記載の触覚メタデータ生成

50

装置。

【請求項 4】

請求項 1 から 3 のいずれか一項に記載の触覚メタデータ生成装置と、
触覚刺激を提示する触覚提示デバイスと、
前記触覚メタデータ生成装置から得られる第 1 及び第 2 の触覚メタデータを基に、予め
定めた駆動基準データを参照し、前記触覚提示デバイスを駆動するように制御する制御ユニ
ットと、
を備えることを特徴とする映像触覚連動システム。

【請求項 5】

コンピュータを、請求項 1 から 3 のいずれか一項に記載の触覚メタデータ生成装置とし
て機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、映像から人物オブジェクトを抽出し、動的な人物オブジェクトに対応する触
覚メタデータを生成する触覚メタデータ生成装置、生成した触覚メタデータを基に触覚提
示デバイスを駆動制御する映像触覚連動システム、及びプログラムに関する。

【背景技術】

【0002】

放送映像等の一般的なカメラ映像の映像コンテンツは、視覚と聴覚の 2 つの感覚に訴え
る情報を提供するメディアである。しかし、視覚障害者や聴覚障害者に対しては視聴覚情
報だけでは不十分であり、番組コンテンツの状況を正確に伝えることができない。そのた
め、テレビを持っていない、若しくは持っても視聴しない障害者も多い。そこで、映
像コンテンツに対し、視覚・聴覚以外の“触覚”で感じられる情報を提示することで、視
覚又は聴覚の障害者もテレビ放送を理解できるシステムの構築が望まれる。

【0003】

また、視覚・聴覚の感覚を有する健常者にとっても、また、触覚刺激を提示すること
により放送番組の視聴時の臨場感や没入感の向上が期待できる。特に、スポーツコンテ
ンツにおける人物の動きは重要な情報であり、これを触覚刺激で提示することにより、コン
テンツ視聴における臨場感が高まる。

【0004】

例えば、野球映像を視聴する際、ボールがバットに当たるタイミングで触覚提示デバイ
スを介して視聴者に刺激を与えることで、バッターのヒッティングの感覚を疑似体験で
きる。また、視覚に障害のある方々に触覚刺激を提供することで、スポーツの試合状況を理
解させることにも繋がると考えられる。このように、触覚は映像視聴における第 3 の感覚
として期待されている。

【0005】

特に、スポーツはリアルタイムでの映像視聴が重要視されるため、映像に対する触覚刺
激の提示は、自動、且つリアルタイムで行われる必要がある。そこで、プレーの種類、タ
イミング、状況などに関する選手の動きに同期した触覚刺激の提示が、触覚を併用した映
像コンテンツの映像視聴に効果的な場合が多い。そして、視覚又は聴覚に障害を持つ方々
にもスポーツの状況を伝えることが可能となる。

【0006】

このため、触覚を併用した映像コンテンツの映像視聴を実現するには、その映像コンテ
ンツから人物オブジェクトの動きを抽出し、抽出した人物オブジェクトの動きに対応した
触覚情報を触覚メタデータとして生成することが必要になる。

【0007】

しかし、従来の触覚メタデータの生成法では、触覚を併用した映像視聴を実現するとし
ても、触覚提示デバイスにより、どのようなタイミングで、またどのような刺激をユーザ
に提示するかを示す触覚メタデータを、映像と同期した態様で人手により編集する必要が

10

20

30

40

50

あった。

【0008】

収録番組の場合、人手で時間をかけて触覚メタデータを編集することが可能である。しかし、生放送映像に対して触覚提示デバイスによる刺激提示を連動させるには、事前に触覚情報を編集することができないことから、リアルタイムで映像コンテンツの映像解析を行い、触覚メタデータを生成することが要求される。

【0009】

近年、スポーツ映像解析技術は、目覚ましい成長を遂げている。ウィンブルドンでも使用されているテニスのホークアイシステムは、複数の固定カメラ映像をセンサとしてテニスボールを3次元的に追跡し、ジャッジに絡むIN/OUTの判定を行っている。また2014年のFIFAワールドカップでは、ゴールラインテクノロジーと称して、数台の固定カメラの映像を解析し、ゴールの判定を自動化している。更に、サッカースタジアムへ多数のステレオカメラを設定し、フィールド内の全選手をリアルタイムに追跡するTRACABシステム等、スポーツにおけるリアルタイム映像解析技術の高度化が進んでいる。

【0010】

一方で、動的な人物オブジェクトとして選手の姿勢を計測するには、従来、マーカー式のモーションキャプチャー方式を用いた計測が一般的である。しかし、この方式は、選手の体に多数のマーカーを装着する必要があるため、実試合には適用できない。そこで、近年では、選手の体に投光されている赤外線パターンを読み取り、その赤外線パターンの歪みから深度情報を得る深度センサを用いることで、マーカーレスでの人物姿勢計測が可能になっている。また、マーカー式ではなく、光学式のモーションキャプチャー方式を応用した種々の技術が開示されている（例えば、特許文献1，2参照）。

【0011】

例えば、特許文献1では、立体視を用いた仮想現実システムにおいて他者の模範動作映像を表示することにより使用者に対して動作を教示する際に、光学式のモーションキャプチャー方式により、計測対象者の骨格の3次元位置を計測する装置が開示されている。また、特許文献2には、光学式のモーションキャプチャー方式を利用してプレイヤーの動作を測定し、測定したデータとモデルのフォームに関するデータとに基づいて同プレイヤーのフォームを評価するトレーニング評価装置について開示されている。しかし、これらの技術は、モーションキャプチャー方式を利用するため、実際の試合に適用できず、汎用的なカメラ映像から人物のプレー動作を計測することは難しい。

【0012】

また、モーションキャプチャー方式によらず、一人又は二人が一组となってバドミントンの試合やバドミントン練習を撮影したカメラ映像のみから、人物の動きをシミュレートする装置が開示されている（例えば、特許文献3参照）。特許文献3の技術では、撮影したカメラ映像から、ショットなどの動作を検出するものとなっているが、専用に設定したカメラによる撮影映像から処理することを前提としており、汎用的な放送カメラ映像から人物のプレー動作を計測することは難しい。

【0013】

ところで、近年の深層学習技術の発達により、深度センサを用いずに、従来では困難であった深度情報を含まない通常の静止画像から人物の骨格位置を推定することが可能になっている。この深層学習技術を用いることで、通常のカメラ映像から静止画像を抽出し、その静止画像に含まれる選手の姿勢を自動計測することが可能となっている。

【先行技術文献】

【特許文献】

【0014】

【特許文献1】特開2002-8063号公報

【特許文献2】特開2002-253718号公報

【特許文献3】特開2018-187383号公報

【発明の概要】

10

20

30

40

50

【発明が解決しようとする課題】**【0015】**

上述したように、触覚を併用した映像コンテンツの映像視聴を実現するには、その映像コンテンツから人物オブジェクトの動きを抽出し、抽出した人物オブジェクトの動きに対応した触覚情報を触覚メタデータとして生成することが必要になる。

【0016】

しかし、従来技術では、リアルタイムで映像コンテンツの映像解析のみから、触覚メタデータを生成することが困難である。即ち、映像のみから触覚メタデータを生成する場合には、カメラ映像からリアルタイムで人物オブジェクトの動きを解析する必要がある。リアルタイムのスポーツ競技では、その競技に影響を与えることは好ましくないため、マーカー装着によるモーションキャプチャー方式や、撮影距離に制限のある深度センサなどを用いずに、撮影条件に制限の無い汎用的な放送カメラ映像のみから触覚メタデータを生成することが望ましい。

10

【0017】

つまり、スポーツを撮影する通常のカメラ映像のみから、自動、且つリアルタイムで人物オブジェクト（選手等）の動きに関する触覚メタデータを生成する技法が望まれる。

【0018】

また、人物オブジェクトの動きを高精度に検出するために、人物以外の動オブジェクト（例えば、バドミントン競技であればシャトル、ラケット）を参考する技法も考えられるが、参考とする人物以外の動オブジェクトが存在しない競技（例えば、柔道やレスリング等）においても、人物オブジェクトの動きを高精度に検出する技法が望まれる。

20

【0019】

尚、近年の深層学習技術の発達により、深度センサを用いずに、従来では困難であった深度情報を含まない通常の静止画像から人物の骨格位置を推定することが可能になっているが、これに代表される骨格検出アルゴリズムは基本的に静止画単位で骨格位置を検出するものである。このため、スポーツを撮影する通常のカメラ映像のみから、自動、且つリアルタイムで人物オブジェクト（選手等）の動きに関する触覚メタデータを生成するには、更なる工夫が必要になる。

【0020】

本発明の目的は、上述の問題に鑑みて、映像から人物オブジェクトを自動抽出し、動的な人物オブジェクトに対応する触覚メタデータを同期して自動生成する触覚メタデータ生成装置、生成した触覚メタデータを基に触覚提示デバイスを駆動制御する映像触覚連動システム、及びプログラムを提供することにある。

30

【課題を解決するための手段】**【0021】**

本発明の触覚メタデータ生成装置は、映像から人物オブジェクトを抽出し、動的な人物オブジェクトに対応する触覚メタデータを生成する触覚メタデータ生成装置であって、入力された映像について、現フレーム画像を含む複数フレーム分の過去のフレーム画像を抽出する複数フレーム抽出手段と、前記現フレーム画像を含む複数フレーム分のフレーム画像の各々について、骨格検出アルゴリズムに基づき、各人物オブジェクトの第1の骨格座標集合を生成する人物骨格抽出手段と、前記現フレーム画像を含む複数フレーム分のフレーム画像の各々について、前記第1の骨格座標集合を基に探索範囲を可変設定し、各人物オブジェクトの骨格の位置及びサイズと、その周辺画像情報を抽出することにより人物オブジェクトを識別し、人物IDを付与した第2の骨格座標集合を生成する人物識別手段と、前記現フレーム画像を基準に、前記複数フレーム分のフレーム画像における前記第2の骨格座標集合を時系列に連結し、人物オブジェクト毎の骨格の軌跡を示す軌跡特徴量の集合として骨格軌跡集合を生成する軌跡特徴量生成手段と、前記現フレーム画像を基準に、当該複数フレーム分のフレーム画像における骨格軌跡集合を基準とした各人物オブジェクト間の距離、及び人物オブジェクト毎の各関節のオブティカルフロー量から各人物オブジェクトの人物動きの状況変化量を算出するとともに、骨格検出の成否、人物骨格の重心の

40

50

移動量、及び前記探索範囲を表す骨格外接矩形のアスペクト比の変化量を算出し、これらを要素とする特徴ベクトルを、動作状況を表す動作状況特徴量として生成する動作状況計測手段と、前記骨格軌跡集合の軌跡特徴量と、当該骨格軌跡集合の軌跡特徴量に対応する動作状況特徴量とを基に、機械学習により、触覚提示デバイスを作動させる衝撃提示用の情報を検出する人物動作認識手段と、前記現フレーム画像に対応して、前記人物動作認識手段から得られる当該触覚提示デバイスを作動させる衝撃提示用の情報を含む第1の触覚メタデータを生成し、フレーム単位で外部出力する第1のメタデータ生成手段と、前記現フレーム画像に対応して、前記動作状況計測手段から得られる各人物オブジェクトの人物動きの状況変化量を示す動作状況提示用の情報を含む第2の触覚メタデータを生成し、フレーム単位で外部出力する第2のメタデータ生成手段と、を備えることを特徴とする。

10

【0022】

また、本発明の触覚メタデータ生成装置において、前記人物識別手段は、前記探索範囲として、最大で人物骨格の全体を囲む人物探索範囲に限定し、最小で人物骨格のうち所定領域を注目探索範囲として定めた絞り込みによる可変設定を行い、状態推定アルゴリズムで得られる人物の骨格の状態遷移推定値に基づいて、少なくとも前記注目探索範囲を含むように前記探索範囲を決定して、当該人物オブジェクトを識別する処理を行う手段を有することを特徴とする。

【0023】

また、本発明の触覚メタデータ生成装置において、前記現フレーム画像を含む複数フレーム分のフレーム画像の各々を用いて隣接フレーム間の差分画像を基に動オブジェクトを検出し、各差分画像から検出した動オブジェクトのうち全ての人物オブジェクトの前記骨格軌跡集合を用いて特定の動オブジェクトを選定し、各差分画像から得られる特定の動オブジェクトの座標位置、大きさ、移動方向を要素とし連結した動オブジェクト情報を生成する動オブジェクト検出手段を更に備え、前記人物動作認識手段は、前記動オブジェクト情報を基に、全ての人物オブジェクトの前記骨格軌跡集合のうち前記触覚提示デバイスを作動させるための骨格軌跡集合を選定し、選定した骨格軌跡集合の軌跡特徴量と、その選定した骨格軌跡集合に対応する動作状況特徴量とを基に、機械学習により、人物オブジェクト毎の衝撃提示用のタイミング及び速さを示す情報を検出することを特徴とする。

20

【0024】

また、本発明の映像触覚連動システムは、本発明の触覚メタデータ生成装置と、触覚刺激を提示する触覚提示デバイスと、前記触覚メタデータ生成装置から得られる第1及び第2の触覚メタデータを基に、予め定めた駆動基準データを参照し、前記触覚提示デバイスを駆動するよう制御する制御ユニットと、を備えることを特徴とする。

30

【0025】

更に、本発明のプログラムは、コンピュータを、本発明の触覚メタデータ生成装置として機能させるためのプログラムとして構成する。

【発明の効果】

【0026】

本発明によれば、映像から人物オブジェクトを自動抽出し、動的な人物オブジェクトに対応する触覚メタデータを同期して自動生成することができる。そして、人物オブジェクトに生じる「衝撃の種類とタイミング」を表す第1の触覚メタデータだけでなく、人物オブジェクトに係る連続的な「動作状況」を表す第2の触覚メタデータをも自動生成できるようになる。これにより、スポーツ映像のリアルタイム視聴時での触覚刺激の提示が可能となり、更には、人物オブジェクトの重なりやオクルージョンが生じやすい柔道等の試合映像での詳細な触覚提示も可能となる。つまり、視覚・聴覚への情報提供のみならず、触覚にも訴えることで、視覚や聴覚に障害を持つ方々へもスポーツの状況を分かりやすく伝えることが可能となる。さらに、視覚・聴覚の感覚を有する健常者の方々にとっても、従来の映像視聴では伝えきれない臨場感や没入感を提供することができる。

40

【0027】

特に、スポーツ映像視聴に際し、各選手の識別、位置座標、及びチーム競技であればそ

50

の分類、並びに、触覚提示デバイスを作動させる種類（衝撃強度の種類）とタイミング（速さを含む）、並びに動作状況を示す情報を含む触覚メタデータを生成することで、触覚提示デバイスにより、動作の状況変化とともに、プレーの種類、タイミング、強度などに関する触覚刺激をユーザに提示できるようになる。これにより、触覚情報を用いたパブリックビューイング、エンターテインメント、将来の触覚放送などのサービス性の向上に繋がる。また、スポーツ以外でも、工場での触覚アラームへの応用や、監視カメラ映像解析に基づいたセキュリティシステムなど、様々な用途に応用することも可能になる。

【図面の簡単な説明】

【0028】

【図1】本発明による一実施形態の触覚メタデータ生成装置を備える映像触覚連動システムの概略構成を示すブロック図である。

10

【図2】本発明による一実施形態の触覚メタデータ生成装置の処理例を示すフローチャートである。

【図3】本発明による一実施形態の触覚メタデータ生成装置における人物骨格抽出処理に関する説明図である。

【図4】（a）は1フレーム画像を例示する図であり、（b）は本発明による一実施形態の触覚メタデータ生成装置における1フレーム画像における人物骨格抽出例を示す図である。

【図5】（a）、（b）は、それぞれ本発明による一実施形態の触覚メタデータ生成装置における人物骨格抽出処理に関する人物オブジェクトの探索範囲の処理例を示す図である。

20

【図6】本発明による一実施形態の触覚メタデータ生成装置における軌跡特徴量の説明図である。

【図7】本発明による一実施形態の触覚メタデータ生成装置における動オブジェクト検出のために生成する差分画像例を示す図である。

【図8】（a）乃至（c）は、それぞれ本発明による一実施形態の触覚メタデータ生成装置における動作状況検出処理に関する人物動きの状況変化量の説明図である。

【図9】本発明による一実施形態の映像触覚連動システムにおける制御ユニットの概略構成を示すブロック図である。

【発明を実施するための形態】

30

【0029】

（システム構成）

以下、図面を参照して、本発明による一実施形態の触覚メタデータ生成装置12を備える映像触覚連動システム1について詳細に説明する。図1は、本発明による一実施形態の触覚メタデータ生成装置12を備える映像触覚連動システム1の概略構成を示すブロック図である。

【0030】

図1に示す映像触覚連動システム1は、カメラや記録装置等の映像出力装置10から映像を入力し、入力された映像から人物オブジェクトを自動抽出し、動的な人物オブジェクトに対応する触覚メタデータ（第1の触覚メタデータと第2の触覚メタデータの2種類）を同期して自動生成する触覚メタデータ生成装置12と、生成した触覚メタデータを基に、本例では2台の触覚提示デバイス14L、14Rと、各触覚提示デバイス14L、14Rを個別に駆動制御する制御ユニット13と、を備える。

40

【0031】

まず、映像出力装置10が出力する映像は、一例として柔道競技をリアルタイムで撮影されたものとしてディスプレイ11に表示され、ユーザUによって視覚されるものとする。

【0032】

柔道競技は、二人の選手が組み合っ、「抑え込み」や「投げ」などの技を競うスポーツであり、各人物に衝撃が生じた瞬間や各人物の動きの状況変化を触覚提示デバイス14

50

L, 14Rにより触覚刺激としてユーザUに提示することで、より臨場感を高め、また視聴覚障害者にも試合状況を伝えることが可能である。

【0033】

特に、柔道競技では、映像上で選手同士の重なりやオクルージョンが多数生じるため、各選手に生じる衝撃の種類に応じたタイミングと速さ以外にも、各選手の押し引きなどの組み合わせ、投げ等に係る動作状況を連続的に触覚提示できるようにすることで、視覚や聴覚の障害者にも試合の緊迫感を伝えることができ、また臨場感を高めることができる。

【0034】

そこで、ユーザUは、左手HLで触覚提示デバイス14Lを把持し、右手HRで触覚提示デバイス14Rを把持して、本例では映像解析に同期した振動刺激が提示されるものとする。制御ユニット13は、触覚メタデータ生成装置12から得られる各人物オブジェクトOp1, Op2に生じる衝撃の種類に応じたタイミングと速さを示す衝撃提示用の情報を含む第1の触覚メタデータと、各人物オブジェクトOp1, Op2に係る連続的な「動作状況」を表す動作状況提示用の情報を含む第2の触覚メタデータを基に、各人物オブジェクトOp1, Op2に対応付けられた2台の触覚提示デバイス14L, 14Rの触覚提示を個別に制御する。ただし、制御ユニット13は、1台の触覚提示デバイスに対してのみ駆動制御する形態でもよいし、3台以上の触覚提示デバイスに対して個別に駆動制御する形態でもよい。また、限定するものではないが、本例の制御ユニット13は、映像内の人物オブジェクトOp1(選手)の動きに対応した振動刺激は触覚提示デバイス14Lで、人物オブジェクトOp2(選手)の動きに対応した振動刺激は触覚提示デバイス14Rで提示するように分類して制御するものとする。

【0035】

触覚提示デバイス14L, 14Rは、球状のケース141内に、制御ユニット13の制御によって振動刺激を提示可能な振動アクチュエーター142が収容されている。尚、触覚提示デバイス14L, 14Rは、振動刺激の他、電磁気パルス刺激を提示するものでもよい。本例では、制御ユニット13と各触覚提示デバイス14L, 14Rとの間是有線接続され、触覚メタデータ生成装置12と制御ユニット13との間も有線接続されている形態を例に説明するが、それぞれ近距離無線通信で無線接続されている形態としてもよい。

【0036】

触覚メタデータ生成装置12は、複数フレーム抽出部121、人物骨格抽出部122、人物識別部123、軌跡特徴量生成部124、動オブジェクト検出部125、人物動作認識部126、第1のメタデータ生成部127、動作状況計測部128、及び第2のメタデータ生成部129を備える。

【0037】

複数フレーム抽出部121は、入力された映像について、現フレーム画像を含むT(Tは2以上の整数)フレーム分の過去のフレーム画像を抽出し、人物骨格抽出部122及び動オブジェクト検出部125に出力する。

【0038】

人物骨格抽出部122は、現フレーム画像を含むTフレーム分のフレーム画像の各々について、骨格検出アルゴリズムに基づき、各人物オブジェクト(以下、単に「人物」とも称する。)Op1, Op2の骨格座標集合 P^n_b (n:検出人数、b:骨格ID)を生成し、現フレーム画像を含むTフレーム分のフレーム画像とともに、人物識別部123に出力する。

【0039】

人物識別部123は、現フレーム画像を含むTフレーム分のフレーム画像の各々について、骨格座標集合 P^n_b を基に探索範囲(詳細は後述する。)を可変設定し、各人物の骨格の位置及びサイズと、その周辺画像情報を抽出することにより人物を識別し、人物IDを付与した骨格座標集合 P^i_b (i:人物ID、b:骨格ID)を生成し、軌跡特徴量生成部124に出力する。

【0040】

10

20

30

40

50

軌跡特徴量生成部 124 は、現フレーム画像を基準に、T フレーム分のフレーム画像における骨格座標集合 P^i_b を時系列に連結し、人物毎の骨格の軌跡を示す軌跡特徴量の集合として骨格軌跡集合 T^i_b (i : 人物 ID、 b : 骨格 ID) を生成し、動オブジェクト検出部 125、人物動作認識部 126、及び動作状況計測部 128 に出力する。

【0041】

動オブジェクト検出部 125 は、現フレーム画像を含む T フレーム分のフレーム画像の各々を用いて隣接フレーム間の差分画像を基に動オブジェクトを検出し、各差分画像から検出した動オブジェクトのうち軌跡特徴量生成部 124 から得られる全ての人物の骨格軌跡集合 T^i_b を用いて特定の動オブジェクトを選定し、各差分画像から得られる特定の動オブジェクトの座標位置、大きさ、移動方向を要素とし連結した動オブジェクト情報を生成し、人物動作認識部 126 に出力する。

10

【0042】

動作状況計測部 128 は、現フレーム画像を基準に、T フレーム分のフレーム画像における骨格軌跡集合 T^i_b を基準とした各人物オブジェクト間の距離、及び人物オブジェクト毎の各関節のオプティカルフロー量から各人物オブジェクトの人物動きの状況変化量 $K(t)$ を算出して第 2 のメタデータ生成部 129 に出力するとともに、骨格検出の成否 $D(t)$ 、人物骨格の重心の移動量 (縦 $G_v(t)$ 、横 $G_h(t)$)、及び当該探索範囲を表す骨格外接矩形のアスペクト比の変化量 $S(t)$ を算出し、これらの $K(t)$ 、 $D(t)$ 、 $G_v(t)$ 、 $G_h(t)$ 、 $S(t)$ を要素とする特徴ベクトルを、動作状況を表す動作状況特徴量 M^i_b (i : 人物 ID、 b : 骨格 ID) として生成し、人物動作認識部 12

20

【0043】

人物動作認識部 126 は、動オブジェクト情報を基に、全ての人物の骨格軌跡集合 T^i_b のうち、触覚提示デバイスを作動させるための骨格軌跡集合 T^i_b を選定し、選定した骨格軌跡集合 T^i_b の軌跡特徴量と、その選定した骨格軌跡集合 T^i_b の軌跡特徴量に対応する動作状況特徴量 M^i_b とを基に、機械学習 (サポートベクターマシン、又はニューラルネットワーク等) により、現フレーム画像内の各人物の識別、位置座標 (及び、本例では柔道競技としているため対象外となるが、チーム競技であればそのチーム分類)、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す衝撃提示用の情報を検出し、第 1 のメタデータ生成部 127 に出力する。

30

【0044】

第 1 のメタデータ生成部 127 は、現フレーム画像に対応して、人物動作認識部 126 から得られる、現フレーム画像内の各人物の識別、位置座標 (及び、本例では柔道競技としているため対象外となるが、チーム競技であればそのチーム分類)、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す衝撃提示用の情報を含む第 1 の触覚メタデータ (衝撃提示用) を生成し、フレーム単位で制御ユニット 13 に出力する。

【0045】

第 2 のメタデータ生成部 129 は、現フレーム画像に対応して、動作状況計測部 128 から得られる各人物オブジェクトの人物動きの状況変化量 $K(t)$ を示す動作状況提示用の情報を含む第 2 の触覚メタデータ (動作状況提示用) を生成し、フレーム単位で制御ユ

40

【0046】

以下、より具体的に、図 2 を基に、図 3 乃至図 8 を参照しながら、触覚メタデータ生成装置 12 における触覚メタデータ生成処理について説明する。

【0047】

(触覚メタデータ生成処理)

図 2 は、本発明による一実施形態の触覚メタデータ生成装置 12 の処理例を示すフローチャートである。そして、図 3 は、触覚メタデータ生成装置 12 における人物骨格抽出処理に関する説明図である。また、図 4 (a) は 1 フレーム画像を例示する図であり、図 4 (b) は触覚メタデータ生成装置 12 における 1 フレーム画像における人物骨格抽出例を

50

示す図である。図 5 (a) , (b) は、それぞれ本発明による一実施形態の触覚メタデータ生成装置 1 2 における人物骨格抽出処理に関する人物オブジェクトの探索範囲の処理例を示す図である。図 6 は、触覚メタデータ生成装置 1 2 における軌跡特徴量の説明図である。図 7 は、本発明による一実施形態の触覚メタデータ生成装置 1 2 における動オブジェクト検出のために生成する差分画像例を示す図である。そして、図 8 (a) 乃至 (c) は、それぞれ本発明による一実施形態の触覚メタデータ生成装置 1 2 における動作状況検出処理に関する人物動きの状況変化量の説明図である。

【 0 0 4 8 】

図 2 に示すように、触覚メタデータ生成装置 1 2 は、まず、複数フレーム抽出部 1 2 1 により、入力された映像について、現フレーム画像を含む T (T は 2 以上の整数) フレーム分の過去のフレーム画像を抽出する (ステップ S 1) 。

10

【 0 0 4 9 】

続いて、触覚メタデータ生成装置 1 2 は、人物骨格抽出部 1 2 2 により、現フレーム画像を含む T フレーム分のフレーム画像の各々について、骨格検出アルゴリズムに基づき、各人物オブジェクト Op_1 , Op_2 の骨格座標集合 P^n_b (n : 検出人数、 b : 骨格 ID) を生成する (ステップ S 2) 。

【 0 0 5 0 】

近年の深層学習技術の発展により、通常の画像から人物の骨格位置を推定することが可能となった。OpenPoseやVisionPose (NextSystem社) に代表されるように、骨格検出アルゴリズムをオープンソースで公開しているものも存在する。そこで、本例の人物骨格抽出部 1 2 2 は、VisionPoseを用いて、図 3 に示すように、フレーム画像毎に人物の骨格 30 点を検出し、その位置座標を示す骨格座標集合 P^n_b を生成する。

20

【 0 0 5 1 】

VisionPoseでは、図 3 において、 P^n_{11} : “ 頭 ”、 P^n_{12} : “ 鼻 ”、 P^n_{13} : “ 左目 ”、 P^n_{14} : “ 右目 ”、 P^n_{15} : “ 左耳 ”、 P^n_{16} : “ 右耳 ”、 P^n_{17} : “ 首 ”、 P^n_{18} : “ 背骨 (肩) ”、 P^n_{19} : “ 左肩 ”、 P^n_{20} : “ 右肩 ”、 P^n_{21} : “ 左肘 ”、 P^n_{22} : “ 右肘 ”、 P^n_{23} : “ 左手首 ”、 P^n_{24} : “ 右手首 ”、 P^n_{25} : “ 左手 ”、 P^n_{26} : “ 右手 ”、 P^n_{27} : “ 左親指 ”、 P^n_{28} : “ 右親指 ”、 P^n_{29} : “ 左指先 ”、 P^n_{30} : “ 右指先 ”、 P^n_{31} : “ 背骨 (中央) ”、 P^n_{32} : “ 背骨 (基端部) ”、 P^n_{33} : “ 左尻部 ”、 P^n_{34} : “ 右尻部 ”、 P^n_{35} : “ 左膝 ”、 P^n_{36} : “ 右膝 ”、 P^n_{37} : “ 左足首 ”、 P^n_{38} : “ 右足首 ”、 P^n_{39} : “ 左足 ”、 及び、 P^n_{40} : “ 右足 ”、 についての座標位置と、各座標位置を図示するような線で連結した描画が可能である。

30

【 0 0 5 2 】

このVisionPoseの骨格検出アルゴリズムに基づき、図 4 (a) に示す柔道競技の 1 フレーム画像 F に対して、人物の骨格抽出を行ったフレーム画像 F_a を図 4 (b) に示している。図 4 (a) に示すフレーム画像 F には、各人物オブジェクト Op_1 , Op_2 (選手) のみが映り込んでいる様子を示しているが、その他の人物オブジェクトである審判の動オブジェクトが映り込むことや、別のスポーツ競技であれば人物以外の動オブジェクト (バドミントン競技であればラケットやシャトル等)、或いは観客等のオブジェクト (実質的には、静オブジェクト) が写り込むことがある。しかし、VisionPoseの骨格検出アルゴリズムを適用すると、選手及び審判の人物オブジェクトの人物についてのみ人物の骨格抽出を抽出することができる。本例では、図 4 (b) に示すように、人物オブジェクト Op_1 , Op_2 にそれぞれ対応する骨格座標集合 P^1_b , P^2_b を推定して生成することができる。図 4 (b) から理解されるように、柔道競技においても、比較的精度よく各人物の骨格を推定できる。尚、骨格検出アルゴリズムは、静止画単位での推定に留まるので、触覚メタデータ生成装置 1 2 は、後続する処理として、人物の識別を行い、各人物の骨格位置の推移を軌跡特徴量として定量化し、時間軸を考慮した高精度な動作認識を行う。

40

【 0 0 5 3 】

続いて、触覚メタデータ生成装置 1 2 は、人物識別部 1 2 3 により、現フレーム画像を

50

含むTフレーム分のフレーム画像の各々について、骨格座標集合 P^n_b を基に探索範囲を可変設定し、各人物の骨格の位置及びサイズと、その周辺画像情報を抽出することにより人物を識別し、人物IDを付与した骨格座標集合 P^i_b (i :人物ID、 b :骨格ID)を生成する(ステップS3)。

【0054】

前述した人物骨格抽出部122により、現フレーム画像を含むTフレーム分のフレーム画像の各々について、骨格座標集合 P^n_b として、1以上の人物の骨格の検出が可能となる。しかし、各フレーム画像の骨格座標集合 P^n_b では、「誰」の情報は存在しないため、各人物の骨格を識別する必要がある。この識別には、各フレーム画像における各骨格座標集合 P^n_b の座標付近の画像情報を利用する。即ち、人物識別部123は、骨格座標集合 P^n_b を基に、各人物の骨格の位置及びサイズと、その周辺画像情報(色情報、及び顔又は背付近のテクスチャ情報)を抽出することにより、人物を識別し、人物IDを付与した骨格座標集合 P^i_b (i :人物ID、 b :骨格ID)を生成する。

【0055】

例えば、柔道では白と青の道着で試合が行われるが、各骨格座標集合 P^n_b の骨格の位置付近の画像情報として、フレーム画像Fにおける色情報を参照することで、選手の識別が可能になる。また、バドミントン競技では、コート縦に構えた画角で撮影される場合に、各骨格座標集合 P^n_b の骨格の位置がフレーム画像Fにおける画面上側であれば奥の選手、画面下側であれば手前の選手、として識別することができる。

【0056】

従って、人物骨格抽出部122における骨格検出アルゴリズムは静止画単位での推定に留まるが、骨格座標集合 P^n_b を基に動オブジェクトとしての人物を認識することができ、各骨格位置の推移を軌跡特徴として扱うことで、時間軸を考慮した高精度な動作認識が可能となる。

【0057】

尚、前述した人物骨格抽出部122では、選手以外にも審判や観客など、触覚刺激の提示対象としない他の人物の骨格を検出してしまうことも多い。審判は選手と別の衣服を着用することが多いため、色情報で識別できる。また、観客は選手に比べて遠くにいることが多いため、骨格のサイズで識別が可能である。このように、各競技のルールや撮影状況を考慮し、人物識別に適切な周辺画像情報(色情報、及び顔又は背付近のテクスチャ情報)を設定することにより、触覚刺激の提示対象とする選手の識別が可能となる。

【0058】

ところで、本実施形態の人物識別部123は、各人物の重なりやオクルージョンにも対応するため、フレーム画像単位で探索範囲(人物探索範囲 R^i 及び注目探索範囲 R^b_i)を可変設定する。例えば、図5(a)に示す人物オブジェクトOp1, Op2(選手)と、人物オブジェクトOp3(審判)について、人物骨格抽出部122により各骨格座標集合 P^n_b (図示略)の抽出が行われると、人物識別部123は、フレーム画像単位で人物探索範囲 R^i 及び注目探索範囲 R^b_i を可変設定することができる。この探索範囲 R^i は、図5(a)において、人物ID(i)ごとに設定し、フレーム画像の画像座標上での人物の位置座標、及び人物の大きさ(幅及び高さ)を有するものとして外接矩形で表している。また、各人物の腰領域(P^n_{22} , P^n_{23} , P^n_{24})を囲む領域を注目探索範囲 R^b_i として表している。

【0059】

より具体的には、本実施形態の人物識別部123は、各フレーム画像で人物の探索範囲を、最大で人物骨格の全体を囲む人物探索範囲 R^i に限定し、最小で人物骨格のうち所定領域(本例では腰領域(P^n_{22} , P^n_{23} , P^n_{24})を囲む領域)を注目探索範囲 R^b_i として定めた絞り込みによる可変設定を行い、状態推定アルゴリズムで得られる人物の骨格の状態遷移推定値に基づいて、少なくとも注目探索範囲 R^b_i を含むように探索範囲を決定して、当該人物オブジェクトを識別する処理を行う。これにより、例えば図5(b)に示すように各人物の動作が変化した場合やフレーム画像に対する相対的な人物の大

きさが変化した場合でも、他の人物の誤認識を防ぎ、また処理速度も向上できる。特に、柔道のように識別対象の人物の重なりが激しく、背景も複雑な映像から精度よく選手を識別するには探索範囲の利用が有効である。

【0060】

つまり、本実施形態の人物識別部123は、各選手及び審判の人物オブジェクトの Op_1 、 Op_2 、 Op_3 における各骨格座標集合 P^n_b のうち、色識別を可能とする所定範囲（本例では腰領域（ P^n_{22} 、 P^n_{23} 、 P^n_{24} ）の色（青、白、茶色））を注目探索範囲 Rb^i として予め定めているので、検出した複数の人物の骨格座標集合 P^n_b が重なる場合には注目探索範囲 Rb^i に絞って探索することで、各フレーム画像で精度よく人物を抽出・追跡できる。尚、背景に解析対象以外の骨格を検出する場合もあるため、解析対象の人物の骨格には、人物ID（ i ）を付与して判別することで、追跡対象の人物の骨格座標 P^i_b を識別できる。

10

【0061】

そして、探索範囲（人物探索範囲 R^i 及び注目探索範囲 Rb^i ）の広さや形の決定は、カルマンフィルタやパーティクルフィルタなどの状態推定アルゴリズムで得られる人物の骨格の状態遷移推定値に基づいて、少なくとも注目探索範囲 Rb^i （本例では、各人物の腰領域）を含むように決定する。

【0062】

そして、探索範囲（人物探索範囲 R^i 及び注目探索範囲 Rb^i ）の安定検出時には範囲を狭め、検出が不安定な際には範囲を広げることができ、例えば、人物ID（ i ）ごとに人物の骨格の状態遷移推定値に基づいて定めた探索範囲を設定し、その状態遷移推定値が直前フレームから所定値以内であれば安定とし、そうでなければ不安定とすることや、状態推定アルゴリズムで得られる人物の骨格の状態遷移推定値に基づいて、 T フレーム分の時間窓間に、検出に成功した割合を計算し、その割合が所定値以上であれば安定とし、当該所定値を下回った場合に不安定とすることで、探索範囲を可変設定することができる。

20

【0063】

続いて、触覚メタデータ生成装置12は、軌跡特徴量生成部124により、現フレーム画像を基準に、 T フレーム分のフレーム画像における骨格座標集合 P^i_b を時系列に連結し、人物毎の骨格の軌跡を示す軌跡特徴量の集合として骨格軌跡集合 T^i_b （ i ：人物ID、 b ：骨格ID）を生成する（ステップS4）。

30

【0064】

ここで、骨格軌跡集合 T^i_b の生成にあたって、まず、任意のフレーム画像における骨格座標集合 P^i_b を $P^i_b(t)$ とし、現フレーム画像を $t=0$ として現フレーム画像における骨格座標集合 P^i_b を $P^i_b(0)$ で表し、過去 T フレームのフレーム画像における骨格座標集合 P^i_b を $P^i_b(T)$ で表す。つまり、軌跡特徴量生成部124は、現フレーム画像のフレーム番号を $t=0$ として、過去 T フレームまでのフレーム番号を $t=T$ で表すと、現フレーム画像を基準に、 $t=0, 1, \dots, T$ の各フレーム画像 F を用いて、人物毎の骨格の軌跡を示す軌跡特徴量の集合として骨格軌跡集合 T^i_b を生成することができる。尚、骨格軌跡集合 T^i_b は、人物探索範囲 R^i 及び注目探索範囲 Rb^i のそれぞれについて分けて軌跡を算出できる。

40

【0065】

尚、骨格軌跡集合 T^i_b の生成に用いる骨格座標は、必ずしも図3に示す30点全てを用いる必要はなく、予め定めた特定の骨格軌跡のみを使用して、処理速度を向上させる構成とすることもできる。また、骨格軌跡集合 T^i_b としては、骨格座標集合 P^i_b の座標表現そのものを連結したものとしてもよいが、人物毎の骨格の軌跡を示すものであればよいことから、各競技のルールや撮影状況を考慮し、軌跡特徴を表わすのに適切な情報（動き量や移動加速度等）に変換したものとしてもよい。

【0066】

例えば、骨格軌跡集合 T^i_b は、各骨格の移動量の二階微分を作成し、加速度に相当する値に変換するのが好適である。そこで、骨格座標集合 P^i_b の軌跡として、加速度に相

50

当する骨格軌跡集合 T^i_b で表すことで、後段の人物動作認識部 126 における動作認識の精度を向上させることができる。

【0067】

まず、式(1)に示すように、隣接する画像フレーム間で、対応する骨格座標集合 $P^i_b(t)$ 、 $P^i_b(t+1)$ の位置座標の差(ユークリッド距離)を取り、その移動量 $D^i_b(t)$ を求める。

【0068】

【数1】

$$D^i_b(t) = \sqrt{(P^i_b(t), x - (P^i_b(t+1), x))^2 + (P^i_b(t), y - (P^i_b(t+1), y))^2} \quad (1)$$

10

【0069】

ここで、 $P^i_b(t), x$ は $P^i_b(t)$ における x 座標、 $P^i_b(t), y$ は $P^i_b(t)$ における y 座標を表す。

【0070】

$D^i_b(t)$ は、各座標点の速度に相当する特徴量となるが、式(2)に示すように、更にその差の絶対値をとることで、加速度に相当する特徴量 $A^i_b(t)$ が得られる。ここで、 $\text{abs}()$ は、絶対値を返す関数である。

【0071】

【数2】

$$A^i_b(t) = \text{abs}(D^i_b(t) - D^i_b(t+1)) \quad (2)$$

20

【0072】

この加速度に相当する特徴量 $A^i_b(t)$ を用いて各人物の動作を追跡した軌跡を示す骨格軌跡集合 T^i_b を生成することができ、図6には、或るフレーム画像における人物オブジェクト $Op1$ 、 $Op2$ にそれぞれ対応する骨格座標集合の軌跡特徴量 T^1_b 、 T^2_b を分かり易く描画したフレーム画像 Fb を示している。

【0073】

続いて、触覚メタデータ生成装置 12 は、動作状況計測部 128 により、現フレーム画像を基準に、 T フレーム分のフレーム画像における骨格軌跡集合 T^i_b を基準とした各人物オブジェクト間の距離、及び人物オブジェクト毎の各関節のオプティカルフロー量から各人物オブジェクトの人物動きの状況変化量 $K(t)$ を算出して第2のメタデータ生成部 129 に出力するとともに、骨格検出の成否 $D(t)$ 、人物骨格の重心の移動量(縦 $Gv(t)$ 、横 $Gh(t)$)、及び当該探索範囲を表す骨格外接矩形のアスペクト比の変化量 $S(t)$ を算出し、これらの $K(t)$ 、 $D(t)$ 、 $Gv(t)$ 、 $Gh(t)$ 、 $S(t)$ を要素とする特徴ベクトルを、動作状況を表す動作状況特徴量 $M^i_b(i: \text{人物ID}, b: \text{骨格ID})$ として生成し、人物動作認識部 126 に出力する(ステップ S5)。

30

【0074】

例えば柔道では、組手の状態から相手の隙を伺い、急激に投げの動作に入ることが多い。この閑から急に至る状況を触覚刺激として提示することで、映像コンテンツ視聴における臨場感を高めることができる。 T フレーム分のフレーム画像における骨格軌跡集合 T^i_b を基準とした各人物オブジェクト(選手)間の距離、及び人物オブジェクト(選手)毎の各関節のオプティカルフロー量から各人物オブジェクト(選手)の人物動きの状況変化量 $K(t)$ を算出できる。

40

【0075】

より具体的には、人物動きの状況変化量 $K(t)$ を以下に示す式(3)のように定義する。この人物動きの状況変化量 $K(t)$ は、第2のメタデータ生成部 129 により、第2の第2の触覚メタデータ(動作状況提示用)として外部出力される。

【0076】

【数 3】

$$K(t) = F(t) / d(t) \quad (3)$$

ここに、 $d(t) = |P_g^1(t) - P_g^2(t)| / L$

$$F(t) = \sum_{i=1}^2 \{ \sum_{b=1}^{30} |P_b^i(t) - P_b^i(t-1)| / 30 \} / 2$$

【0077】

ここで、 $d(t)$ は各人物オブジェクト（選手）間の距離を表し、 $F(t)$ は、現フレーム画像の時刻 t と直前フレームの時刻 $t-1$ における人物オブジェクトごとの骨格30点のオプティカルフローの平均値を表す。 $d(t)$ を算出する際の g は、重心（本例では、腰の位置である $P_{22}^n, P_{23}^n, P_{24}^n$ の中央値座標）を表している。従って、 $P_g^1(t)$ は、現フレーム画像の時刻 t における $i=1$ の人物IDの人物オブジェクト $Op1$ の重心位置を表し、 $P_g^2(t)$ は、そのフレーム画像の時刻 t における $i=2$ の人物IDの人物オブジェクト $Op2$ の重心位置を表す。また、 L は双方の人物オブジェクト $Op1, Op2$ の骨格30点のうち両選手の首から腰までの距離の平均値であり、 L で正規化することにより各選手の撮影サイズによる影響を排除できる。

【0078】

一方、動作状況を表す動作状況特徴量 M_{ib}^i (i : 人物ID、 b : 骨格ID) は、それぞれ式(4)、式(5)、式(6)、式(7)に示すように、骨格検出の成否 $D(t)$ 、人物骨格の重心の移動量（縦 $Gv(t)$ 、横 $Gh(t)$ ）、及び骨格外接矩形（人物探索範囲）のアスペクト比の変化量 $S(t)$ を算出し、人物動きの状況変化量 $K(t)$ に加えて、 $D(t)$ 、 $Gv(t)$ 、 $Gh(t)$ 、 $S(t)$ を要素として生成する。

【0079】

【数 4】

$$D^i(t) = \begin{cases} 1: \text{if 骨格検出成功} \\ 0: \text{if 骨格検出失敗} \end{cases} \quad (4)$$

【0080】

【数 5】

$$Gv^i(t) = P_b^i(t).y - P_b^i(t-1).y \quad (5)$$

【0081】

【数 6】

$$Gh^i(t) = P_b^i(t).x - P_b^i(t-1).x \quad (6)$$

【0082】

【数 7】

$$S^i(t) = R^i(t).height / R^i(t).width - R^i(t-1).height / R^i(t-1).width \quad (7)$$

【0083】

ここで、 $R^i(t)$ は、時刻 t のフレーム画像 $F(t)$ における人物オブジェクト $Op1$ の当該探索範囲を表す骨格外接矩形を表し、その左上座標 (x, y) と、幅 ($width$) 及び高さ ($height$) の情報を有し、“ $aaa.bbb$ ” は、 aaa の bbb 成分を表す。

【0084】

柔道の「投げ」動作時は、急に重心が下がり、画像上で横長のアスペクト比になることが多い。また不自然な体勢になることも多く、しばしば骨格検出に失敗する。さらに、関節の移動量が全体的に増加する。そこで、人物動きの状況変化量 $K(t)$ に加えて、 $D(t)$ 、 $Gv(t)$ 、 $Gh(t)$ 、 $S(t)$ の要素からなる動作状況を表す動作状況特徴量 M_{ib}^i を、後段の人物動作認識部126における機械学習の識別器に用いることで、例えば「投げ」動作を精度よく検出できる。

10

20

30

40

50

【0085】

続いて、触覚メタデータ生成装置12は、動オブジェクト検出部125により、現フレーム画像を含むTフレーム分のフレーム画像の各々を用いて隣接フレーム間の差分画像を基に動オブジェクトを検出し、各差分画像から検出した動オブジェクトのうち軌跡特徴量生成部124から得られる全ての人物の骨格軌跡集合 T^i_b を用いて特定の動オブジェクトを選定し、各差分画像から得られる特定の動オブジェクトの座標位置、大きさ、移動方向を要素とし連結した動オブジェクト情報を生成する（ステップS6）。図7に示す差分画像Fcに示されているように、人物オブジェクトOp1'、Op2'が検出できていることが分かる。

【0086】

10

後段の人物動作認識部126では、骨格軌跡集合 T^i_b を用いて人物の動作認識を行うことが可能であるが、人物（選手）の動作は多種多様であり、誤検出や検出漏れが発生するケースも少なくない。そこで、動オブジェクト検出部125は、現フレーム画像を含むTフレーム分のフレーム画像の各々を用いて、解析対象とする人物の動オブジェクトの位置や動きに関する情報を抽出する（尚、バドミントン競技などの人物以外のラケットやシャトル等の動オブジェクトも参考とすることができるときは、その動オブジェクトも利用できる。）。この情報を利用することで、後段の人物動作認識部126は、動作認識の精度をより向上させることができる。

【0087】

20

続いて、触覚メタデータ生成装置12は、人物動作認識部126により、動オブジェクト情報を基に、全ての人物の骨格軌跡集合 T^i_b のうち、触覚提示デバイスを作動させるための骨格軌跡集合 T^i_b を選定し、選定した骨格軌跡集合 T^i_b の軌跡特徴量と、その選定した骨格軌跡集合 T^i_b の軌跡特徴量に対応する動作状況特徴量 M^i_b とを基に、機械学習（サポートベクターマシン、又はニューラルネットワーク等）により触覚提示デバイス14R、14Lを作動させる衝撃提示用の情報を検出する（ステップS7）。衝撃提示用の情報には、現フレーム画像内の各人物の識別、位置座標（及び、本例では柔道競技としているため対象外となるが、チーム競技であればそのチーム分類）、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す情報が含まれる。

【0088】

30

機械学習（サポートベクターマシン、又はニューラルネットワーク等）時には、事前に学習用の軌跡特徴量を作成して学習させておく。例えば、サポートベクターマシンを利用するときは、衝撃を表す瞬間の軌跡特徴量を正例、それ以外の軌跡特徴量を負例として学習することで、人物動作認識部126は、触覚提示デバイス14R、14Lを作動させるタイミング及び速さを示す情報を動作認識として検出することが可能となる。更に、人物動作認識部126は、選定した骨格軌跡集合 T^i_b の軌跡特徴量から、動作認識の精度を高めるとともに、どの選手がどのような衝撃を受けているか等、現フレーム画像内の各人物の識別、位置座標、（及びチーム分類）の情報も検出することも可能である。

【0089】

40

更に、動作状況特徴量 M^i_b を機械学習の識別器に加えることで、柔道などでの投げ動作などの動きに応じて生じる衝撃を認識することが可能となる。従って、人物動作認識部126は、その動作認識において、人物以外の動オブジェクトの情報を参照する技法もあるが、本例では人物以外の動オブジェクトを参照できない場合でも、高精度に柔道競技等の人物を認識するため、動作状況特徴量 M^i_b を用いるようにしている。この動作状況特徴量 M^i_b を加味することで、人物の動きの検出精度や動きの種類の判別を向上させている。

【0090】

50

ただし、これらの選定した骨格軌跡集合 T^i_b の軌跡特徴量及び動作状況特徴量 M^i_b の各特徴量の変化は、その全特徴量が同時刻に変化するように発生するとは限らない。そこで、 $T=15$ フレームなど一定時間の時間窓を設けた特徴量とする。一定時間内の各特徴量を用いてSVMなどで機械学習を行うことで、各特徴量の変化に時間的なズレが生じ

ても頑健な識別器を構成することができる。

【0091】

最終的に、触覚メタデータ生成装置12は、第1のメタデータ生成部127により、現フレーム画像に対応して、各人物の識別、位置座標、（及びチーム分類）、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す衝撃提示用の情報を含む第1の触覚メタデータ（衝撃提示用）を生成し、フレーム単位で制御ユニット13に出力する（ステップS8）。

【0092】

また、触覚メタデータ生成装置12は、第2のメタデータ生成部129により、現フレーム画像に対応して、各人物オブジェクトの人物動きの状況変化量を示す動作状況提示用の情報を含む第2の触覚メタデータ（動作状況提示用）を生成し、フレーム単位で制御ユニット13に出力する（ステップS9）。

【0093】

そして、触覚メタデータ生成装置12は、映像出力装置10から映像のフレーム画像が入力される度に、ステップS1乃至S9の処理を繰り返す。

【0094】

人物動きの状況変化量 $K(t)$ による第2の触覚メタデータに応じて制御ユニット13が触覚提示デバイス14L、14Rを制御することで、各選手が組み合った際の近接時、且つ投げようと体が大きく動いた際に対応する触覚提示デバイスを大きく振動させることができる。制御ユニット13は、第1の触覚メタデータについては衝撃を表す単発的な刺激を与えるに留まるが、動作状況を示す第2の触覚メタデータを併せて取得し、対応する触覚提示デバイスを制御することで、映像コンテンツの状況を連続値で恒常的に提示することができ、ユーザUは、常に各人物オブジェクトOp1、Op2の動作状況を把握することができるようになる。

【0095】

人物動きの状況変化量 $K(t)$ を可視化した例を図8に示している。図8(a)の左図は、人物オブジェクトOp1、Op2（選手）と人物オブジェクトOp3（審判）の動作状況“組み合い”の様子が見て取れる現フレームのフレーム画像Fであり、図8(a)の右図は、過去のフレーム画像の時刻tpから現フレーム画像の時刻tcまで（15フレームの経過期間）の最大値100で正規化した人物動きの状況変化量 $K(t)$ の大きさを表している。図8(a)の右図から理解されるように、“組み合い”の動作状況時（投げの動作状況時も同様）には人物動きの状況変化量 $K(t)$ が時間経過とともに大きくなり、各人物の動作量が増えていることが判別できる。

【0096】

また、図8(b)の左図は、人物オブジェクトOp1、Op2（選手）と人物オブジェクトOp3（審判）の動作状況“抑え込み”の様子が見て取れる現フレームのフレーム画像Fであり、図8(b)の右図は、過去のフレーム画像の時刻tpから現フレーム画像の時刻tcまで（15フレームの経過期間）の最大値100で正規化した人物動きの状況変化量 $K(t)$ の大きさを表している。図8(b)の右図から理解されるように、“抑え込み”の動作状況時には人物動きの状況変化量 $K(t)$ が時間経過とともに小さくなり、各人物の動作量が減少していることが判別できる。

【0097】

また、図8(c)の左図は、人物オブジェクトOp1、Op2（選手）と人物オブジェクトOp3（審判）の動作状況“待て”の様子が見て取れる現フレームのフレーム画像Fであり、図8(c)の右図は、過去のフレーム画像の時刻tpから現フレーム画像の時刻tcまで（15フレームの経過期間）の最大値100で正規化した人物動きの状況変化量 $K(t)$ の大きさを表している。図8(c)の右図から理解されるように、“待て”の動作状況時（試合開始前、両選手の距離が離れている際も同様）には人物動きの状況変化量 $K(t)$ が時間経過を経ても小さく、各人物の動作量がほとんどないことが判別できる。

【0098】

10

20

30

40

50

また、人物動きの状況変化量 $K(t)$ は動作状況特徴量 M^i_b に含まれる要素であり、この動作状況特徴量 M^i_b を用いることで、式(4)に示すようなif-thenルールを用いても各人物の動作認識を高精度に行うことができるようになる。即ち、図8(c)に示すように、選手2人と審判を独立に認識し、各人の関節の推移を軌跡特徴量から判別することで、審判が手を挙げて「待て」の動作を自動認識することが可能となる。触覚メタデータ生成装置12が、「待て」の動作に応じた第1及び第2の触覚メタデータを触覚提示デバイス14L, 14Rに出力することで、触覚提示デバイス14L, 14Rを利用するユーザUは、連続した一定量の刺激提示からほぼゼロの刺激提示に強制的に示すことが可能となり、「待て」の前後の動作で、試合開始前であるかのような試合状況を誤解させるおそれも少なくなる。従って、本実施形態の触覚メタデータ生成装置12は、スポーツ映像のリアルタイム視聴時でも触覚刺激を人物の動きに応じて提示することが可能となる。

10

【0099】

(制御ユニット)

図9は、本発明による一実施形態の映像触覚連動システム1における制御ユニット13の概略構成を示すブロック図である。制御ユニット13は、メタデータ受信部131、解析部132、記憶部133、及び駆動部134-1, 134-2を備える。

【0100】

メタデータ受信部131は、触覚メタデータ生成装置12から第1の触覚メタデータ(衝撃提示用)及び第2の触覚メタデータ(動作状況提示用)を入力し、解析部132に出力する機能部である。第1の触覚メタデータは、現フレーム画像内の各人物の識別、位置座標、(及びチーム競技であればそのチーム分類)、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す情報を含む。第2の触覚メタデータは、図8に例示した動作状況の情報を含む。

20

【0101】

解析部132は、触覚メタデータ生成装置12から得られる第1及び第2の触覚メタデータを基に、予め定めた駆動基準データを参照し、駆動部134-1, 134-2を介して、対応する各触覚提示デバイス14L, 14Rの振動アクチュエーター142を駆動するように制御する機能部である。例えば、解析部132は、一方の選手が組合から投げ動作に移行するときは、第2の触覚メタデータにおける動作状況に応じた振動提示に加えて、第1の触覚メタデータにおける人物の識別、位置座標、(及びチーム分類)、並びに、触覚提示デバイスを作動させるタイミング及び速さから、予め定めた駆動基準データを参照して、触覚提示デバイス14Lの振動アクチュエーター142の作動タイミング、強さ、及び動作時間を決定して駆動制御する。

30

【0102】

記憶部133は、第1及び第2の触覚メタデータに基づいた駆動部134-1, 134-2の駆動を制御するための予め定めた駆動基準データを記憶している。駆動基準データは、第1及び第2の触覚メタデータに対応付けられた触覚刺激としての振動アクチュエーター142の作動タイミング、強さ、及び動作時間について、予め定めたテーブル又は関数で表されている。また、記憶部133は、制御ユニット13の機能を実現するためのプログラムを記憶している。即ち、制御ユニット13を構成するコンピュータにより当該プログラムを読み出して実行することで、制御ユニット13の機能を実現する。

40

【0103】

駆動部134-1, 134-2は、各触覚提示デバイス14L, 14Rの振動アクチュエーター142を駆動するドライバである。

【0104】

このように、本実施形態の触覚メタデータ生成装置12を備える映像触覚連動システム1によれば、映像から人物オブジェクトを自動抽出し、動的な人物オブジェクトに対応する触覚メタデータを同期して自動生成することができるので、触覚提示デバイスと映像を連動させることができるようになる。そして、人物オブジェクトに生じる「衝撃の種類とタイミング」を表す第1の触覚メタデータだけでなく、人物オブジェクトに係る連続的な

50

「動作状況」を表す第2の触覚メタデータをも自動生成できるようになる。これにより、スポーツ映像のリアルタイム視聴時での触覚刺激の提示が可能となり、更には、人物オブジェクトの重なりやオクルージョンが生じやすい柔道等の試合映像での詳細な触覚提示も可能となる。つまり、視覚・聴覚への情報提供のみならず、触覚にも訴えることで、より詳細に、視覚や聴覚に障害を持つ方々へもスポーツの状況を分かりやすく伝えることが可能となる。さらに、視覚・聴覚の感覚を有する健常者の方々にとっても、従来の映像視聴では伝えきれない臨場感や没入感を提供することができる。特に、スポーツ映像視聴に際し、各選手の識別、位置座標、（及びチーム分類）、並びに、触覚提示デバイスを作動させるタイミング及び速さを示す情報を含む触覚メタデータを生成することで、より精度よく、より詳細に、1台以上の触覚提示デバイスにより、プレーの種類、タイミング、強度などに関する触覚刺激をユーザに提示できるようになる。

10

【0105】

尚、上述した一実施形態の触覚メタデータ生成装置12をコンピュータとして機能させることができ、当該コンピュータに、本発明に係る各構成要素を実現させるためのプログラムは、当該コンピュータの内部又は外部に備えられるメモリに記憶される。コンピュータに備えられる中央演算処理装置（CPU）などの制御で、各構成要素の機能を実現するための処理内容が記述されたプログラムを、適宜、メモリから読み込んで、本実施形態の触覚メタデータ生成装置12の各構成要素の機能をコンピュータに実現させることができる。ここで、各構成要素の機能をハードウェアの一部で実現してもよい。

20

【0106】

以上、特定の実施形態の例を挙げて本発明を説明したが、本発明は前述の実施形態の例に限定されるものではなく、その技術思想を逸脱しない範囲で種々変形可能である。例えば、上述した実施形態の例では、主としてバドミントン競技の映像解析を例に説明したが、柔道や卓球、その他の様々なスポーツ種目、及びスポーツ以外の映像にも広く応用可能である。例えば、触覚情報を用いたパブリックビューイング、エンターテインメント、将来の触覚放送などのサービス性の向上に繋がる。また、スポーツ以外の例として、工場での触覚アラームへの応用や、監視カメラ映像解析に基づいたセキュリティシステムなど、様々な用途に応用することも可能である。従って、本発明は、前述の実施形態の例に限定されるものではなく、特許請求の範囲の記載によってのみ制限される。

30

【産業上の利用可能性】

【0107】

本発明によれば、映像から人物オブジェクトを自動抽出し、動的な人物オブジェクトに対応する触覚メタデータを同期して自動生成することができるので、触覚提示デバイスと映像を連動させる用途に有用である。

【符号の説明】

【0108】

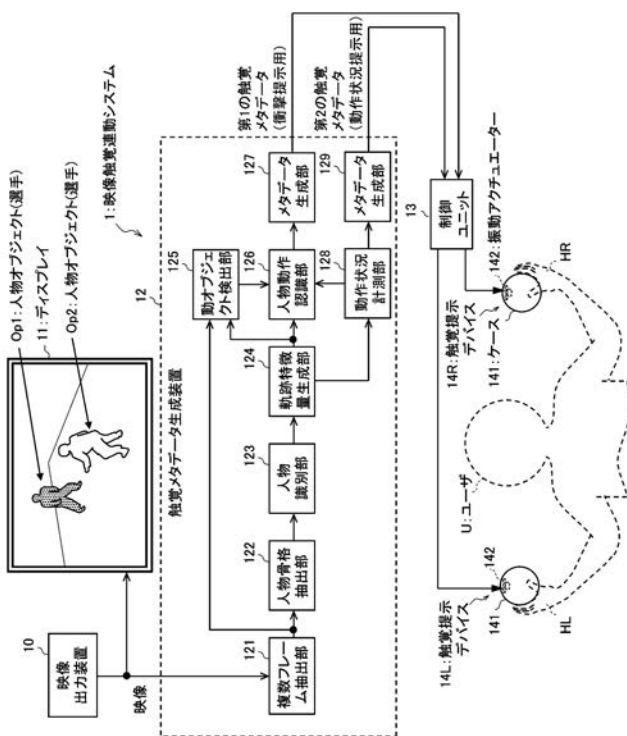
- 1 映像触覚連動システム
- 10 映像出力装置
- 11 ディスプレイ
- 12 触覚メタデータ生成装置
- 13 制御ユニット
- 14 L, 14 R 触覚提示デバイス
- 121 複数フレーム抽出部
- 122 人物骨格抽出部
- 123 人物識別部
- 124 軌跡特徴量生成部
- 125 動オブジェクト検出部
- 126 人物動作認識部
- 127 第1のメタデータ生成部
- 128 動作状況計測部

40

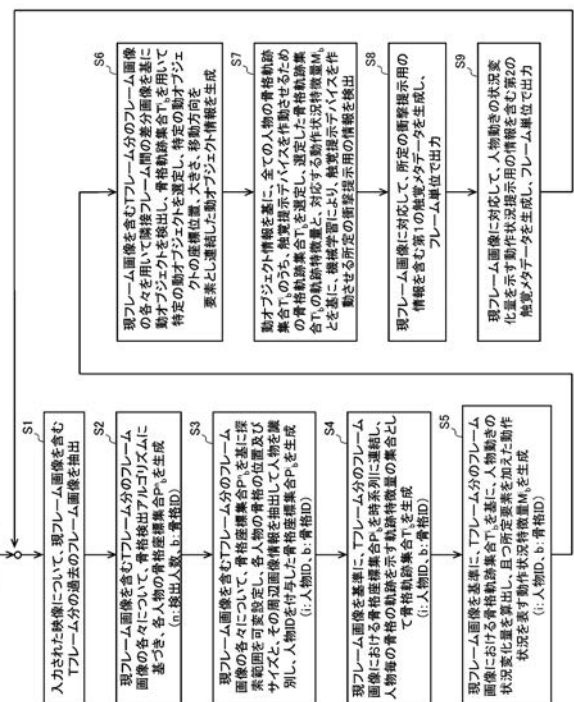
50

- | | |
|-----------------------|---------------|
| 1 2 9 | 第 2 のメタデータ生成部 |
| 1 3 1 | メタデータ受信部 |
| 1 3 2 | 解析部 |
| 1 3 3 | 記憶部 |
| 1 3 4 - 1 , 1 3 4 - 2 | 駆動部 |
| 1 4 1 | ケース |
| 1 4 2 | 振動アクチュエーター |

【 図 1 】

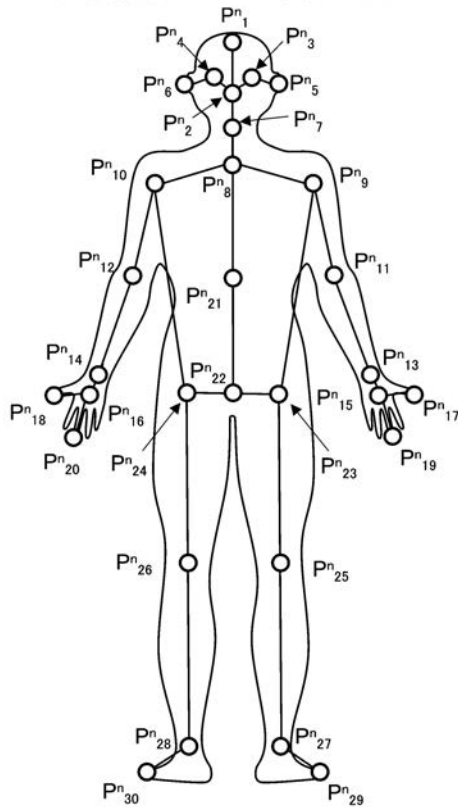


【圖 2】



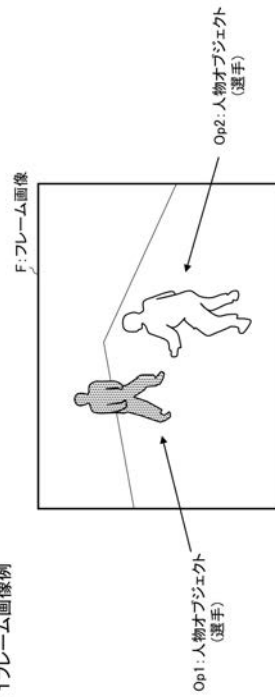
【図3】

〔人物骨格抽出における骨格点の例〕



【図4】

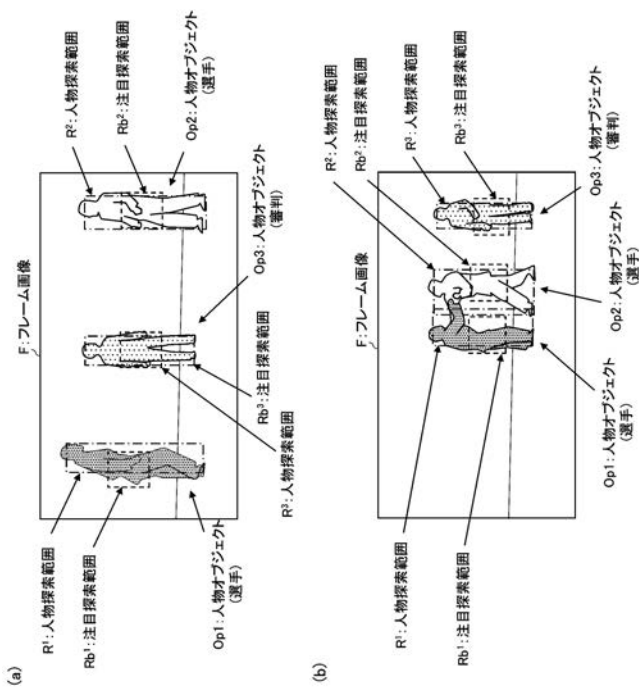
(a) 1フレーム画像例



(b) 1フレーム画像における人物骨格抽出例

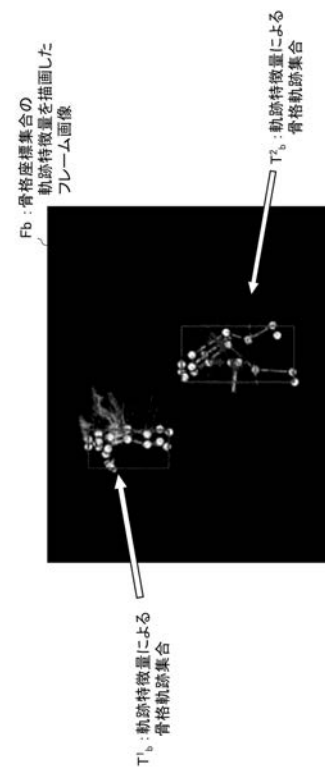


【図5】

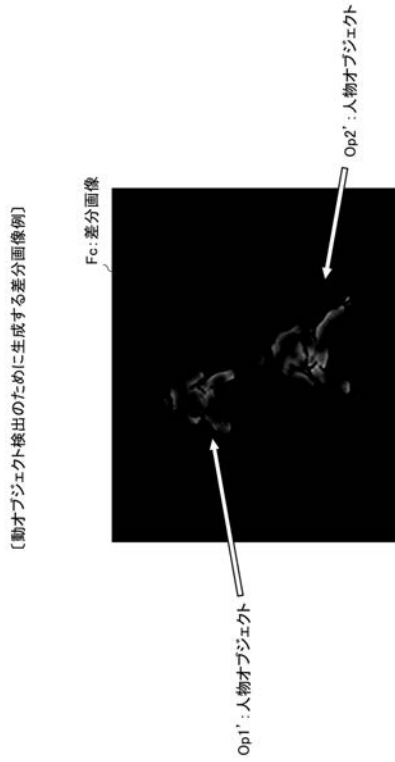


【図6】

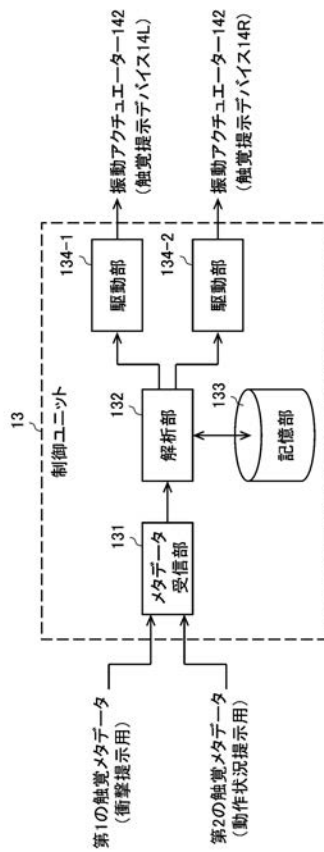
〔軌跡特徴量に対応する人物骨格画像例〕



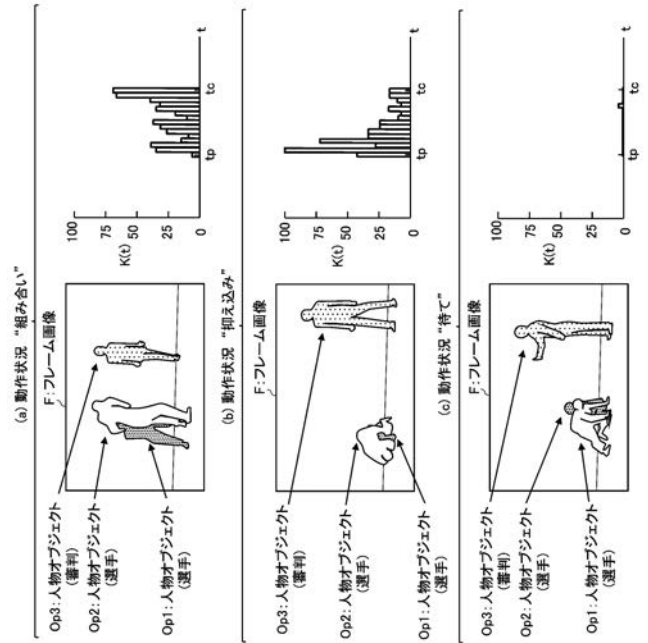
【図 7】



【図 9】



【図 8】



フロントページの続き

(72)発明者 佐野 雅規

東京都世田谷区砧一丁目 1 0 番 1 1 号 日本放送協会放送技術研究所内

Fターム(参考) 5L096 AA01 AA06 CA04 DA01 FA32 FA60 FA64 FA66 FA69 GA08
GA38 GA55 HA04 HA11 JA11 KA04