様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成２５年５月３０日現在

機関番号：１４３０１
研究種目：若手研究（B）
研究期間：2011～2012
課題番号：23700170
研究課題名（和文）ダイナミック　コンピュータ　ビジョン

研究課題名（英文）Dynamic Computer Vision (DCV)

研究代表者
TUNG, Tony（タントニー）
京都大学・学術情報メディアセンター・助教
研究者番号：30586061

研究成果の概要（和文）：
人間の振る舞いやインタラクションを，それらが持つ動的な特徴およびそれらの織りなすタイミング構造によって記述する，Dynamic Computer Vision（DCV）と呼ばれる新たなコンセプトの提案を目指している．
本研究では，グループ対話における複数人のマルチモーダルインタラクションを計測する新たなシステムを設計した．
今後，ここで得られたデータを DCV の枠組みで記述することで，多くの研究分野における更なる発展，新たな研究の展望・応用がもたらされると期待される．

研究成果の概要（英文）：
We have aimed to introduce Dynamic Computer Vision (DCV), a novel concept to understand human behavior and interaction using timing structure of dynamic features. Particularly, we designed a novel system that handles multiple people multimodal interaction in group communication. We believe human behaviors can be captured using DCV and that this model will make a significant breakthrough in many research areas and bring new research perspectives and applications.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	3,400,000	1,020,000	4,420,000

研究分野：情報学
科研費の分科・細目：
キーワード：コンピュータビジョン, インタラクション, ３次元ビデオ映像

１．研究開始当初の背景

話すテンポや頷き、ちゅうちょなどの微妙なジェスチャにおけるダイナミクスは、スムーズで自然なコミュニケーションを具現化し、複雑な事象を特徴付けするのに、重大な役割を果たしている。時間や行動の表示に対しては、過去十数年にわたり注目が向けられている。近年では、人間の言語的・非言語的コミュニケーションのモデル化において、有望な結果が得られてきた。例えば、意図的な笑顔と無意識な笑顔とを区別するために、表情のハイブリッドライナーダイナミカルシステムの時間構造が利用されていたり、唇の動作モデルや話者の検出のために、音の時間構造が利用されたりしている。または、頭部と目のジェスチャ認識に対して、潜在的・動的を区別するモデルが利用されてきた。

しかし、これらの研究成果が出ているにもかかわらず、コミュニケーションのダイナミクスと関連する研究は、単一視点からの映像の流れを利用した顔のジェスチャにのみ重点が置かれており、複数人間間のインタラクションのダイナミクスや、複数の性質を理解・利用する人間の行動においては、複数視

点ビデオカメラを含むシステムを利用した研究が見られない。

２．研究の目的

コンピュータビジョン分野では、ジェスチャ認識など１人の動作の分析・認識については数多くの研究がなされてきたが、複数人物間で創発されるインタラクションの分析についてはほとんど研究がなされていない。そのためインタラクションのダイナミクスに焦点を当てた本研究は、大きな新規性・独創性を有するものと考えている。特に本研究では、インタラクションを機械学習のフレームワーク内で学習段階から認識段階にかけてアクションとリアクションの関係、つまり原因と結果の関係に焦点を当ててモデル化する。

本研究では、動的システムの階層的ネットワークを用いて複雑なイベントをモデル化する。そして、random subspace classifier を用いてアクション・リアクションペアの分類をおこなう。

今後は、多視点映像入力に対するアルゴリズムの開発を行うとともに、インタラクションの間の分析に焦点を当て、グループコミュニケーションの場の状況理解へと展開を図る計画である。

３．研究の方法

ポスター発表などのグループコミュニケーションにおける参加者の意識・無意識的な動作・仕草を検出・分類し、action-reaction 動作の持つダイナミクスを Hybrid Dynamical System (HDS)を用いて分析する（Fig. 1）。具体的には、予め学習され action-reaction 動作のパターンに基づいて、計測された action 動作によって誘引される reaction 動作を予測するアルゴリズムを開発し、グループコミュニケーション多視点映像撮影システムで得られた映像データに対して適用し、その有効性を評価する。

この中で特に本研究では新しいインタラクションのモデルを用いて分析を行う予定である。アイディアのポイントは複数 HDS で仕草や頭、腕、上体など全身のモーションダイナミクスを分析する点である。モーションの検出のために適応 variational optical flow を用いることとし、HDS の random subspace classifier による階層分類体系を用いた分析を行う。

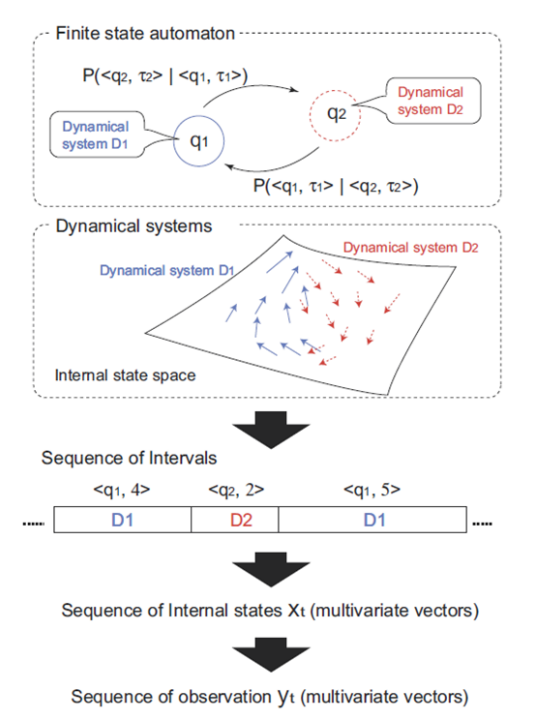

Figure 1. HDS モデル

まず、手や頭といった人体における特定の部位を含む領域のオプティカルフローを獲得する。ここでの人体部位は、顔検出といった既存のコンピュータビジョン技術によって獲得する。提案モデルでは、アクションが起こる時区間を HDS によって獲得し、同時刻に起こる部位のアクション、およびそれらの相互依存関係を表現する。

グループコミュニケーションを例にあげると、ある聞き手は話者や他の聞き手が起こす刺激に反応する。このような刺激や反応といった異なるイベント間のタイミング構造によって、複数信号間の依存関係を表現できることが期待される。

それから HDS の random subspace classifier 一段階層分類体系を用いて分析する。この分類器を動的マッチング手法と統合することで、実時間での認識を行うことができる。

想定するポスター発表のグループコミュニケーションの状況：話者１人（話・説明、手の仕草、頭の動きなど)、聴衆３人（頷き・頭の動きなど)。
主たるインタラクションは話者と聴衆の間でなされるが、聴衆と他の聴衆の間でのインタラクション、例えば他人のうなずきを真似するなどの仕草も対象とする。

多視点映像撮影システムデザイン：キャリブレーション済みの同期撮影カメラ
PointGrey の Grasshopper ビデオカメラ

(1280*960@30fps)４台、マイクロソフト XBOX キネクトセンサー２台、パソコン PC INTEL COREi7 3GBRAM１台、FireWire800 ケーブル、SSD RAM 64GB２台、50 型プラズマディスプレイ(Fig. 2)。

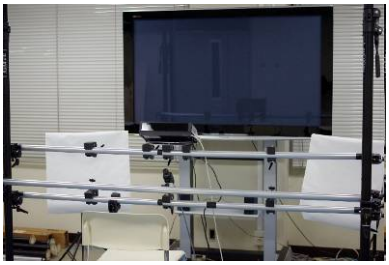

Figure 2. システム

インタラクションシステムを実現するためにはリアルタイム処理が不可欠である。そのため新しいインタラクションモデルを含むフレームワークとして、ポスター発表の様な現実の場面でも適用することが出来、さらにグループミーティングにおけるインタラクションを支援する拡張も可能となるものを開発する。また、多数のカメラを用いて多視点からの映像を撮影することで、参加者が多数のインタラクションも支援することができることを目指す。

提案するインタラクションモデルは、異なる種類の信号を扱うことができる。特に本研究では、マイクを用いて音声信号を獲得する。発話といった音声信号と表情変化や手の動きといったジェスチャはしばしば同期するため、提案モデルによってそのインタラクションを表現することができる。すなわち、このようなマルチモーダルシステムによって、複数人間の言語・非言語コミュニケーションを解析することが可能となる。

４．研究成果

ポスター発表などのグループコミュニケーションにおける参加者の意識・無意識的な動作・仕草を検出・分類し、action-reaction 動作の持つダイナミクスを Hybrid Dynamical System (HDS)を用いて分析した.

具体的には、予め学習された action-reaction 動作のパターンに基づいて、計測された action 動作によって誘引される reaction 動作を予測するアルゴリズムを開発し、多視点グループコミュニケーション撮影システムで得られた映像データに対して適用し、その有効性を評価した。我々は，複数 HDS で仕草や頭、腕、上体など全身のモーションダイナミクスを分析する新たなインタラクションのモデルを導入し、映像データの分析を行った。

開発したシステムは、複数の被験者が大型ディスプレイの非常に近くに立ち、一般的な深度センサの計測可能範囲外にある状態を想定した設定となっている。
これは国際会議などでのポスター発表を想定したものである。
当該システムはスマートポスターシステムとして平成２４年３月に国際会議 IEEE ICASSP 2012 にてデモ発表された。システムは実時間での複数の映像・音声の同期撮影を実現し、顔検出もほぼ実時間で処理となっている。

また上記のマルチモーダルインタラクションモデルは、その基礎アルゴリズムを国際会議 ECCV-Workshop2012 にて発表した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１件）

① Tony Tung and Takashi Matsuyama: Topology Dictionary for 3D Video Understanding,IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.34, No. 8., pp. 1645-1657, 2012.8.(Selected as Spotlight Paper)

〔学会発表〕（計２件）

① Tony Tung and Takashi Matsuyama: Invariant Surface-Based Shape Descriptor for Dynamic Surface Encoding, Asian Conference on Computer Vision (ACCV'12), Lecture Notes in Computer Sciences (LNCS), Springer, Daejeon, Korea, 2012.11.7 (acceptance rate: 231/869, oral: 31, poster: 200)

② Tony Tung, Randy Gomez, Tatsuya Kawahara, and Takashi Matsuyama: Group Dynamics and Multimodal Interaction Modeling using a Smart Digital Signage,European Conference on Computer Vision (ECCV2012), Ws/Demos, Lecture Notes in Computer Sciences (LNCS), Springer, Part I, Vol. 7583, pp. 362-371, Florence, Italy, 2012.10.7

〔図書〕（計１件）

① Takashi Matsuyama, Syohei Nobuhara, Takeshi Takai, and Tony Tung: 3D Video

and Its Applications, Springer, 2012.6.

〔その他〕
ホームページ等
Webpage of Tony TUNG
http://tonytung.org/

６．研究組織
(1)研究代表者
TUNG,Tony（タントニー）
京都大学・学術情報メディアセンター
助教
研究者番号：30586061