

Title	異機種端末に適応した個人化ビデオの自動生成
Author(s)	CHEN, Fan
Citation	科学研究費助成事業研究成果報告書: 1-5
Issue Date	2014-06-04
Type	Research Paper
Text version	publisher
URL	http://hdl.handle.net/10119/12181
Rights	
Description	研究種目: 若手研究(B), 研究期間: 2011~2013, 課題番号: 23700110, 研究者番号: 10576432, 研究分野: 情報学・総合領域, 科研費の分科・細目: メディア情報学・データベース

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 4 日現在

機関番号：13302

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700110

研究課題名(和文)異機種端末に適応した個人化ビデオの自動生成

研究課題名(英文)Autonomous Production of Personalized Videos for Heterogeneous Terminal Devices

研究代表者

CHEN Fan (CHEN, FAN)

北陸先端科学技術大学院大学・情報科学研究科・助教

研究者番号：10576432

交付決定額(研究期間全体)：(直接経費) 3,200,000円、(間接経費) 960,000円

研究成果の概要(和文)：複数撮影視点からの情報統合により、高精度な追跡や映像解析ができるほか、映像の表現方式にも高い自由度を与えた。我々はスポーツテレビ及び映像監視システムを応用事例とし、語義上の内容選択、時間軸上の要約及び空間軸上の画面構成の三つ階層から、異機種端末に適応した個人化ビデオの自動生成を検討した。追跡された対象を群相互作用により個体群に分ける。個体群の間に関覧対象を遷移し、配信内容を選択する。更に、ユーザーの好みと端末の性能に対する最適な再生速度及び画面構成を自動的に決定する。研究期間中で特許一点、論文誌論文五篇、解説論文一篇、国際会議論文十一篇(全査読付き)を発表し、展示会に二回出展した。

研究成果の概要(英文)：Multi-view information fusion not only enables accurate video analysis (e.g. object tracking and scene understanding), but also provides high flexibility to content presentation. Given sport video broadcasting and intelligent video surveillance as two use cases, we discussed the production of personalized videos for better presentation on various terminal devices, by considering content remixing from three different levels, i.e. semantic story organization, temporal summarization and spatial camerawork planning. After dividing tracked objects into mutually interactive groups, we organize the story by shifting the browsing target between interactive groups. The story is then rendered with adaptive fast-forwarding and virtual zooming to better satisfy the conditions defined by users preferences and device limitations. We have applied for 1 patent, published 5 reviewed journal papers, 1 tutorial paper, and 11 reviewed international conference papers, and attended two exhibitions.

研究分野：情報学・総合領域

科研費の分科・細目：メディア情報学・データベース

キーワード：個人化ビデオの自動生成 カメラ視点の自主決定 自動要約システム 物体追跡及び事象理解

1. 研究開始当初の背景

情報端末の多様化が急速に広がっている日本では、大画面テレビでブルーレイ映画を鑑賞する人もいるし、低解像度携帯端末でワンセグ番組を楽しむ人もいる。鑑賞者の背景と鑑賞設備の性能によって、同じ映像を鑑賞しても、満足度は大きく変わる。高画質のニュースやスポーツ動画を低解像度端末で縮小表示すると、重要な内容(例えば、株価/天気情報、選手情報など)が見づらくなる例が挙げられる。映像監視システムの場合にも、制御室での大画面閲覧と巡回中警備員の手持ち端末の両方に配信内容を適応的に調整すべきである。更に、鑑賞時間や通信状況などの制限があるため、全動画内容の受信を望ましくない場合もある。

端末の性能制限を解消するほかに、動画コンテンツの増加及び普及に伴い、各ユーザーの好みをそれぞれ満足する個人化コンテンツを提供できる配信サービスに対する要求が高まってきた。学校での試合動画を配信する際に、各保護者が自分の子供をより注目したいだろう。なお、このような中小規模配信システムには、映像の手動編集に関わるコストが高いため、自動編集システムが望まれる。もう一方、大規模な配信システムにも、自動編集が作業負担の低減、コストの削減及び大量なユーザーの同時対応に有利である。例として、映像監視システムが挙げられる。大量なカメラを監視する作業員の負担を低減し、人為的要素による誤判断を抑えるため、与えられた監視タスクに関連する映像内容を膨大な録画データから自動的に抽出し、分かりやすい要約にまとめる必要がある。

それゆえ、異種端末に向け、ユーザーの関心のある部分を抽出し、決められた解像度で快適に鑑賞できる個人化動画コンテンツを提供することは、ユーザーを中心とした配信サービスの必然的な発展であり、コンテンツ・プロバイダーが近い将来に解決すべき課題である。

撮影・計算設備の性能改善及び映像解析手法の進展が個人化映像配信サービスの実現に可能性を与えた。特に、スポーツ試合の撮影や映像監視システムなど制御される撮影環境には、カメラの配置関係や照明条件などを適切に管理することにより、高精度な映像解析結果が得られる。更に、カメラの急速な低価格化を背景に、多視点撮影の現実性が高まってきた。複数視点からの情報統合により、オクルージョンによる曖昧さを抑え、高精度な追跡や解析ができ、映像の表現方式にも高い自由度を与えた。

2. 研究の目的

本研究は、制御される撮影環境において、(自動的に)抽出された運動物体と事件情報に基

づき、カメラワークの自動決定とビデオ内容の自動要約の二つの問題を解決することで、高度な個人化ビデオの自動生成手法の提案を目指す。配信内容の時間制限に対して、重要事件を抽出し、時間軸上でストーリーを再構成することは、自動要約として知られ、よく研究されている。それに比べ、本研究で考える個人化動画の自動編集には、時間軸上の内容再構成のほかに、最適なカメラ視点及び運動パターンの決定による空間軸上での内容再組織も検討する。

具体的には、下記の特色のある個人化映像配信システムを実現したい。

(1) バーチャルカメラの運動パターンの自動決定。カメラ画面の振る舞いを自然に実現するために、カメラ操作の基本原則に基づき、バーチャルカメラの運動を検討する。

(2) 物体移動の軌跡に基づき、動的スケジューリングによるストーリー構成。カメラワークの連続性や完備性を保証するため、連続移動の物体軌跡を基本要素とし、要約問題を時空間制約上で軌跡描写の最適スケジューリング問題に変換し、軌跡間の時空間関係からストーリー構成を行なう。

(3) 視聴者とのインタラクション。視聴者が指定した物体や閲覧要求に従い、個人化ビデオを動的に再構成できる手法を検討したい。

最後には、個人化映像配信システムの実現に必要なとされる映像解析機能の改善・改良も本研究の目的に含まれる。

3. 研究の方法

本研究は三年間の研究プロジェクトとして計画された。検出、追跡及び映像理解などコンピューター・ビジョン的解析に加え、我々は映像情報の効率的表現方式を含むマルチメディア解析にも着目した。最初の二年間には、カメラワークの自動決定とビデオ内容の自動要約の二つの中心タスクに集中した。三年目では、インタラクションを考慮したオンライン生成など機能の追加及びデモをするための最適化調整を行った。特に、スポーツ試合と映像監視システムを応用事例として個人化配信システムの実現を議論した。

研究プロジェクトは五つのタスクに分けられ、順次改良で推進してきた。

(1) 実験データの準備。公開用データベースのほかに、多カメラ環境での撮影とデータ収集を行い、本プロジェクトのデータセットを収集した。

(2) 物体追跡と事件検出。特に複数視点からの情報統合による映像解析を議論した。

(3) バーチャル視点の運動決定。画面の完備性(多くの物体を含む)、詳細度(各物体を詳しく描写する)、連続性及びオクルージョンの解消を総合的に考慮し、最適視点の決定基を定義する。

(4) 動的スケジューリングによるストーリー構成。各カメラの位置関係、軌跡の相関性、及び事件発生の時間上の関連性などによって内容構成の効果を評価する。この評価関数の最大化により、最適な描写方式を決定する。
 (5) インタクションを考慮したオンライン生成。視点決定とストーリー構成に基づき、ユーザーとのインタクション方式を定義するほか、実時間処理のためのシステム最適化を行う。

最初年度で研究プロジェクトのホームページを立ち上げ、本研究の最新進展と共に関連データ及び論文を公開した。

4. 研究成果

本研究には、異種端末向けの個人化映像配信システムの実現を目指して、映像解析(人物追跡や事象理解など)から映像の内容構成まで一連の研究を行い、それぞれに新たな手法を提案した。

(1) 配信システム。異なる視点・時間の映像間で閲覧内容を飛ばす際に、理解の容易さが物語の連続性に大きく影響される。我々は語義上の内容選択、時間軸上の要約及び空間軸上画面構成の三つの階層から提示内容を決し、個人化配信内容の編集に係る手作業(早送り、カメラ切り替えなど)を自動化した(図1)。内容の連続性と理解のしやすさを意識し、関連する相互作用間に関覧内容の遷移を限定し、語義上で表現すべき内容を選択する。選ばれた内容に対して、時間軸上で再生速度の調整[TMM13]、空間軸上でカメラの選択及びデジタルズームを行い[TMM11]、最終的に出力映像を生成する。



図1 語義及び時空間上で配信内容を自動的に決定する。

(2) 人物追跡。人物追跡の目的は人物の位置を連続的に特定することにある。この問題は、更に空間的位置を特定する検出問題と時間的連結を行う追跡問題に分けられる。実に存在した物体のみ、各視点の画面に矛盾なく逆投影できる。これが複数視点人物検出手法の基本原理である。我々がボトムアップの検出手法と非実時間追跡による追跡を実現した(図2)。幾何変換(図3)¹により、人物の立つ方向を前景画像の縦方向に統一し、画素距離を

¹ 同じ幾何変換は三次元再構築にも使われる [APSIPAASC11]。

物理距離に正規化する。身長分の前景画素を累積することにより前景画像を統合し、信頼度マップを算出する。信頼度マップ上の極大点を順次に検出し、人物の位置を特定する。ハンガリーのアルゴリズムにより検出結果を連結し、得られた軌跡断片を高い階層で長い軌跡にまとめる。確率的追跡手法に比べ、階層的追跡手法は効率的に計算できるほか、不定長で中断した軌跡断片の連結に有効である。室内監視映像での実験結果として、複数人物を精度よく追跡できた(>90%) [ROMAN13]。

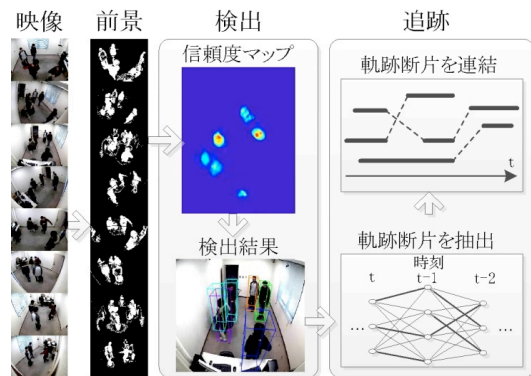


図2 我々が実現した物体追跡の流れ



図3 画像を統合するための幾何校正

(3) 事象理解。事象理解は映像の適応符号化、記憶域と転送コストの節約、映像の要約技術及び機械の自動制御など幅広い応用に使われる重要な基盤技術である。複雑な事件を理解するためには、各人物の独自の行動を分析するほか、人物と周辺環境(人物や背景物)との相互作用も考慮すべき。

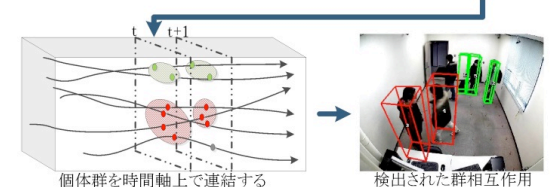
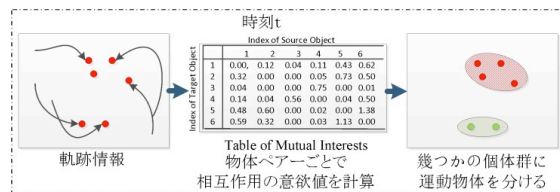


図4 群相互作用の検出による異常検知 [ICASSP13]

均一速度で直線的移動を好むや移動方向の前方に重視するなど歩行者の行為特徴に関する観察に基づき、群相互作用の自動検出による異常検知手法が提案された(図4) [ICASSP 13]。異常の有無に加え、異常に関わ

る人物及び期間も決定される。人物間の速度・位置関係から、お互いに相互作用を行う意欲を評価する。フレームごとで強い意欲を持つ個体群を抽出する。物体追跡の考えを用いて、個体群を時間軸上で連結し、安定的に相互作用を行う個体群を抽出する。群相互作用の検出により、約 80%の異常事件を検出できることが示された[ICASSP13]。ノイズのある軌跡に対する手法の頑健性も議論された。

(4) 個人化配信内容の再構成。個人化内容再構成には、内容の関連性、物語の連続性、閲覧上の快適さを同時に考慮に入れるべき。ビデオ内容に含まれる対象、時間軸上の物語の構成及びカメラ画面の構成を描写方式の三要素として定義し、入力データの表現上最も適切な描写方式を推定することで、個人化内容の生成問題を確率推論による最適化問題に変換する。物語の完備性(関連する内容を含めること)、連続性(急な画面遷移やストーリーの突然中断が起きないこと)及び繊細さ(対象を描写するために必要以上の解像度及び再生時間を保証すること)から基準値を定義し、最適な配信内容を求める。

内容の連続性を保証するため、監視映像の要約には早送りがよく使われる。24 時間の映像を 10 分間にまとめるなど高圧縮比の要約タスクには、再生速度が理解できる範囲を超え、ちらつきによる不快感が生じやすい。それを抑えるため、早送り及び内容の飛ばしを適応的に切り替える要約手法が提案された[TMM13]。決められた閲覧時間を各映像内容に分配することで、時間軸上の要約を資源配分問題として解決した(図5)。

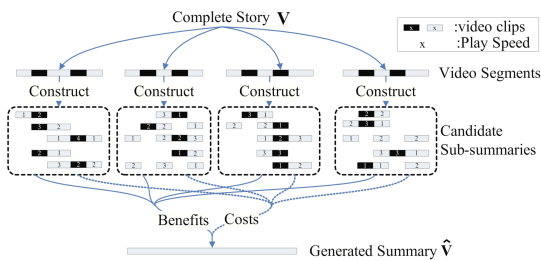


図5 資源配分による時間軸上内容再構成

空間軸上の画面構成にも、人物の投影後の大きさ、端末で表示するための画像ダウンサンプリングによる画質低下、及びオクルージョンによる可視率低下を総合的に評価する必要がある(図6) [TMM11]。砂箱実験での検証結果に示されたように、カメラ画面構成に関する総合的評価基準にはダウンサンプリングとオクルージョンの両方とも少ない斜め上方向が好まれている(図7)。前述手法で生成された閲覧用監視映像の性能を更に一連の主観・客観実験により検証した。

論文発表のほかには、特許申請及び展示会出展なども行った。

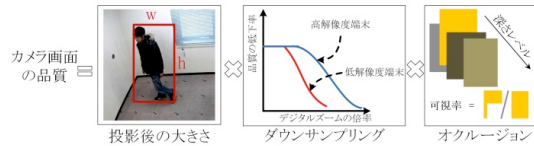


図6 閲覧上の快適さを意識した画面構成

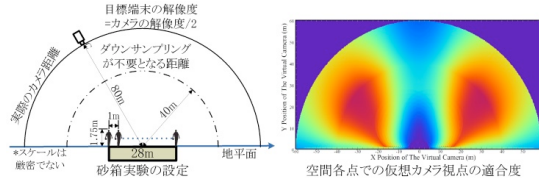


図7 砂箱実験による画面構成基準の検証

F. Chen, "監視システムにおける監視画像の統合、閲覧及び配信方法," Innovation Japan 2013, Aug.29-30, 2013, Tokyo, Japan.

F. Chen, "異種端末向けの個人化ビデオ配信サービス," 北陸技術交流テクノフェア 2011, Oct.20-21, 2011, Fukui, Japan.

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6件)

(1) Fan Chen, Christophe De Vleeschouwer, and Andrea Cavallaro, "Resource Allocation for Personalized Video Summarization," IEEE Transactions on Multimedia, 査読有, Vol.16, 2014, pp.455-469 (doi:10.1109/TMM.2013.2291967).

(2) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "Extraction of Discriminative Patterns from Skeleton Sequences for Accurate Action Recognition," Fundamenta Informaticae, 査読有, Vol.130, 2014, pp.247-261 (doi:10.3233/FI-2014-991).

(3) Fan Chen, "監視映像の解析技術の研究—監視映像の理解から内容の表現方式まで," 検査技術, 日本工業出版, 査読無, 3月号, 2014, pp.1-6.

(4) Fan Chen and Christophe De Vleeschouwer, "Personalized Summarization of Broadcasted Soccer Videos with Adaptive Fast-forwarding," Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 査読有, Vol. 124, 2013, pp.1-11, (doi:10.1007/978-3-319-03892-6).

(5) Fan Chen, Damien Delannay, Christophe De Vleeschouwer, "An Autonomous Framework to Produce and Distribute Personalized Team-Sport

Video Summaries: A Basketball Case Study,"
IEEE Transactions on Multimedia, 査読有,
Vol.13, 2011, pp.1381 - 1394
(doi:10.1109/TMM.2011.2166379).

(6) Fan Chen, Christophe De Vleeschouwer,
"Formulating Team-Sport Video Summarization as a Resource Allocation Problem," IEEE Transactions on Circuits System and Video Technology, 査読有, Vol.21, 2011, pp.193- 205
(doi:10.1109/TCSVT.2011.2106271).

[学会発表] (計 11 件/全査読付き)

(1) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "Automatic Extraction of Semantic Action Features," The 9th International Conference on Signal Image Technology & Internet-based Systems (SITIS 2013), Dec.2-5, 2013, Kyoto, Japan.

(2) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "An Apriori-like Algorithm for Automatic Extraction of the Common Action Characteristics," The IEEE Visual Communications and Image Processing (VCIP) 2013, Nov. 17-20, 2013, Kuching, Sarawak, Malaysia.

(3) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "An Apriori-like Algorithm for Automatic Extraction of the Common Action Characteristics," The 10th IEEE RIVF International Conference on Computing and Communication Technologies (RIVF) 2013, Nov. 10-13,2013, HANOI,Vietnam.

(4) Fan Chen, "Hot-spot Detection by Group Interaction Extraction from Trajectories," IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) 2013, Aug. 26-29, 2013, Gyeongju, Korea.

(5) Fan Chen and Andrea Cavallaro, "Detecting Group Interactions by Online Association of Trajectory Data," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2013, May.26-31, 2013, Vancouver, Canada.

(6) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "Automatic Extraction of Common Action Characteristics," IEEE Symposium on Signal Processing and Information Technology (ISSPIT) 2012, Dec. 12-15, 2012, Saigon, Vietnam.

(7) Fan Chen, and Christophe De Vleeschouwer, "Partial Motion Trajectory Grouping Through Rooted Arborecence," IEEE

International Conference on Image Processing (ICIP) 2012, Sep. 27-Oct.2, 2012, Orlando, Florida, USA.

(8) Than Thang Tran, Fan Chen, Kazunori Kotani and Bac Le Hoai, "Extraction of Discriminative Patterns from Skeleton Sequences for Human Action Recognition," The 9th IEEE RIVF International Conference on Computing and Communication Technologies (RIVF) 2012, Feb.27-Mar.1,2012, Ho Chi Minh City ,Vietnam.

(9) Itaru Tsuchida, Fan Chen, Junko Izawa and Kazunori Kotani, "High Quality Free Viewpoint Synthesis Using Multiview Images with Depth Information," IEEE International Symposium on Multimedia (ISM) 2011, Dec.5-7, 2011, Dana Point, CA, USA.

(10) Fan Chen, and Christophe De Vleeschouwer, "Robust Volumetric Reconstruction from Noisy Multi-view Foreground Occupancy Masks," the APSIPA Annual Summit and Conference (APSIPA ASC) 2011, Oct. 18-21, 2011, Xi' an, China.

(11) Fan Chen, and Christophe De Vleeschouwer,"Automatic Summarization of Broadcasted Soccer Videos with Adaptive Fast-forwarding," Int'l Conf. on Multimedia & Expo(ICME) 2011, Jul. 11-15, 2011, Barcelona, Catalonia, Spain.

[産業財産権]

○出願状況 (計 1 件)

名称：映像監視システム、映像処理サーバ及び監視映像の閲覧・配信方法

発明者：陳 凡

権利者：北陸先端科学技術大学院大学

種類：特許

番号：特願 2013-067773

出願年月日：2013年3月28日

国内外の別：国内

[その他]

ホームページ等

HP：<http://www.jaist.ac.jp/project/prime-proj/>

6. 研究組織

研究代表者

陳 凡 (CHEN FAN)

北陸先端科学技術大学院大学・情報科学研究科・助教

研究者番号：10576432