4C-03

[1]深層学習を用いた瞬き確率推定による ハイライト映像の自動抽出

○中野　珠実†　　阪田　篤哉‡　　岸本　章宏‡
大阪大学大学院生命機能研究科†
大阪大学大学院情報科学研究科‡

1. はじめに

近年,膨大な数のスポーツ映像が YouTube などのソーシャルコミュニケーションサイトにアップロードされている.映像の量が多くなればなるほど,TV のニュース報道のように、注目すべき重要なハイライトシーンだけを選りすぐった映像を視聴したい,というニーズが高まる.しかし,映像の中から重要なシーンを探しだし、それを編集するという作業は,多大な時間と労力を要する.そのため, マルチメディア処理の分野においてハイライトシーンを自動的に抽出する方法の開発が盛んに行われてきた[2].

これまでのハイライト検出は,映像の物理的な特徴解析に基づく教師無し学習手法が主流である[1, 3].しかし,この手法では,人間の映像に対する関心度を真に反映したハイライトの抽出ができていない可能性がある.そのため、表情や声、心拍数など映像視聴時の人間の情動・覚醒反応を基にハイライト検出する教師あり手法も開発されてきた[4, 5]。しかし、情動・覚醒反応は比較的ゆっくりとした時間で反応が変化するうえに、人間の関心度を直接反映していない可能性も高い。

そこで, 本研究は、高い時間精度で人間の関心度を鋭敏に表す客観的指標として、人間の自発的な瞬き行動に着目した.これまで我々は、同じ映像を視聴している人々の瞬きが 0.2 秒という高い時間精度で同期して発生していること、特に、ハイライトシーンでは一斉に抑制される一方,出来事の切れ目など暗黙裡の映像の切れ目で揃って発生することを発見した[6].つまり,人々の瞬きの発生確率は,人間が映像の文脈の中で,どのようなシーンに最も注意・関心を持っているのかの客観的かつ時間精度の高い指標として利用することができる.

そこで,映像の各フレームにおける瞬きの発生確率を深層ニューラルネットに学習させることで,新規の映像に対する瞬きの発生確率の推定値を算出し,その指標を基に映像のハイライトシーンを自動抽出する方法を提案する.

Blink Probability Estimation with Deep Learning for Automatic Video Highlight Extraction

†Tamami Nakano, Graduate School of Frontiers Bioscience, Osaka University
‡ Atsuya Sakata, Akihiro Kishimoto, Graduate School of Information Science and Technology, Osaka University

2. 映像に対する瞬き発生確率の推定

スポーツは個人競技からチーム競技まで様々な種類があるが,本稿では数分間にわたる連続したパフォーマンスの中に,複数のハイライト・イベントが織り込まれているフィギュアスケートのシングル競技を対象に選んだ.

まず,フィギュアスケート競技の映像(30 FPS)のデータセットを作成し,各映像を 12 人以上の参加者に実験室で視聴してもらった.その時の参加者の瞬きの発生タイミングを赤外光カメラ(Tobii Pro Spectrum, 120Hz)で検出し,映像の各フレームにおける瞬き発生確率（そのフレームが提示された瞬間に,目を閉じていた人数の割合）を算出した.

つぎに,スケートの動きの特徴を抽出しやすくするために,映像の各フレームにおける演技者の関節位置の 2 次元座標を OpenPose[7]を用いて推定した(図１左上).そして,数秒間にわたる 18 個の関節の時間変化を 2 次元画像にして,多層 1 次元畳み込みニューラルネットに入力し,その最終フレームにおける瞬き発生確率を推定させた(図１右上).ニューラルネットワークの学習は,推定値と実際の値の間の平均平方二乗誤差の最小化により最適化した.

ニューラルネットワークの性能評価は,推定された瞬き発生確率の時系列変化のパターンが,実際の値の時間変化と相関している度合いを比べた.また,ジャンプ,スピンといったフィギュアスケートにおける特徴的なハイライト・イベント前後での瞬き発生確率の時間変化のパターンの

相似性も評価した.

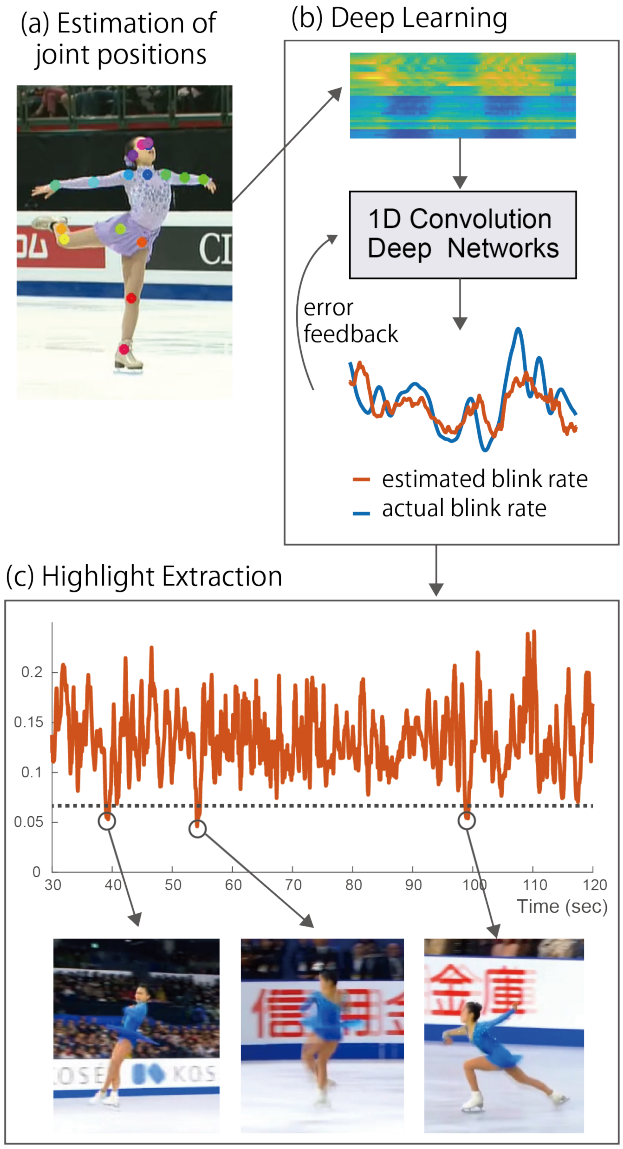

図1： 提案システムの概要
(a) 演技者の関節の位置座標の同定 (b) 瞬き確率を推定するニューラルネットワークモデル (c) 神経回路網が推定した瞬き確率の時系列データからハイライトシーンを自動検出

3. 瞬き推定値に基づくハイライトの抽出

我々の先行研究から,人々の瞬きが同時に抑制されているときは,皆が揃って高い関心を持っている重要な映像シーンであることがわかっている[6].そこで,ニューラルネットワークにより推定された瞬き発生率が著しく抑制されている時間帯を一定の条件の下で同定し,ハイライトシーンと定義した(図1下).さらに,自動的に抽出されたシーンがハイライトに値するかを,人間の評価者に評価させることで,ハイライトの抽出に成功しているかを検証した.

4. 結びに

フィギュアスケートの演技者の運動情報をニューラルネットワークに入力し,その瞬間の瞬き発生確率を推定させることで,人間の関心度に基づく映像のハイライトシーンを自動抽出する手法を開発した.瞬き情報は外部から容易にセンシングできるため,この手法を使えば,より人間の情報処理スタイルにマッチした映像のハイライトシーンの自動抽出を手間のかかる作業なしに行うことが可能になる.

今後はこの手法をフィギュアスケート以外のスポーツの映像や,日常的な人間の行動を撮影した映像にも適用できるのかを検証する必要がある.

謝辞

本研究の一部は JST さきがけ 11027, JSPS 科研 18H04084, 18H05522 の助成をうけたものである.

参考文献

[1] de Avila, S. E. F., Lopes, A. P. B., da Luz, A. and Araujo, A. D. VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. *Pattern Recogn Lett*, 32, 1 (Jan 1 2011), 56-68.

[2] Truong, B. T. and Venkatesh, S. Video abstraction: a systematic review and classification. *ACM Trans. Multimedia Comput. , Commun. Appl.*, 3, 1 (2007), 1-37.

[3] Wang, Z. K., Yu, J. Q., He, Y. F. and Guan, T. Affection arousal based highlight extraction for soccer video. *Multimed Tools Appl*, 73, 1 (Nov 2014), 519-546.

[4] Chakraborty, P. R., Tjondronegoro, D., Zhang, L. and Chandran, V. *Automatic identification of sports video highlights using viewer interest features*. ACM, City, 2016.

[5] Joho, H., Staiano, J., Sebe, N. and Jose, J. M. Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents. *Multimed Tools Appl*, 51, 2 (2011), 505-523.

[6] Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T. and Kitazawa, S. Synchronization of spontaneous eyeblinks while viewing video stories. *P Roy Soc B-Biol Sci*, 276, 1673 (Oct 22 2009), 3635-3644.

[7] Cao, Z., Simon, T., Wei, S. E. and Sheikh, Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *Proc Cvpr Ieee* (2017), 1302-1310.