

# 深層学習を用いた高精度リアルタイム人物映像解析技術

Accurate Real-Time Video Analysis System to Recognize Road Relay Race Teams in Live Images Using Deep Learning

山地 雄土 YAMAJI Yuto 小林 大祐 KOBAYASHI Daisuke 佐藤 誠 SATO Makoto

近年、スポーツ中継では、ICT（情報通信技術）を活用した戦況把握や、放送映像へ重畳表示するCG（コンピューターグラフィックス）の自動生成などの技術導入が始まっている。

東芝は、駅伝競走（以下、駅伝と略記）番組制作の効率化を目的に、走行中の選手のチームをリアルタイムに自動識別する駅伝中継映像解析システムを開発した。駅伝は、選手同士の重なりが多く、日照条件の変化も大きいので、自動識別が困難なタスクである。開発したシステムは、選手同士が重なっても見えやすいユニフォーム領域を識別の対象とし、様々な日照条件下でのロゴや色について深層学習したモデルを用いることで、高精度にチームを識別する。実際の駅伝中継放送で実証実験を行った結果、適合率98.1%と実用的な性能が得られ、人手で行っていた記録作業を大幅に省力化できた。

The field of sports broadcasting has recently seen an accelerating movement toward the introduction of functions utilizing information and communication technologies (ICTs) to grasp the situation of sports events and automatically create computer graphics (CG) superimposed on video images.

With the aim of improving the efficiency of live TV production of road relay races, Toshiba Corporation has developed a real-time video analysis system that automatically recognizes the team of each runner. In a road relay race, it is difficult to recognize runners' teams because of the frequent overlapping of runners and constantly changing outdoor lighting conditions. By using a deep learning model, our newly developed system makes it possible to recognize runners' teams based on their uniforms, which can be readily tracked even when runners overlap, and to identify each team with a high degree of accuracy based on learned uniform logos and colors under various outdoor lighting conditions. A demonstration experiment on this system at an actual road relay race confirmed that it achieves a practical team recognition accuracy of 98.1% in live broadcasting and contributes to the reduction of conventional manual confirmation work.

## 1. まえがき

近年のスポーツ中継では、ICTを活用した戦況把握や、放送映像へ重畳表示するCGを、自動生成する技術などの導入が始まっている。東芝は、スポーツ中継をターゲットとした映像解析技術の開発を進めており<sup>(1)</sup>、戦況把握の省力化や、選手の動きに追従して選手名などの情報を付与する新しい表現方法、解析したメタ情報を用いた過去映像からのシーン検索、などの機能の実現を目指している。

ここでは、駅伝の番組制作の効率化を目的とした映像解析技術について述べる。駅伝の番組制作では、戦況把握や、放送波に重畳するキャプションを作成するため、中継映像を常時目視しながらチーム名と走行位置をタグ付けする作業が行われている。しかし、目視作業が長時間に及ぶため、集中力が続かず、記録ミスが起こる問題があった。そこで、当社は、人手作業を省力化するため、中継映像に映る選手を自動的に検出・識別する映像解析技術を開発した（図1）。

駅伝は、屋外競技で日照変化も大きく、密集して走行す

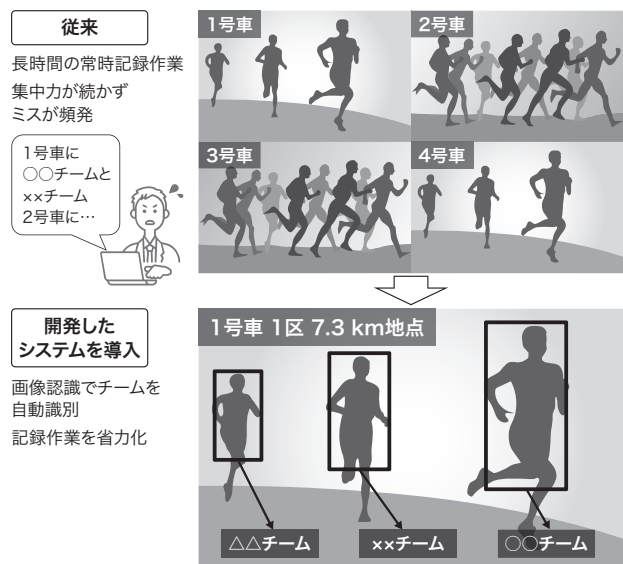


図1. 駅伝中継映像解析技術による記録作業の省力化

従来は、四つの中継映像を目視してチーム順位などの記録作業をしていたが、開発したシステムではチームを自動的に検出・識別することで記録作業を省力化した。

Reduction of manual confirmation of each team using real-time video analysis system for road relay races

ることあるため、画像認識にとっては難度の高いタスクである。また、多方面からの選手撮影が可能な屋内競技と異なり、中継車カメラで撮影された映像だけで認識する必要があるため、高い識別性能が要求される。

## 2. 駅伝中継映像解析技術の概要

画像から選手を特定する手段として、顔やゼッケンを認識する方法がある。しかし、顔は、後ろ向きや、横向き、遠くの選手など見えないことが多く、ゼッケンは、歪み（ゆがみ）や、他選手との重なりによる隠れなどがあり、認識が難しい。

そこで、開発したシステムでは、顔やゼッケンよりも大きく映り、一部が隠れてもチームの判別が可能なユニフォームを識別の対象とした。中継車は、走行中の区間を把握しており、区間ごとの走者リストもあるため、チームを識別することで個人を特定できる。また、ユニフォームは、数年単位でしか変更されないため、直近の過去映像から学習データを大量に収集することができ、高い識別性能を実現できる。

図2に、駅伝中継映像解析システムの概要を示す。このシステムは、画像を1s間バッファリングし、1s以内に解析した結果を出力する。まず、CNN（Convolutional Neural Network）を用いた一般的な物体検出手法である、SSD（Single Shot Multibox Detector）<sup>(2)</sup>を用いて人物を検出し、各検出矩形（くけい）を時間方向に対応付ける追従処理を行う。ここで、追従処理には、人同士の重なりによりロバストなDIET（Dynamic Integration of Extended Tracklets）<sup>(3)</sup>を用いた。次に、ユニフォームや、襷（たすき）の色、ロゴの見え方などを学習した50層のResidual Network<sup>(4)</sup>を用いて、選手のチームを識別する。そして、解析結果は、

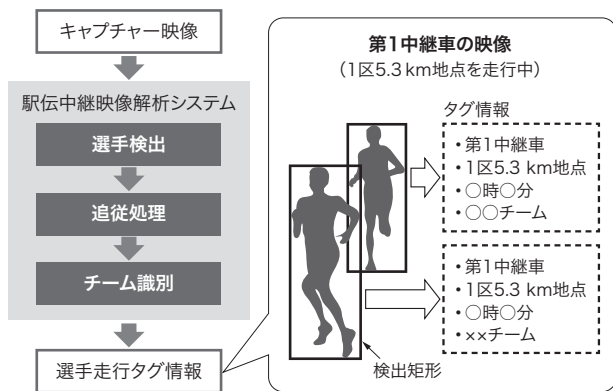


図2. 駅伝中継映像解析システムの概要

中継映像から選手を検出・追従してチームを識別し、中継車の情報と組み合わせることで、選手の走行情報をリアルタイムに出力する。

Outline of real-time video analysis system

時刻や中継車の走行位置情報などとともに映像に対するタグ情報として記録し、逐次情報を更新しながら番組制作現場で共有される。

## 3. 識別処理の高速化と高性能化

生放送のスポーツ中継での制作支援として解析システムを活用するには、リアルタイムでの動作が必須である。また、誤識別した結果の修正作業をできるだけ発生させないために、高い識別性能も必要となる。開発したシステムは、以下の四つの手法を導入することで、高速かつ高性能なチーム識別を実現している。

### 3.1 検出結果からの観客フィルタリング

中継映像には、選手以外にも大勢の観客が映り込んでいる。検出した観客の識別処理も行くと、選手と似た服装の人物を誤識別するおそれがあり、また、大量の検出結果が識別対象となることで、処理速度が低下する。そこで、中継カメラが選手と並走しながら、選手が中央に映るという映像の絵作りの特性を利用する。図3に示すように、観客は画像端を流れるように動き、選手と観客では追従軌跡に差が生じる。したがって、追従した検出矩形の移動距離が長い場合は、観客として判定し、除去する。

### 3.2 複数枚画像を用いたチーム識別

選手同士の重なりや動きによって、フレームごとに選手の見え方が異なるため、フレーム単位ではチーム識別を誤るこ

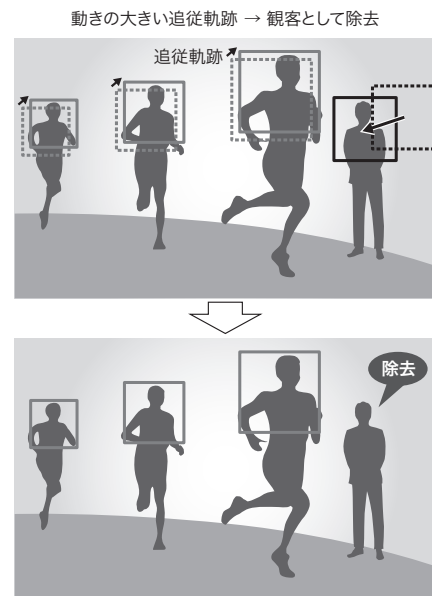


図3. 観客フィルタリング手法

映像上では、選手と観客の追従軌跡に差が生じるので、動きの違いから観客の検出矩形を除去する。

Method of filtering out roadside spectators from recognition targets

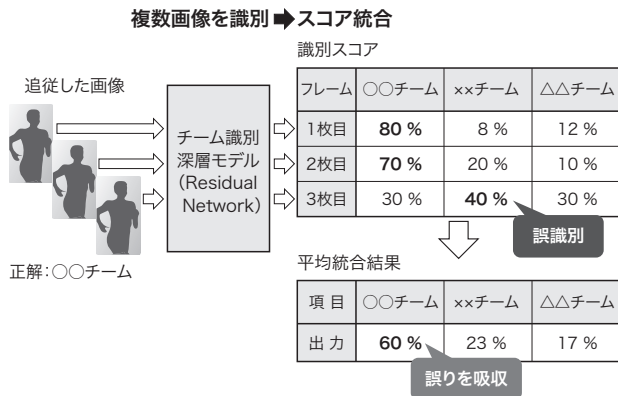


図4. 複数枚画像を用いたチーム識別手法

フレーム単位では、チーム識別を誤ることがある(3枚目)が、スコアを平均化することで、誤りを低減できる。

Method of team recognition using multi-frame images

とがある。そこで、追従した複数のフレームで識別スコアを算出し、そのスコアの平均を求めてから識別することで、見え方の変化に対してロバスト化が図れる(図4)。

### 3.3 顔と上半身検出を併用した追従性能の向上

選手の未検出が増えると追従が困難になり、追従結果を用いている観客フィルタリングや複数画像によるチーム識別の性能が劣化する。そこで、より高精度に追従するために、顔と上半身を同時に検出して併用する。顔は、上半身に比べて選手同士が重なっても隠れることが少ないが、遠くの選手の顔は小さく映るため検出できない。逆に、上半身は、顔に比べて遠くの選手でも検出できるが、選手同士が重なると検出率が低下する。このため、顔と上半身の異なる範囲を同時に検出することで未検出を低減し、どちらかが検出されていれば追従を継続することで、途切れを減らすことができる(図5)。検出は、処理時間を削減するために、顔クラスと上半身クラスの検出を同時に学習し、一度の検出処理で顔と上半身の検出結果が得られるようにした。

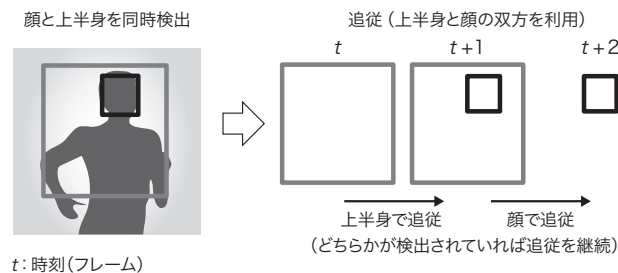


図5. 顔と上半身の検出を併用した選手の追従手法

顔と上半身を同時に検出して補充しながら追従することで、未検出を低減し、追従の途切れを減らすことができる。

Method of tracking runner by detecting both face and upper body

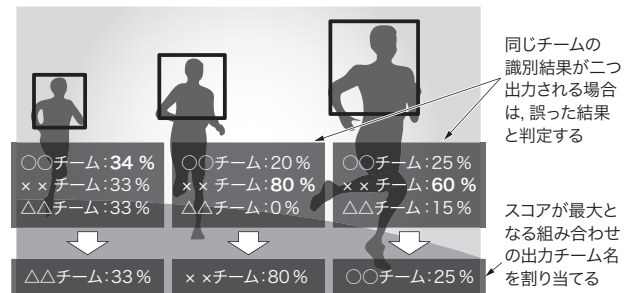


図6. 出力するチーム名の割り当て最適化手法

中継以外では、一つの中継映像中に同じチームは存在しないので、出力チームの割り当てを最適化することで、誤識別を低減できる。

Method of optimizing output team combination

### 3.4 中継車ごとの出力チーム名割り当ての最適化

駅伝では、類似するユニフォームのチームが複数存在するため、識別を誤ることがある。そこで轡をつなぐ中継以外では、一つの中継映像中に同じチームは存在しないことを利用して、識別スコアに対し、出力するチームの割り当てを最適化することで、識別誤りを低減できる。この問題は、出力するチームに対し、識別スコアの合計が最大となる組み合わせを求める問題として解くことができる(図6)。組み合わせスコアの最大化は、2部グラフのマッチング問題であり、ハンガリアンアルゴリズム<sup>(5)</sup>を用いて解を求める。

## 4. 実証実験

実際の駅伝中継放送で、開発したシステムの実証実験を、日本テレビ放送網(株)と共同で行った。合計10h以上の検証を行い、選手同士の重なりによる見え方の変化や、朝日などの日照変化、動きのぼけによるノイズなど、識別が難しい条件でもロバストな識別が実現できていることを目視確認した。また、3minごとに正解を教示した217枚の画像で定量評価した結果、システムの識別出力が正解している割合を示す適合率は98.1%であり、高い精度で動作することが確認できた。

処理時間としては、1sの映像に対して平均0.96sで結果を出力し、遅延なく動作することが確認できた。更に、システム出力を併用しながら記録作業していた番組制作オペレーターからは、「従来よりも大幅に作業量が削減され、非常に効率的であった」との意見もあり、システムの有効性が確認できた。

### 5. 駅伝中継映像解析結果の応用

駅伝中継映像解析技術は、手作業でのタグ付け作業の省力化だけでなく、幅広い応用が可能である。

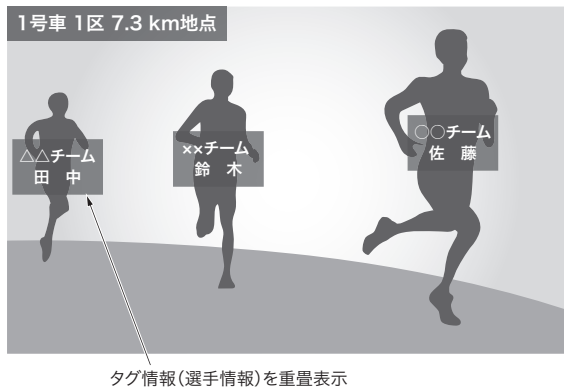


図7. 解析結果を応用したCG表示

選手の位置情報がリアルタイムに検出できるので、対象選手付近に情報をCG表示するような、今までにない表示が可能となる。

CG display showing runner information

### 5.1 チーム識別と検出位置を利用した自動CG生成

これまで、選手の名前や所属チームを表示するCGは、人手作業でタグ付けされた情報を基に生成されていた。また、タグ情報には、画像上の選手位置が含まれておらず、一覧での表示だけであった。これに対し、駅伝中継映像解析システムは、選手を検出することで位置情報をリアルタイムに取得できるため、対象選手の胸元にCGを表示するといった、今までに実現できなかった表示方法が可能になる(図7)。また、放送波向けだけでなく、実況・解説者にCG重畳した映像を提示することで、選手情報の読み間違いを防ぐことができる。

### 5.2 解析結果を利用した過去映像検索とシーン抽出

番組制作では、中継映像以外に、過去映像を編集した映像を放映することもある。しかし、過去映像には、どの時間帯にどの選手が映っているかというタグ情報が付いていないため、意図した映像を探すのに労力が掛かっていた。これに対し、駅伝中継映像解析システムでは、映っていたチームが時間帯に対応した状態で自動的に出力されるため、過去映像の検索が容易となる。また、検出した矩形サイズの遷移を利用すれば追い抜きのシーンを、識別したチーム数の情報を利用すれば密集して走行しているシーンを抽出できるなど、解析データの応用の幅は広い。特に、追い抜きは、注目度が高いシーンであるため、生放送で番組制作担当者に提示することで、見逃しを防ぐことも可能である。

(注1) ほかに、画像センシング技術研究会の第24回画像センシングシンポジウム(SSII2018)デモンストレーション賞、公益財団法人放送文化基金の第44回放送文化基金賞(個人・グループ部門)放送技術、一般社団法人日本映画テレビ技術協会の第71回(2017年度)技術開発奨励賞、一般社団法人日本民間放送連盟の平成30年日本民間放送連盟賞 技術部門最優秀。

## 6. あとがき

当社は、難度の高いスポーツ映像を対象とし、画像認識の基盤技術を強化することを目的に、中継映像だけを用いてリアルタイムかつ高精度に駅伝の選手を識別する、駅伝中継映像解析システムを開発した。また、今回開発した技術は、2018年度に、一般社団法人映像情報メディア学会の第45回技術振興賞進歩開発賞(現場運用部門)など、幾つかの賞を受けた(注1)。

今後は、人物映像解析の基盤技術として得られた知見を取り込み、スポーツエンターテインメント分野はもとより、産業分野への幅広い応用も検討していく。

## 謝辞

この研究の一部は、(株)日テレITプロデュースの協力を受けて進めました。有益なコメントを頂いた関係各位に感謝の意を表します。

## 文献

- (1) 大内一成, ほか. 密集領域での動作を理解するためのハイブリッド型映像解析. 東芝レビュー. 2017, 72, 4, p.30-34. <[https://www.toshiba.co.jp/tech/review/2017/04/72\\_04pdf/a08.pdf](https://www.toshiba.co.jp/tech/review/2017/04/72_04pdf/a08.pdf)>, (参照 2019-05-20).
- (2) Liu, W. et al. "SSD: Single shot multibox detector". European conference on computer vision (ECCV 2016) Proceedings Part1. Amsterdam, Netherlands, 2016-10, Springer. 2016, p.21-37.
- (3) Pham, Q. V. et al. "DIET: Dynamic Integration of Extended Tracklets for Tracking Multiple Persons". 22nd International Conference on Pattern Recognition. Stockholm, Sweden, 2014-08, IEEE. 2014, p.1206-1211.
- (4) He, K. et al. "Deep residual learning for image recognition". Proceedings of 2016 IEEE conference on computer vision and pattern recognition (CVPR 2016). Las Vegas, NV, 2016-07, IEEE. 2016, p.770-778.
- (5) Munkres, J. Algorithms for the assignment and transportation problems. Journal of the society for industrial and applied mathematics. 1957, 5, 1, p.32-38.



山地 雄士 YAMAJI Yuto  
研究開発本部 研究開発センター  
メディア AI ラボラトリー  
Media AI Lab.



小林 大祐 KOBAYASHI Daisuke  
研究開発本部 研究開発センター  
メディア AI ラボラトリー  
映像情報メディア学会会員  
Media AI Lab.



佐藤 誠 SATO Makoto  
日本テレビ放送網(株)  
技術統括局技術開発部  
映像情報メディア学会会員  
Nippon Television Network Corp.