

Human Body Pose Estimation Framework for Team Sports Videos Integrated with Tracking-by-Detection

January 2016

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Engineering



Keio University

Graduate School of Science and Technology
School of Integrated Design Engineering

Masaki Hayashi

主 論 文 要 旨

報告番号	甲 乙 第	号	氏 名	林昌希
主 論 文 題 目： Human Body Pose Estimation Framework for Team Sports Videos Integrated with Tracking-by-Detection (Tracking-by-Detection に統合されたチームスポーツ映像向け人物姿勢推定フレームワーク)				
(内容の要旨) 戦術分析を目的とした、チームスポーツ映像からの複数選手の追跡技術が盛んに研究されている。試合中の各選手の位置情報は最も基本的でかつ重要な情報であり、サッカーなど実際のプロスポーツの現場でも活用されている。この追跡した選手軌跡に加えて、各選手の「姿勢」まで自動的に推定できるようになれば、行動認識や注視解析などより高次の戦略解析への展開が期待できる。しかし従来の動画像を用いた人物姿勢推定技術は、限定的な姿勢パターンでのみ通用し、姿勢変化が非常に多いスポーツ映像中の選手に対しては、ロバストな姿勢推定が実現されていない。 本論文では、チームスポーツ映像向けの人物姿勢推定フレームワークを、標準的な人物追跡手法である Tracking-by-Detection に統合された形で提案する。単眼の入力映像から選手の 1 フレームの頭部領域を追跡したのちに、下半身姿勢(両膝両足 4 つの関節の位置)の推定モジュールと、上半身姿勢(身体の方向、背骨の姿勢)の推定モジュールが、それぞれ独立に姿勢を推定する。また、それら 2 つのモジュールを統合した手法も提案する。 本フレームワークは、時系列情報を用いることなく、追跡された 1 フレームの位置合わせ済み人画像のみから姿勢推定を行う。従って、動画から追跡された窓画像群、もしくは 1 枚の入力画像から検出された窓の、どちらに対しても姿勢を推定できる利点を有する。また、全身領域もしくは半身領域の Global な HOG(Histograms of Oriented Gradients)特徴量のうち、Random Forests で選択した特徴を推定に用いることで、Local な歪みや姿勢変化、未知の姿勢画像パターンに頑健である特徴も有する。 第 1 章では本論文の導入である。スポーツ映像からのデータ化と解析の現状を元に、提案手法が切り開く可能性について述べる。チームスポーツの文脈から「姿勢」の各種類とその意味を定義し、提案するフレームワークの要旨と貢献をまとめる。 第 2 章では、先行研究として古典的なモーションキャプチャと、近年の機械学習を基にした姿勢推定の発展を紹介する。 第 3 章では、下半身姿勢推定手法を提案する。窓中心が骨盤中心位置となる人物領域追跡の後、その追跡窓に対して関節位置をグリッド構造特徴(HOG 特徴)のグリッド位置として学習・推定する Label-Grid 識別器を用いた姿勢推定を提案する。実験では、従来手法では推定しづらい横向き視点での姿勢推定精度も検証することで、提案手法の有効性を示す。 第 4 章では、上半身姿勢推定手法を提案する。追跡した頭部領域中心に位置合わせされた上半身 HOG 特徴から、頭部中心に対する相対的な骨盤中心の回帰を行う Poselets-Regressor の提案を行う。また、推定した背骨の傾き範囲ごとの画像群でそれぞれ学習した 5 つの識別器を用いた、上半身方向の推定手法も提案する。実験により、従来の歩行者向け手法では精度よく推定しづらい、上半身が傾いている時の体方向推定が、提案手法によって頑健に行えることを示す。 第 5 章では、上記 2 モジュールの統合により、頭部領域追跡後に Poselets-Regressor を 2 回用いることで、頭部中心に対する相対的な骨盤中心位置、および骨盤中心位置に対する相対的な下半身関節位置、を推定する手法を提案する。実験を通じて、統合版が更に高解像度で下半身の関節位置を推定できることを示す。また HOG 特徴領域の追跡による位置合わせを行う本手法が、その仕組みから想定される通りの挙動をしていることを確かめるための追加実験も行う。 第 6 章では、本論文を総括した上で、今後の課題と展望を述べる。				

SUMMARY OF Ph.D. DISSERTATION

School Integrated Design Engineering	Student Identification Number	SURNAME, First name HAYASHI, Masaki
Title Human Body Pose Estimation Framework for Team Sports Videos Integrated with Tracking-by-Detection		
Abstract <p>Multitarget tracking in team sports videos has been widely studied for the purpose of analyzing tactics. Since the basic and most important information estimated from those tracking systems are the player locations, such systems have been applied in professional team sports, such as soccer. In addition to the player trajectories, knowledge of their poses would provide higher-level cues for analyzing sports videos. If the poses of the players can be automatically recognized from team sports videos, more-detailed analysis of the sports plays, such as recognizing actions or understanding the focus of attention, could be achieved. However, there are no robust methods for estimating the poses of team sports players, because the previous techniques only work for restricted patterns.</p> <p>In this thesis, a novel human pose-estimation framework for team sports videos is proposed; the framework is integrated with the standard tracking-by-detection approach, and it is able to estimate most of the poses of the players in such videos. After tracking either the player windows or the position of the head of a player in a monocular input video, two independent modules are applied: the first estimates the lower-body pose (the locations of the four lower-body joints), and the second estimates the upper-body pose (upper-body orientation and spine pose). An integrated version of those two pose-estimation modules is also presented. Each chapter empirically shows the proposed module can estimate more types of poses than the previous methods, while achieving competitive results to the previous methods.</p> <p>The framework does not use any temporal information, but only the per-frame player window from a monocular camera. It can use both tracked windows from the videos, or it can use the detected windows from a single image. Moreover, it can disregard the effect of local deformations and local changes in pose (including unknown poses in the training dataset), because it is based on randomized features, trained by random forests, and obtained from the histograms of oriented gradients (HOG) features within the global (whole- or half-body) region.</p> <p>Chapter 1 provides an introduction to this thesis. This chapter introduces the methods of data analysis that are currently used in professional sports; the chapter defines the types of poses and their meaning in a team-sport context, and the proposed framework is summarized.</p> <p>Chapter 2 discusses previous work on pose-estimation techniques.</p> <p>Chapter 3 proposes a lower-body pose-estimation module that uses label-grid classifiers to estimate the locations of the joints on a grid and the grid-structured features within a tracked player window for which the center is aligned with the pelvis location.</p> <p>Chapter 4 proposes an upper-body pose-estimation module that uses poselets-regressors to estimate the location of the pelvis relative to the head center, by using the HOG features in the upper-body region, which the head tracker aligns with the head center. This chapter also proposes a method to estimate the orientation of the body by selecting from five classifiers that are trained independently using only the images that have a spine angle within a shared range.</p> <p>Chapter 5 proposes an integrated version of the modules presented in Chapters 3 and 4; it uses the poselets-regressor twice: once to estimate the location of the pelvis center relative to the head center, and the second time to estimate the location of each of the lower-body joints relative to the pelvis.</p> <p>Chapter 6 presents the conclusions and future works.</p>		