Title	A Study on Human Actions Representation and Recognition
Author(s)	Sheikh Mohammad Masudul Ahsan
Issue Date	2016-03-25
URL	http://hdl.handle.net/10228/5660
Rights	

氏　　名	Sheikh Mohammad Masudul Ahsan（バングラデシュ）
学位の種類	博　士（工学）
学位記番号	工博甲第４０９号
学位授与の日付	平成２８年３月２５日
学位授与の条件	学位規則第４条第１項該当
学位論文題目	A Study on Human Actions Representation and Recognition （人の行動の表現と認識に関する研究）
論文審査委員	主　査　准教授　タン　ジュークイ 教　授　森江　　隆 教　授　黒木　秀一 教　授　金　　亨燮 教　授　石川　聖二

学　位　論　文　内　容　の　要　旨

In recent years, analyzing human motion and recognizing a performed action from a video sequence has become very important and has been a well-researched topic in the field of computer vision. The reason behind such attention is its diverse applications in different domains like robotics, human computer interaction, video surveillance, controller-free gaming, video indexing, mixed or virtual reality, intelligent environments, etc. There are a number of researches performed on motion recognition in the last few decades. The state of the art action recognition schemes generally use a holistic or a body part based approach to represent actions. Most of the methods provide reasonable recognition results, but they are sometimes not suitable for online or real time systems because of their complexity in action representation. In this thesis, we address this issue by proposing a novel action representation scheme.

The proposed action descriptor is based on a basic idea that rather than detecting the exact body parts or analyzing each action sequence, human action can be represented by a distribution of local texture patterns extracted from spatiotemporal templates. In this study, we use a novel way of generating those templates. Motion History Image (MHI) merges an action sequence into a single template. However, having the problem in overwriting old information by a new one in the MHI, we use a variant named Directional MHI (DMHI) to diffuse the action sequence into four directional templates. And then we use the Local Binary Pattern

(LBP) operator with a unique way, i.e. a rotated bit arranged LBP, to extract the local texture patterns from those DMHI templates. These spatiotemporal patterns form the basis of our action descriptor which is formulated into a concatenated block histogram to serve as a feature vector for action recognition. However, the extracted patterns by LBP tends to lose the temporal information in a DMHI, therefore we take linear combination of the motion history information and texture information to represent an action sequence. In this case, we generate the DMHI histogram in a similar fashion as that of LBP histogram. To keep the feature vector dimension in a limit, rather than concatenation, we take the component-wise addition of DMHI histogram and LBP histogram as a feature vector. We found in the experiment that the histogram of spatiotemporal texture individually gives better result than DMHI histogram. However, mixing a certain amount of temporal information (DMHI histogram) with the texture information provides even better results. We also use some variants of the proposed action representation that include the shape or pose information of the action silhouettes as a form of histogram.

We show that, by effective classification of the proposed histograms, i.e., an action descriptor, robust human action recognition is possible. We demonstrate the effectiveness of the proposed method along with some variants of the method over two benchmark datasets; the Weizmann dataset and the KTH dataset. Our results are directly comparable or superior to the results reported over these datasets by other researchers. Higher recognition rates found in the experiment suggest that, compared to complex representation, the proposed simple and compact representation can achieve robust recognition of human activity for practical use. The proposed method is simple and compact in a sense that we do not need to detect any interest points or shape in every frame and do not need to create any space-time model to represent an action. The proposed method simply extracts the patterns lying on a temporal template and constructs a distribution of those patterns to represent an action. Besides the recognition rate, due to the simplicity of the proposed technique, it is also advantageous with respect to computational load.

After all, the originalities of the proposed methods are that (i) we use an idea that a human action descriptor can be formulated as a distribution of spatiotemporal texture patterns, (ii) to realize the idea, we use the LBP to extract the patterns from a DMHI template, (iii) we introduce a new way of creating a LBP images by using the rotated arrangement of LBP bits for different DMHIs, and finally (iv) we use a novel framework that combines DMHI, LBP and SVM for actions representation and recognition.

During the experiment, we have used two classifiers; k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM). We found that the SVM classifier performs better than the k-NN classifier. We obtained the best average recognition rate of 95.4% and 95.6% for the Weizmann and the KTH dataset, respectively.

The proposed method can be applied to some real life applications like gaming or human computer interaction without using any controller such as a mouse, a trackball, a joystick, etc. The potential of the proposed method can also be applied to other related domains like a patient's activity monitoring system or automatic labeling of video sequences in a video dataset.

学 位 論 文 審 査 の 結 果 の 要 旨

人の動作や行動をカメラとコンピュータを用いて自動認識するシステムは，安全で安心して暮らせる社会，文化的に豊かな社会，また特に高齢者や障害者にとって活動しやすい社会を実現するために近年特に注目され，活発な研究・技術開発が行われている．例えば，安全運転支援のための歩行者の検出とその行動予測，不審な行動・挙動を行う人物の検出と追跡，エンターテインメント分野における人の動作やジェスチャーによるゲームのコントロール，また高齢者や障害者等の必要に応じた行動支援等への応用がある．

ビデオ映像に基づく人の行動認識に関する研究は，これまでに数多く行われてきたが，この分野は発展途上にあり，実用に供することのできる手法はまだ実現していない．人の行動認識のためには，まず行動をコンピュータ処理できる形で表現する必要がある．そこで，本論文では著者は，人の行動を表現する新しい方法と，それを用いた行動の認識法を提案している．

本論文では，著者はまず，研究の背景について述べ，人の行動認識研究の重要性・その応用分野の広さについて説明している．また従来研究について考察し，それらが主として人の手足の検出や体全体の形状の検出を前提とする方法であり，コンピュータによる人の行動認識法として，そのような研究の限界について議論している．

次に著者は，提案法について詳述している．提案法は，人の形状を利用する従来法と異なり，行動による人の動きを，ビデオ映像の連続フレームからひとつに集約した画像，すなわち時空間画像で表現する．また，主として二種類の時空間画像を利用する．第一は，動きを表す連続するフレームを時間が経過するほど濃度値を下げながら重ね合せる動作履歴画像（Motion History Image：MHI）を上下左右の四方向に分離して表現する方向性動作履歴画像（Directional Motion History Image：DMHI），第二は，DMHIを回転桁配置型局所二値パターン（Rotated bit arranged Local Binary Pattern: RLBP）を用いて表現した画像（RLBP Image）である．どちらの画像も矩形領域に分割（ブロ

ック化）され，各ブロックはヒストグラムで表現される．

これらの時空間画像をベースに，著者は主に二つの行動表現法を提案している．すなわち，第一の方法は，RLBP 画像と動きの特徴的断片画像フレームに基づく特徴ベクトルを用いた行動表現法，また第二の方法は，DMHI と RLBP 画像，それに動きの範囲を示す動作エネルギー画像（Motion Energy Image）に基づく特徴ベクトルを用いた行動表現法である．人の様々な行動をこれらの方法で表現し，サポートベクターマシン（SVM）を用いて認識を行う，というのが著者が提案する，コンピュータを用いた人の行動の表現と認識法である．

次に著者は，人の行動に関する二種類の標準映像データベースを用いて，提案法の性能を実験的に評価している．カメラが固定された状態で撮影されたビデオ映像からなる Weizmann データセットを用いた実験では第一の方法が有効で，95.4%の認識率を挙げている．これは同じデータセットを用いた他の諸研究と比較してほぼ同等の高い認識率である．また，１フレーム当たりの処理時間は 35.8ms で，ほぼ実時間処理である．また，手持ちカメラで撮影されたビデオ映像からなる KTH データセットを用いた実験では第二の方法が有効であり，95.6%の認識率を挙げている．これは同じデータセットを用いた諸研究の中で最も高い認識率である．また，１フレーム当たりの処理時間は 45.9ms で，これも高速処理である．以上の二つの実験により，提案法が，コンピュータによる人の行動認識において，認識率および処理時間の両面で優れた性能を持つ手法であることを示している．

最後に著者は本研究をまとめ，今後の課題について言及している．

以上のように本論文は，コンピュータによる人の行動の表現と認識に関する新しい方法を提案し，標準映像データベースを用いた実験によって，その有効性を示している．提案する行動表現・認識法は，従来の諸手法のように動作を行う身体形状を直接用いるのではなく，動きの連続フレームをひとつに集約した時空間画像で行動を表し，その濃淡表現および RLBP という画像表現を用いるという点において従来法とは異なる方法である．このように提案法は独創性があり，かつ性能も高く，人の行動認識研究の実用化への大きな一歩を与えるものであり，コンピュータビジョン，ロボットビジョン，画像計測分野への貢献は大きい．

なお，本研究に関して，審査委員および公聴会における出席者から，処理時間，特徴量を結合するときのパラメータの意味するもの，RLBP の意味と計算法，ジョギングとスキッピング動作の誤認識問題，実環境への応用可能性等，種々の質問がなされたが，いずれも著者からの適切な説明によって質問者の理解が得られた．

以上より，本審査委員会は，学位論文及び最終試験の結果に基づき慎重に審査した結果，本論文が，博士（工学）の学位に十分値するものであると判断した．