Robotics and Autonomous Systems 78 (2016) 29–35

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Action recognition from only somatosensory information using spectral learning in a hidden Markov model

Wataru Takano *, Junya Obara, Yoshihiko Nakamura

Mechano-Informatics, University of Tokyo, 7-3-1, Hongo, Bunkyoku, Tokyo, 113-8656, Japan

HIGHLIGHTS

- This paper proposes an approach to classifying somatosensory information in the human full body into the action categories.
- The muscle activities in the human full body are estimated from captured motions, ground reaction forces, and EMG data.
- The discrete hidden Markov models to be optimized by spectral learning are adopted for the action classifiers.

ARTICLE INFO

Article history:
Received 14 July 2015
Accepted 4 January 2016
Available online 19 January 2016

Keywords:
Action classification
Somatosensory
HMM

ABSTRACT

Human action classification is fundamental technology for robots that have to interpret a human's intended actions and make appropriate responses, as they will have to do if they are to be integrated into our daily lives. Improved measurement of human motion, using an optical motion capture system or a depth sensor, allows robots to recognize human actions from superficial motion data, such as camera images containing human actions or positions of human bodies. But existing technology for motion recognition does not handle the contact force that always exists between the human and the environment that the human is acting upon. More specifically, humans perform feasible actions by controlling not only their posture but also the contact forces. Furthermore these contact forces require appropriate muscle tensions in the full body. These muscle tensions or activities are expected to be useful for robots observing human actions to estimate the human's somatosensory states and consequently understand the intended action. This paper proposes a novel approach to classifying human actions using only the activities of all the muscles in the human body. Continuous spatio-temporal data of the activity of an individual muscle is encoded into a discrete hidden Markov model (HMM), and the set of HMMs for all the muscles forms a classifier for the specific action. Our classifiers were tested on muscle activities estimated from captured human motions, electromyography data, and reaction forces. The results demonstrate their superiority over commonly used HMM-based classifiers.

1. Introduction

It is expected that robots will be integrated into our everyday lives and, in this context, they will be required to understand our intended actions in order to establish smooth human–robot communication. Research on human action recognition focusing on classifying a video containing a human action into a specific category has been intensively conducted. The video includes only superficial color images or kinematic motions. It is consequently difficult to estimate either the force with which a human body acts on its external environment or the internal state in the human body and to recognize the human action by using these estimates. More specifically, robots cannot understand whether a human lifts heavy or light luggage from observing only the human's motion, and the robots need to perceive other modal information, such as muscle activity or the object that the motion acts on.

The estimation of human muscle activity constitutes a fundamental research problem in the field of biomechanics, and has been tackled using measurements from electromyograph (EMG) sensors or computations of kinematics and dynamics. Lloyd et al. developed an EMG-driven musculoskeletal model of the human knee and predicted the torque in the knee joint with a Hill-type muscle model [1]. Yamane et al. proposed an algorithm to estimate muscle forces in the human full body from captured motion, EMG signals, and reaction forces on the force plate. This algorithm estimates the joint torques from the motion and the reaction forces

* Corresponding author. Tel.: +81 3 5841 6378; fax: +81 3 3818 0835.
E-mail addresses: takano@ynl.t.u-tokyo.ac.jp (W. Takano), obara@ynl.t.u-tokyo.ac.jp (J. Obara), nakamura@ynl.t.u-tokyo.ac.jp (Y. Nakamura).

http://dx.doi.org/10.1016/j.robot.2016.01.001
0921-8890/

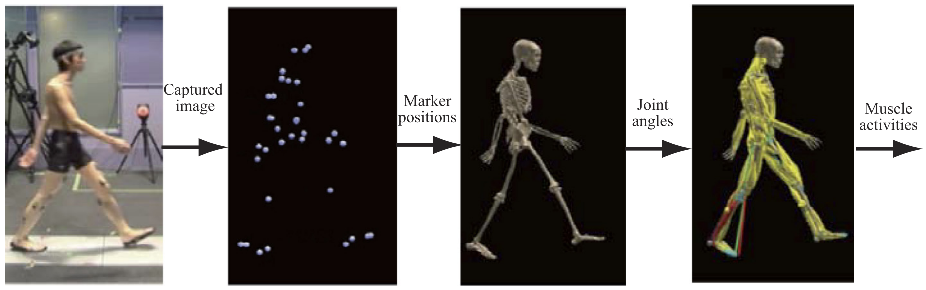

Fig. 1. An optical motion capture system detects markers attached to a performer. The angles of all joints are estimated from the marker positions by inverse kinematics. The joint torques are subsequently estimated from the joint angles and reaction forces acting on the body by inverse dynamics. The muscle tensions generating the joint torques are computed from the joint torques and EMG data. The muscle tensions are converted to the resultant muscle activity.

through inverse kinematics and inverse dynamics computations, and optimizes the muscle tensions generating the estimated joint torques using EMG signals [2]. Applying this algorithm for a large number of muscles is time consuming, but Murai et al. developed a method to estimate the muscle forces in real time by grouping similar functional muscles [3]. These musculoskeletal models were extended to include the nervous system by combining them with a neural network, and were used for the analysis of reflex mechanisms [4,5]. These techniques allow partial somatosensation to be estimated.

The robotics research has been targeting the learning framework for motion representation, which is often referred to as "imitation learning" [6] or "programming by demonstration" [7]. The learning algorithm encodes the motion as a sequence of configurations, such as joint angles, into model parameters, and the model can recover the joint angles so that the robots can perform almost the same motion as they learned [8–11]. In addition to motion generation, the model can be re-used for motion recognition. These frameworks have been extended to handle other modal data, such as visual or linguistic information, and the robot can consequently manipulate objects or understand human actions in linguistic forms [12–16].

This research inspires us to combine the estimation of somatosensation with the representation of human motions in order to develop robots that understand human behavior in a deep way. This paper introduces the extension of a stochastic representation to a large number of muscle activities in human full body motion patterns, and evaluates the performance of this representation for motion recognition. Human full body motion is expressed by a sequence of vectors, each of whose elements is a muscle activity, and which do not include positions or joint angles, and the sequence is encoded into hidden Markov model (HMM), The muscle activities in the human full body are high dimensional, and a sequence of the activities of each muscle is encoded into its specific Markov chain. The integration of Markov chains of all the muscles allows for classifying the somatosensation into its relevant motion category. Additionally, we adopt a discrete HMM and apply the spectral learning algorithm to encode the sequence into the HMM, since the sequence of muscle activities looks impulsive and the complex. The proposed approach was tested on human full body motions in captured video, EMG data, and contact forces, and the test demonstrated that the proposed approach outperformed the commonly used HMM for motion recognition.

2. Estimation of muscle activity

Here we briefly explain the estimation of the activity of all the muscles in the full body from captured motion, EMG data, and contact forces [2]. Fig. 1 shows the pipeline for the conversion of the measured data into muscle activity. The human musculoskeletal model consists mainly of three kinds of elements: bones, muscles, and tendons. The bones, muscles, and tendons are represented by solid links, wires that actively generate forces, and wires that passively generate forces, respectively. Inverse kinematics computation converts the captured motion data $\boldsymbol{p}$ into joint angles $\boldsymbol{\theta}$ according to the skeletal model, and inverse dynamics computation estimates the joint torques $\boldsymbol{\tau}$ from the derived joint angles and contact forces. The relation between the joint torques $\boldsymbol{\tau}$ and their equivalent muscle tensions $\boldsymbol{f}$ is given by

$$\boldsymbol{\tau} = \boldsymbol{J}^T \boldsymbol{f}, \tag{1}$$

where $\boldsymbol{J}$ is the Jacobian matrix of the muscle lengths with respect to the joint angles. Unique muscle tensions cannot be derived from Eq. (1) since the dimension of the muscle tension vector $\boldsymbol{f}$ is larger than that of the joint torques $\boldsymbol{\tau}$. The unique muscle tensions are found by minimizing the cost function

$$z = \frac{1}{2}\left(\boldsymbol{\tau} - \boldsymbol{J}^T\boldsymbol{f}\right)^T \boldsymbol{W}_\tau \left(\boldsymbol{\tau} - \boldsymbol{J}^T\boldsymbol{f}\right) + \frac{1}{2}\left(\boldsymbol{f} - \boldsymbol{f}_E\right)^T \boldsymbol{W}_E \left(\boldsymbol{f} - \boldsymbol{f}_E\right), \tag{2}$$

where $\boldsymbol{f}_E$ contains the reference muscle tensions estimated from the EMG data, and $\boldsymbol{W}_\tau$ and $\boldsymbol{W}_E$ are positive weight matrices. An inequality constraint $-\boldsymbol{f}_{\max} \leq \boldsymbol{f} \leq 0$, where $\boldsymbol{f}_{\max}$ is a vector of the maximum muscle tensions, is added so that the muscle tensions cause only contraction. The optimal muscle tensions for the cost function in Eq. (2) are dynamically and biologically valid. A sequence of muscle activities $\boldsymbol{x}$, defined as the ratio of muscle tension to maximum muscle tension, is thus obtained from the captured motion, EMG data, and contact forces.

3. Learning of muscle activity

Sequences of muscle activities are encoded into specific discrete HMMs. A discrete HMM can be defined by the compact notation $\lambda = \{\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{T}, \boldsymbol{O}, \boldsymbol{\Pi}\}$. For this, $\boldsymbol{X} = \left\{{}^1X, {}^2X, \ldots, {}^nX\right\}$ is a set of n possible distinct symbols X to be generated by the HMM; $\boldsymbol{H} = \left\{{}^1H, {}^2H, \ldots, {}^mH\right\}$ is a set of nodes; $\boldsymbol{T} = \left\{T_{ij}\right\}$ is a matrix whose entries T_{ij} are the probabilities of transitioning from node ${}^jH$ to node ${}^iH$; $\boldsymbol{O} = \left\{O_{ij}\right\}$ is a matrix whose entries O_{ij} are the probabilities of generating the symbol ${}^iX$ from the node ${}^jH$; and $\boldsymbol{\Pi} = \{\pi_i, \pi_2, \ldots, \pi_m\}$ is a vector whose entries p_i are the probabilities of starting at node ${}^iH$. The probabilities $\boldsymbol{T}, \boldsymbol{O}$, and $\boldsymbol{\Pi}$, are commonly optimized using algorithms, such as the Baum–Welch algorithm [17] or Viterbi training [18], so that the obtained HMM is the one most likely to generate the training data. These commonly used algorithms need to fix the structure of the HMM and are not very effective in dealing with high-dimensionality data.

This study focuses on encoding high-dimensionality muscle activity into an HMM. The complexity of muscle activity can vary greatly depending on the muscles and the motion patterns: several muscles may remain inactive while muscles specific to a motion