Neurocomputing 116 (2013) 291–300

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

A gesture recognition system with retina-V1 model and one-pass dynamic programming ☆

Takashi Kuremoto *, Yasuhiro Kinoshita, Liang-bing Feng, Shun Watanabe, Kunikazu Kobayashi, Masanao Obayashi

Graduate School of Science and Engineering, Yamaguchi University, Tokiwadai 2–16-1, Ube, Yamaguchi 755–8611, Japan

ARTICLE INFO

Available online 9 October 2012

Keywords:
One-Pass DP
Gesture recognition
Skin color analysis
Motion estimation

ABSTRACT

Dynamic Programming (DP) algorithm has been studied from 1940s and successfully applied to pattern recognition fields such as continuous speech recognition, hand writing recognition, gesture recognition and so on. In this paper, we propose a novel hand gesture recognition system which includes three kinds of image processing: skin area segmentation, motion estimation by a retina-V1 model, and a gesture discrimination algorithm of One-Pass Dynamic Programming (One-Pass DP). A *HSV-RGB* filter is used to extract skin area in the color image, and the simple motion of hand area is estimated in eight directions by a retina-V1 model which is a computational model of primary visual cortex. Then the motions are used to compose 40 basic templates of gestures. In other words, hand gestures are considered as combinations of templates of simple motions, and One-Pass DP is used to recognize the pattern of gestures. Experiments dealt with individual and compound gestures were executed by online processing, and the results confirmed the effectiveness of the proposed system.

1. Introduction

Computer vision (CV) research has a long history since latter half of last century. However, gesture of a human, i.e., a series of images capturing the movement of body or a part of body (hand, arm, etc.), is still difficult to be recognized by a robot in the real world [1–6]. The key points of gesture recognition techniques are simply considered as two problems:

1. How to estimate the motion accurately?
2. How to classify the time series patterns of the movements?

For the first problem, optical flow computation methods such as first-order derivatives, region-based matching, energy-based or phase-based methods, etc., have been proposed earlier [7]. Tohyama and Fukushima constructed a neural network model to extract the motion appeared in the layer of retina [8–10]. Their model includes neural layers of retina, V1, MT and MST, where local velocity, absolute velocity, relative velocity and type of motion are calculated and output respectively. However, these researches are limited in the estimation of the motion or the kinds of motions without extending to gesture recognition.

For the second problem, various approaches such as 3D model-based methods [1], Self-Organizing Map (SOM) [2,3], and Hidden Markov Model (HMM) [4–6], have been proposed in the past decades, however, a distressed problem among these methods is the high computation cost of stochastic modeling with a large scale image processing.

Meanwhile, Dynamic Programming algorithm (DP), proposed by Bellman [11] to deal with Operation Research (OR) problems such as optimal stock, optimal consumption, shortest path problem and so on in more early time. And DP has been popularly applied to the field of speech recognition [12–15]. Yoshiike and Takefuji proposed a hand gesture recognition system using DP matching and with preprocessing of noise motion removal, a gesture of waving of left hand was recognized successfully in the real scenes.

The algorithm of DP, in a word, is to match a time series data of input signal (such as aural signals or image sequences) to a time series data of templates (such as speeches or gestures) by evaluating the distances of orders or mismatches between them. One-Pass DP, a kind of Viterbi search to decide the optimal path of continuous signals, is shown that it works well in speech recognition with its low cost of computation by [14,15].

In this paper, we adopt the One-Pass DP into a novel gesture recognition system to real online processing of gesture recognition.

☆ A part of this work was supported by Grant-in-Aid for Scientific Research (JSPS 20500207, 20500277, 23500181).

* Corresponding author. Tel.: +81 836 859 520; fax: +81 836 859 501.
E-mail addresses: wu@yamaguchi-u.ac.jp (T. Kuremoto), koba@yamaguchi-u.ac.jp (K. Kobayashi), m.obayas@yamaguchi-u.ac.jp (M. Obayashi).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.03.027

Templates of different motions are pre-defined, and then they are composed of kinds of gestures which may be used to realize Human–Computer-Interaction (HCI) such as instruction learning or sign language recognition. The element of motion is similar to the phoneme of speech in the recognition process, and it is estimated by a primary visual cortex model (retina-V1 model) which is a part of the neural network model of Tohyama and Fukushima [8]. The element motions are used to compose templates of gestures and gestures of a hand (fist) with translation or rotation movements and their compositions were recognized in the experiment. Methods for online processing were also considered and investigated in the experiment.

The rest of this paper is as follows. Section 2 shows the structure of a gesture recognition system proposed here at first, and then describes the preprocessing of extraction of skin area, and optical flow estimation by the V1-retina model. One-Pass DP and its application to gesture recognition are introduced in Section 2 too. Online recognition experiments and results are reported in Section 3, and Section 4 gives a conclusion.

2. A gesture recognition system

The processing flow chart of a gesture recognition system using One-Pass DP algorithm is shown in Fig. 1. A series of color images captured by CCD camera (or Web camera) is input to the system at first. Skin area segmentation using a method proposed in [16,17] is executed then. Motion estimation uses the concept of the neural network of [8–10] and gesture recognition is realized by the application of One-Pass DP method to the time series patterns of different motions of the hand area. The details of each processing are described in this section.

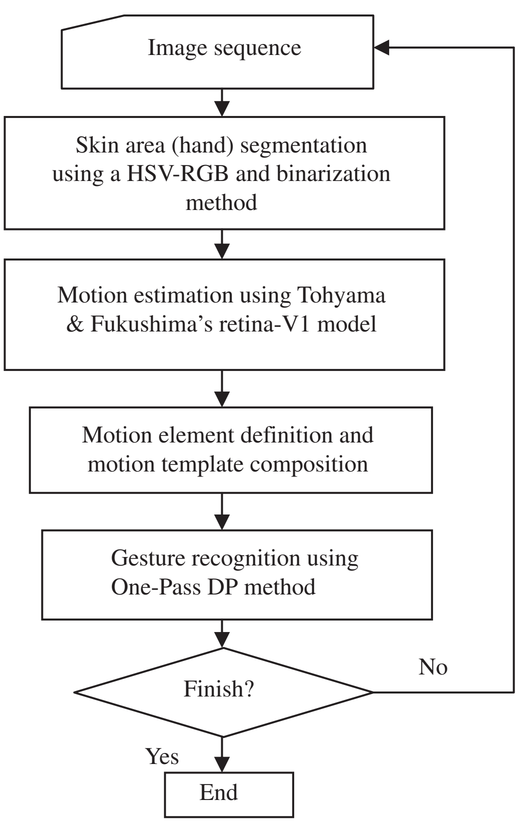

Fig. 1. Flow chart of a gesture recognition system.

2.1. Skin area segmentation

Skin area, i.e., a hand or a fist which is considered as an subject yields a human instructor's gestures here, is segmented with a *HSV-RGB* filter and binarization method proposed by our previous work of static image processing [16,17]. Human skin area estimation is also a high hurdle theme of computer vision and Khan et al. surveyed nine kinds of methods recently [18]. As we know, skin of yellow race human being in the color image is highly related to the value of Hue (H), and Saturation (S) [19] in the bright enough environment. We found the thresholds of H and S and that of Red (R) in *RGB* to estimate skin areas which are shown in Fig. 2 [16,17]. The illumination condition is in the room of fluorescent lights (around 500 lx).

Additionally, CCD or CMOS cameras capture images with *RGB* format, and it is possible to transform *RGB* to *HSV* according to color conversion algorithm. Fig. 3 shows an example of skin area segmentation processing where noise elimination and hole filling processing are also performed after the binarization of skin area estimation. From the result of the extraction of *a fist*, the effectiveness of the skin area segmentation is confirmed as shown in Fig. 3.

2.2. Motion estimation

Motion, which is also called "optical flow", is able to be estimated in many kinds of methods during a long history of Computer Vision (CV) research. A neural network model which can calculate the velocity of the optical flow with the connection of retina-V1 layers, and categorize (recognize) it in the Middle Temporal area (MT) and the Middle Superior Temporal area (MST) layers is proposed by Tohyama and Fukushima [8–10]. However, the neural network model stresses the function realization of neural circuit of visual cortex, i.e., it is designed to be aware of the types of motion, such as rotation, expansion and reduction of objects without dealing to spatio-temporal change of motion. We adopt the concept of motion vector calculation part of this model, i.e., retina-V1 layers into the gesture recognition system in this research. Fig. 4 shows its structure which is a part of the neural network model in [8]. Each frame of images captured by the camera is considered as the layer of retina of the model, and each layer of V1 expresses the different motion elements such as the basic motion in different directions. Deep estimation of motion category is executed in MT and MST layers in [8–10], however, not used here.

To illustrate how a receptor in V1 responses to the stimuli of a fired small area of retina, Fig. 5 is depicted. The receptor field is expressed as a square just for the convenient of calculation in a Euclidean coordinates, though it is usually an elliptical shape in visual cortex models.

1) When $H, S \in [0, 360]$ degree,
if $10 \le S < 15$, then $H > 350$;
if $15 \le S < 20$, then $H > 330$;
if $20 \le S < 30$, then $H > 300$ or $H < 40$;
if $30 \le S < 50$, then $H > 250$ or $H < 30$;
if $50 \le S < 70$, then $H > 230$ or $H < 30$;
if $70 \le S < 150$, then $H > 220$ or $H < 40$;
if $S < 10$ or $150 \le S < 360$, then $H > 300$ or $H < 40$.

2) When $R, G, B \in [0, 255]$,
$30 < R < 250$.

Fig. 2. Thresholds of H and S are shown in (1) which are used to estimate skin area, additionally, R of *RGB* defined in (2) is also used.