Pattern Recognition 61 (2017) 139–152

Contents lists available at ScienceDirect

Pattern Recognition

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

An adaptive local binary pattern for 3D hand tracking

Joongrock Kim [a], Sunjin Yu [b], Dongchul Kim [c], Kar-Ann Toh [a], Sangyoun Lee [a,*]

[a] Department of Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, South Korea
[b] Department of Broadcasting and Film, Cheju Halla University, 38, Halladaehak-ro, Jeju-si, Jeju-do, South Korea
[c] Department of Computer Science, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, South Korea

ARTICLE INFO

Article history:
Received 5 June 2013
Received in revised form
25 July 2016
Accepted 26 July 2016
Available online 27 July 2016

Keywords:
3D hand tracking
Hand gesture recognition
Human computer interaction
Natural user interface

ABSTRACT

Ever since the availability of real-time three-dimensional (3D) data acquisition sensors such as time-of-flight and Kinect depth sensor, the performance of gesture recognition can be largely enhanced. However, since conventional two-dimensional (2D) image based feature extraction methods such as local binary pattern (LBP) generally use texture information, they cannot be applied to depth or range image which does not contain texture information. In this paper, we propose an adaptive local binary pattern (ALBP) for effective depth images based applications. Contrasting to the conventional LBP which is only rotation invariant, the proposed ALBP is invariant to both rotation and the depth distance in range images. Using ALBP, we can extract object features without using texture or color information. We further apply the proposed ALBP for hand tracking using depth images to show its effectiveness and its usefulness. Our experimental results validate the proposal.

1. Introduction

A natural user interface (NUI) for human–computer interaction (HCI) has gained much attention recently where its main goal is to recognize natural motion without any previous learning [1–5]. Among those existing NUI techniques, hand gesture recognition is recognized as among the most effective ways to convey user's intention to the computer [6,7].

Existing 2D image based hand gesture recognition systems use a variety of cue information such as color and texture [8–11]. However, the performance of these systems is affected by external environments such as illumination and complex background. In addition, the motion space for gesture recognition is limited by the cameras inability to perceive any change in visual depth.

Recently, 3D depth sensors which provide depth images with 3D scene information in real-time have been introduced into the commercial market (e.g. [22,23]). In terms of environmental illumination variation and background complexion, depth images taken by 3D depth sensor are not affected by these external factors as compared to that in RGB images. In addition, not only can the 3D shape of an object be analyzed directly, but also any change in the depth of an object can be perceived via depth images [24–26].

Although the 3D depth images based system can overcome limitations in 2D image based system, few literatures on NUI using 3D information can be found [19–21,31]. The conventional feature extraction methods such as local binary pattern (LBP) [27,28] cannot be directly applied to extract useful features due to the lack of texture information in depth images. In addition, many feature extraction methods such as PCA and LDA [29,30] are not invariant to depth scene.

In this paper, we propose a novel feature extraction technique, called adaptive local binary pattern (ALBP), for depth image based applications. Contrasting to the conventional LBP which is only rotation invariant, the proposed ALBP is invariant to both rotation and depth distance in range images. Using ALBP, we can extract object features without using texture or color information. Moreover, the proposed ALBP can verify detected features without the need of a classifier as that in typical applications of LBP. Consequently, we apply the proposed ALBP for hand tracking using depth images to show its effectiveness and its usefulness.

The paper is organized as follows. In the next section, we briefly review related existing hand tracking methods. In Section 3, we introduce the proposed ALBP and ALBP based hand tracking using depth images. Section 4 presents an extensive performance comparison with state-of-the-art hand tracking methods. Finally, our conclusion is given in Section 5.

* Corresponding author.
E-mail addresses: jurock@yonsei.ac.kr (J. Kim), sjyu@chu.ac.kr (S. Yu), dckim@msl.yonsei.ac.kr (D. Kim), katoh@yonsei.ac.kr (K.-A. Toh), syleee@yonsei.ac.kr (S. Lee).

http://dx.doi.org/10.1016/j.patcog.2016.07.039
0031-3203/

Table 1
Solutions for main problems of the object tracking.

Problem		Similar color objects	Complex background	No motion	Distance variation
Solution	Color			√	
	Model	√		√	
	Motion	√	√		
	Hybrid	√	√	√	
	3D	√	√	√	√

2. Related works

2.1. Previous hand tracking methods

A variety of information has been attempted to make reliable hand detection and tracking for HCI. Generally, two types of information have been adopted: 2D image based and 3D depth image based.

Table 1 summarizes those cue information adopted in existing tracking schemes and the corresponding problems they tackle. As seen in the table, one of the most commonly used approaches for 2D hand tracking is to use color information [8]. Although the color information is a relatively reliable cue for hand tracking, it has difficulty when the background object has color similar to skin color. Therefore, it is frequent to find motion information being combined with color information in order to deal with the background noise such as face behind the hand [9–11]. Also, pre-defined 2D and 3D hand models are frequently found in hand tracking to tackle variations in illumination, pose and occlusion [12–14]. Moreover, hybrid methods combining several hand tracking algorithms are currently a focused research topic [15–18] (Table 2).

Recently, 3D depth image including distance information has been acquired from a time-of-flight (ToF) camera or Kinect sensor in real-time [22,23]. This promotes wide applications of 3D tracking and gesture recognition such as interactive TV and games. In [19], a click event is used as initialization for activation of a hand tracking algorithm. In [20], an approach for head and hand tracking including hand posture is proposed based on range or depth images. In addition, a simple method for hand detection and tracking using Kalman filter in depth image from Kinect is proposed in [21] (Table 3).

Apart from the above problems in object tracking, there are other challenges particular to our hand tracking application. Firstly, the hands can be occluded by other objects such as a person standing right in front of the sensor. Secondly the location of hands may not be tracked accurately due to its fast movement. Thirdly, the hand may hold an object where tracking can be affected. We will discuss about the process which can deal with such situations in Section 3.

2.2. Local binary patterns based hand tracking

Local binary patterns (LBP) is an effective method to extract powerful features for texture classification [27,28]. Particularly, it has been successfully adopted for object tracking. To improve the tracking performance, additional features such as color and motion are combined with LBP feature. Although color feature based tracking methods such as MeanShift [51,52,54] CamShift [57] and skin color detection [53] are used to detect regions of interest, they are much affected by ambient lighting condition. Hence, LBP feature which has robustness to light variations is added to color features in object tracking [51,52,57,53,54]. LBP texture is combined with 3D rotation model [58] and motion features [56].

Table 2
State-of-the-arts literature survey of LBP based object tracking.

Method	LBP feature	Additional features
[51]	LBP histogram	MeanShift + particle filter
[52]		MeanShift
[53]		Skin color + particle filter
[54]		MeanShift + color histogram + expectation–maximization (EM)
[55]		Color + particle filter
[56]		Color + motion
[57]	LBP texture	CamShift
[58]		3D rotation model

2.3. Depth image based hand tracking

Many depth image based hand gesture recognition technologies have been proposed in the literatures [40–47] for touchless interfaces with devices such as TV, PC and gaming devices. In these applications, it is important to detect accurate hand location to achieve reliable recognition performance. Ever since the introduction of range camera such as Kinect sensor and ToF camera, it becomes possible to reach a level of recognizing natural hand motions precisely in 3D space.

Many researchers have been active in developing hand detection and tracking techniques utilizing depth images acquired from range camera. At the early first stage, simple depth threshold methods on depth images had been adopted for hand detection and tracking [40–42]. Since the distance information can be used to analyze the depth images, it is easy to detect objects within a certain range. Although such approaches are simple and fast, the tracking results are noisy and not precise. At the second stage, a combination of depth threshold with other features such as skin color, contour and size information was proposed to make up for weakness of depth threshold method alone [43–45]. However, such a combined method usually requires additional computational resources. Moreover, it has yet to achieve desired reliable performance. On the other hand, many hand gesture recognition algorithms have adopted open library such as Kinect SDK and OpenNI and show relatively reliable performance [46,47]. However, it is not documented which type of algorithms has been adopted for hand detection and tracking.

PrimeSense has provided two kinds of software, OpenNI and NITE [32]. First, OpenNI is the driver to capture the depth data from Kinect. And, NITE is the middleware module based on OpenNI for hand tracking and body tracking. Since NITE has many

Table 3
State-of-the-arts literature survey of hand detection and tracking using depth images.

Method	Hand detection		Hand tracking
[40]	Depth threshold		Closest pixel + mayor axis
[41]			Depth threshold (bounding box)
[42]			
[43]	Depth threshold	Skin color detection	
[44]		Contour features	No tracking
[45]		Size filtering	Depth threshold + size filtering
[46]	Kinect SDK		Kinect SDK
[47]	OpenNI		OpenNI