Vision-based Human Behavior Recognition and Behavior understanding: A Survey

Yonggang Zhang, Yingdong Ma

Inner Mongolia University, West road No. 235, Hohhot city, Inner Mongolia province, China

zyg0416@163.com, csmyd@imu.edu.cn

Keywords: Behavior recognition, Behavior understanding, Computer vision;

Abstract. Behavior recognition and behavior understanding has been a popular topic in computer sc ience. It attracted increasing interest in computer vision research community. Vision-based behavior Analysis has been applied broadly in many fields such as intelligent surveillance, perceptual user in terface, and virtual reality. This article analyses the up-to-date development of motion representatio n, the processing procedure of behavior analysis, and the commonly used methods of this field.

Introduction

Behavior recognition and behavior understanding is a key problem in computer vision with several i mportant applications including intelligent surveillance, perceptual user interface, motion analysis, and virtual reality [1,2,3]. It can be simply understood as the problem of data information classificat ion.

The university of Maryland implemented a real-time visual surveillance system W4 [4], which ca n not only localize and segment people body part, but also realize the multi-human tracking based o n the establishment of the appearance model.

In recent years, research on vision-based human behavior recognition and understanding has ma de tremendous progress. Wang ling introduced the general framework of human motion analysis [2]. Gavrila [1] study the movement of human body part. The development in this area has been introduc ed, which includes system initialization, target tracking, human pose estimation and recognition. A brief summary of some research challenges was also presented in his work. Most well-known intern ational journals such as IJCV, CVIU, PAMI, IVC, TIP and important academic conferences such as ICCV, CVPR, ECCV, etc. accept the human behavior recognition and understanding as an importan t research topic.

Shape Based Motion Representation Method

Human body model based methods built and parameterize human shape models first. Human motion recognition is then achieved by analyzing the parameterized models. These methods are often used to represent fine action, such as gestures, head movement, etc. Feng and Perona[5] presented a 2-D model to describe the human body pose. The 2-D model consists of ten rectangular modules to adapt the variety of human body pose.

The human motion is represented by the ratio of human body pixels and the total pixels within ea ch grid. Veeraraghavan [6] used the mark point on human contour to compute human motion. For a finite set of marked points, they analyze motion based on the Kendall shape theory [7]. This method has more accurate motion description. Moreover, it has small feature space dimensions and low calc ulation complexity.

Motion feature Based Motion Representation Method

Motion feature based motion representation method implement target classification by using the peri odic motion of human body. These methods extract visual features from a video frame and match si milar areas in consecutive image frames.

A typical motion feature based approach is the optical flow method which is built upon the fact t hat the motion characteristics of the target change over time. The optical flow algorithm computes

motion trajectories of pixels on moving objects to describe the movement of the target. Cutler and D avis [8] track the movement of a specific target to identify the man's head movement. Meyer [9] trac ks moving target by calculating the displacement vector of the optical flow field.

Behavior Recognition and Understanding

Behavior recognition can be regarded as a problem of data classification. That is, a process which m atches the target data with a specific model to identify the target behavior. Behavior recognition usu ally consists of four main components: input target data, model establishment, recognition events, an d the recognition results. There are several methods: template matching method, the state space met hod,2D and 3D behavior understanding and so on.

Template Matching

Template matching methods extract target features from the input image sequence and translate into a set of static shape models. Template matching approach presents several drawbacks. The significa nt one is that, in view of the way of generating template, new abnormal behavior can not be discover ed. In order to recognize abnormal behaviors from motion patterns, Khalid proposed a method to filt er anomalous activities[10].In this method, he believed that normal behaviors possess high correlati on between each other , thus abnormal activities can be detected through the comparison with norm al behaviors recorded in video sequences.

State Space Approaches

State space approach sets up profiles for normal behaviors. The activities deviated from these profil es are treated as anomalous. In other words, state space approach constructs a graphical model using aset of normal patterns to establish a classifier that can discriminate between normal and abnormal behaviors .A sequence of motion and static posture can be regarded as a traversal through all of diff erent states. The typical state space methods for behavior recognition and understanding are hidden markov models (HMM), dynamic Bayesian networks (DBNS), and neural networks.

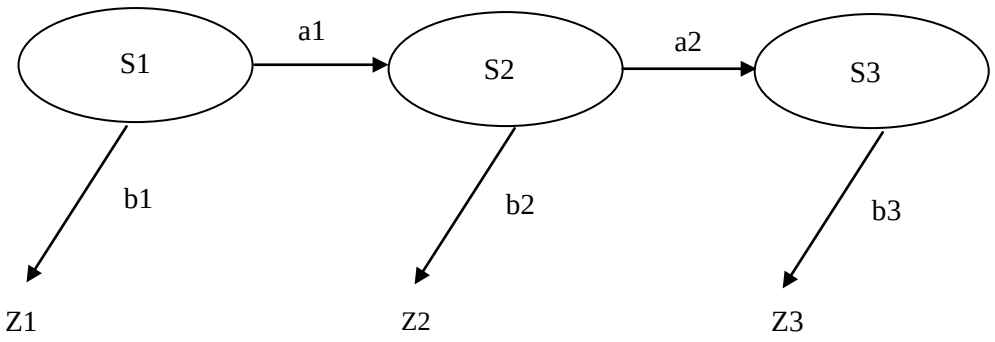

Figure. 1, Hidden markov model.

The HMM is a matching method with well developed theories [11]. It is a double stochastic proc ess. Since Yamato[12] introduced the HMMs to solve the behavior recognition problem,HMM meth od and its improved model has gradually become one of the mainstream approaches.For example, Yi n and Meng proposed a self-adaptive Hidden Markov Model(SA-HMM) based framework for abnor mal behavior recognition[13].Different from normal behaviors need to train from a large data set , th is framework can learn on-line from current data set and generated new models for abnormal activiti es.Nguyen [14] analyze the human motion by using the hierarchical hidden HMMs.

HMMs algorithm is efficient for single event sequence but not suitable for handling multiple eve nts. The dynamic Bayesian networks (DBNS), an extension method of HMMs, has been developed. DBNS algorithm reflects probability relationship of variables based on the network topological struc ture. Gong and Xiang [15] applied the DBNS to model the behavior of the crowd. The model was th

en be utilized to recognize the action of specific crowd.

Neural network was originally proposed by psychologists, neuroscientists, and mathematicians. It is aimed at developing and testing the mathematical model of neurons (basic working units of huma n brain).

2D and 3D behavior understanding

The aim of human segmentation in behavior understanding is to detect regions corresponding to stat i-c or moving humans,which provides a basis for later processing such as behavior understanding a nalysis.

In the discussion of related work, we focus on previous work in sign language recognition usin g 2D and 3D data as input.

HMM offers the advantage of being able to segment a data stream into its constituent signs implicit ly. In addition, it can deal with temporal and shape variance while preserving the order in hand mo vement [16]. Consequently, most work on sign language recognition is based on HMM. For exampl e Starner and Pentland presented an extensible system to recognize sentence level American Sign Language (ASL) [17]. This system used a view-based approach with a single camera to extract 2 D hands motion features as the input of HMM.Afterward, studies on 3D virtual environment and th e developments of various 3D input devices encourage adding the 3D interaction between human and computer to the user interface design. In order to reduce dimensional complexity of rotation al variance and global translation in 3D space, they project the 3D position sequence to the 2D coor dinates.

The Problems and the Future Development

Nowadays, research on behavior recognition and behavior understanding focuses on standard postur es, such as walking, running, jumping, squatting and sitting [18]. In recent years, some new technol ogies have been applied to improve behavior recognition performance. For example, the machine lea rning algorithm is applied to train a human action model for better recognition rate.

The second direction is the combination of behavior identification with biometric identification. This combination is very important in intelligent monitoring, entrance guard system, and intelligent robot interaction. In these applications, a system is not only required a preliminary recognition of tar get behavior, but also need to collect the biological characteristics of the target for further recognitio n.

Conclusion

Human behavior recognition and understanding has become an important research direction in thefi eld of computer vision. It has a wide range of applications in the smart surveillance, perceptual inter face, motion analysis, and virtual reality.

In this paper, the current research status and the commonly used methods of human behavior reco gnition and understanding are reviewed. Three main aspects are covered, including motion feature e xtraction, general behavior recognition process, and typical approaches of behavior understanding. Finally, the existing problems and the future development directions of behavior recognition and un derstanding are discussed.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 6 1461039, and the Program of Higher-level talents of Inner Mongolia University, 30105-125130.

References

[1] Gavrila D. The visual analysis of human movement: a survey, J. Computer Vision and Image Un derstanding, 1999, 73(1):82-98.

[2] Wang Liang, Hu Wei ming, Tan tie niu. Analysis of human motion vision, J .Chinese journal of computers,2002,25(3):225-237.

[3] Collins R, Lipton A and kanade T.introduction to the special section on video survelliance.IEEE Trans pattern Analysis and Machine Intelligence, 2000, 22(8):745-746.

[4] Haritaoglu I, Harwood D and Davis L.W4:real-time survelliance of people and their activities, J IEEE Trans Pattern Analysis and Machine Intelligence, 2000, 22(8):809-830.

[5] Feng X, Perona P. Human action recognition sequence of movelet codewords, A. In Proc. 3D P VT, C. Italy: IEEE Computer Society Press, 2002.717-723.

[6] Veeraraghavan A,Chowdhury A R, Chellappa R. Matching shape sequences in video with appli cations in human movement analysis ,J. IEEE Trans.PAMI,2005,27(12):1896-1909.

[7] Kendall D G, Barden D, Carne T K, Le H.Shape and Shape Theory, M. New York: Wiley, 1999.

[8] Cutler R and Davis L. Robust real-time periodic motion detection, analysis, and applications. IE EE Trans Pattern Analysis and Machine Intelligence, 2000, 22(8):781-796.

[9] Meyer D, Denzler J and Niemann H. Model based extraction of articulated objects in image seq uences for gait analysis. In:proc IEEE International Conference on Image Processing, Santa Bar bara, California 1997, 78-81.[18]

[10] S.Khalid and A.Naftel, "Classifying spatiotemporal object trajectories using unsupervised learni ng of basis function coefficients," in 2005 Proc. of the third ACM international workshop on Vi deo surveillance & sensor networks.New York, NY,USA: ACM Press,2005,pp.45-52.

[11] Rabiner L. An introduction to hidden Markov models, J. IEEE ASSP Magazine, 1986, 3(1):4-8.

[12] Yamato J, Ohya J, Ishii k.Recognizing human action in time-sequential images using Hidden Ma rkov model, A. In Proc CVPR, C. IEEE, 1992.379-385

[13] J.Yin and Y.Meng,"Abnonnal Behavior Recognition Using Self- Adaptive Hidden Markov Mod els," Lecture Notes in Computer SCience, vol. 5627, pp. 337-346,2009.

[14] Nguyen N T, Phung D Q, Venkatesh S, Bui H.Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model, A. In Proc CVPR, C. San Diego: IEE E computer Society Press, (2005).955-960.

[15] Gong S, Xiang T. Recognition of group activities using dynamic probabilistic networks, A. In P roc ICCV, C. Washington: IEEE Computer Society Press, 2003.742-749.

[16] N.Robertson and I.Reid,"A general method for human activity recognition in video,"Computer Vision and Image Understanding（CVIU）, pp. 232-248,2006.

[17] T. Starner and A. Pentland, "Real-time American sign language recognition from video using hi dden markov models," in Con! Rec. 1995 IEEE International Conference on Computer Vision, pp. 265-270.

[18] Jin Ning, Mokhtarian F. Human motion recognition based on statistical shape analysis, C .Proc of IEEE Conference on Advanced Video and Signal based Surveillance.2005.