한국컴퓨터정보학회 동계학술발표대회 논문집 제16권 제2호 (2008. 12)

휴먼 모션 분석을 통한 이벤트 검출 및 인식

최요환*, 이창우*
*군산대학교 컴퓨터정보공학과
e-mail: {emilg3, leecw}@kunsan.ac.kr

Vision-based human motion analysis for event recognition

Yao Huan Cui*, Chang Woo Lee*
*Dept of Computer Informatioin Engineering, Kunsan National University

요 약

최근 컴퓨터비전 분야에서 이벤트 검출 및 인식이 활발히 연구되고 있으며, 도전적인 주제들 중 하나이다. 이벤트 검출 기술들은 많은 감시시스템들에서 유용하고 효율적인 응용분야이다. 본 논문에서는 사무실 환경에서 발생할 수 있는 이벤트의 검출 및 인식을 위한 방법을 제안한다. 제안된 방법에서의 이벤트는 입장(entering), 퇴장(exiting), 착석(sitting-down), 기립(standing-up)으로 구성된다. 제안된 방법은 하드웨어적인 센서를 사용하지 않고, MHI(Motion History Image) 시퀀스(sequence)를 이용한 인간의 모션 분석을 통해 이벤트를 검출할 수 있는 방법이며, 사람의 체형과 착용한 옷의 종류와 색상, 그리고 카메라로부터의 위치관계에 불변한 특성을 가진다. 에지검출 기술을 HMI 시퀀스정보와 결합하여 사람 모션의 기하학적 특징을 추출한 후, 이 정보를 이벤트 인식의 기본 특징으로 사용한다. 제안된 방법은 단순한 이벤트 검출 프레임웍을 사용하기 때문에 검출하고자 하는 이벤트의 설명만을 첨가하는 것으로 확장이 가능하다. 또한, 제안된 방법은 컴퓨터비전 기술에 기반한 많은 감시시스템에 적용이 가능하다.

키워드 : Event recognition, ROI, Motion analysis, MHI.

I. Introduction

Visual surveillance systems usually attempt to detect, recognize, and track objects of interest from image sequences. In addition, those systems try to understand and describe the objects' behaviors. Numerous event detection and recognition methods have been proposed for visual surveillance systems. For example, in [7], unusual events on stairs are automatically detected and recognized from video data to rapidly find and analyze the events of interest within large quantities of video data. Falling-down event detection has been conducted to detect human falls in real time for the enhanced safety [6]. Michael Spann proposed event detection in the traffic system to detect tailgating for intelligent car park[8].

For humans, it is easy to identify events even though they have much similarity in some actions. But for computers, it is difficult to distinguish those events and also to detect one of such complex events as easily as humans can do. For recognizing those events, the proposed method, first utilizes a little domain knowledge that assumes the door's position is known. In the proposed method, six events are recognized by a combination of edge detection, MHI, and geometrical characteristics of the human body shape.

II. Proposed approach

An event can be considered as a series of continuous actions. Different events always consist of different actions. There must be an event occurring when the door is open regardless of whether this event is entering or exiting. In this paper, entering and exiting events are defined as a combination of the door's status and human movements. In addition the geometric features extracted from the human motion analysis also are used to recognize sitting down and standing up events effectively. With our own definitions, a brief overview of the proposed method is illustrated in Figure. 1.

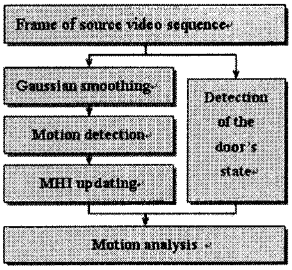

Figure 1. Overview of the framework

2.1. Detection of the door's state

In order to detect the door's state, the frames representing the change of the door's state are detected with a canny edge detector and Probabilistic Hough Transform (PHT). A region of interest (ROI) is set in the higher position of the door manually to decrease the calculation and avoid occlusion by the human motion. Additionally, Gaussian smoothing algorithm is applied to the ROI for denoising.

According to the moving direction in experiments, the moving doorframe is considered as the most right vertical line among the detected vertical lines. The threshold values used in the PHT are determined empirically, in which the minimum line length is 28 pixels while the height of ROI is 30. The maximum gap between line segments lying on the same line is 1. It was considered as a vertical line when the horizontal difference of the two points in the detected line is less than 3. After detecting the change of the door state, we determine the door state as follows. If the size of the door decreases to less than 2 pixels for two consecutive frames, the door is closed; otherwise it is open. If the change of the door state is not detected, the former state is preserved.

2.2. Motion detection and MHI updating

Nearly every visual surveillance system starts with motion detection. Motion detection aims at segmenting regions corresponding to moving objects from the rest of an image [3]. Temporal differencing was used to detect motion to be prepared for MHI updating. The MHI is a static image template where the pixel intensity is a function of the recency of motion in a sequence [1]. It is constructed by layering motion regions over time with an update rule in Equation (1).

$$mhi_i(x,y)=\begin{cases} timestamp & if \quad silh(x,y)\neq 0 \\ 0 & if \quad silh(x,y)=0 \; and \; mhi_{i-1}(x,y)<timestamp-duration \\ mhi_{i-1}(x,y) & otherwise \end{cases} \quad (1)$$

The binary sequence silh(x, y) indicating a region of motion is extracted from the original image sequence by image differencing. Timestamp is the current time, and the duration means the maximal duration of motion track in the same units as the timestamp. The useful MHI offers not only the moving human's current position but also the historical movement information during certain duration in a single image used in the following motion analysis step.

2.3. Motion analysis

Similar with the method described in [4], after getting the MHI, the whole moving region is enclosed by a minimum bounding box as shown in Figure 2, which we call as global bounding box.

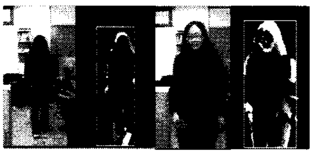

Figure 2. Global bounding box located in MHI

MHI provides important information, i.e. the direction of motion. With the aim of seeing clearly, most recently moving pixels are colored with magenta and the other moving pixels are colored with cyan as shown in Figure 3 (b). In Figure 3 (c), the outer blue rectangle represents the global bounding box; while the inner red one denotes the most recently moving region in MHI. The relationship between the positions of the two bounding boxes is defined as the geometrical characteristics for our event recognition, referred as geometric patterns.

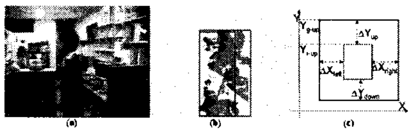

Figure 3. (a) sample frame (b) colour MHI (c) geometric pattern

Several variables are selected as features from the geometric pattern, and their variation rules are considered to describe human behaviors for event recognition.

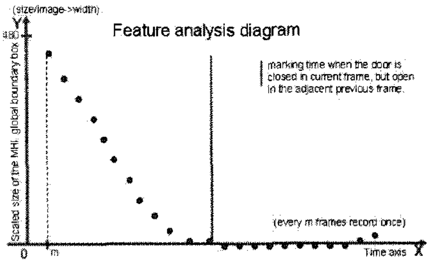

Figure 4. Example of geometrical characteristics

$Y_{g\text{-}up}$ is the y value of the upper position of the global bounding box. $Y_{r\text{-}up}$ is the y value of the upper position of recently moving region bounding box. The distances between these two bounding boxes are calculated in pixels for each direction as ΔY_{up}, ΔY_{down}, ΔY_{left}, and ΔY_{right}. In Figure 4, tracking the size of the global bounding box produces one feature for the detection of events. Similar analyses are constructed for the variables $Y_{g\text{-}up}$, $Y_{r\text{-}up}$, or ΔY_{up}, respectively.

2.4. Definition of events

The features obtained from the motion analysis provide the information necessary to recognize the events. We define events as follows.

1. **Entering: No movement, door opens, person approaches.**
2. ***Exiting*: Person leaves, door opens, person disappears.**
3. ***Sitting-down*: $Y_{g\text{-}up}$ varies from level condition and continuous falling condition, in the same time, $Y_{r\text{-}up}$ rises and ΔY_{up} falls continuously.**
4. ***Standing-up*: $Y_{g\text{-}up}$ varies from continuous rising condition to level condition, in the same time, $Y_{r\text{-}up}$ rises continuously and ΔY_{up} stays zero level condition.**

Whether entering or exiting, the door's state changes are in the same order in most instances as follows: closed, open, and then closed. However, the significant difference between entering and exiting events is observed when human activity differs but the door's states are unchanged. In the proposed method, the changing laws of the bounding box's size and the relation of bounding boxes' position in an MHI are selected as the features to distinguish the entering/exiting events. The diagram of the scaled global bounding box's size is constructed, and the detection of leaving and approaching pattern is defined to robustly distinguish them. For a human walking straight, the leaving and approaching patterns are illustrated in Figure 5. In the approaching pattern, the absolute distance values in four directions are all smaller than a certain threshold, while they are greater in the leaving pattern.

To determine whether it is entering or exiting, we define these events as follows. If the leaving pattern is detected when the door is closed and the average global bounding box size is decreased to zero, then we define it as an exiting event. On the other hand, if the size is almost zero when the door is closed and the approaching pattern is detected, then it is an entering event.

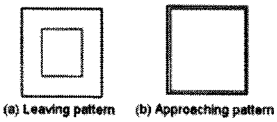

Figure 5. Defined leaving and approaching pattern

For standing up and sitting down events, three patterns of $Y_{g\text{-}up}$, $Y_{r\text{-}up}$ and ΔY_{up} are constructedasthespecificfeaturesfor each event. The time is considered as the candidate time at which sitting down or standing up event occurs when the value of $Y_{g\text{-}up}$ varies between the level condition and the continuous rising or falling condition. Furthermore, unique features of the changing rule in $Y_{r\text{-}up}$ and ΔY_{up} value are combined to determine the event type. Therefore, if $Y_{g\text{-}up}$ varies from the level condition to continuous falling conditionand at the same time, $Y_{r\text{-}up}$ rises and ΔY_{up} falls continuously, sitting down event occurs; if $Y_{g\text{-}up}$ varies from the continuous rising condition to the level condition and at the same time, $Y_{r\text{-}up}$ rises continuously and ΔY_{up} stays at the zero level condition, standing up event occurs.

III. Experiment results

We tested the proposed method on video sequences with 640 x 480 pixel resolution. We captured video sequences of ourselves using an EVI-D70 SONY camera set on the table about 1.5 m above the floor. The test platform was Windows XP Professional running on a desktop PC with Pentium IV 3 GHz CPU and 1 GB of memory.

The MHI extracted from the video sequences with a fixed duration, τ of 3 seconds in our experiment is shown in Figure 6 (b). By getting the motion information from an MHI, the whole moving region is enclosed with a global bounding box. The leaving and approaching patterns detected are presented in Figure 6 (c).

The size variation of the global bounding box in our test videos as described in Section 2.3 is shown in Figure 6(d). The dot indicates the scaled size of the global bounding box at a corresponding time frame. Also the marking times are shown as green vertical lines in the figure below. The red lines indicate the coordinate system. Before the first marking time, the y value decreases and the leaving pattern is detected. After the marking time however, the y value is almost zero for a while, and thus we determined that the exiting event occurs. Meanwhile, the entering event happens after the second marking time y values are increasing and the approaching pattern is found.

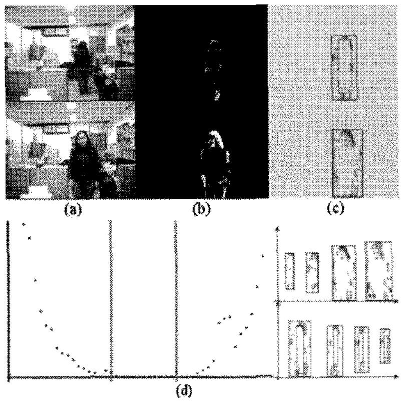

Figure 6. (a) Video frame (b) MHI (c) geometrical characteristic (d) diagram and sample geometric patterns

Corresponding to the geometric patterns of the same video sequence, the three variables (Y_{g-up}, Y_{r-up}, ΔY_{up}) are plotted as curves in Figure 8. The four vertical dashed lines indicate the sitting down and standing up events where the value of Y_{g-up} changes between level and continuously rising or falling. The horizontal axis represents the time reference. When Y_{g-up} varies steady, Y_{r-up} falls rises continuously, and when ΔY_{up} rises simultaneously for a space, sitting down event occurs; when both Y_{g-u}p and Y_{r-up} rise continuously and the ΔY_{up} becomes zero simultaneously for a space, then standing up event occurs.

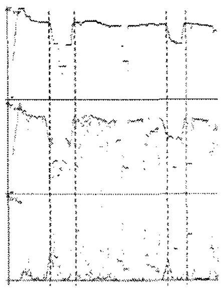

Figure 8. Variation graph of Y_{g-up}, Y_{r-up} and ΔY_{up}

The proposed method also works on other relative video sequences performed by different persons (dressed in different clothes), including 35 entering/exiting events and 32 sitting down/standing up events respectively. However, the experimental results show that the proposed method is invariant to body shape, clothing and position. The experiment accuracy is shown in Table 3. The false recognition rates are caused by the obstacles occluded such as chairs.

Table 3. Events recognizing accuracy

Event type	entering	exiting	sitting down	standing up
Accuracy	91.4%	94%	90.6%	93.7%

IV. Conclusions

In this paper, motion detection and geometrical characteristics of the human shape with MHI sequences were proposed as events detection and recognition method for surveillance systems. The features extracted from the human shape in MHI sequences describe the human behaviors uniquely for event recognition. Moreover, the proposed framework is also easy to extend for further development, in which it is simple without any model construction, probability calculation and training work. The proposed method is widely applicable to many applications such as for people counting field for security management. In the future, more investigations are required to describe various human activities for visual surveillance applications.

Reference

[1] Intel open source Computer Vision Laboratory http://www.cse.ohi-state.edu/~jwdavis/CVL/Research/MHI/mhi.html
[2] Osama Masoud, Nikos Papanikolopoulos; "A Method for Human Action Recognition" Image and Vision Computing, Volume 21, Number 8, 1 August 2003.
[3] Weiming Hu; "A Survey on Visual Surveillance of Object Motion and Behaviors" IEEE Systems, Man, and Cybernetics Society, AUGUST 2004
[4] Ji Tao, Mukherjee Turjo; "Fall Incidents Detection for Intelligent Video Surveillance" Information Communications and Signal Processing 2005 Fifth International Conference;
[5] A. Bobick and J. Davis; "The recognition of human movement using temporal templates" In IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 23, pages 257 - 267, March 2001.
[6] Jean Meunier; "Fall Detection from Human Shape and Motion History using Video Surveillance"; 21st International Conference on AINAW'2007.
[7] Jasper Snoek; "Automated Detection of Unusual Events on Stairs", the 3rd Canadian CRV Conference on 07-09 June 2006.
[8] Georgios Diamantopoulos, Michael Spann; "Event detection for intelligent car park video surveillance" Real-Time Imaging Volume 11, Issue 3, June 2005.
[9] Douglas Ayers, Mubarak Shah; "Recognizing Human Action in a Static Room"; Fourth IEEE Workshop on Applications of Computer Vision (WACV'98).
[10] Douglas Ayers, Mubarak Shah; "Monitoring human behavior from video taken in an office environment"; Image Vision Computing Volume 19, Issue 12, 1 October 2001, Pages 833-846.