Toward Real-Time Extraction of Pedestrian Contexts with Stereo Camera

Kei Suzuki\(^1\), Kazunori Takashio\(^2\), Hideyuki Tokuda\(^{1,2}\), Masaki Wada\(^3\), Yusuke Matsuki\(^3\), Kazunori Umeda\(^3\)
\(^1\)Graduate School of Media and Governance, Keio University,
\(^2\)Faculty of Environment and Information Studies / CREST, JST
\(^3\)Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University / CREST, JST

Email: suzuk@ht.sfc.keio.ac.jp, kaz@mkg.sfc.keio.ac.jp
wada@sensor.mech.chuo-u.ac.jp, matsuki@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

Abstract—We extract the mood of disquiet on street corners in real-time with stereo video camera systems. Last year we proposed a novel stereo measurement algorithm to detect moving people, which was focusing on moving region in video data. In this paper, we report our prototype of probabilistic inference engine that can detect contexts of individual pedestrian and groups of pedestrians. We demonstrated that the real-time extraction of higher-level pedestrian contexts using the Bayesian Network model was effective for extracting several pedestrians’ context.

Index Terms—distributed camera system, stereo vision, real-time context analysis, pedestrian context, bayesian network model.

I. INTRODUCTION

The goal of this research project is to detect existence of suspicious individuals and uneasy atmosphere by using stereo cameras set up at street corners. In recent years, with the advance of video camera technology and the evolution of machine learning, many projects on extraction of individual pedestrian’s activities from video data have been conducted [1],[2],[3]. Our project aims not only to detect such individual contexts but also extract contexts of pedestrian group in real-time. By using a stereo camera system that focuses on moving regions, we can make real-time 3D measurement of more than one pedestrian’s region. However, the more people in one video camera window exists, the more computation the system needs to process to execute real-time extraction of higher-level pedestrian contexts. In order to keep the complexity reasonable, the prioritized observation range should be narrowed dynamically.

This paper describes our prototype system of extracting the individual and group contexts in real-time. This prototype infers the pedestrian contexts by using the Bayesian Networks. Our stereo camera system detects moving region data at the frequency of 18 times/sec or less. Since the average speed of target pedestrians is about 1.56 m/sec, this detection rate of the moving region is enough to trace.

In this paper, we first present stereo camera system that focusing on moving region. We then discuss the pedestrian contexts and describe the usage of Bayesian Network. Real-time extraction of pedestrian context show the design of our prototype system. Finally, we discuss the experiments, and current limitations and potential improvement for this research approach.

II. STEREO CAMERA SYSTEM THAT FOCUSING ON MOVING REGIONS

Basic algorithm of the stereo camera system used in our prototype is shown in Fig.1. In prior technique on stereo image analysis, a parallax image is obtained by matching right and left camera images. In our method, a parallax image is obtained as follows. First, moving regions of both right and left cameras are extracted independently. Then, moving regions of two images are matched by using the subtraction processing to extract the moving region.

Fig.3(a) and Fig.3(b) shows parallax images in the scene shown in Fig.2 generated by our method and the prior method respectively. We use the simple background subtraction method for generating the moving region in Fig.3(a). In comparison with Fig.3(b), Fig.3(a) denotes that the target areas of stereo image matching are narrowed down to only the moving regions. Therefore, erroneous decision in a phase of stereo image matching can be suppressed. Especially this method is reasonable for the monitoring application because it can generate a necessary and sufficient result in each monitoring scene.

The outputs of our stereo camera system are as follows: observed moving regions’ feature (i.e., distances, center of gravity coordinates, height and width), timestamp and label number. In our prototype, the time stamp is measured by millisecond accuracy. The same label number indicates same moving object. With the planar dimension of moving region, we can easily recognize whether observed region is a cat or a person, an adult or a children, or count the number of people in the area.

III. PEDESTRIAN CONTEXTS

In this section, we discuss pedestrian contexts, and describe the usage of Bayesian Network for extraction the pedestrian context. Then, we show some concrete contexts to be extracted by our prototype system and discuss Bayesian Network model to be used.

A. Features of pedestrian contexts

First, we describe typical features of pedestrian moving at a street corner. Acquaintances act as a group, and they move and keep a close distance. On the other hand, unrelated people...
generally act separated. However, when they are crossing each other, they may be so close that their shoulders are hit. Moreover, when a pedestrian squats down suddenly, it is difficult to recognize whether he/she is taking rest or falls down from only one event at that time. Therefore, in order to extract accurate pedestrian contexts we have to observe a sequence of such changes of moving region over a certain period, and analyze them as a time-series data. For example, when a pedestrian who has been moving and stopped suddenly, the system detects he/she was falling down, that is the situation has high probability of tumble. In this prototype, the inference engine of the pedestrian contexts uses sequences of moving region data divided by different intervals for each context.

This prototype classifies the pedestrian contexts into individual contexts and group contexts. As the single person moving in a camera, we can find a person who is walking in usual speed and a person who trots along, etc. As the group of pedestrians, a harmless crowd (companions or people having same intentions), an accidental crowd (group that moves in the same direction incidentally or the temporary crowd), disquiet crowd should be considered.

B. Bayesian Networks approach

As the model for analyzing the time-series data of pedestrians’ moving region, we adopted the Bayesian Networks[5],[4]. The Bayesian Networks is useful to show integrated probability distributions effectively by using graph structures, and has an advantage that it can visually describe dependencies among events. Moreover, compared with traditional rule base models, it has an advantage of appropriate probabilistic inferences can be executed even with uncertain factors. For the analysis of the sequence of moving region data, the Bayesian Networks that can deal with uncertain and complex probability distributions is appropriate, since the system has to work well even under the situations where the system is temporarily unable to obtain any data by the influence of the light source, or the situation where the much noise enters to observation data.

From the view point that how to infer the posterior probability, the Bayesian Network algorithm can be classified into two categories, the exact inference algorithm and the approximate inference one[7]. For the real-time context analysis, we use the latter scheme that has excellent response performance. The approximate inference algorithm infers posterior probabilities for each probability variable approximately by the computer simulation. As the inference engine can output results with arbitrary accuracy, it is possible to keep the inference time reasonable, which satisfies both of real-time performance and
C. Bayesian Network for target pedestrian contexts

The prototype system generates individual pedestrian contexts and group pedestrian contexts. It can extract the tumble situation as an individual pedestrian context and can recognize the friendly group as a group context.

1) Context of individual pedestrian: Fig.4 shows a Bayesian Network model to extract “tumble” situation as the individual pedestrian context. Each ellipse corresponds to the probability variables that indicate the moving region data. Here, the “tumble” denotes to tumble context, and the xs1, ys1, and z1 corresponds to average value of x, y speed, and z coordinates in one second respectively. The xs2, ys2, and z2 indicate the past one second data of xs1, ys1, and z1. When the pedestrian has been tumbled, it is generally observed that the pedestrian’s moving speed is dropping suddenly. And as he/she is falling down, the z coordinate value indicates also dropping suddenly. These probability variables are used to define the feature of these tumble situation. The probability variables corresponding to moving region data have direct dependencies with the variable “tumble”, which are shown as direction links connecting between the two probability variables. Each dependence relationship is weighted by “conditional probability distribution table”.

2) Context of group or mobs: Fig.5 shows the Bayesian Network model for extracting “companion group” context. The “companion group” corresponds to this context, and the vs1, dist1, and angle1 correspond to the similarity of speed vector, the distance, the angle of between speed vectors of two pedestrians in one second. The vs2, dist2 and angle2 correspond to each one second before the marked data, vs1, dist1 and angle1. The vs3, dist3 and angle3 denote to each one second before the marked data respectively. When pedestrians construct a companion group, it is generally observed that walking speed and direction of the pedestrians are very similar, and they are close together. Moreover, basically they adjoin each other, thus it is highly possible that a pedestrian who walks back or ahead of the group is stranger.

IV. REAL-TIME EXTRACTION OF PEDESTRIAN CONTEXTS

We employed a sliding window algorithm for the moving region data’s time series analysis. To realize the prioritized observation, we developed the priority queue that contains the label numbers. Each label number in this queue is attached to the observation target to extract the context according to this priority queue.

The context inference by Bayesian Networks requires the heaviest processing load in our system. If the number of observation targets increases, the system requires more time to extract context from all the targets. To meet with the frame rate of the stereo camera, we applied the event mechanism. The context analysis is executed only when this event is called. In this prototype, we defined two kind of events: tumble event and companion event. For the tumble event of individual pedestrian, the system checks the alteration of the z coordinate. If there is a sudden change in the value of z coordinate, the system fires the tumble event. For the companion event of group of pedestrians, the system checks the distance between pedestrians. If the distance between pedestrians falls below the certain value, the system fires the companion event.

In this prototype, two kinds of worker threads were implemented. One is an event detection thread of each pedestrian context, and the other is the inference thread. The event detection thread refers the priority queue, and the system sequentially generates the event by analyzing the data of the registered pedestrian. When an event is detected, the system calls the inference thread which corresponds to that event. The event detection thread shares the buffer of moving region information. The inference thread infers the context by using Bayesian Network engine.

We implemented our prototype system by using the C++ programming language. The graphical user interface (GUI) for our system is developed by using the QT Library.

V. EXPERIMENTS

We tested our prototype system by using the stereo camera system. We utilized Bumblebee2 (Color, f=3.8mm, XGA) of
Point Grey Research inc.) as the stereo camera. The vision size was 320 pixels x 240 pixels at the this experiment. The inference engine is activated in a personal computer which has Intel’s Core2Duo E6850(2.6GHz 2 core) CPU and Windows Vista operating system.

Fig.6(a) shows the experimental scene and Fig.6(b) shows the subtraction image by using our system. In this experiment, although the person to be detected is considerably far and small, the system succeeded in detecting the person and calculates the distance. From this experiment, we could show that our system can be implemented for the general monitoring purpose. The processing speed of the system was very near to 18fps the maximum frame rate of Bumblebee2 stereo camera.

Fig.7 shows the GUI of our prototype system. In this experiment, the moving region data was sent from the stereo camera system at the speed of 7 times/sec. Because we conducted this experiment in outdoor environment, the noise data such as the natural sunlight was added to the vision data. The Bayesian networks is appropriate model for this kind of uncertain environment. Even though the required time for inferring context by the Bayesian network engine was 58 milliseconds, since the prototype system employs the event trigger model, the context extraction was realized in real-time. The prototype system could extract the companion and tumble context from experimental data. We could confirm the validation of Bayesian networks and our prototype system from this experiment. The system could show the effectiveness of extracting two or more pedestrians’ context in real-time.

VI. CONCLUSION

We presented an approach for extracting pedestrian contexts in real-time that uses the stereo camera system focusing on moving regions. By using the Bayesian Networks, our prototype system can extract the tumble situation as individual pedestrian context, and the companion group as group context.

Future works includes revision of software architecture for handling increase of observation targets and that of stereo cameras at street corners. And we will discuss the basic performance of our method by comparing with other technical approaches.

ACKNOWLEDGMENT

This research has been conducted as part of Core Research for Evolutional Science and Technology of Japan Science and Technology Agency.

REFERENCES