Intelligent Monitoring of Pedestrian Behavior by Image Understanding

Handri Santoso
Nagaoka University of Technology
Graduate School of Engineering
Nagaoka, Niigata 940-2188 Japan
E-mail: bondry@alice.nagaokaut.ac.jp

Kazuo Nakamura
Nagaoka University of Technology, Department of
Management and Information Systems Science
Nagaoka, Niigata 940-2188 Japan
E-mail: nakamura@kjs.nagaokaut.ac.jp

Abstract—**The important aspect that should be taken in traffic design planning is pedestrian safety. One of safety factors, which haven't sufficiently been considered yet, is the pedestrian behavior itself. This paper proposes a conceptual method of human behavior monitoring when crossing road by image understanding. The method is based on information fusion of image features obtained from pedestrian behavior using Choquet Integral Agent Network for realizing linguistic understanding. At the front end image processing is required to extract features of pedestrian image and estimate basic attributes and states of pedestrian body parts, such as skin color of face, hair color, height, central body position, etc. By information fusion mechanism, multiple input data are aggregated to estimate macroscopic information, i.e. macroscopic attributes and states of pedestrian as an individual such as age classification, gender, pedestrians flow states, etc.**

Keywords: Pedestrian behavior, Image processing, Choquet integral agent network, Macroscopic information.

I. INTRODUCTION

Pedestrian activity can be considered to be the product of two distinct components — the configuration of the street network and the location of particular attractions (shops, offices, public buildings etc.) on that network. In order to explore the influence of each of these, it is first necessary to observe and record the movement of pedestrians in city streets. It is not just the city planner who is interested in pedestrian movements in town centers. The retail industry has a particular interest since retailers aim to locate their shops in areas which can attract a lot of passing trade. While planners and retailers are the most obvious groups, it is clear that others (such as the emergency services) have an interest in understanding the way which people move in an urban setting. The need to understand the way in which people move through towns leads to the desire to predict pedestrian movement for, e.g., identifying the likely impacts of pedestrian crossing in the a city street, identifying the optimum location for a new shop, or assigning and allocating staff to manage a street festival [5]. In monitoring the movement of people in street, first, it is also important to understand behavior of pedestrian itself. With understanding the behavior of pedestrian in the street, we can estimate type of pedestrian attribute, like elder people, adult or children, man or women, good or bad mood pedestrian, worker or tourist, local community or foreigner, etc. This information can become valuable for the planner or advertising agencies when they have planned to build public facilities or invest new shop in that area.

This paper aims at proposing a conceptual cognitive framework to comprehend collective pedestrian behavior in the street. The basic idea of the proposal is introducing an approach as interactive agents each of which derives basic human feature by image understanding. More specifically each agent behaves as human attribute factor under mutual interactions among autonomous agents.

In this paper, first, general consideration is to extract basic attributes and state of pedestrian body parts from video frames by image processing. And then a conceptual cognitive framework is proposed to realize pedestrian behavior understanding. After the proposal several aspects of uncertainty and flexibility to be considered in the modeling are discussed. Finally possibility of soft computational approaches to manage with such uncertainty and flexibility is shown for realizing linguistic understanding of pedestrian behavior owing to several required level.

II. FEATURES OF COLLECTIVE PEDESTRIAN BEHAVIOR

A. General Consideration on Features of Pedestrian Behavior

As the basis of understanding pedestrian behavior by a cognitive framework several features of pedestrian are discussed below.

1) Pedestrian is a collection of various types which can be classified as male, female, children and cyclist.
2) Pedestrian behavior also is affected by various condition of natural / social environment, e.g., population, culture and habitual, community, area or town activity, etc.
3) Image of pedestrian behavior consists of several features, e.g. position, color, shape, etc.
4) Weather condition such as cloudy, sunny, rainy or snow will give significant effect for intensity changes and noise during image extraction.

B. Feature Extraction by Image Processing

The main module in the tracking unit is using a digital video camera, to capture images and save the image into mini DV video tape. The images are obtained from video camera directly connected to computer through USB connection and processed by computation using MATLAB program [3].

Figure 1, shows flowchart extracting pedestrian image features on basic states using conventional image processing.

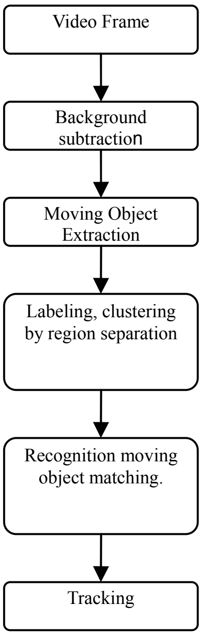
Video Frame
Background subtraction
Moving Object Extraction
Labeling, clustering by region separation
Recognition moving object matching.
Tracking

Figure 1 Basic image feature extraction

The simplest method to detect changes between two registered frames would be to analyze the frame difference (FD) image, which is given by

$$FD_{k,r} = s(x,k) - s(x,r) \qquad (1)$$

where $\boldsymbol{x}=(x_1, x_2)$ denotes pixel location and $s(\boldsymbol{x}, k)$ stands for the intensity value at pixel $\boldsymbol{x}$ in frame k. The *FD* image shows the pixel-by-pixel difference between the current image $\boldsymbol{k}$ and the reference image $\boldsymbol{r}$. In order to distinguish the nonzero differences that are due to noise from those that are due to local motion, segmentation can be achieved by thresholding the *FD* as

$$z_{k,r}(x) = \begin{cases} s(x,k) & \textit{if } |FD_{k,r}(x)| > T \\ 0 & \textit{otherwise} \end{cases} \qquad (2)$$

where T is an appropriate threshold. Here, $z_{k,r}(x)$ is called foreground pixel extracted, which is equal to "***s(x,k)***" for changed regions and "0" otherwise [8].

The results from the background difference sometimes contain noise and disunity. Preprocessing standard opening and closing morphology is performed before object detection to reduce the noise and increase the unity of the connected component of objects. After the moving objects are separated from background, the pixel that contain the picture of pedestrian and other moving objects are detected in the binary image. A simple but powerful tool for identifying and labeling the various objects in a binary image is a process called region labeling, blob coloring or connected component identification as shown in Figure 2.. It is useful since once they are individually labeled, the object can be separately manipulated, displayed, or modified. Region labeling seeks to identify connected group of pixels in a binary image that all have the same binary value [2]. The simplest such algorithm accomplished this by scanning the entire image (left to right, top to bottom), searching for occurrences of pixel of the same binary value and connected along the horizontal or vertical direction. A record of connected pixel groups is maintained in a separate label array having the same dimension, as the image is scanned. After an object in the binary image is detected, all pixels in the corresponding location of color image are captured. A bounding box can be drawn to mark detected object and object descriptors are calculated based on both the binary and color images. A number of features are calculated based on the binary object, color object and the contour object. Chain coding of the contour object gives the perimeter, object width and height. A binary object produces the area and centroids coordinate of object. Statistical descriptors such as color mean, variance, skewness, kurtosis, as well as area, perimeters, compactness, center of gravity, and moments, etc. are also calculated. Each object detected is assumed one feature point that contains the aforementioned descriptors. Area of an object is the number of black pixels in the region bounded by the box. The coordinate of the moving object is calculated based on the center of gravity. Each black mask of binarized image has a corresponding pixel in the color image that can be measured as color object descriptor. The descriptors of each basis color, RGB, are measured and the averages of the three components represent the color descriptors. Mean color, color standard deviation and center of gravity of color are quantified as color descriptors.

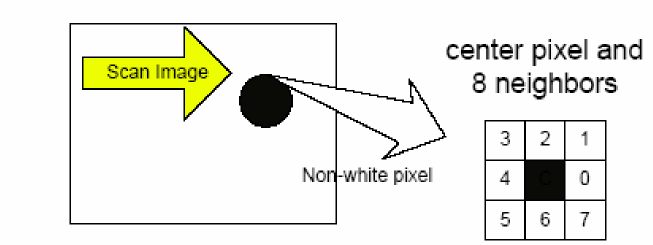

Figure 2 Object detection principle

C. *Pedestrians Tracking*

Tracking is needed for determining the object correspondences between frames [6] [7]. In our approach, the tracked feature is the object centroid which is a quite steady fixation point regardless of the small fragmentation of the blobs from time to time. Tracking enhances the structural information perceived from the moving objects, which improves the classification results. We can also estimate the apparent speed of the objects which is often a useful feature in classification between pedestrians and cyclists. The feature vector is therefore augmented with the median of the apparent speed of the tracked object.

Tracking is initiated every time a new moving object is determined in the scene and the features found fulfill the redefined criteria. This object is compared with the option regions appearing in the next few frames within a search window. If the structural features are alike, a Kalman filter based tracker is started. A simple tracking scheme is used where the horizontal x_i and the vertical coordinate's y_i of the object motion vector at time i are assumed to be independent of each other, which makes it possible to decompose the motion into two separate single coordinate models:

$$\begin{bmatrix} x_{i+1} \\ \dot{x}_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ \dot{x}_i \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} \upsilon_{x,i} \qquad (3)$$

$$\begin{bmatrix} y_{i+1} \\ \dot{y}_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_i \\ \dot{y}_i \end{bmatrix} + \begin{bmatrix} T^2/2 \\ T \end{bmatrix} \upsilon_{y,i} \qquad (4)$$

where T is the time step between frames, $\dot{x}$ and $\dot{y}$ are components of velocity vector. It is assumed that the target is moving with constants velocity, but the is subject to zero-mean white Gaussian acceleration error $\upsilon_{x,i}$ and $\upsilon_{y,i}$ [4].

The tracking model used leads to a computationally very efficient steady-state Kalman filter formulation. It means that the error covariance matrices as well as the Kalman gain attain constant values after a few frames. By recording these values in advance, and using them in filtering, it becomes possible to restrain the computation required to only state vector updating. First, the state of the tracker is predicted using the dynamic model of equations. (3), (4). The new observations are selected based on their geometric distances from the predicted coordinates and also based on their structural similarities with respect to previous observations. The best match is used as a new measurement of the object location. This new information is then integrated by the Kalman filter to the state variables.

If the features tracked are unreliably found the tracking may diverge or produce erroneous results. Due to occasional errors in feature extraction, or occlusions caused by foreground objects, the measurements are not always available for each frame [4]. In those cases tracking should not be terminated, because the discontinuity in the path may cause false increment of the people counter. This problem is solved, at least in most cases, by keeping the tracker alive for a while without new observations.

III. UNDERSTANDING PEDESTRIAN BEHAVIOR FROM SEQUENTIAL IMAGES

A. Conceptual Knowledge for Understanding Pedestrian Behavior

Human vision is a complicated process that requires numerous components of the human eye and brain to work together. The initial step of this fascinating and powerful sense is carried out in the retina of the eye. Specifically, the photoreceptor neurons (called photoreceptors) in the retina collect the light and send signals to a network of neurons that then generate electrical impulses that go to the brain. The brain then processes those impulses and gives information about what we are seeing.

When human eye recognize the objects like pedestrians, some aspect which pedestrians are doing, wearing and bringing, it will give perception, evaluation and judgment about the pedestrian itself, as shown in Figure 3. In some case our intention to pedestrian will be different, if some unexpected incident happen in the street or unusual object capture by our eye .

By linguistic understanding of pedestrian behavior image, we develop the artificial intelligent system base on human knowledge as illustrate in Figure 4. In this concept we distinguish between elder and young people, or children and adult by aggregating pedestrian features extraction like clothes color, posture, hair color, walking speed, skin color, height, etc.

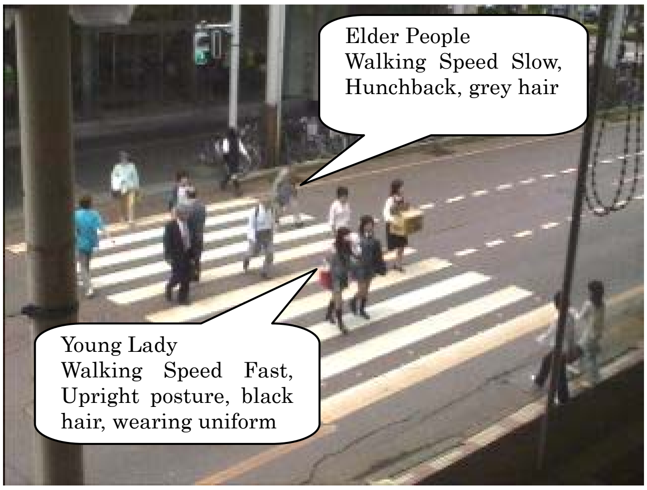
Elder People Walking Speed Slow, Hunchback, grey hair
Young Lady Walking Speed Fast, Upright posture, black hair, wearing uniform

Figure 3 Linguistic understanding of pedestrian's image

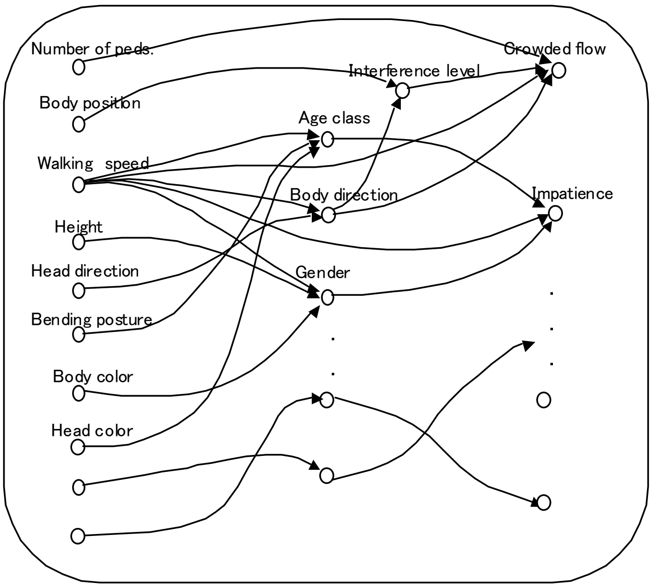

Figure 4 Concepts of network structures for intelligent monitoring systems.

Development of intelligent monitoring system is not only to estimate the pedestrian behavior but also to give information about unsafe condition cause by changing of environment condition, e.g. rain or snow.

B. Framework of the Model Concepts

Based on extraction of image feature, basic human state attributes are given as input, for a basic fusion layer. This research tries to estimate human behaviors by Information Fusion. After recognizing objects and measuring basic states of pedestrian, they are set as inputs of each agent in Choquet Integral Agent Network (CHIAN) [1]. CHIAN is a soft computing approach using flexible modeling. Explanation of CHIAN is following. CHIAN is a quasi-hierarchical network of units making a Choquet integral of multiple inputs in the interval [0, 1] with assigned fuzzy measures. Figure 5 shows flow of the CHIAN information fusion. This computational mechanism has a feature realizing much more flexible information fusion comparing to Neural Network. The skill based and ruled based cognitive knowledge could be embedded simultaneously and hierarchically in the mechanism. CHIAN knowledge for deriving linguistic motion features from physical motion features are made in consideration of human knowledge about nature of the motions.

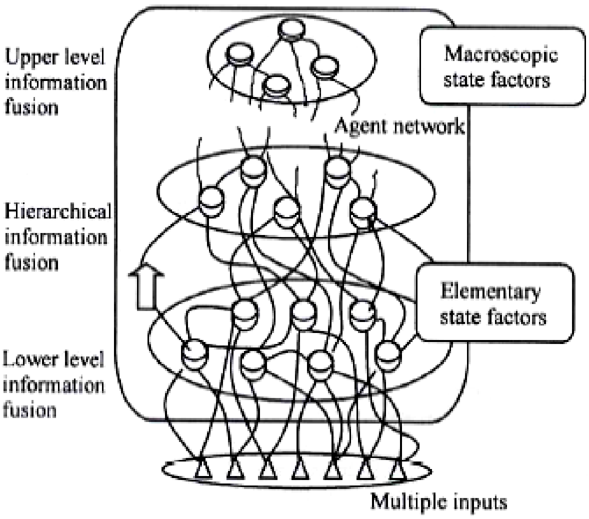
Upper level information fusion
Macroscopic state factors
Agent network
Hierarchical information fusion
Elementary state factors
Lower level information fusion
Multiple inputs

Figure 5 Flow of the CHIAN information fusion

Using CHIAN, it is tried to estimate the simple state of pedestrian behaviors, i.e. pedestrian's classification, gender, and body direction. The obtained features of pedestrian behavior image are processed by information fusion system, where basic attributes and states of pedestrian are gotten as the outputs of agents in the basic fusion layer. Each agent has its own meaning for understanding pedestrian behavior states. Information fusions by CHIAN extract emergent knowledge by hierarchical aggregation of multiple basic information as the meaning full factor, i.e. crowded and impatience degree.

As illustrated in Figure 4, to estimate aged class at the agent in basic fusion layer, normalization is perform for qualitative input data , i.e. walking speed, bending posture, and head color, then they are integrated by Choquet integral mechanism which result in normalize output value, degree of elderness [1]. The same step we did for impatience degree, where the agent gets the normalized data from qualitative input and other agent, walking speed, age class, and gender, then they are integrated by Choquet integral mechanism.

The mathematical expression is describe as follow

Let $S=(s_1, s_2, s_3, .., s_m)$ be collection of input channels for an agent and let a set function $\mu : 2^S \rightarrow [0, 1]$ be a fuzzy measure satisfying the properties;

(i) boundedness : $\mu(\phi)=0, \mu(S)=1,$

(ii) monocity : $A \subseteq B \subseteq S \Rightarrow \mu(A) \leq \mu(B)$

Then, setting the real valued function $f : S \rightarrow [0, \infty]$ as an integrated function, i.e. multiple input values.

Choquet integral with respect to fuzzy measure μ is defined by

$$x = C(f,\mu) = (c)\int f(s)d\mu$$

$$= \sum_{k=1}^{m} (f(s_k) - f(s_{k-1}))\mu(A_k) \quad \text{(5-a)}$$

$$= \sum_{k=1}^{m} f(s_k)(\mu(A_k) - \mu(A_{k+1})) \quad \text{(5-b)}$$

where $\bullet_{(k)}$ be a permutation of indices of elements of S so that

$$f(s_{(0)}) = 0, \qquad 0 \leq f(s_{(1)}) \leq f(s_{(2)}) \leq ... \leq f(s_{(m)}),$$
$$A(k) = \{S_{(k)}, S_{(k+1)}, ..., S_{(m)}\}.$$

As illustrated in above, we can embed knowledge in linguistic expression using CHIAN for estimating human behavior.

The complete result of this research will be shown during presentation

IV. CONCLUSION

In this paper, a conceptual cognition framework was proposed to comprehend collection pedestrian behavior on the walkway. We examine that pedestrian behavior can be collection of some image features. By extracting basic features of pedestrian behavior through image processing, and then aggregating by information fusion system, we can estimate pedestrian behavior according to several understanding level.

REFERENCES

[1] K.Nakamura, "A scheme for information fusion by Choquet integral agent networks," *Eighth IFSA Congress*, pp.954-958, 1999.

[2] Al Bovik, "Handbook of Image & Video Processing," Academic Press, 2000

[3] R. C. Gonzalez, R. E. Woods, and S. L. Eddins "Digital Image Processing using MATLAB," Pearson Prentice Hall, 2004.

[4] J. Heikkila, and O. Silven, "A Real-time System Monitoring of Cyclist and Pedestrians," Science Direct Journal on Image and Vision Computing, Vol. 22, pp 563-570, 2004.

[5] T. Schelhorn, D.O'Sullivan, M. Haklay, M. Thurstain, " Streets: An Agent-Based Pedestrian Model", Paper 9, Centre for Advanced Spatial Analysis,

[6] K. Kitai, "Automatic Tracking System of Pedestrains Using Image Processing Method," J. Archit. Plann.

Environ. Eng., AIJ. No. 493, pp 195-200, Mar., 1997.

[7] M. Tsujimoto, K. Shida, and K. Tatebe, “Application of Digital Image Processing Techniques to Analysis of Pedestrain Behavior,” J. Archit. Plann. Environ. Eng., AIJ. No. 436, June, 1992.

[8] T. Kardi, H. Inamura, Y. Takeyama “Tracking System to Automate Data Collection of Microscopic Pedestrian Traffic Flow,” Proc. 4th Eastern Asia Society for Transportation Studies Congress, Vol.3 no 1, pp.11-25, Oct. 2001.

[9] C. Stauffer, W. E. L. Grimson, “Learning Patterns of Activity Using Real-Time Tracking,”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, August 2000.

[10] Y. Ricquebourg, P Bouthemy, ”Real-Time Tracking of Moving Persons by Exploiting Spatio-Temporal Image Slides,” IEEE Trans On Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, August 2000.

[11] C.J Pai, H.R Tyan, Y.M Liang, H.Y Liao, S.W Chen, ”Pedestrians Detection and Tracking at Crossroads,” Journal of Pattern Recognition Society, 37, pp. 1025 – 1034, 2004