Obtaining People's Global Trajectories Considering Tracking Failure in Each Camera

Shinpei Takeda Ikuhisa Mitsugami Koh Kakusho Michihiko Minoh
Academic Center for Computing and Media Studies, Kyoto University
{*takeda, mitsugami, kakusho, minoh*}*@mm.media.kyoto-u.ac.jp*

Abstract

In these days, a huge number of cameras have been installed in various places in our daily living environments. The "Sensing Web" is a concept that the camera images should be opened and shared considering the privacy problem. Based on this concept, this paper proposed a method to obtain the global trajectories of people using the surveillance cameras .

1. Introduction

A human's trajectory is thought to be useful information for many fields; marketing, security, city planning, and so on. In a commerce facility, for example, analysing tendency of each customer's pathway can help to consider marketing plans because his/her trajectory is expected to indicate what kind of shops he/she likes. It may be possible to find automatically a customer who gets ill and a suspicious person by observing their trajectories.

There have been a lot of approaches to obtain the trajectory of humans, in which various kinds of sensors are used. Most of them use cameras to detect and track humans in captured images. Some studies[1] use infrared sensors distributed in the environment, but their performances are not so good because their spatial resolution is very low and they are usually located sparsely. Methods by laser sensors[2] also exist, but the size of such device is quite large so that they are not suitable to be installed in our daily environments. As mentioned here, each of these approaches has different features. However, they have a serious problem in common; these sensors have to be newly installed to the environment.

On the other hand, surveillance cameras have been already installed in various places. If these cameras can be utilized for the purpose above, we do not need to install them newly. However, it is usually unable because the surveillance camera images contain a lot of privacy information so that they are captured and stored only for their purposes and never opened to other systems or organizations. Considering this restriction, we proposed the "Sensing Web,"[3] which is a novel framework for multiple kinds of sensors located in the real world. In this framework, each sensor (including camera, of course) is not allowed to output obtained raw data directly to the network; the raw data is filtered by a privacy filtering module, transformed to a privacy-free and structured information, and then opened to the network. From the viewpoint of the network, therefore, the combination of a sensor and the filtering module is regarded as a new type of sensors which output the privacy-free and structured information, so that it is called a "sensor node."

This paper thus aims at a system for obtaining the people's trajectories using existing surveillance cameras based on the "Sensing Web" framework. Each sensor node would detect and track the people, so as to output their trajectories in its observation area. These local trajectories would be then integrated to the global trajectories in a integration server. In this system, the tracking accuracy in each sensor node is very important. One of the most serious problems that decrease the accuracy is the mutual occlusions of the people with similar appearances. In these cases, it is essentially impossible to track them correctly only by a single camera. And, as a result, multiple sensor nodes may output different trajectories for the same people, so that the integration server could not judge which are the correct global trajectories. In this study, we tackle this problem by proposing a new protocol and an integration method; each sensor node should output not only the obtained trajectories but also their reliabilities. It means that when trajectories estimated by a sensor node may be incorrect because of mutual occlusion of the people the node would output an alert with the trajectory data to the integration server. This is very simple idea but very important for the system covering a wide area. Experimental results show its effectiveness.

2. Sensing Web

2.1. Concept of Sensing Web

In these days, a huge number of cameras have been installed in various places in our daily living environments: stations, streets, malls, etc. Some of those cameras constitute networks for exchanging their data in order to attain the purpose for which they are installed, more efficiently. In this study, we refer to these networks as Ubiquitous Sensor Networks (USNs). Each USN is installed by some institutions including a local government, a transit company, a security company, and so on, for some specific purposes: traffic control, building management, video surveillance, etc. The sensor data obtained from the USN is used only for the purpose by the institution exclusively. However, such sensor data can actually be used for various purposes other than their original one, because the data include raw real-time information of the real world. If the sensor data were opened to the public so that anyone can use the data for their own purpose, similar to the Web, the data could serve as a new worldwide social information infrastructure that supplies the information different from that supplied by current Web. In this study, we call this new social information infrastructure the Sensing Web [3].

2.2. Sensing Web Framework

According to the concept, each sensor is not allowed to output obtained raw data directly to the network. The raw data is filtered by a privacy filtering module, transformed to privacy-free and structured information, and then opened to the network as shown in Figure 1. This information contains peoples' positions, sizes, properties like estimated ages and genders. Color histograms which can be obtained by appearances of people are thought to be also contained in such information. However, on choosing these features, we have to well consider whether each of them are really free from privacy issues. We may have to consult lawyers.

3. People Tracking in a Sensor Node

3.1. People Detection Based on Background Subtraction

Considering that the camera is fixed so that it observes the same place, the background subtraction is

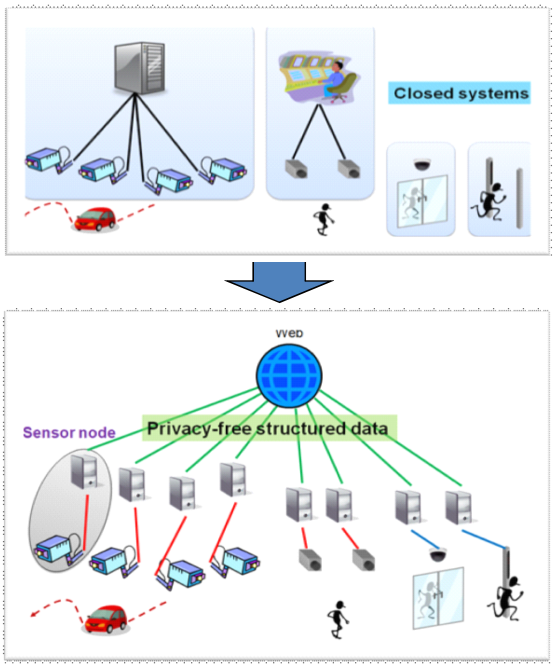


Figure 1. Framework of the Sensing Web.

thought to be reasonable for the human region extraction. However, as known well, the simple method is very weak to changes of illumination so that it is not practical especially for outdoor scenes. Thus, we have proposed a new method[4] which is robust to sudden changes of illumination because it utilize the correlations between the pixel value variations, which are obtained by long-term observation. Figure 2 shows the results of the method. It is confirmed that only people are detected even when the illumination changes suddenly.

3.2. People Tracking Based on Particle Filtering

The people tracking, which means keeping the same label assigned to a certain person with time, is executed by the particle filtering method[5, 6]. In this study, the color histogram of each person as well as his/her region detected in the image by the background subtraction is evaluated in the tracking process, so that the tracking works quite well even when more than one person appear in the images, as shown in Figure 3(a).

3.3. Problem of Tracking in a Node

Although the people tracking works well as described in the previous section, it still suffers from the case that people whose appearances are similar pass close to each other. These people may and may not

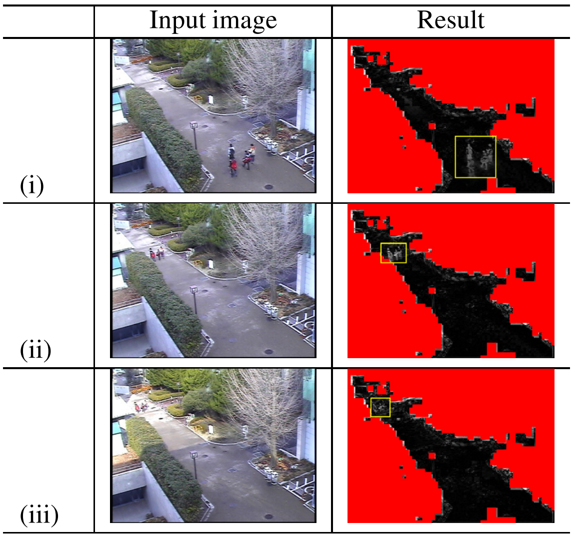
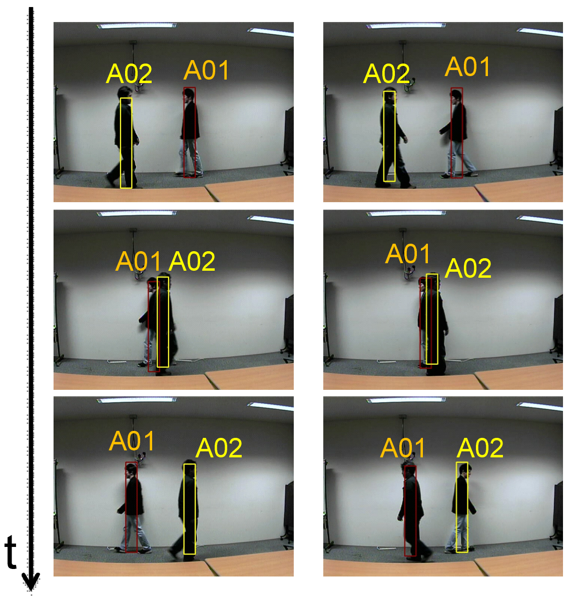


Figure 2. Background subtraction results by the proposed method.



(a) Success (b) Failure

Figure 3. People Tracking.

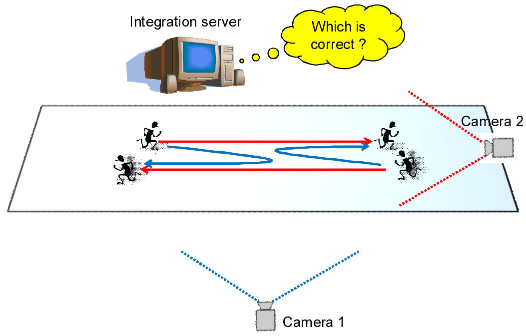


Figure 4. Failure in the trajectory integration.

be tracked correctly. In fact, such situations are expected to happen quite frequently as shown in Figure 3(b), because our cloths often look similar to other's ones around us. It is also one of the reasons that the size of each person is not large in the images so that the differences about detailed features cannot be obtained. This problem is essentially impossible to solve as long as using information of only a single camera.

4. Obtaining Global People Trajectories Considering Failure in Each Node

4.1. Integration of Partial Trajectories from Sensor Nodes

Each trajectory obtained from each sensor node is just a part of the whole trajectory of a person and do not have the privacy information. The integration server thus integrates the partial trajectories to obtain the global one of each person by focusing on the overlapped observed areas of the cameras.

When multiple cameras partially share their sights, people in the shared sight are expected to be tracked more correctly than by a single camera because the viewpoints are different from each other; even when the mutual occlusion occurs in a camera, the other camera(s) may capture them without the occlusion. In fact, there have been several studies which use multiple cameras for people tracking robust to their mutual occlusion motivated by this fact.

In the case of the Sensing Web, however, such methods cannot be applicable. This is because each sensor node tracks people independently from other sensor nodes in order to transform the raw sensory data to the privacy-free structured data by itself. Therefore, the integration server cannot judge which trajectories are cor-

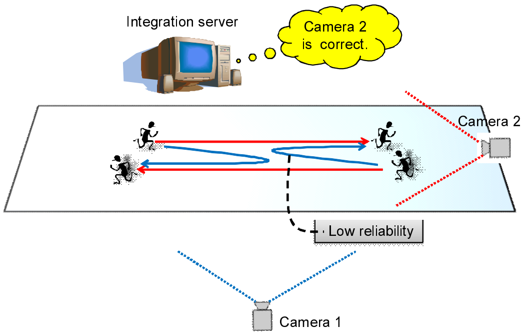


Figure 5. Choice of the correct trajectory.

rect, as shown in Figure 4. As a result, the server cannot estimate the global trajectories.

4.2. Proposed Method to Avoid Conflict of Trajectories

In this study, we tackle this problem by proposing a new protocol and an integration method; each sensor node should output not only the obtained trajectories but also their reliabilities. In the case of the particle filtering method we have implemented in this study, the possibility of tracking failure can be estimated by observing the conditional density distributions corresponding to the tracked people; if the distributions are close to each other in the state parameter space. When a sensor node detects the low reliability, it outputs its alert as well as the trajectory data. The integration server, which collects the trajectory data from all the sensor nodes in a certain area, is thus able to know which node and which trajectory is less reliable.

When receiving a alert from a node, the server adopts only reliable trajectories, and at the same time records a transformation from the label of the less reliable one to the reliable label of the other trajectories. The transformation is stored in a table, so that the following data is transformed according to this table. The table would be updated iteratively every time the alert is output. Note that this transformation is just for the labels attached to the trajectories to maintain the consistency about the labels of the trajectories. In other words, each trajectory itself is not modified at all.

Once achieved the consistency about the labels, the integration of the partial trajectories of a certain person is executed. It is thought that there are several approaches to realize this integration; just a mean of the trajectories, estimation by the Kalman filter[7], or a more sophisticated way that takes into consideration the distances from a person to the cameras.

5. Conclusion

This paper proposed a method to obtain the global trajectories of people using the surveillance cameras. In order to utilize the cameras existing in the environment, we have proposed the Sensing Web framework and this proposed method is designed according to the framework.

Future work contains experiments for evaluating the performance and the stability of the proposed method. We plan to conduct the experiments in a shopping mall in Kyoto.

References

[1] S.Honda, K.Fukui, K.Moriyama, S.Kurihara, M.Numao: “Multi-Person Tracking with Infrared Sensor Network,” The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 2006.

[2] K.Katabira, T.Suzuki, K.Nakamura, H.Zhao, R.Shibasaki, Y.Nakagawa: “Tracking Pedestrians and Visualization of the Crowds-Flow using Multiple Single-Row Laser Range Scanners,” IPSJ SIG Notes. CVIM, Vol.2007, No.31, pp.229–236, 2007.

[3] M. Minoh, K. Kakusho, N. Babaguchi, T. Ajisaka “Sensing Web Project - How to handle privacy information in sensor data,” 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems, 2008.

[4] T.Yashiro, I.Mitsugami, K.Kakusho, M.Minoh: “Foreground Region Extraction Using Correlation between Sequences of Pixel Value,” 11th Meeting on Image Recognition and Understanding, 2008.

[5] M.S.Arulampalam, S.Maskell, N.Gordon, T.Clapp: “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” IEEE Transactions on Signal Processing 50 (2), pp.174–188, 2002.

[6] M.Isard and A.Blake: “CONDENSATION – conditional density propagation for visual tracking,” International Journal on Computer Vision, 29, 1, pp.5–28, 1998.

[7] G.Welch and G.Bishop: “An Introduction to the Kalman Filter,” SIGGRAPH2001, Course 8, 2001.