TRACKING PERSONS USING PARTICLE FILTER FUSING VISUAL AND WI-FI LOCALIZATIONS FOR WIDELY DISTRIBUTED CAMERA

T. Miyaki†, T. Yamasaki‡and K. Aizawa‡

†Dept. of Frontier Informatics and ‡Dept. of Information and Communication Engineering
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

ABSTRACT

This paper describes an object tracking scheme employs sensor fusion approach which is composed of visual information and location information estimated from Wi-Fi signals. Location information is calculated by a set of received signal strength values of beacon packets from Wi-Fi access points (APs) around the targets. Different from the conventional approaches which use another kind of sensors, our approach can cover wider areas both indoor and outdoor with lower cost because of characteristics of Wi-Fi signals. Particle filter is applied to combine these two different kinds of sensory input to track the target continuously. Wi-Fi observation model is involved in a conventional visual particle filtering scheme in order to evaluate importance weights of each particle. By using multiple modality, robust tracking performance is achieved even if reliability of one sensory input declines. In this paper, we present experimental results applied to outdoor surveillance camera environment.

Index Terms— object tracking, sensor fusion, Wi-Fi, video surveillance, distributed camera network

1. INTRODUCTION

Visual surveillance system is getting more important because of our growing concern about security. So, fundamental image processing techniques such as object detection and tracking is demanded to get higher performance. In widely distributed multimedia sensor networks, object tracking with multiple cameras is one of the key features [1]. But it is difficult to achieve more accurate and stable tracking performance by only using visual information. To make realize this objective, sensor fusion techniques are proposed in many application areas which use another kind of sensors (e.g., global positioning system (GPS), pressure sensors on the floors to detect foot steps, laser-range scanners [2], etc.) with video information from cameras. Because of the characteristics of such sensors, these solution often tends to restrict target area of the surveillance application. GPS cannot be used indoors because radio waves from satellites do not penetrate into buildings and pressure sensors and laser-range scanners cannot be set in wide areas because of the cost problem. A location estimation system which can be used easily in wide area is strongly needed for surveillance applications, even if its accuracy of the estimated location is lower than that of existing positioning systems. Recently, city-wide public wireless LAN services have become increasingly popular. A lot of companies (e.g., Google, EarthLink, FON, etc.) provide wide-area public wireless network services in metropolitan area. The method has great possibility of making use of all these infrastructure to achieve a city-wide positioning system.

The authors have reported a joint tracking algorithm [3] which has two phases: detecting phase based on Wi-Fi location estimation and tracking phase based on conventional visual tracking algorithm. And the algorithm switches these two phases depending on the estimated position of the target. But this approach cannot handle the target occluded by another objects in a camera shot area or camera handover between which has disjointed shot areas. In this paper, mixture particle filter is applied to combine these two different kinds of sensory input to track the target continuously. Wi-Fi observation model is involved in a conventional visual particle filtering scheme to evaluate importance weights of each particle. By using multiple modality, robust tracking performance is achieved even if reliability of one of the sensory inputs declines.

Following section describes system architectures, our proposed mixture tracking algorithm with Wi-Fi location estimation and the experimental results of the object tracking we did on the our multimedia sensor network. Our experimental results demonstrated that the target was successfully identified and tracked over multiple disjoint area.

2. DISTRIBUTED CAMERA SYSTEM

To make realize our mixture tracking algorithm, distributed camera network system is installed around the campus and station. We installed 12 camera nodes, which are composed of three cameras, a processing unit, and a wireless LAN interface on 12 power poles along the street in front of our campus[3]. And also the station area, total six cameras with

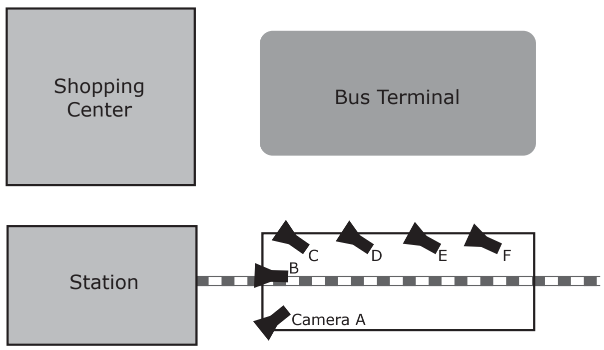

Fig. 1. Wireless LAN supported area. todo

Fig. 2. Distributed camera setup (Camera D in Fig.1)

three processing units are installed. The captured image size is 640x480. The cameras are set so that the view range overlaps with each other. Data transfer from the sensor nodes to the central control server is achieved using the HOTSPOT [4] service provided by NTT Communications, which is a public wireless LAN connection service available in public space. As a result, the system is easy-to-implement and low cost. Fig.2 shows an installed sensor node at the station area. This system is also used as a public infrastructure for urban network system.

3. TRACKING ALGORITHM

3.1. Wi-Fi Location Estimation

There are two types of positioning systems over wireless networks. One is called time difference of arrival (TDOA), the other is called received signal strength identifier (RSSI) model. The former approach needs accurate time synchronization between multiple Wi-Fi APs. Because of this reason, dedicated wireless systems are required, which tends to yield higer cost. RSSI model is easier to use. The client scans signal strength from multiple APs and match the data with location database which has a set of estimated location of APs. As client-side does not need processing power, small mobile terminals such as Wi-Fi enabled mobile phones or PDAs are enough to estimate location of the user. Several approaches about this location estimation technique are reported already [5, 6].

In RSSI models, only the beacon packets are needed for RSSI scanning, thus clients does not need to be able to connect with the networks and all the APs around the targets can be used as positioning infrastructure. Our system uses PlaceEngine[7] which is based on RSSI model as a location estimation engine. One of the authors participated in the system development. It works both indoor and outdoor, and estimated accuracy is often comparable with that of existing positioning systems. In this system, client software scans 802.11 beacon packets and submits these information (MAC address of the Wi-Fi AP (48bit), service set indicator (SSID) and received signal strength indicator (RSSI) for every scanned APs) to location database server which stores a number of pairs of MAC address and its location (latitude-longitude coordinates) via wireless network. The server estimates the user's location from the scanned Wi-Fi signal strengths by the modified version of the Centroid algorithm [5]. The algorithm is based on the relationship function between RSSI and distance derived from actual signal scanning values of 800 APs whose geographical positions are known in advance. Using this $dist(RSSI)$ function and observed Wi-Fi signal data,

$$[(AP_1, RSSI_1), (AP_2, RSSI_2), ...(AP_n, RSSI_n)]$$

position of the client is estimated by

$$pos = \frac{1}{W}\sum_{i=1}^{n}\frac{1}{dist(RSSI_i)}pos(AP_i) \tag{1}$$

where

$$W = \sum_{i=1}^{n}\frac{1}{dist(RSSI_i)} \tag{2}$$

Then estimated location information is returned to the client with in the form of latitude and longitude with time-stamp.

The accuracy of Wi-Fi location estimation algorithms depend on an initial training phase. This involves "war driving" around the neighborhood with a Wi-Fi equipment and an attached GPS device. The Wi-Fi card periodically scans its environment to discover wireless networks while the GPS device records the latitude-longitude coordinates of the "war driver" when the scan was performed.

3.2. Pedestrian Tracking by Multiple Cameras

Particle filtering algorithm is suitable for pedestrian tracking on video sequences. It should be done in a distributed way

Fig. 3. Screen shot of PlaceEngine client software.

at the processing nodes because of the limited network bandwidth. The result of the tracking have to be handed over from one camera to the adjacent camera.

Particle filter (sequential Monte-Carlo method) is well known for its good object tracking performance in computer vision application. We modeled the target by a simple three dimensional cylindrical object. Images captured from cameras are degenerated to two dimensional planar images. By doing simple camera calibration before capturing, captured images can be mapped into three dimensional X-Y-Z coordinate. The likelihood of each particle is evaluated on this X-Y coordinates approximately same as the ground plane. For recursive estimation of the particle (state vector X) in the state space at discrete time t, two models below are employed. One is the camera observation model Yc_t and the other is dynamic model of the target between X_{t-1} and X_t which is shown in 3 and 4, respectively.

$$Yi_t = H(X_t) + \omega_t \tag{3}$$

$$X_t = F(X_{t-1}) + \rho_t \tag{4}$$

where H is perspective transform, F is dynamic function of the target, and ω,ρ are random vector represents "observation noise".

3.3. Mixture Tracking

Based on these two algorithms shown in above, particle filtering and Wi-Fi location estimation, we propose the pedestrian tracking algorithm which can be used in wide-area and can be one of solutions to initial object detection problems in traditional particle filtering approach.

These two algorithms has complemental characteristics. Particle filtering on video images can accurately detect and track the position of targets by no means inferior to GPSs, but cannot cover the areas outside the shot scene and hard to identify targets one by one. On the other hand, Wi-Fi location estimation system can be used in everywhere as long as the client equipment can receive at least one Wi-Fi AP independent of indoor or outdoor, but accuracy of estimated location is easily affected by environment (eg., number of APs, database of location engine, etc.).

Our proposed algorithm combines these characteristics by using mixture of observation model derived from cameras and Wi-Fi. The camera observation model is already described in 3. Using time series of Wi-Fi estimation output from 1, We can denote Wi-Fi observation model Yw_t with noise υ_t

$$Yw_t = X_t + \upsilon_t \tag{5}$$

Overall observation model of the target is calculated by weighted summation of 3 and 5. By doing this, the system can continuously tracking the target even if he/she is occluded by other objects or hidden because of disjoint camera shot areas.

4. EXPERIMENTAL RESULTS

4.1. Implementation

We implemented the proposed method on the wireless multimedia sensor networks introduced in section 2. Equipments of environmental side are composed of a PC (Intel Pentium 4) and four channel video capture board (Matrox Morphis) for each camera node. The target side only needs a Wi-Fi enabled PDA, we used SHARP WS003SH (Intel PXA270 416MHz), which runs PlaceEngine program, Wi-Fi RSSI based location estimation software, every 10 seconds. The GPS receiver used in initial training phase of positioning system is GlobalSat BU-353 (SiRFstarIII chip)

4.2. Initial Training of Wireless Positioning System

Before the experiment, the Wi-Fi location estimation system needs to be calibrated for target areas by making initial registration of training dataset. We registered Wi-Fi signal strength data at 20 different points in the camera node area (approximately 100 m^2). This data is collected using a PC with a GPS receiver and a PDA for Wi-Fi RSSI data which are correctly time synchronized. GPS values captured in this initial training phase are filtered by their precision support parameters, dilution of precision (DOP) with static threshold. On the estimation phase, the maximum error was 18m when the GPS data was used as a reference (in this experiments, ground truth positions were hard to define.) Fig. 4 shows comparison of the two positioning system, Wi-Fi based location estimation and GPS.

4.3. Target Tracking

In this experiment, we evaluated the sequence in which the target walked into camera-covered-area (CCA) from the outside of CCA. The target had PDA which ran Wi-Fi location engine every 10 seconds. The camera nodes received the User ID and the color feature information of the target from PDA when the area of the target predicted by Wi-Fi estimation goes over camera boundary. As shown in Fig. 5, we confirmed that the proposed algorithm is successfully accomplished in the case of non overlapping cameras.

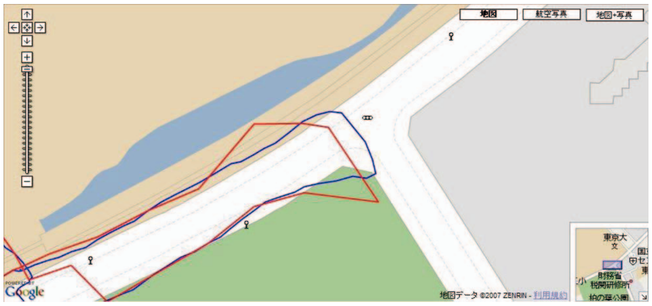

Fig. 4. Result of location estimation in front of the campus area. (red: Wi-Fi RSSI based location estimation, blue: ground truth derived from filtered GPS output.)

In addition, time series filtering approach like particle filtering has the significant weakness at the initial detection phase. Our mixture tracking approach can be a solution for this problem because Wi-Fi location estimation covered area is wider than the camera is.

5. CONCLUSION

The architectures of tracking system fusing visual and Wi-Fi localizations using mixture particle filter approach is described. In this system, the observation model of particle filtering consists from two parts: one is conventional camera observation model based on color feature of the target, and the other is location estimation system based on Wi-Fi RSSI which PDAs receive from Wi-Fi APs. With this mixture tracking algorithm, two algorithms which have complemental characteristics: one can cover large areas and the other can detect location of the target with visual information, are seamlessly and effectively co-operate in order to realize continuous object tracking. From experimental results, we observed that Wi-Fi location estimation system and a visual tracking algorithm are successfully combined and the target was continuously tracked between non overlapping cameras. We also showed evaluation result of Wi-Fi location estimation accuracy and demonstrated that the accuracy is comparable to GPS.

Fig. 5. Target handover between non overlapping cameras (up: camera B, down:camera D in Fig.1)

6. REFERENCES

[1] R. Cucchiara, "Multimedia surveillance systems," in *Proc. of ACM VSSN*, 2005, pp. 3–10.

[2] H. Zhao and R. Shibasaki, "A novel system for tracking pedestrians using multiple single-row laser-range scanners," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 35, pp. 283–291, 2005.

[3] T. Miyaki, T. Yamasaki, and K. Aizawa, "Visual tracking of pedestrians jointly using wi-fi location system on distributed camera network," in *Proc. of IEEE ICME*, 2007.

[4] HOTSPOT, "http://www.hotspot.ne.jp/en/," .

[5] Y.C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *Proc. of ACM MobiSys*, 2005, pp. 233–245.

[6] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, "Place Lab: Device Positioning Using Radio Beacons in the Wild," in *Proc. of IEEE Pervasive*, 2005, pp. 116–133.

[7] J. Rekimoto, A. Shionozaki, T. Sueyoshi, and T. Miyaki, "PlaceEngine: a WiFi location platform based on realworld folksonomy," in *In Proc. of Internet Conference*, 2006, pp. 95–104, (written in Japanese).