Understanding Human-Place Interaction from Tracking and Identification of Many Users

Donghoon Lee and Songhwai Oh
CPSLAB, ASRI
Department of Electrical and Computer Engineering
Seoul National University, Seoul, Korea
Email: {donghoon.lee, songhwai.oh}@cpslab.snu.ac.kr

***Abstract*—This paper considers the problem of understanding human-place interaction, such as relationships among many users in a space and interactions between users and their surroundings, from trajectories of users in a common space. The discovered information can be applied to provide a number of services. For example, we can determine the optimal arrangement of items in a store or at an exhibition to maximize the profit or attention and systematically manage the pedestrian traffic. Users in a space is detected and tracked by a vision-based multi-target tracking algorithm and trajectories of users are identified by combining visual information and accelerometer readings from users' smartphones. We demonstrate that trajectories of users can be used to reveal a number of useful information about the users and the space, such as spatial occupancy of individual users, intimacy between users, objects of interests, and a common interest of users.**

***Index Terms*—Multi-target tracking, Identification, Smartphones, Spatial occupancy, Intimacy, Common interest**

I. Introduction

The motivation of this paper comes from the observation that the psychological state and purchase behavior of a customer is deeply related to the path of a customer in a shop. The relationship between the shopping path of a customer and her psychological state has been demonstrated in [5] based on satisfaction of customers with respect to their shopping paths and disposition of goods. Furthermore, the customer's movement is stronger drivers of money spent than the number of customers in a store [1]. Therefore, by analyzing the indoor pedestrian traffic, we can shape the movement of customers to maximize the performance of a store.

Visual information from cameras can be used to track multiple users in a complex scene and it provides better localization of multiple users compared to other methods. In [11], Wireless LAN is used to monitor traffic flow in a store. A similar study is performed in [3] using RFID. Both approaches require additional infrastructures and do not provide accurate localization of individual users in a store. In [6], human subjects are detected using visual information based on the skin color and shape of a head and they are tracked using background subtraction. They analyzed arrival time patterns of people and flaws in the arrangement of customer service facilities. However, simple detection and tracking methods used in the paper are not suitable for real complex scenes and the provided analysis is mostly relied on the number of customers. Lastly, in all previous cases, identification of individual users is not performed and user-specific information cannot be collected.

In this paper, we are interested in discovering high-level information about users and their interactions with each other and surroundings based on trajectories of multiple users in a common space. We track multiple users with a camera mounted on the wall. Tracking multiple human subjects using a camera is a difficult problem due to issues, such as occlusions, lighting condition changes, and the deformable shape of a human body, to name a few. We address this difficult problem of data association using a multi-target tracking algorithm which first detects tracklets and then combines tracklets using track-level data association [4]. A tracklet is a segment of a track with no association ambiguity, so it is safe to assume that all tracklets have correct associations. Given the tracking results, characteristics of the space such as highly occupied areas [7], traffic model [12], and entrances and exits [9] can be discovered.

Identification of multiple users in a complex scene using visual information alone is a challenging task due to frequent occusions, changes in lighting conditions, changes in human pose, and the limited resolution of a camera. To address this difficulty, we use smartphones as additional sensors. Each user carries a smartphone and each smartphone sends accelerometer readings to a server which processes visual information from a camera. Unlike simple approaches, such as [8], [10], we can reliably identify the owner of a trajectory for a long duration of time by matching visual tracking results and the movement information from accelerometer readings. A smartphone offers a better approach for identification compared to other existing identification methods such as fingerprint scan, iris scan, and face recognition, which are highly intrusive. The use of smartphones for identification in a space, such as a museum or an exhibition, can be easily accepted by users who wants to receive personalized services.

The proposed system can be used to understand human-place interaction. As an example, we have demonstrated in experiments that the following information can be extracted from identified trajectories of users: spatial occupancy of individual users, intimacy between users, objects of interests, and a common interest of users.

The remainder of this paper is organized as follows. Sec-

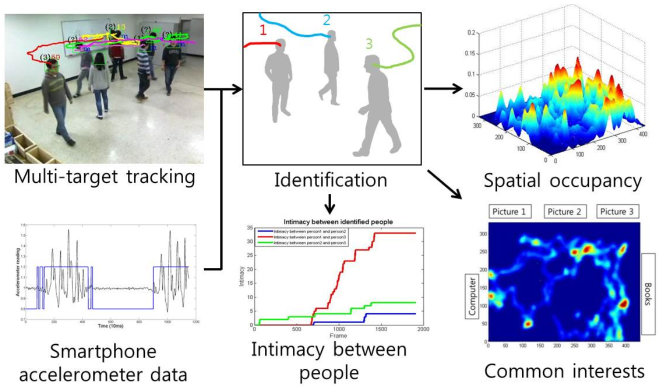

Fig. 1. People are moving in the space with their smartphones. Trajectories of people are extracted using a camera and identified via smartphone accelerometer readings. We can infer various information from the identified trajectories, such as spatial occupancy of individual users, intimacy between users, objects of interests, and a common interest of users.

tion II describes multi-target tracking and identification of many users. Section III describes a few examples of analyzing human-place interaction. Experimental results are provided in Section IV.

II. Multi-Target Tracking and Identification

We consider a general environment in which a camera is mounted on the wall higher than the people's height and watching over the space of interest, where people are freely interacting with others and objects. People move in and out of the space at any given time and entrances and exits are not specified. Our goal is to generate a reliable tracking and identification of all users in the space and infer characteristics of trajectories. The overview of the method is shown in Figure 1. The proposed framework is based on the multi-target tracking and identification algorithm described in [4]. For the full description of multi-target tracking and identification algorithms, see [4]. The algorithms are briefly introduced in this paper for the completeness.

A. Multi-Person Tracking

A person is found by detecting her head using the HOG descriptor [2] and a linear support vector machine classifier. A head is detected because it is less occluded than other body parts and can represent the position of a person reliably.

Multi-person tracking consists of two parts, tracklet generation and delayed track-level association. In this paper, a tracklet is defined as an association between detections of two consecutive frames. We calculate Farneback optical flow between a pair of detections and, based on the count of optical flows, a tracklet is generated. Then they are associated by the Hungarian algorithm based on the appearance model and motion model. The appearance model measures the Bhattacharyya coefficient between two track's HSV color histograms, and the motion model measures the smoothness of the tracks.

B. Multi-Person Identification

Identification is performed by matching motion patterns collected from smartphone accelerometers and from multi-target tracking results. Since accelerometer readings can be influenced by an unknown orientation of a smartphone and the digital compass of a smartphone can be easily corrupted by nearby electronic devices, we determine the most basic information from accelerometers, that is, whether the user is moving or not. It can be determined by checking the ratio between the mean and variance of accelerometer reading magnitudes in certain time intervals.

Similarly, we convert each trajectory from section II-A into a binary vector which indicates the movement of the tracked person. The conversion considers the distance between the camera and the tracked object. It means that a near object has to move more than a distant object to be declared as a movement. It can be determined by checking the ratio between the distance between detections and the size of the detection. At last, a matching between two measured motion patterns is found by a data association technique with a novel distance measure to build a cost matrix that considers the different lengths of two binary vectors. In this way, users are matched with their smartphones and the uniqueness of the smartphone yields the identity of user.

III. Understanding Human-Place Interaction

In this section, we show how we can infer various characteristics of human-place interaction using identified trajectories.

1) Spatial Occupancy: By dividing the ground plane into grids, we count the number of visits at each grid by users. We can calculate visit counts for all trajectories $\mathcal{T}$ or for each identified person's trajectory, τ. Based on the distribution of occupancy, we can measure the spatial density of people averaged over time that reflects their tendencies of movement. Let $N(x,y)$ be the occupancy density at grid $\mathbf{x}=(x,y)$ and let $\mathbf{d_i}=(d_x,d_y)$ be the i-th detected location of τ. Considering detection noises and possible calibration errors, each detection increases the occupancy density using a Gaussian kernel and, for a trajectory τ, the occupancy density is updated as follows:

$$N(x,y)=\sum_{\mathbf{d_i}\in\tau}\frac{1}{2\pi|\Sigma|}\exp\left(-\frac{1}{2}(\mathbf{d_i}-\mathbf{x})^T\Sigma^{-1}(\mathbf{d_i}-\mathbf{x})\right),$$

where Σ is a 2×2 diagonal matrix whose diagonal entries are a variance of the localization error. By taking a sum of occupancy densities of all trajectories, we can compute the occupancy density of all uses.

2) Intimacy Between Users: To measure the intimacy between users, we check the similarity of their trajectories in terms of spatio-temporal closeness. We first denote D as the maximum distance that an interaction between two users is possible and Θ as the maximum angle between two users' directions when they are walking together. Let $I(u,v)$ be an intimacy measure between user u and v, owners of identified trajectories $\tau^{(u)}$ and $\tau^{(v)}$, respectively. In addition, let d_t^u and d_t^v be detected locations of user u and v at time t, respectively.

Then $I(u, v)$ is increased by one if the following *Intimacy Condition* is satisfied at time t.

Intimacy Condition: (1) $||d_t^u - d_t^v|| < D$ and they are standing still; or (2) $||d_t^u - d_t^v|| < D$ and they are moving with $\angle(d_t^u, d_t^v) < \Theta$, where $\angle(d_t^u, d_t^v)$ is the angle between a line formed by d_t^u and d_{t-1}^u and a line formed by d_t^v and d_{t-1}^v.

It means that the users within the distance who walk along the similar direction or stand still are assumed interacting and their intimacy level is increased.

3) Objects of Interests and A Common Interest: If we have a prior knowledge of positions of objects in the space, we can estimate each identified person's interests in various objects by comparing her trajectory with the object's position. Comparison is done by the time each user spent and the distance from the object because it is a difficult task to determine if the user is using the object only from visual information. Since we have already calculated the occupancy density, interests about different objects can be easily inferred. Also, by considering the intimacy measure $I(u, v)$ obtained above, we can find the place where the $I(u, v)$ is high, i.e., the place where a group of users are visiting together. It can indicate their commonly interested places or objects as well.

IV. Experiments

We analyze human-place interaction using identified trajectories. We have performed various experiments that contain 21,848 frames with 12 scenarios with up to eight moving people. Among them, we picked two scenarios for explanation due to the space limitation. The first scenario has 1,924 frames and the second scenario has 1,800 frames. Scenarios contain various actions of users such as walk, run, jump, dance, talk, eat, play with an iPad, writing on a white board, and reading a book. In each scenario, there are three users carrying smartphones. Based on the extracted trajectories, inferred characteristics of human-place interaction are described below.

A. Spatial Occupancy

The resulting occupancy density for scenario 1 is shown in Figure 2. The top row of the figure is the result for all users in the space, which shows that people were generally occupied in the right half of the space. There are high peaks for person 1 and person 3 (side view), showing that they have spent for a long duration at the same locations. The peak of person 2 is relatively low which means he was moving rather than staying. The movement of person 2 indicates that he has different preferences about the space with respect to other people because his movement is tends to be the left side of the space. Finally, we can see unidentified people's movement in the final row. From the result, we can make a rough guess that the bottom right of the ground plane could be an entrance or exit because people are dense at an entrance or exit and then spread in the space. We can apply these results to the psychological state of people and the profit of a store, which are related to trajectories of customers. If we consider this space as a store, in the view of psychological satisfaction, person 2 would have low satisfaction because he had a long walk before finding a place of interest. In the view of the profit of a store, the owner has to spread goods from the right side to the left. It will expose more goods to the customers and raise the chance of purchase because currently most popular goods are concentrated on the right side of the store.

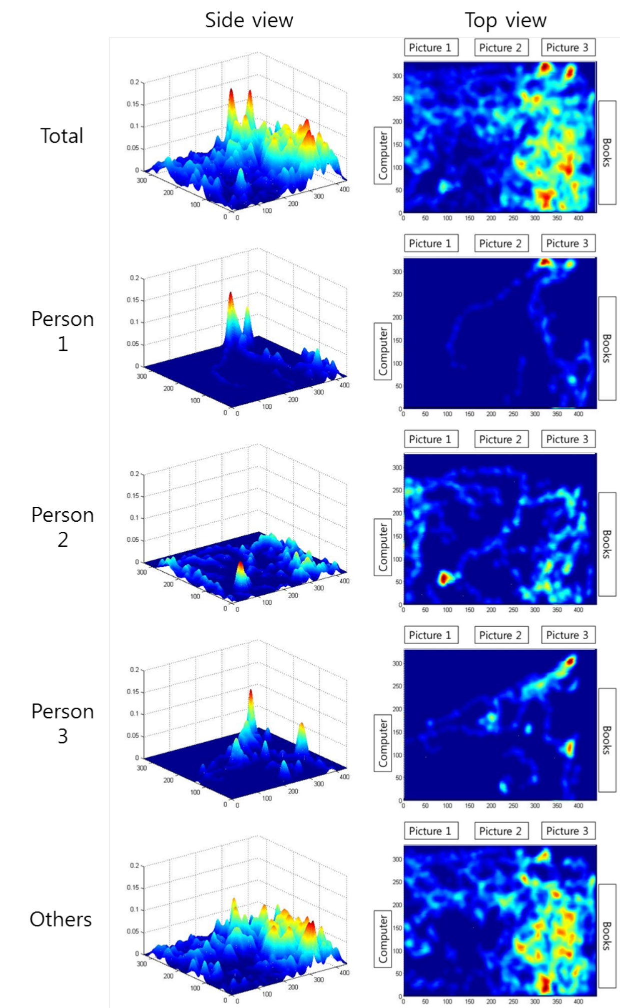

Fig. 2. The result of spatial occupancy for the first scenario. Red peak means that relatively more people visited at the grid than the blue area. Among the three identified people, person 2 shows different movements with respect to person 1 and person 3. Also, by combining spatial occupancy with the prior knowledge of the objects in the space, we can infer the objects of interests.

B. Intimacy Between Users

Intimacy between users measured by the *Intimacy Condition* are shown in Figure 3, which shows measured intimacy of three identified people over time. For the first scenario, we can guess the result of intimacy from Figure 2, since the closeness is one element for measuring intimacy. As we can expect, intimacy between person 1 and person 3 is the highest because they were standing close for a long time. In fact, during the experiment, they were debating an issue on the white board. Also we can check that the intimacy between person 2 and others are low. It is because person 2 is moving around so there

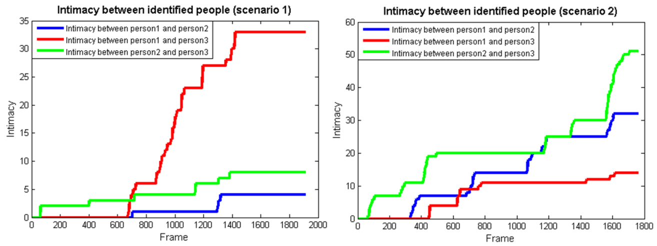

Fig. 3. The result of $I(u, v)$ for the both scenarios. Since the intimacy is measured between two people, we have three intimacy measure among three people. As we can expect from Figure 2, intimacy between person 1 and person 3 dominates others.

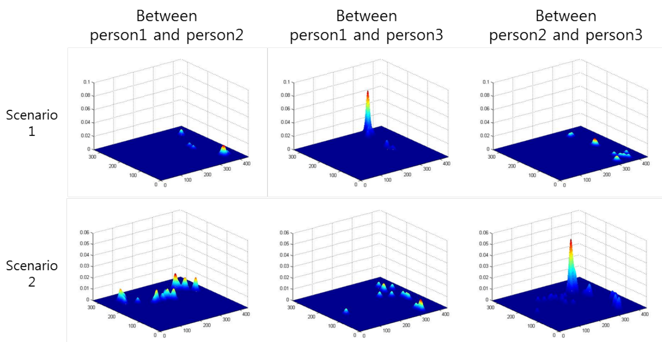

Fig. 4. The results of each scenario show the place where people are getting close each other. We can relate the place with the object so that what object they are commonly interested.

is an insufficient time for cumulating intimacy with others. The result of intimacy can be used as a stronger clue for social network analysis than co-occurrence of people which was used in [13] because intimacy is measured by considering not only co-occurrence but also direction and movement of people.

C. *Objects of Interests and Common Interest*

Objects of interests for each person are depicted in the top view column of Figure 2. It says that the people who visited the space are generally interested in 'Picture 3' and 'Books'. Person 1 is highly interested in 'Picture 3'. Person 2 is interested in 'Computer'. Person 3 is interested both in 'Books' and 'Picture 3'. By combining the occupancy density for each scenario, we can find the place where people usually meet as shown in Figure 4. The first row of the result shows that person 2 is not much correlated with others and it can be inferred from Figure 2 because person 2 shows different peak locations while others showing the same high peak at the right side of the space. Therefore, we can confirm that the peak position in the center image of the first row of Figure 4 is a commonly interested place for both person 1 and person 3. The corresponding object 'Picture 3' would be their common interest with high probability.

V. Conclusions

We have demonstrated that users in a common space can be reliably tracked and identified using a camera and smartphones. Tracking is done by associating tracklets considering the appearance and motion models. User identification is achieved by matching movement information of visual data and accelerometer readings from smartphones. From trajectories of users, we have extracted information about human-place interaction such as spatial occupancy, intimacy between users, objects of interests, and a common interest of users. Spatial occupancy is measured by counting the number of visits for each region of the space. With the prior knowledge of objects in the space, we can infer each user's interests about different objects and a common interest of a group of users. Intimacy is measured by comparing the similarity among identified trajectories. Since we already use smartphones for identification, we can provide personalized location-based services through users' smartphones. We envision that a number of useful smartphone applications can be developed based on the proposed framework.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0027155).

References

[1] I.-D. Anic, S. Radas, and L. K. Lim. Relative effects of store traffic and customer traffic flow on shopper spending. *The International Review of Retail, Distribution and Consumer Research*, 20(2):237–250, 2010.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of the IEEE Computer Vision and Pattern Recognition*, 2005.

[3] I.-C. Jung, Y. S. Kwon, and Y.-S. Lee. A sequence pattern matching approach to shopping path clustering. In *Proc. of the International Conference on Industrial Engineering and Operations Management*, 2012.

[4] D. Lee, I. Hwang, and S. Oh. OPTIMUS: Online persistent tracking and identification of many users for smart spaces. *Machine Vision and Applications, submitted*, 2013.

[5] A. J. Newman and G. R. Foxall. In-store customer behaviour in the fashion sector: some emerging methodological and theoretical directions. *International Journal of Retail & Distribution Management*, 31(11):591–600, 2003.

[6] A. J. Newman, D. K. Yu, and D. P. Oulton. New insights into retail space and format planning from customer-tracking data. *Journal of Retailing and Consumer Services*, 9(5):253–258, 2002.

[7] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert. Density-aware person detection and tracking in crowds. *Proc. of the IEEE Computer Vision and Pattern Recognition*, 2011.

[8] O. Shigeta, S. Kagami, and K. Hashimoto. Identifying a moving object with an accelerometer in a camera view. In *Proc. of the IEEE Intelligent Robots and Systems*, 2008.

[9] C. Stauffer. Estimating tracking sources and sinks. In *Proc. of the IEEE Computer Vision and Pattern Recognition Workshop*, 2003.

[10] T. Teixeira, D. Jung, G. Dublon, and A. Savvides. Identifying people in camera networks using wearable accelerometers. In *Proc. of the ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2009.

[11] V. Uotila and P. Skogster. Space management in a DIY store analysing consumer shopping paths with data-tracking devices. *Facilities*, 25(9/10):363–374, 2007.

[12] S. Yoon, H. Oh, D. Lee, and S. Oh. PASU: A personal area situation understanding system using wireless camera sensor networks. *Personal and Ubiquitous Computing*, 17(4):713–727, 2013.

[13] T. Yu, S.-N. Lim, K. Patwardhan, and N. Krahnstoever. Monitoring, recognizing and discovering social networks. In *Proc. of the IEEE Computer Vision and Pattern Recognition*. IEEE, 2009.