IMPROVING LOCATION ACCURACY BY COMBINING ZIGBEE AND VISUAL POSITIONING TECHNIQUES

YI-LIU CHAO, LI-WEI CHAN, JANE HSU, YI-PING HUNG

Graduate Institute of Networking and Multimedia
Department of Computer Science and Information Engineering
National Taiwan University

Abstract:

This paper presents an indoor localization system which improves localization by combining ZigBee radio and visual positioning techniques. Radio positioning is capable of locating people in a large area, while visual positioning only monitors the sub-areas covered by camera field-of-views. However, when a crowd of people appear, radio positioning could be completely unstable while visual positioning can still obtain good observations. In this work, we take advantage of the two systems, and develop a preferable indoor localization system. The contributions of the work include: (1) a data association method which is proposed to decide the correspondences between the observations from the visual and radio positioning systems, and (2) A refinement method which utilizes the association results not only to preserve the location accuracy within people cluster, (3) but also to calibrate the radio map for better radio positioning.

The system presented is applicable to a variety of existing radio and visual positioning techniques. The experimental results show that the visual supported radio localization system produces 41.1% improvement in positioning accuracy with people clusters in various situations.

Keywords:

Radio Positioning; Visual Positioning; Data Fusion

1. Introduction

In recent years, mobile positioning has drawn increasing interest due to the progress of the light-weight hand-held computing devices. It is desirable that the mobile device applications provide appropriate information according to the circumstances encountered by the user. Emergency services and context-aware services[1] are examples for this kind of applications. As the location is a useful cue to the user's circumstance, a positioning system is the foundation of the intelligent mobile applications.

The approach most usually adopted in indoor positioning problems is positioning using radio-based sensors such as WiFi and Zigbee devices. In a radio positioning system, beacons are set up in the environment and tags are carried by the user. Since the signal strength decays as the distance increases, we can estimate the user location according to the received signal strength (RSS) of the user-carried tag. However, signal strength is also affected by the other factors such as variations in atmosphere and changes in the environment structure. Moreover, because radio signals can be blocked by human bodies, signal strength is entirely unstable in people cluster situation[2]. As a result, the uncertainty in signal strength results in inaccurate location estimates in radio positioning systems.

In the proposed system, the visual positioning technique is integrated for accurate location estimates. Visual positioning is suitable to support radio positioning system for two reasons. First, they are complementary in several aspects. Radio positioning provide large range location estimates, say 2000 m^2, while visual positioning monitors small sub-areas. However, visual positioning provides location estimates with high accuracy which is not achievable for radio positioning. Second, the two systems have different weakness. radio positioning could be completely unstable with people cluster while visual positioning can obtain some definite observations within (i.e. the location of the nearest person to the camera). Otherwise, visual positioning is influenced by lighting condition but radio positioning isn't. Thus, taking advantages of the the two techniques should annihilate the weaknesses of individual system and improving the localization system in terms of positioning performance as well as the system stability.

1.1. Related Work

A data fusion positioning system is considered more stable and robust since it survives situations defeating individual systems. For outdoor mobile positioning, global Positioning System (GPS) coupled with inertial navigation system (INS) has long been investigated[3]. For indoor situations, Evennou *et al.*[4] proposed a localization system with similar approach using WiFi radio positioning and INS. Hightower *et al.*[5] take the laser range finder and RFID sensors to provide accurate positioning estimate with some extend of observation identification.

Ming *et al.*[6] proposed localization system framework for outdoor situation where no device is carried by the target to be located. Simulation is done to evaluate the effectiveness of the proposed system. They integrate a visual positioning system and a non-imaging sensor system which is similar to the radio positioning system mentioned before. The non-imaging sensor system locates the target with a simple algorithm and the visual positioning result is used to continuously calibrate the non-imaging sensor system. However, the simple algorithm of the non-imaging system limits accuracy of the location estimates. Although this is a quite primitive work, it gives a sense that visual positioning is helpful to improve radio positioning system.

1.2. System Architecture

The proposed system integrates the radio and visual positioning systems for localization accuracy and stability. The observations taken by the two systems are quite different. Radio positioning measures the signal strength with identification of the user while visual positioning "watches" the user's appearance without his identification. Thus a data association method is needed to decide the correspondences of the observations. The proposed association method considers the trajectories and the appearances of the observations. The appearances of the radio observations are obtained from previous associated visual observations.

The major components of the system are shown in Figure 1. The blocks on the left are the radio positioning and visual positioning. The adopted methods for the two systems are described in Section 2 and Section 3. After the observations of the two are available, the association method shown in the middle determines the correspondences between the observations (Section 4). The association results and the location estimates from the two systems are provided for the proposed radio positioning refinement presented in Section 5 which brings the final accurate location estimates. In our system, the association results also helps the visual positioning system with the color calibration component which is also described in Section 3.

2. Radio Positioning

In the radio positioning system, radio beacons are installed on the ceilings and the radio tags are carried by the users. In our system, both the beacons and the tags are the NTU Taroko nodes [1]. The Taroko nodes follow the ZigBee protocol which is based on the IEEE 802.15.4 standard. It is a short-range radio technology for small radio device with low power consumption and supports instantaneous connections. Therefore, communication between the user-carried tags is also established in our system.

After the received signal strength (RSS) measurements of the tags are available, we adopted the well-known fingerprinting algorithm proposed by Bahl *et al.*[7] to estimate user locations. The fingerprinting algorithm divides into the off-line learning phase and the real-time positioning phase. In the off-line phase, at every position in the concerned area, the signal strength of each beacon is recorded in the radio map. In the real-time phase, the RSS measurements of the user-carried tag are compared with the records in the radio map. We find out the k nearest neighbors in signal space, that is, the k most signal-similar records in the radio map. Then the position average of the k neighbors results in the location estimate of the user.

3. Visual Positioning

In the visual positioning system, cameras are set up and fixed in the environment. We adopted the method proposed by Senior[8] which performs single camera tracking algorithm on the input images of the cameras. The adopted algorithm is a multiple non-rigid target tracking algorithm for static cameras. Given a sequence of input frames, the algorithm detects and tracks people as they appear and move in the camera's field-of-view. It handles the occlusion problem and determines the depth order of the people occluding each other. Confident location estimates such as estimates of non-occluded people are considered as valid observations and are taken into the following of the system. The results of the tracking algorithm are the regions occupied by people in the image plane. After that, the homography transformation[9] converts the position of the region in the image plane into the location estimates

[1] the website of the Taroko nodes. http://www.chnds.com.tw/index_e.html

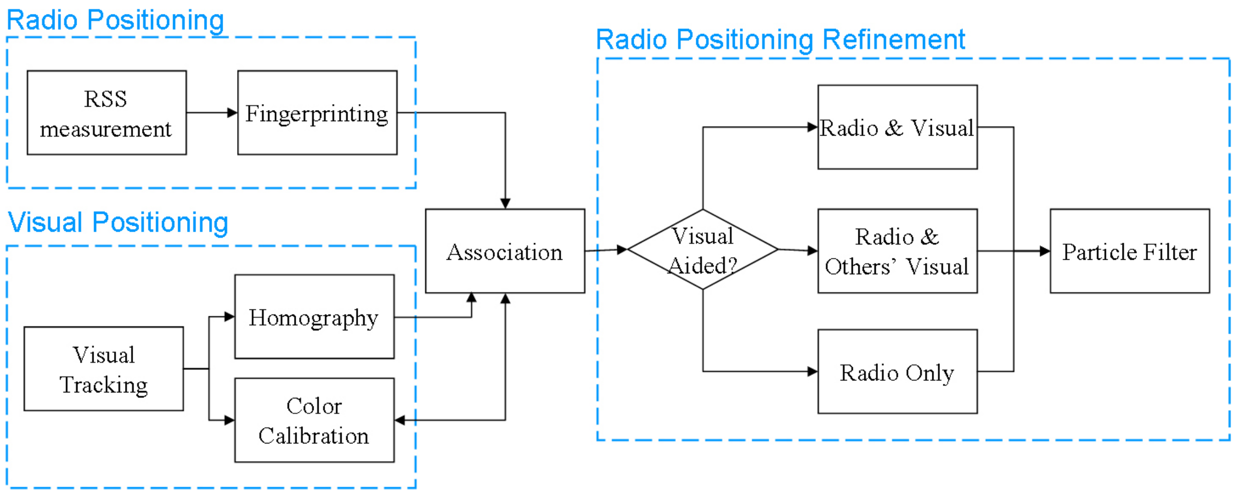

Figure 1: Block diagram of the combined localization system.

of the person in the map coordinate.

Since the association method (Section 4) takes the appearances of the observations into account, color calibration is performed on the appearances to remove the effect caused by different lighting conditions of the cameras. The adopted color calibration method is proposed by Gilbert *et al.*[10]. A color transformation matrix is constructed to model the deviation of the colors observed by two cameras. The transformation matrix is originally the identity matrix and is formed incrementally. As one person passes through two camera views alternatively, his appearances in the input images can be taken to update the color transformation matrix. In our system, the above situation can be detected if there are two visual observations associated with the same identifiable radio node. With better color transformation matrix, the association method can decide the correspondences of the observations more precisely and more quickly.

4. Association Method

Given the radio and visual positioning results, the proposed association method decides the correspondences between the radio and visual observations considering both the trajectories and the appearances.The main part of the proposed association method is the construction of the association matrix described below. The association matrix is based on the trajectory similarity matrix and the appearance similarity which are presented in Section 4.2 and Section 4.3.

4.1. Association Matrix

At every time frame t, we construct the association matrix $A(t)$ which evaluates the similarities between the observations in both trajectories and appearances. The rows of the association matrix correspond to the valid visual observations and the columns correspond to the radio nodes adjacent to them. The element at the ith row and the jth column of $A(t)$ represents the probability that the ith visual observation and the jth radio node originate from the same user. Regarding both the trajectories and the appearances, the association matrix is defined as

$$A(t) = \alpha_a K(t) + (1 - \alpha_a) P(t) \tag{1}$$

where $K(t)$ and $P(t)$ are respectively the trajectory similarity matrix and the appearance similarity matrix described in the following sections. The ith visual observation and the jth radio node are associated if the corresponding element in $A(t)$ exceeds the threshold and is larger than all the other elements in the ith row or in the jth column of $A(t)$.

4.2. Trajectory Similarity Matrix

The trajectory similarity matrix compares the trajectories of the observations. The trajectories are considered similar if they are adjacent to each other continuously for a period of time. Thus, at each time frame, the adjacency matrix $\Delta K(t)$ is constructed considering the distance between the observations at the current time frame. The definition of the elements in $\Delta K(t)$ is:

$$\Delta K_{ij}(t) = \begin{cases} 1, & \text{if } d(pos_{v_i}, pos_{r_j}) < \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where pos_{v_i} and pos_{r_j} are respectively the location estimates of the ith visual observation and the jth radio node. The function $d(x, y)$ represents the Euclidean distance between x and y. Then the adjacency matrix $\Delta K(t)$ is updated into the trajectory similarity matrix $K(t)$ by:

$$K(t) = \alpha_k K(t-1) + (1 - \alpha_k)\Delta K(t) \quad (3)$$

where $K(t-1)$ is the trajectory similarity matrix at the last time frame and $K(0)$ is the zero matrix since no information is available at the beginning. If two observations are close enough for consecutive times, the corresponding element in $K(t)$ would grow and approximate to 1, which properly characterizes the trajectory similarity of the observations.

4.3. Appearance Similarity Matrix

Likewise, the appearance similarity matrix $P(t)$ compares the appearances of the observations. With the calibrated color histograms of the observations, we construct the appearance adjacency matrix $\Delta P(t)$ by calculating the Bhattacharyya distances between the color histograms. Since the observations would only be associated when they are adjacent in map coordinate, the Bhattacharyya distance, $d_B(h_{v_i}, h_{r_j})$, between the color histograms h_{v_i} and h_{r_j} belonging to the ith visual observation and the jth radio node is only computed when $\Delta K_{ij}(t) = 1$. Thus, the definition of the elements in $\Delta P(t)$ is:

$$\Delta P_{ij}(t) = \begin{cases} 1 - d_B(h_{v_i}, h_{r_j}), & \text{if } \Delta K_{ij}(t) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Similarly, $\Delta P(t)$ is updated into the appearance similarity matrix $P(t)$ by:

$$P(t) = \alpha_p P(t-1) + (1 - \alpha_p)\Delta P(t) \quad (5)$$

The value of $P_{ij}(t)$ approximates to $1 - d_B(h_{v_i}, h_{r_j})$ after a period of time. If the histogram distance is large, $P_{ij}(t)$ would be low and the observations won't be associated.

5. Radio Positioning Refinement

The radio positioning refinement is based on the location estimates from both radio and visual positioning systems and also the correspondences decided by the association method. The information provided can be arranged as

$$\{\langle r_i, pos_i^r, pos_i^v \rangle | i{=}1\ldots m\} \cup \{\langle r_i, pos_i^r \rangle | i{=}m{+}1 \ldots R\} \quad (6)$$

where r_i is the ith radio node and pos_i^r is r_i's radio location estimate. For the associated radio client, pos_i^v is r_i's visual location estimate obtained from the associated visual observation. Only a subset of the radio nodes are associated and hence $0 \le m \le R$. In the following, the visible radio node refers to the radio node associated with some visual observation.

We maintain a particle filter[11] for each radio node r_i to integrate all the information provided. A particle filter contains a set of weighted particles denoted as $\{\langle s_n, \pi_n \rangle | n = 1, ..., N\}$ which approximates the probability density function of r_i's location. s_n, the state of the nth particle, represents the location of r_i and π_n is the weight of the nth particle. In out system, the prediction stage in the particle filter algorithm is the same as in [11]. In the measurement stage, the particles' weights are updated according to the available information of r_i. The method used in the measurement stage is described in the following.

It is possible that there's no user in the camera views. In this situation, only radio location estimates are available. Therefore, the weight of r_i's nth particle is updated by:

$$\pi_n = \exp(-d(s_n, pos_i^r)) \quad (7)$$

π_n is larger if the state s_n is closer to the radio location estimate pos_i^r of r_i. In other situations, there are some radio nodes associated with visual observations. We propose different methods for the visible radio nodes and the other ones and present them in the following sections.

5.1. Enhancement for Visible Radio nodes

Since the visual location estimates are much more accurate than the radio ones, we take care of the visible radio clients first at each time frame. The available information for visible radio node r_i includes pos_i^r and pos_i^v, the radio and visual location estimates respectively. The weight of r_i's nth particle is updated by:

$$\pi_n = \exp(-(\alpha_v \cdot d(s_n, pos_i^v)) + (1 - \alpha_v) \cdot d(s_n, pos_i^r))) \quad (8)$$

which takes pos_i^r and pos_i^v into account at the same time. Since pos_i^v is considered much more accurate, the blending rate α_v is chosen that pos_i^v has a higher rate and stands a significant portion in providing location estimate of r_i. The resultant location estimate of r_i is denoted by pos_i. Since we have great confidence in the location estimates of the visible radio nodes, the RSS measurements of them are updated into the radio map. The radio map is now automatically calibrated to better model the signal strength distribution in the environment.

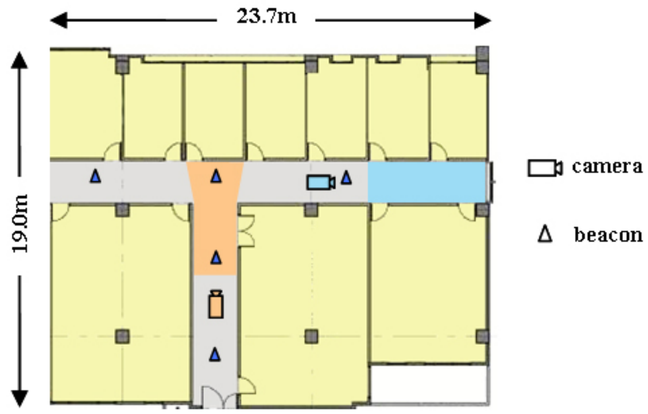

Figure 2: The system setup.

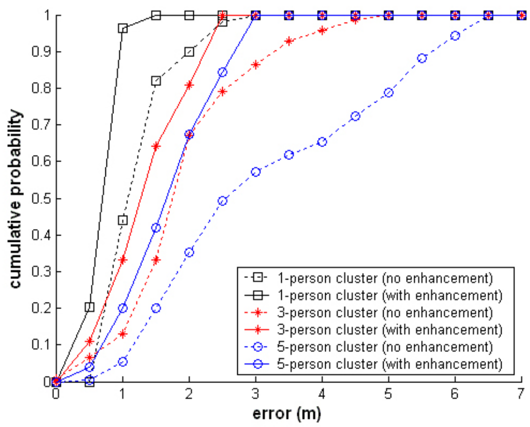

Figure 3: CDFs of the errors using traditional radio positioning and the proposed system with different number of people in clusters.

5.2. Enhancement for Other Radio nodes

Now we figure out the location estimates of the other radio nodes based on the results of the visible radio nodes. By this way, the location estimates of the users not in the camera views are also improved by the visual positioning system. First of all, we perform the fingerprinting algorithm with the calibrated radio map. A refined radio location estimate $pos_i^{r\prime}$ is obtained. Besides, the distances between r_i and the visible radio nodes are considered. In our system, there is also communication between the radio tags since ZigBee supports instantaneous connection. The signal strength between the tags provides the information about the distance between the users. Let $\{r_j|j = 1 \ldots q\}$ be the set of visible radio nodes whose signals are received by r_i. The weight of r_i's nth particle is updated by:

$$\pi_n = \exp(-(\alpha_r \cdot d(s_n, pos_i^{r\prime}) + (1 - \alpha_r) \cdot \frac{1}{q} \sum_{j=1}^{q} |d(s_n, pos_j) - dis(rss_{ij})|)) \quad (9)$$

where pos_j is the resultant refined location estimate of r_j. rss_{ij} is the signal strength received by r_i from r_j. The function $dis(rss_{ij})$ translates the signal strength rss_{ij} into the distance between r_i and r_j.

6. Experimental Results

Our system setup is shown in Figure 2. The RSS measurements and the input images are collected and then reused to evaluate the performance of each method. The cumulative distribution function (CDF) of the errors is used to exhibit the localization accuracy of each case.

Figure 3 shows the CDFs of the errors from the proposed visual enhanced system and the traditional radio positioning system (the fingerprinting method). Situations with different numbers of people in clusters are considered. As the result suggests, the performance of fingerprinting get worse when more people exists. Yet, the proposed system enhances the localization accuracy regardless of the number of people in the cluster. Figure 4 compares the performances in situations where there are different numbers of valid visual observations. In this experiment, the valid visual observations are produced by the people seen

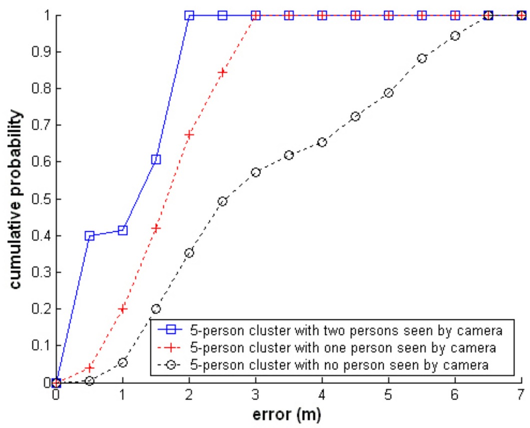

Figure 4: Experimental results in five-person cluster situations with different number of people in camera views.

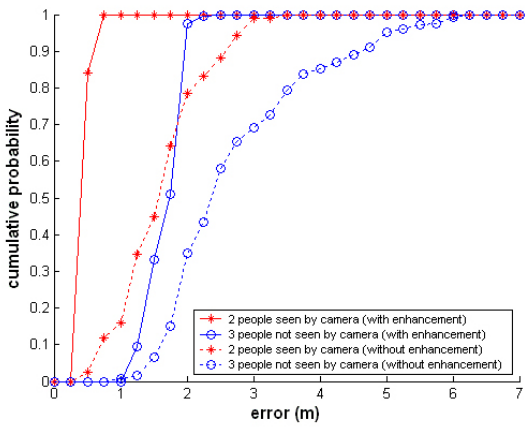

Figure 5: Positioning error CDFs of the users in camera views and the users outside camera views.

by the cameras. As the figure shows, the performance betters when more valid visual observations exists. However, there's still great improvement with only one valid visual observation.

In Figure 5, the positioning error CDFs of people in camera views and those of the other people are shown. We can see that the improvement not only occurs on the people in camera views but also the other people who is not in camera views. That is, the improvement is on the scale of the whole system. Moreover, the visual positioning is indeed much more accurate since the errors of the visible clients are all less than 1 meter and over than 80% errors are less than 0.5 meter.

7. Conclusion and Future Work

We present a hybrid indoor localization system which improves localization by combining radio and visual positioning techniques. The experimental results show that the proposed system improves the localization performance by 41.1% (27.0% $\sim$ 46.2%) comparing to an ordinary radio positioning system. The proposed system is applicable to any other radio positioning system. In our system, visual positioning is only needed on critical spots where people clusters frequently occur. Moreover, there are more cameras available. As a result, it is feasible to integrate visual positioning system.

For future work, some mechanism is required to deal with the miss-association situation since it is unlikely but probably occurs. Moreover, it is desirable that the localization result can be helpful to resolve the weakness of the visual system.

Acknowledgements

This work was partially supported by grants from NSC 96-2218-E-002 -011 and Intel.

References

[1] G. Chen and D. Kotz, "A survey of context-aware mobile computing research," Dept. of Computer Science, Dartmouth College, Tech. Rep. TR2000-381, November 2000.

[2] L.-W. Chan, J.-R. Chiang, Y.-C. Chen, C. nan Ke, J. Y. jen Hsu, and H.-H. Chu, "Collaborative localization: Enhancing wifi-based position estimation with neighborhood links in clusters." in Pervasive, 2006, pp. 50–66.

[3] M. S. Grewal, L. R. Weill, and A. P. Andrews, Global positioning systems, inertial navigation, and integration. New York, NY, USA: Wiley, 2001.

[4] F. Evennou and F. Marx, "Advanced integration of wifi and inertial navigation systems for indoor mobile positioning," EURASIP Journal on Applied Signal Processing, vol. 2006, 2006.

[5] J. Hightower, D. Fox, and G. Borriello, "The location stack," University of Washington, Department of Computer Science and Engineering, Seattle, WA, UW CSE 03-07-01, July 2003.

[6] M. Ding, A. Terzis, I. J. Wang, and D. Lucarelli, "Multi-modal calibration of surveillance sensor networks," in Milcom, 2006.

[7] P. Bahl and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in Proc. of the IEEE Infocom 2000.

[8] A. Senior, "Tracking people with probabilistic appearance models," in 3rd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance PETS 2002.

[9] R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521623049, 2000.

[10] A. Gilbert and R. Bowden, "Tracking objects across cameras by incrementally learning intercamera colour calibration and patterns of activity," in European Conference on Computer Vision, 2006.

[11] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," International Journal of Computer Vision, vol. 29, no. 1, pp. 5–28, 1998.