On Addressing Network Synchronization in Object Tracking with Multi-modal Sensors

Sangkil Jung, Jinseok Lee and Sangjin Hong
Mobile Systems Design Laboratory
Department of Electrical and Computer Engineering
Stony Brook University-SUNY, Stony Brook, NY 11794-2350
Email: {sjung, jselee, snjhong}@ece.sunysb.edu
*Corresponding Author : Sangjin Hong

Received June 15, 2009; revised July 25, 2009; accepted July 28, 2009; published August 31, 2009

Abstract

The performance of a tracking system is greatly increased if multiple types of sensors are combined to achieve the objective of the tracking instead of relying on single type of sensor. To conduct the multi-modal tracking, we have previously developed a multi-modal sensor-based tracking model where acoustic sensors mainly track the objects and visual sensors compensate the tracking errors [1]. In this paper, we find a *network synchronization problem* appearing in the developed tracking system. The problem is caused by the different location and traffic characteristics of multi-modal sensors and non-synchronized arrival of the captured sensor data at a processing server. To effectively deliver the sensor data, we propose a time-based packet aggregation algorithm where the acoustic sensor data are aggregated based on the sampling time and sent to the server. The delivered acoustic sensor data is then compensated by visual images to correct the tracking errors and such a compensation process improves the tracking accuracy in ideal case. However, in real situations, the tracking improvement from visual compensation can be severely degraded due to the aforementioned *network synchronization problem*, the impact of which is analyzed by simulations in this paper. To resolve the *network synchronization problem*, we differentiate the service level of sensor traffic based on Weight Round Robin (WRR) scheduling at the routers. The weighting factor allocated to each queue is calculated by a proposed Delay-based Weight Allocation (DWA) algorithm. From the simulations, we show the traffic differentiation model can mitigate the non-synchronization of sensor data. Finally, we analyze expected traffic behaviors of the tracking system in terms of acoustic sampling interval and visual image size.

Keywords: Object tracking, multi-modal sensor network, visual compensation, network synchronization problem, time-based packet aggregation algorithm, delay-based weight allocation

DOI: 10.3837/tiis.2009.04.001

1. Introduction

In addition to sensor devices passively and single-functionally operating in general sensor networks, more complicated functions such as object tracking and reliable monitoring functions are added into the sensor nodes to construct a navigation and surveillance system. For tracking methods, the authors in [2] propose the time-delay estimation methods that approximate location based on the time delay of arrival of signals at the receivers. On the other hand, direct tracking methods are proposed in [3][4], in which the frequency-averaged output power of a steered beamformer are used. These two traditional tracking methods have the drawbacks under the reverberant indoor environment which frequently generates extraordinary signals. In order to overcome this problem, Particle Filtering(PF)-based state-space approaches are proposed in [5][6][7]. The PF method considers to be a powerful methodology for nonlinear and non-Gaussian signal processing problems [8][9][10][11][12] [13]. However, in many PF-based tracking systems, initial state is not clear, so that PF tracking system may have lost the target object even with a known dynamic model. Moreover, incorrect dynamic model and corrupted observation lead to continuous wrong estimation of the object trajectory which is called trajectory divergence problem. Other than the PF-based models, the tracking systems with visual cameras are also investigated in [14][15][16]. In these research efforts, multiple cameras are used to extract the real position information of the target objects.

If multiple types of sensor data are combined, the weakness of each type of sensor can be compensated by other types and also they can assist each other to have better measurement. Moreover, they are more adaptive and robust in diverse environment as a specific sensor can be less sensitive to a certain condition of environment. The authors in [17] have proposed a mixed method of an acoustic-based PF algorithm and a visual tracking. In the approach, the visual camera mainly performs tracking task and the acoustic sensor assists the task. In general, the visual image processing needs high computational power compared with the PF algorithm calculation. Therefore, the reduction of complexity is critical to the operation of the system.

In order to reduce the computational complexity, we have developed a new tracking system model in [1] where the low computational acoustic-based PF primarily tracks the objects and two visual sensors resolve the unclear initial state and trajectory divergence problems inherent in the PF algorithm. In other words, the acoustic sensor detects two angle components (azimuth angle θ, elevation angle φ) using three dimensional acoustic localizer, and PF algorithm associated with the acoustic sensor obtains the coordinate information of the target object. At this point, the visual images captured from visual sensors are used to correct the tracking errors of acoustic sensors by supportive tasks such as position initialization, detecting of silent movement, and compensation of the deviated tracking from acoustic signal. However, the outcomes in [1] assume that multiple visual sensors capture the tracking space with no time difference, the PF and visual algorithms activate as soon as the acoustic and visual sensors are sampling object information, and the visual compensation results are immediately applying to the next PF state generation. The assumptions are not applicable to real situations since the sampling and calculation points generally locate in different places, and the non-synchronized sampling and data arrival take place in the middle of tracking.

This paper addresses a network synchronization problem caused by the absence of the aforementioned assumptions. After both visual sensors capture the tracking space independently, they need to send the images to a processing server. At the different time, the PF estimates from acoustic sensors also arrive at the server with different end-to-end delivery

delay. The server should determine whether the visual compensation process needs to be performed based on the acoustic and visual sampling times. If the server generates the compensated position estimation based on PF estimates and visual images, it sends feedback data to acoustic sensor to correct the possible estimation errors in the PF calculation. In order to model the developed tracking system, we configure a distributed wireless tracking network in which routers have a role in backbone nodes and acoustic sensors are sampling object information under the communication range of the routers. Both visual sensors are located at appropriate positions to efficiently capture the tracking space with sufficiently different angles. When the routers deliver the acoustic data, they use the proposed time-based packet aggregation algorithm. In the algorithm, a router checks whether the sampling time of the packet is the most recent one after it receives a packet from an acoustic sensor. If it is true, the packet is saved for future aggregation operation. After the router receives new sampling times from all the acoustic sensors, it aggregates all the packets and sends the aggregated packet to a server. Based on the aggregation algorithm, the PF estimates from acoustic sensor are efficiently delivered to a server with removing unnecessary network resources consumed to deliver a number of acoustic data.

From the simulation study, we show the increased tracking accuracy from joint operation of multiple sensor types severely deteriorates when acoustic sensors use short sampling interval, and non-sensor traffic volume flowing the wireless tracking network increases. For the possible solutions of the performance degradation, we propose a traffic differentiation model. The basic idea of the model is that we can solve the skewed end-to-end delivery delay of sensor traffic and non-sensor traffic by adopting different network queues and Weighted Round Robin (WRR) scheduling mechanism at routers. The weight allocated to each queue is obtained by a proposed Delay-based Weight Allocation (DWA) algorithm. In the differentiation model, we first obtain the end-to-end delivery delay of sensor data from multi-modal sensors to a server. Based on ratio of the obtained delay, we allocate normalized scheduling weight to each network queue. From the simulations, we show the differentiation model mitigates the network synchronization problem, so that the tracking system provides the sustainable support of visual sensors to correct the PF estimation errors. In the observation of the tracking system, we identify the successful visual compensation depends on the key parameters: sampling interval of acoustic sensors and the end-to-end delivery delay of multi-modal sensor data. Therefore, in the next work, we investigate the tracking system behaviors in terms of the two key parameters.

Organization of this paper is as follows. In Section 2, we illustrate the details of the developed multi-modal tracking system and define problems to be resolved. The basic environment for network synchronization in object tracking is illustrated in Section 3, and the performance evaluation of network synchronization and traffic differentiation model is shown in Section 4. We conduct the behavior analysis of the tracking system in Section 5 and finally conclude in Section 6.

2. Background and Problem Definition

2.1 Tracking by Particle Filter

Particle Filtering (PF) [13] is a powerful method for sequential signal processing for nonlinear and non-Gaussian problems. It is broadly used in applications that need the tracking and detection of random signals. The algorithm is also based on its operations on representing

relevant densities by discrete random measures composed of particles and weights, and computes integrals by Monte Carlo methods [10]. In the tracking problem based on PF, the measuring outputs from an acoustic sensor are bearings or angles ($\mathbf{Z}_t$) on the grid along the perpendicular coordinates at time t. Based on the angle information, we can get the estimated position $(\tilde{x}_t, \tilde{y}_t)$ and velocity $(\tilde{v}_{x,t}, \tilde{v}_{y,t})$ in the cartesian coordinate system. At the next acoustic sampling time, $t+1$, we can get the next angle ($\mathbf{Z}_{t+1}$) and corresponding outputs based on the previous PF estimates. In this tracking method, we obtain more accurate position estimates as we increase the number of particles.

1) Tracking Problems: In PF-based applications, there are two key problems preventing accurate tracking process. The first one is the initial state problem where initial state may not be reliable and sometimes is not existing. For example, in the beginning of the tracking or when the signal from an object re-appears after silence movement or blocking obstacle, we can consider the cases as the initial state problem. Since the PF application assumes the initial state is clearly given, the PF approximation outputs will show significant deviation from the real object trajectory in the presence of the initial state problem. The trajectory divergence problem is another key problem that appears in many PF applications. The object dynamic model could change in the middle of tracking even with the given initial state. The change with or without the initial state results in the tracking to be diverged. Since the next PF estimation is based on the current state vector, the deviated state vector in current time will lead to further erroneous tracking in the wrong direction.

2) Possible Solutions: The aforementioned tracking problems can be solved by multiple dynamic model [18][19], multiple acoustic sensor detection [20], and audio-visual multi-modal tracking algorithm [17]. Especially, the last multi-modal algorithm has recently been active research domain due to its accuracy and fast implementation. In [17], the visual sensor mainly tracks the object and an acoustic sensor supports the tracking when the object disappears from the visual space. However, in realtime point of view, the complexity of image processing becomes an overhead factor. Therefore, the authors in [1] have adopted a low computing acoustic sensor for the main tracking device and the visual sensor compensates the tracking deviation caused by acoustic-based PF outcomes.

2.2 Tracking by Visual Sensor

Visual localization algorithm is performed to extract the object position estimation from captured visual image. It is based on the parallel projection model [21], which simply approximates the position with a known reference point, $P_r(x_r, y_r)$. Arbitrary point on a camera or an estimate obtained by PF algorithm could be the reference point. The algorithm assumes both cameras can capture the target object at the same time. In the algorithm, the reference point is projected on the viewable planes of both cameras, and the object points appeared in the camera sensing planes are also projected on the viewable planes. Let the projected point of reference point be $P_v^i = (x_v^i, y_v^i)$ and the projected point of sensing plane be $P_s^i = (x_s^i, y_s^i)$, where i is the camera id and takes 1 and 2. Then, we can obtain the distance Δd_i between the projected points as $\Delta d_i = \overline{P_v^i P_s^i}$. If we assume the each viewable plane of camera 1 and 2 forms x and y cartesian coordinate respectively, the estimated object position is obtained by $P_e(x_e, y_e) = (x_r \pm \Delta d_1, y_r \pm \Delta d_2)$.

2.3 Target Application Model

The target application model in our approach is based on the integration of an acoustic sensor and two visual sensors as shown in **Fig. 1**. A three dimensional acoustic localizer is located at an acoustic sensor to get the direction of arrival (DOA) [22]. The localizer detects two angle components (azimuth angle θ, elevation angle φ) from the arrival time difference between embedded adjacent microphones. The PF associated with acoustic sensor mainly obtains the coordinate information of the moving object, and supportive tasks such as position initialization, detecting of silent movement, and compensation of the deviated tracking from acoustic signal are done by two visual sensors. The visual sensors require to be located in appropriate positions to capture the object with sufficient angles which are used in the localization algorithm in Section 2.2. **Fig. 1** shows that four microphones measure interaural time differences of an object. By scaling the speed of wave propagation and the unit dimensions of the microphones array, the θ and φ angles are derived.

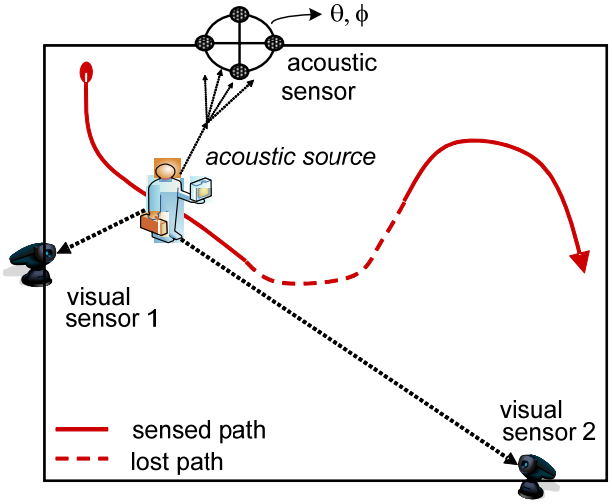


Fig. 1. Target application model for object tracking. It consists of an acoustic sensor and two visual sensors to capture the object information. The dashed line means the lost of the acoustic signal in the middle of object moving.

2.4 Visual Compensation Effect

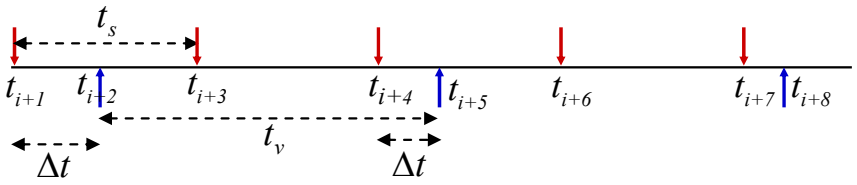
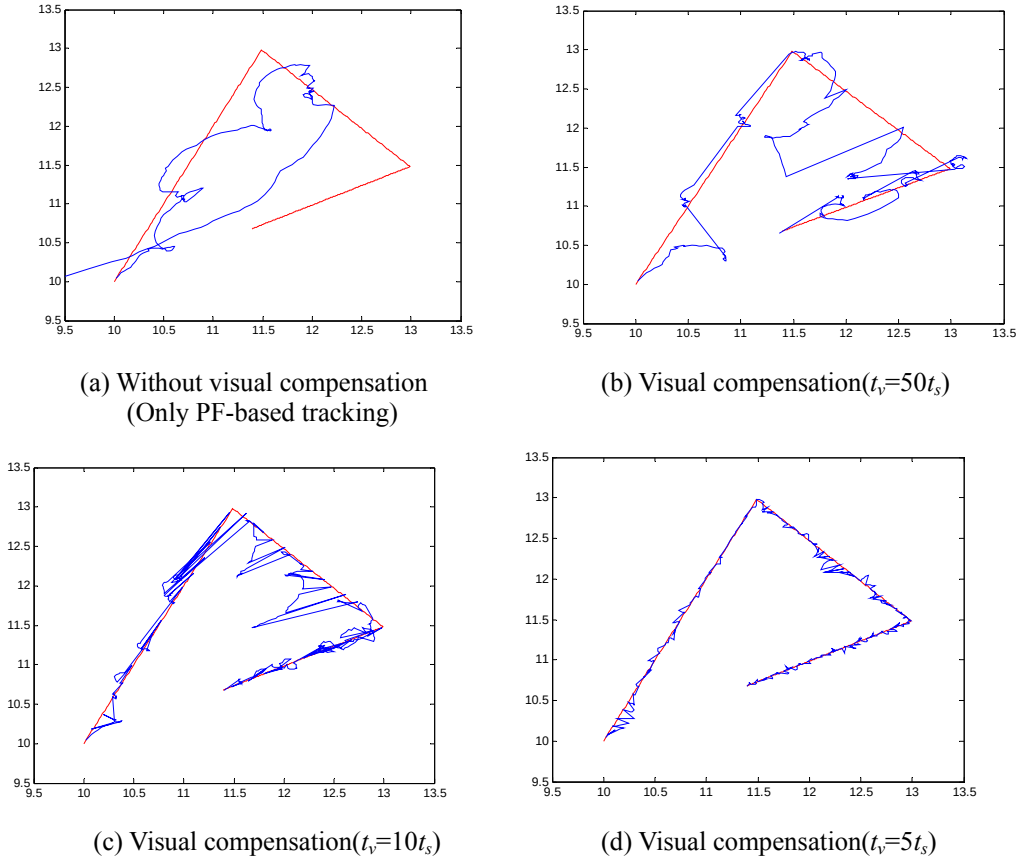


Fig. 2. Sampling time sequence of an acoustic sensor and visual sensors. Red arrow is the acoustic sampling time and the blue is the sampling point of the visual sensors.

In the visual compensation process, the operation of visual sensor is independent of the acoustic sensor′s operation. **Fig. 2** shows an example of sampling time sequence possibly occurring in the application model. The acoustic sensor samples the object signal by t_s interval and the visual sensors capture the tracking space every t_v. In this case, both sampling tasks

have the time difference Δt. If the Δt is in [0, t_s], the visual localization algorithm can be successfully performed to correct the tracking error in the PF-estimate. Here, the visual localization algorithm uses the PF estimate for the reference point. For example, if we consider the compensation at time t_i+5, acoustic estimate at t_i+4 becomes the reference point.

We observe the advantage of the visual compensation assisting the PF-based tracking system in the remainder of this section. The advantage appears when the multi-modal sensors independently operate like **Fig. 2**. For the observation, we use non-linear model with semi-triangular movement where an acoustic sensor is placed at (0,0) of a cartesian coordinate



(a) Without visual compensation (Only PF-based tracking)

(b) Visual compensation(t_v=50t_s)

(c) Visual compensation(t_v=10t_s)

(d) Visual compensation(t_v=5t_s)

Fig. 3. Tracking accuracy when the visual sensor assists the PF-based tracking. Red line represents the real object movement, and the blue line is the trajectory estimate obtained by associating an acoustic sensor with two visual sensors.

Fig. 3 shows the simulation results for the various sampling interval of visual sensors. t_v takes 50, 10, or 5 times longer than the sampling interval of acoustic sensors. The target object is moving along the red line, and the trajectory estimate is indicated by blue line. In **Fig. 3**(a), only PF-based estimate has large deviation from real object movement. However, when we associate two visual sensors with an acoustic sensor creating the PF estimate, we can increase the tracking accuracy in proportion to the visual sampling frequency. When t_v=50t_s, the trajectory estimate roughly follows the object movement with large variation as shown in **Fig. 3**(b). As we increase the visual sampling frequency to 10t_s, the trajectory estimate is almost the same as the object movement as shown in **Fig. 3**(c). In more visual sampling frequency like

t_v=5t_s, the trajectory estimate becomes more accurate as shown in **Fig. 3**(d). Note the processing overhead of PF is a few microseconds as indicated in [23], and the localization algorithm for visual compensation is performed rarely compared with the PF calculation. Therefore, our application model can minimize the overall processing overhead for the tracking task as well as provide the accurate tracking task.

2.5 Network Synchronization Problem in the Application Model

Even if we have mentioned the visual compensation provides significant improvement in tracking accuracy in previous section, it requires some assumptions: (1) two visual sensors capture the tracking space with no time difference, (2) as soon as acoustic and visual sensors are sampling object information, the PF and visual localization algorithms should calculate the position estimation without time delay, and (3) the visual compensation results are immediately applying to the next PF state generation. However, the sampling and the calculation points generally locate in different places, so that there exists a synchronization problem caused by the end-to-end delivery delay of sensor data in the tracking system.

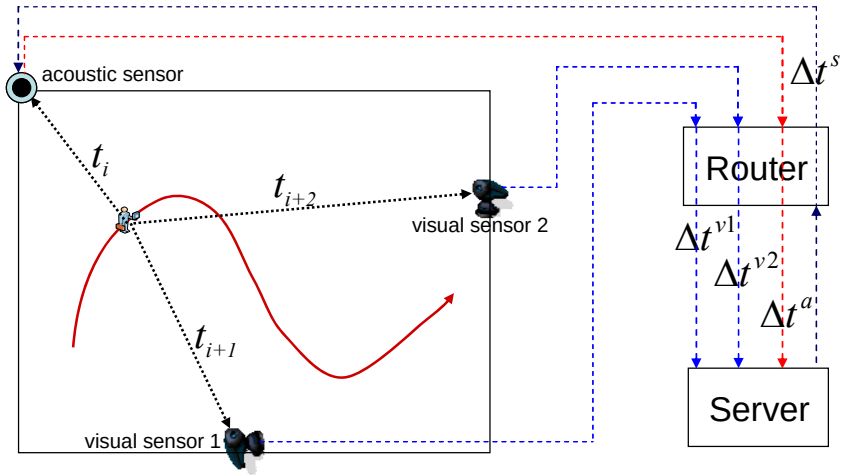


Fig. 4. Network synchronization problem in the tracking model.

Fig. 4 indicates the factors to be considered in the tracking model at the network point of view. At t_i, an acoustic sensor receives the object signal, and the visual sensor 1 and 2 take the image at time t_{i+1} and t_{i+2}, respectively. Based on concepts in the previous section, $t_{i+1} = t_i + \Delta t_{v1}$ and $t_{i+2} = t_i + \Delta t_{v2}$, where Δt_{v1} and Δt_{v2} are sampling time difference between acoustic and visual sensor 1 and 2. If we assume the PF calculation is done at acoustic sensor, and the visual localization is done at a remote computing machine, namely, server, the PF-based position estimate and the visually sampled data need to be sent to the server via network routers. In this situation, the PF estimate requires to arrive at the server with end-to-end delivery delay Δt^a, and the image frames taken by visual sensor 1 and 2 arrive at the server after Δt^{v1} and Δt^{v2} delays. Additionally, the visual compensation estimate at the server needs to be re-sent to the acoustic sensor with delay Δt^s for the adjustment of the next PF calculation. We define a tracking problem caused by the end-to-end delivery delay in visual compensation process as *network synchronization problem*. The independent delivery delay of sampled data in addition to sampling time difference among multi-modal sensors causes the *network synchronization*

problem, so that we address how to tackle the problem in the remainder of this paper.

3. Network Synchronization for Object Tracking

3.1 Configuration of Wireless Tracking Network

In order to support the tracking task, we consider a wireless tracking network connecting the multi-modal sensors and a server. This is a type of distributed wireless network since the algorithm processing points are distributed in the network to support the tracking task.

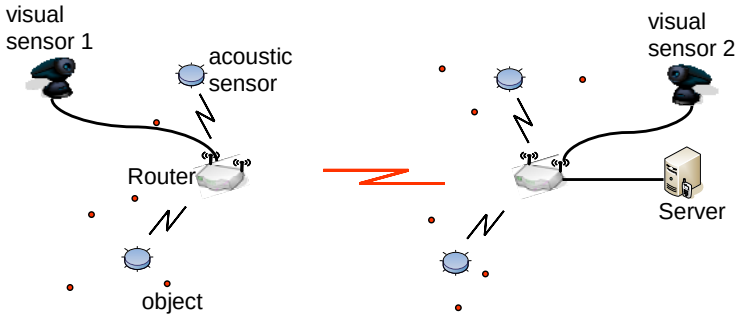


Fig. 5. An example of the wireless tracking network.

Fig. 5 shows an expected configuration of the tracking network. Routers are communicating each other by wireless channel and the last mile router is connected to a server. More than one acoustic sensor sample and send the object information to the router. Two visual sensors are connected to routers and independently send the visual image to the server. The PF calculation is done at the acoustic sensor and the visual localization algorithm having more complexity is performed at the server. The localization algorithm could be performed at the routers. However, as we have indicated in [24], the fully distributed tracking architecture has large end-to-end delivery delay of the visual image since the visual sensors have to send the same image to all the routers, which causes heavy traffic in the network. Note the image size from a visual sensor is relatively larger than the packet from acoustic sensor. For example, when we capture the visual space by IP camera [25], the size of image frame is within the range of 30 to 55 KBytes. Therefore, we adopt the server-based architecture to reduce the duplicate transmission of the same image as well as to use the high computational power of the server.

3.2 Time-based Packet Aggregation of Acoustic Sensor Data

The first problem to be solved in the network synchronization is how to deliver to the server the object information from more than one acoustic sensor in a timely manner. For this problem, we propose a time-based packet aggregation algorithm as described in Algorithm 1. Whenever a packet from an acoustic sensor is coming to the router, the router first checks if the sampling time (t_i) of the packet is the most recent one. The received packet is inserted into a Queue until the router receives packets having the most recent sampling time from all the acoustic sensors. If ti of the packet is older than the previously saved sampling time (T_i), the packet is dropped. If the router receives the packets with the latest time, it makes an aggregated packet (P_a) and sends it to the next hop router. At this point, we need to save the sampling time of the dequeued packet ($T_i = t_i$) for the next comparison of the sampling times. We assume that the sampling point of the acoustic sensors are same, which can be realized by regularly

sending SYN packets from the router to acoustic sensors to adjust the distorted sampling point. By using the aggregation algorithm, we can reduce the network traffic and end-to-end delay of the acoustic sensor data as well as simplify the visual compensation process since the acoustic data can arrive at the server at the same time. Since the sampling interval at acoustic sensors relatively larger than the end-to-end delivery delay of sensor data between acoustic sensors and a router, we can ignore the impact of waiting time to aggregate packets from acoustic sensors.

Algorithm 1: Time-based packet aggregation algorithm to deliver the data packets of more than one acoustic sensors.

P_i: a packet currently arriving at a router. It is originated from acoustic sensor i.
Q_i: a packet being queued in a *Queue*. It belongs to acoustic sensor i.
P_a: an aggregated packet to be sent to the next hop router.
n_a: the number of acoustic sensors in the range of the router.
$P_a = \{\varnothing\}$, $T_i = 0$, i=1...n_a
// Check the sampling time of the incoming packet is the latest one
c=0
for i=*1* to n_a do
    t_i = sampling time of P_i
    if $T_i < t_i$ then
        $c = c$+1
    else drop(P_i)
end
// Make an aggregated packet based on sampling time
if *c is equal to* n_a then
    for i=*1* to n_a do
        dequeue(Q_i)
        t_i = sampling time of Q_i
        $T_i = t_i$
        $P_a = \{P_a \cup Q_i\}$
    end
    send(P_a)
    $P_a = \{\varnothing\}$
else
    enqueue(P_i)
end

3.3 Visual Compensation Considering Network Synchronization

The next problem in the tracking task is to identify visual compensation in network synchronization point of view. To clarify the point, we show a packet flowing example in **Fig. 6** possibly appearing in the system. This traffic pattern could happen in a situation that an object's signal frequently disappears, so that visual image is sent as many as possible to detect the object trajectory. Similar to **Fig. 2**, the acoustic sampling times are denoted by red arrows and blue arrows are for visual image capturing points. We additionally add red points to represent the calculation of the visual localization algorithm at the server side. Since the visual compensation estimates are re-sent to the acoustic sensor for the next PF adjustment, we add

the black arrows to represent the feedback arrival at the acoustic sensor. When a router is ready to send an aggregated acoustic packet to a server, it sends the packet with delivery delay Δt_i^a (i=1,2,3,...). The visual sensor 1 and 2 also send the captured image frame to the server with delay Δt_i^{v1} and Δt_i^{v2}, respectively. After finishing visual localization algorithm, the server sends the compensated PF-estimate to the acoustic sensor. This feedback transmission takes Δt_i^s delay. Note that the delay size of multi-modal sensor data is different in times since the tracking network has a number of delay factors like router capacity and background traffic volume. In other words, the delivery delay takes randomness property. If we compare this example with **Fig. 2**, the visual compensation process needs to be differently interpreted. For example, the t_9 in **Fig. 6** could be a successful visual sampling time under the illustration of **Fig. 2** since it is within t_8 to t_{14}, which means Δt is within the [0, t_s]. However, in this example, the sampling time t_7 of the second visual sensor is not between t_8 and t_{14} due to the different image capturing point of visual sensors. This indicates the first assumption in Section 2.5 does not valid in real situation. Therefore, the image from second visual sensor may give the wrong information to the estimation procedure. In this case, we would better use visual images captured at t_{10} and t_{11} since they more precisely contain the tracking space between t_8 and t_{14}. Even if the visual images seem to capture the tracking space in timely manner, they do not provide good information with the visual compensation process when we consider the delivery delay (Δt_2^s) of feedback data that reaches the acoustic sensor at t_{15}. The arriving point is between t_{14} and t_{21} that is a new acoustic sampling period. For Δt_2^s delay time, the object can have abrupt moving behavior in which case the information at t_{15} can also give the negative information to the PF estimation. Fortunately, we have another feedback arrival at t_{20}, and the feedback gives a right information to the next PF calculation at t_{21}. This complicated situation takes place since the second and third assumptions in Section 2.5 are not applicable to the real environment.

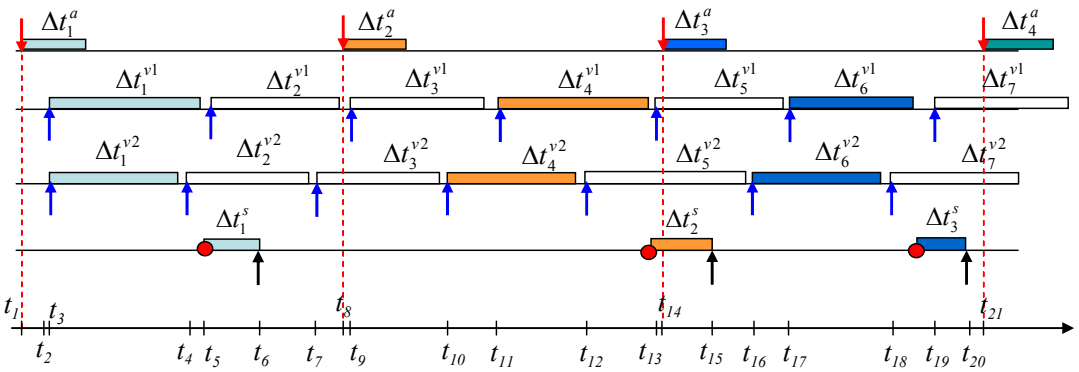


Fig. 6. Packet flowing example appearing in the tracking model.

3.4 Definition of Success and Fail Conditions

To clearly define when the visual localization algorithm can be executed and what is the *success* in the visual compensation process, we make two conditions as follows.

Condition 1: The server sees that the sampling times of both visual sensors are later than the acoustic sampling time of previously arrived acoustic data. At this point, the server performs

the localization algorithm.

Condition 2: The compensated estimate should be feedbacked to acoustic sensors before the next acoustic sampling time.

Fig. 7 shows the message flowing diagram between sensors, router and server, and possible *success* and *fail* cases occurring in the tracking model. In the figure, the messages related with acoustic sensors are directly delivered from source to destination since they are delivered by UDP. On the other hand, the visual sensor data need reliability so that TCP is used for the transferring. Since the visual image size is larger than Maximum Transmission Unit (MTU) size, more than one packet are exchanged between visual sensors and a server. Similar to **Fig. 6**, we use red and blue arrows, and red point to indicate the generation of the acoustic, visual sampling, and the execution point of localization algorithm. We can find out only **Fig. 7**(a) satisfies the *Condition 1* and *Condition 2* at the same time. Note in **Fig. 7**(c) and (d), the final result is not *success* due to the *network synchronization problem* even if the localization algorithm calculation at the server side is successful.

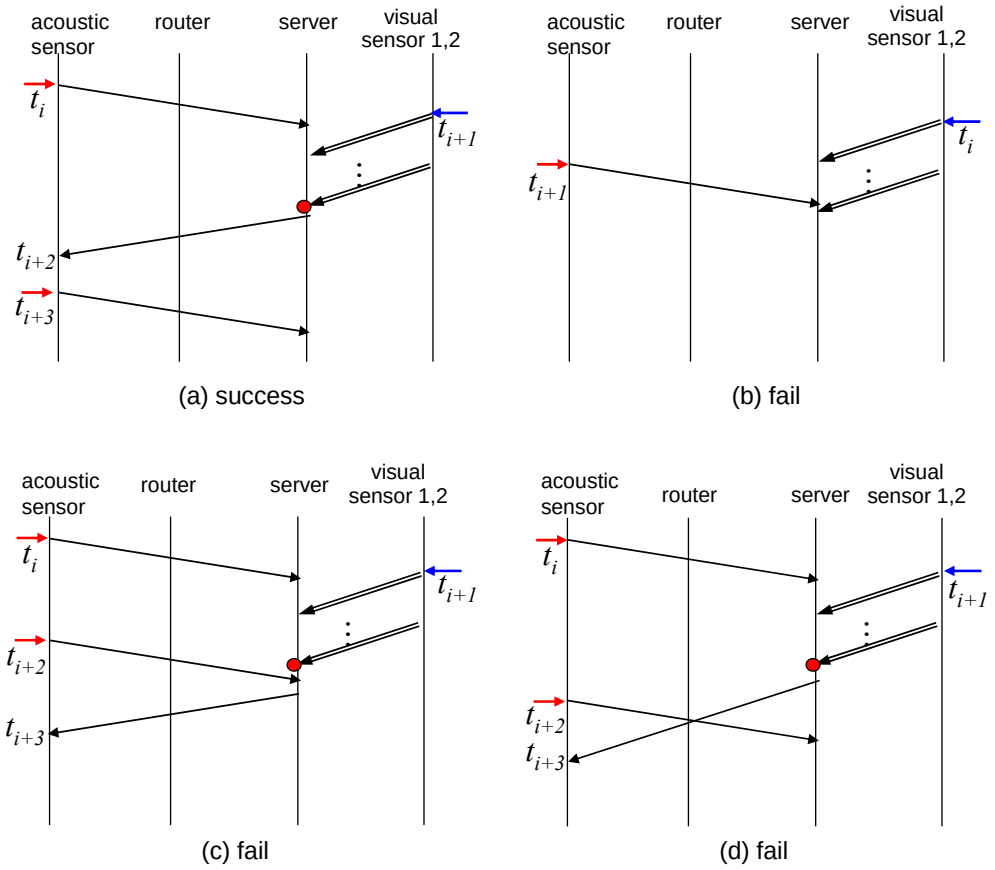


Fig. 7. An example of success and fail cases in the tracking model.

4. Performance Evaluation of Network Synchronization and Traffic Differentiation Model

4.1 Impact of Network Synchronization

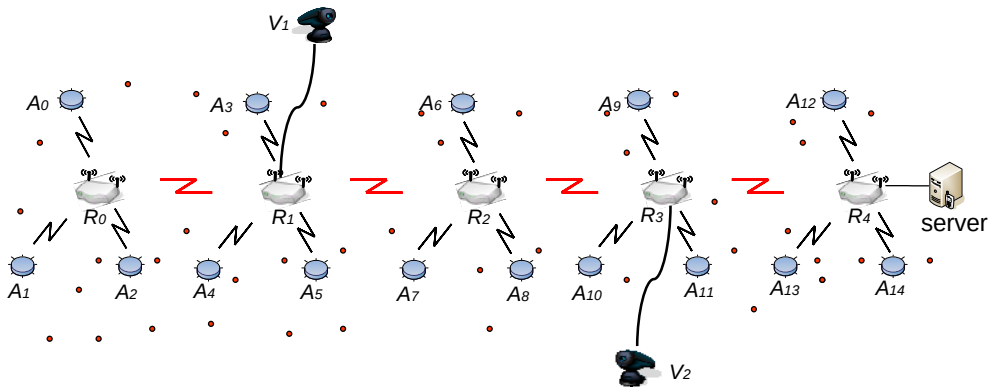
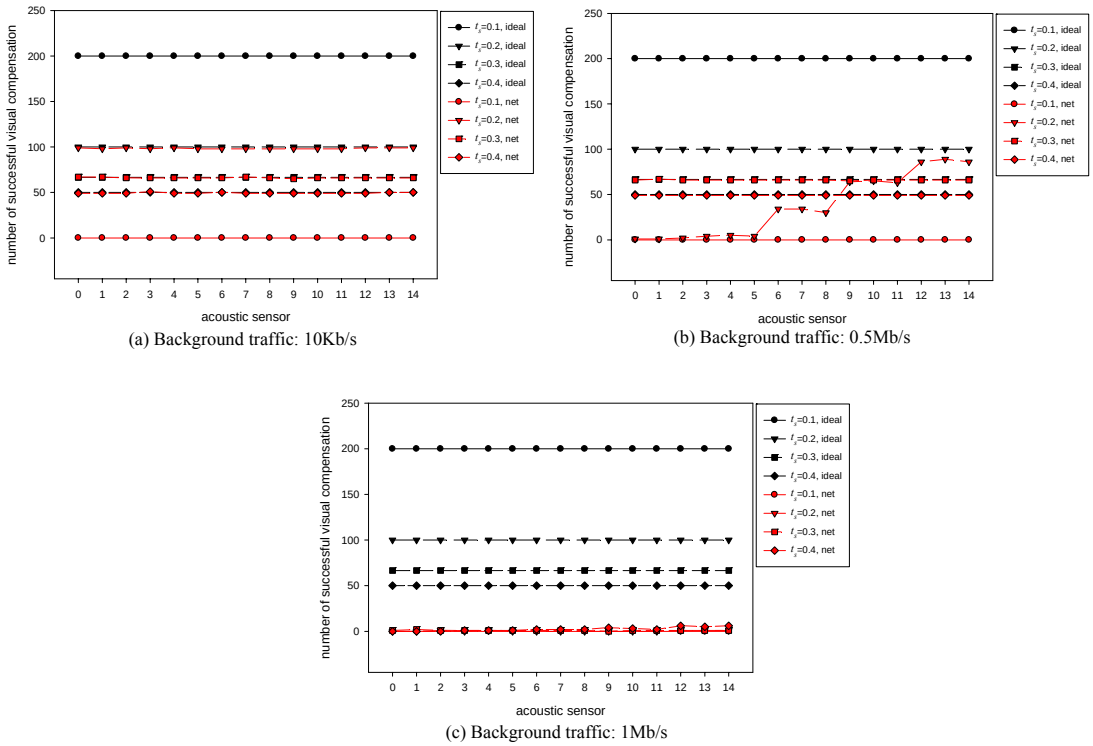


Fig. 8. String scenario for the tracking model. Acoustic senors, visual sensors and routers are denoted by A_i, V_i, and R_i (i =0, 1, ...), respectively.

In this section, we investigate the impact of *network synchronization problem* by simulation study. We use NS-2 simulator to build up a scenario of the developed tracking system as shown in **Fig. 8**. Even if its configuration is simple in terms of routers, the tracking complexity is affected by the number of acoustic sensors and tracking objects. Since the acoustic sensor-router communication delay is separate from the router-router transmission and can be minimized by packet aggregation, we believe it suffices to configure a line of routers for characterizing the impact of network synchronization and traffic pattern analysis of the developed tracking system. Five routers are communicating each other with 54Mb/s 802.11a single channel wireless link with 0.0005 uniformly distributed Bit Error Rate (BER). We turn off the RTS/CTS to reduce the network traffic overhead. Three acoustic sensors are located within the communication range of each router to send and receive data. The wireless channel between acoustic sensor and router is different to the channel of router-router links to reduce the interference. We assume each acoustic sensor is tracking five moving objects. The server is connected to last mile router $R4$, and two visual sensors are attached onto $R1$ and $R3$ to effectively capture the tracking space with different angles. The image frames generated by visual sensors are fixed by 40KByte size and we generate 10Kb/s, 0.5Mb/s, and 1Mb/s non-sensor (background) traffic by Constant Bit Rate (CBR). For the visual sampling interval, we set it up as $t_v = 10t_s$, where the t_s takes various values: 0.1, 0.2, 0.3, and 0.4 seconds. The simulation time is 200 seconds.

Fig. 9 shows how many visual compensation process can be successful when the network synchronization is considered in the tracking system. It plots the number of *success* in visual compensation at each acoustic sensor. The black point lines are for the ideal case achieved based on the **Fig. 2**. For example, if the acoustic sensors are sampling the object signal with t_s =0.1 second, correspondingly t_v =1.0 second, we expect the visual compensation in ideal case will be performed 200 times. However, **Fig. 9**(a) indicates that in real situation represented by red point line, no visual compensation is performed when t_s =0.1. This is due to the end-to-end delivery delay of sensor data, especially, visual images. Note 10Kb/s background traffic could be considered to be equal to a network with only multi-modal sensor traffic. When we measure the end-to-end delivery delays of 40KBytes visual image under even lower 1kb/s background traffic, we obtain 0.117 and 0.035 seconds of delivery delay for visual sensor 1 and 2. Since the visual sensor 1 is far away from the server, its delay is larger than that of the second visual

sensor and we can conjecture the compensation performance mainly depends on the large delay of the visual sensor 1. This result indicates that we can not realize visual compensation when we try to track a fast moving object with t_s =0.1 even in no background traffic. When the background traffic increases to 0.5Mb/s and 1Mb/s, the visual sensors can not assist the PF-based object tracking in small t_s values. Especially, the *network synchronization problem* leads to around zero visual compensation for all the t_s values in 1Mb/s background environment as shown **Fig. 9**(c).



(a) Background traffic: 10Kb/s

(b) Background traffic: 0.5Mb/s

(c) Background traffic: 1Mb/s

Fig. 9. Impact of *network synchronization problem* in the tracking system

4.2 Possible Solutions for Network Synchronization Problem

1) Adjustment of Sampling Time: A simple solution for the network synchronization is achieved from changing the sampling interval of acoustic sensors. For example, in the existence of 1Mb/s background traffic in the previous section, we can eliminate the non-synchronization by increasing the t_s to 0.8 second. This solution makes it satisfied with both *Condition*s mentioned in Section 3.4. However, this solution has difficulty in supporting the visual-assisted tracking for fast moving objects, so that there exists a limitation of tracking accuracy.

2) Traffic Differentiation Model: Network delivery delay, especially, the large image exchanging delay between visual sensors and a server is critical to the network synchronization. From this fact, we propose a sensor traffic differentiation model by using Weighted Round Robin (WRR) scheduling mechanism to be installed into routers. The basic idea of the model is that we can balance the network delay among the multi-modal sensor traffic and non-sensor traffic by using WRR, and eliminate the non-synchronization problem appearing in the tracking network.

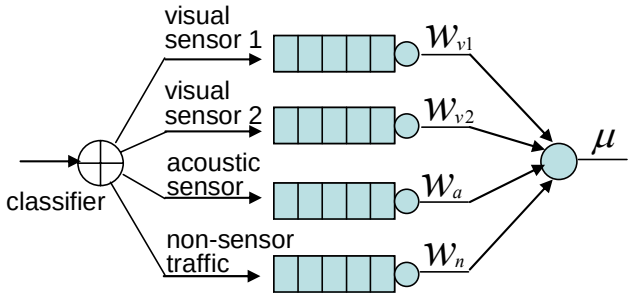


Fig. 10. Reference model for traffic differentiation.

Fig. 10 shows the reference model for the traffic differentiation. It has four separate queues and each queue is assigned to visual sensor 1 and 2, acoustic sensor, and miscellaneous non-sensor traffic. The feedback from a server to acoustic sensors delivering visual localization results is assigned to the second queue. If the service rate of a router is μ, each queue is served by a service weight factor w_{v1}, w_{v2}, w_a, or w_n. Here, we propose a weight allocation algorithm, Delay-based Weight Allocation (DWA) to determine the weight factor. In DWA, we first categorize the traffic into sensor and non-sensor traffic and assign weight into them. Let c_s and c_n be the allocated weights for sensor and non-sensor traffic and $c_s + c_n = 1$. Then, we need to perform fine grained weight allocation of c_s into the three different sensor queues, which is conducted based on the end-to-end delivery delay of each sensor type. The fine grained weight allocation is done by two rounds. In the first round, we need to measure the end-to-end delivery delay from multi-modal sensors to a server. Here, we will follow the notation of **Fig. 4**. Each value can be easily obtained since at the Service side, we can know the generation time and the arrival time of the sensor data. In the second round, we can obtain the weight allocation based on the measured delays. Let's define the total measured delay d_t as:

$$d_t = \Delta t^{v1} + \Delta t^{v2} + \Delta t^{a} \tag{1}$$

Based on the d_t, we get the weight factors as follows:

$$w'_{v1} = c_s \frac{\Delta t^{v1}}{d_t}, w'_{v2} = c_s \frac{\Delta t^{v1}}{d_t}, w'_a = c_s \frac{\Delta t^{a}}{d_t}, w'_n = \frac{c_n}{c_s} d_t \quad , \tag{2}$$

where w'_n is obtained from d_t based on the ratio of the categorized weight allocation. In order to normalize them, we define $w_t = w'_{v1} + w'_{v2} + w'_a + w'_n$, and get the final DWA formula as:

$$w_{v1} = \frac{w'_{v1}}{w_t}, w_{v2} = \frac{w'_{v2}}{w_t}, w_a = \frac{w'_a}{w_t}, w_n = \frac{w'_n}{w_t}, \tag{3}$$

where $w_{v1} + w_{v2} + w_a + w_n = 1$. The accomplished DWA weight factor is now applied for the service differentiation in the next router executions.

Fig. 11 shows the simulation result under the 0.5Mb/s and 1.0Mb/s background traffic environments when traffic differentiation model is applied into the tracking system. We compare the number of *success* of the proposed model with that of normal case in which only one queue serves the sensor and non-sensor traffic. For the categorized weights, we set them up by $c_s : c_n = 0.9 : 0.1$ to support the fast transmission of the sensor data. In the first round, we assign the identical weight for each queue such as $w_{v1} + w_{v2} + w_a + w_n = 0.25$.

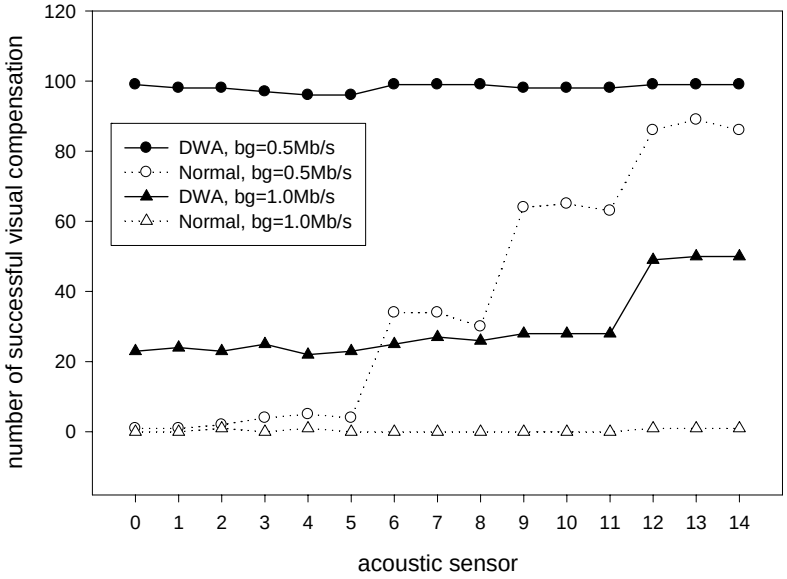


Fig. 11. Simulation results when traffic differentiation model is applied to the tracking system, where t_s = 0.2 and $t_v = 10t_s$.

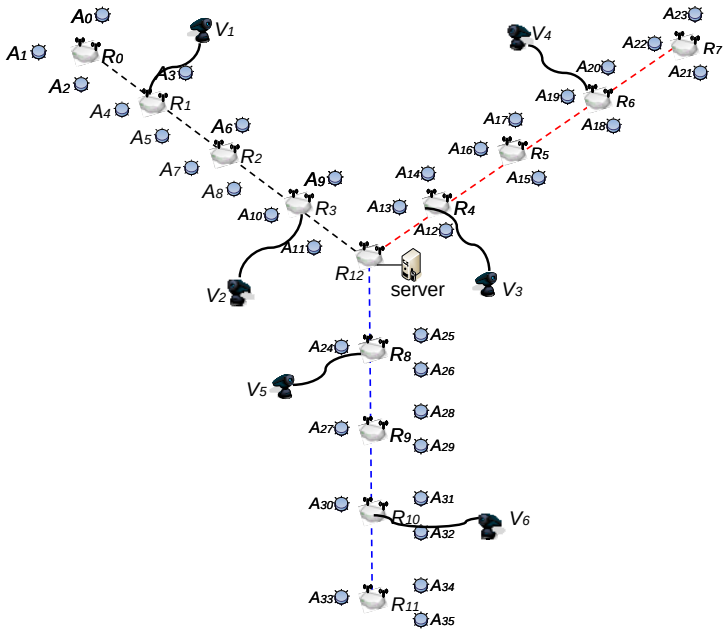


Fig. 12. Star scenario for the tracking model. Each branch uses different channel to reduce the channel interference.

In the second round, the initial weights are changed like w_{v1} =0.732, w_{v1} =0.222, w_a =0.043, and w_n =0.003 by means of DWA calculation. The simulation result is obtained only when t_s =0.2. Note that as indicated in Section 4.1, the tracking system achieves no number of *success* in visual compensation at t_s=0.1 since the pure end-to-end delivery delay of visual sensor 1 is

larger than 0.1 second. Thus, t_s=0.1 case is not plotted in the figure. We also do not plot the cases of t_s=0.3, 0.4 under 0.5Mb/s and 1Mb/s background since we achieve the same number of *success* as the ideal case when we apply the traffic differentiation model into the tracking system. When we remind that there is no *success* in **Fig. 9**(c) in any t_s values, we understand the differentiation model can efficiently mitigate the non-synchronization in network. To save the space, we do not plot the case of 10Kb/s background situation since the result is same as **Fig. 9**(a). When we observe the plot for 0.5Mb/s background traffic case, the differentiation model applying the DWA achieves almost the same as the ideal case even if the normal case which reflects the *network synchronization problem* shows the unbalanced visual compensation. Since the acoustic sensor 0 to 5 are far from the server, their packet end-to-end delivery delays are larger than the delays of other acoustic sensors. Therefore, we obtain the unbalanced curve in normal case. For the 1.0Mb/s background environment, differentiation model also provides more number of *success* than the normal case even if the result is affected by the background traffic volume.

In order to investigate the effect of traffic differentiation model in more complicated scenario, we show the number of successful visual compensation from a star scenario with three branches in **Fig. 12**. Each branch takes the same configuration as **Fig. 8**. Since the heavy interference between sensors and routers, each branch uses different wireless channel for data exchange. The simulation parameter setting is also same as the **Fig. 8**.

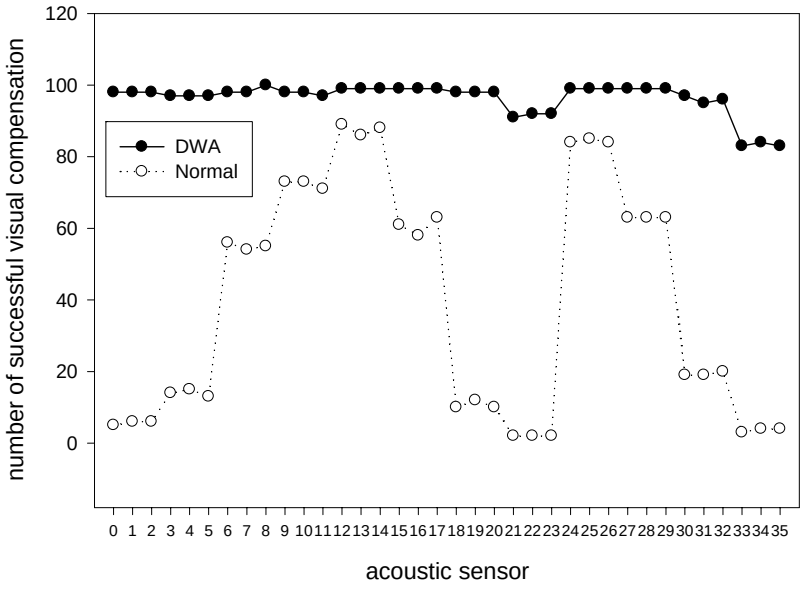


Fig. 13. The number of *success* in the visual compensation when traffic differentiation model is applied: 0.5 Mb/s background. Here, t_s = 0.2 and t_v = 10t_s.

Fig. 13 shows the simulation results under 0.5 Mb/s background traffic in the star scenario. Due to the same reason as **Fig. 11**, we only plot the result from only t_s =0.2. We observe that the DWA provides almost the same as the ideal number of *success* for 36 acoustic sensors. However, the Normal case shows large variation and low successful visual compensation result. For the case of 1.0 Mb/s background traffic, **Fig. 14** contains the plots for t_s =0.2, t_s =0.3, and t_s =0.4. We can find out that the Normal case supports below 10 *success* even in t_s =0.4 due to *network synchronization problem* caused by the complex tracking scenario. However, the

traffic differentiation model efficiently resolves the *network synchronization problem* and provides significant improvement of the *success* in the visual compensation process. When t_s =0.3 and t_s =0.4, the differentiation model provides the same as the ideal number of *success*. However, in DWA, t_s =0.2, the contention for the network resource between large volume of acoustic sensor data and visual sensor data causes large fluctuation among acoustic sensors even with better number of *success* than Normal case.

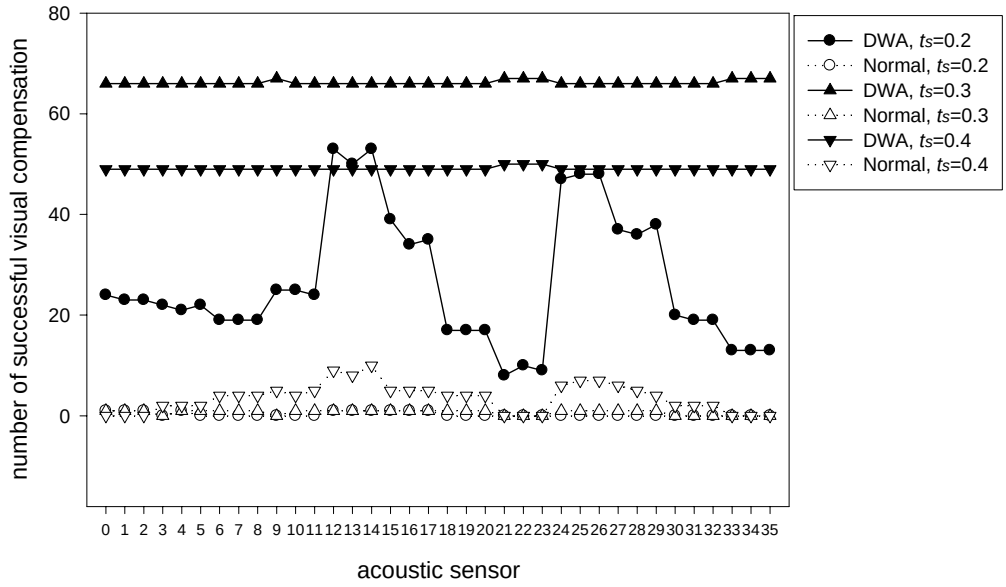
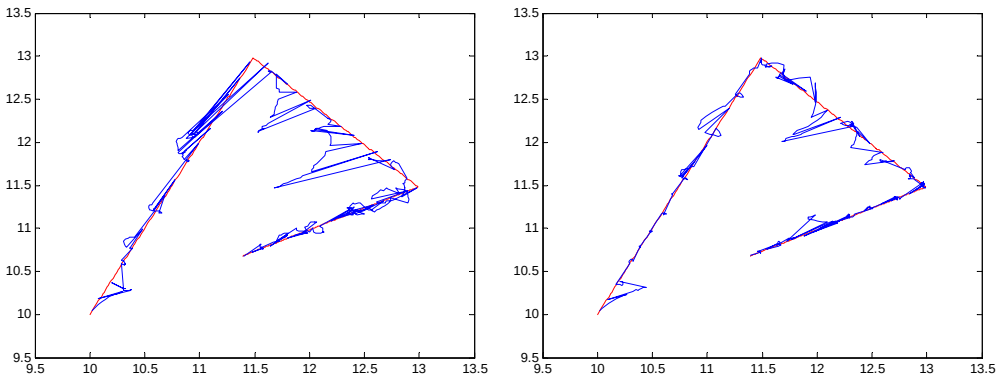


Fig. 14. The number of *success* in the visual compensation when traffic differentiation model is applied: 1.0 Mb/s background. Here, $t_v = 10t_s$.



(a) Acoustic sensor 0 to 11 (b) Acoustic sensor 12 to 14

Fig. 15. Estimated trajectory of a target object under traffic differentiation model.

In order to investigate how the trajectory estimate of object movement changes when the differentiation model is applied, we show the tracking results in **Fig. 15**. This is the result only for DWA,bg=1.0Mb/s case in **Fig. 11**. Since the differentiation model provides the same results as the ideal case in DWA,bg=0.5Mb/s, the trajectory estimation in the case is almost the same as the real object movement. The trajectory estimation in **Fig. 15**(a) reveals that the

differentiation model provides reasonably accurate tracking outcome in acoustic sensor 0 to 11 even with a little bit large deviation from real object movement. Note that the trajectory estimation of acoustic sensor 0 to 11 under non-differentiation model has the result in **Fig. 3**(a) since the normal one queue model achieves around zero *success* in visual compensation. The acoustic sensor 12 to 14 show better trajectory estimation in **Fig. 15**(b) since they are assisted with more *success* of visual compensation.

5. Behavior Analysis of the Tracking System

From the observation of the developed tracking system, we can understand the successful visual compensation mainly depends on the sampling interval of the acoustic sensor and the image delivery delay. Therefore, in this section, we investigate the accuracy of a tracking system in terms of the two parameters.

5.1 Performance Metric

In order to represent the *success* and *fail* cases mentioned in Section 3.4 by a numeric value, we define a new performance metric applicable to the tracking system. We call it Successful Compensation Rate (SCR) and define it as:

$$\text{SCR}=\frac{n_s}{n_t} \tag{4}$$

where n_s is the number of successful compensation that satisfies both conditions in Section 3.4 and n_t is the total number of sampling of an object signal at an acoustic sensor. If total running time of the tracking system is T , $n_t = T/t_s$. If a tracking model achieves large SCR value, the PF algorithm is highly compensated by localization algorithm, so that we can more accurately track the target object. Therefore, the SCR metric can be a gauge to determine the tracking accuracy of the established tracking system. Note the accomplished SCR reflects the *network synchronization problem*.

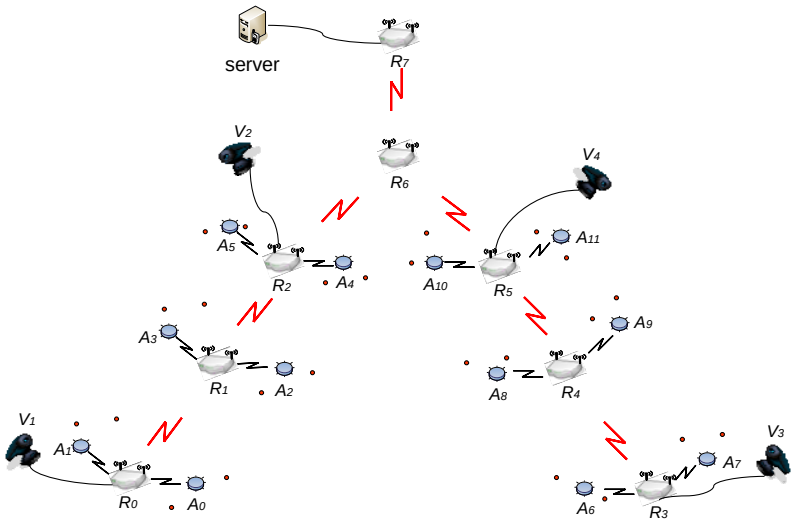
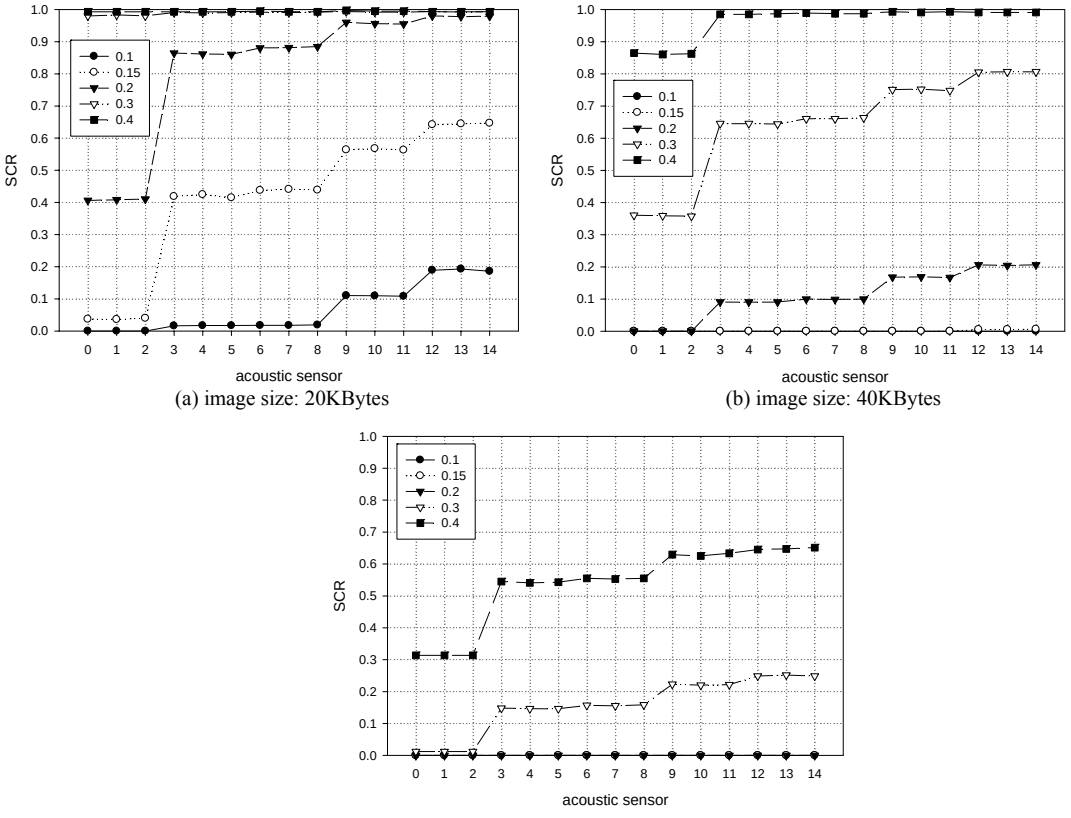


Fig. 16. Tree scenario for the tracking model. Due to the line-of-sight characteristics of visual sensors, we install 4 visual sensors.

5.2 Behavior Analysis

1) Simulation Setup: In order to observe the system behaviors, we perform the simulations based on the scenarios in **Fig. 8** and a more complex tree scenario as shown in **Fig. 16**. Visual sensors generate image frame of 20, 40, and 60 KBytes size. We observe the behaviors based on the acoustic sampling interval of 0.1, 0.15, 0.2, 0.3, and 0.4. For the tree scenario, we set up the same communication link as **Fig. 8**, and only two acoustic sensors work in the communication range of each router. Each acoustic sensor tracks five objects. We assume that both branches of the tree cannot guarantee the line-of-sight characteristics of a visual sensor, so that we install two visual sensors for each branch.

2) Simulation Results: **Fig. 17** shows the simulation results achieved from string scenario. We plot the SCR variations of 15 acoustic sensors. Let's assume the both tracking scenarios need SCR = 0.6 to detect the object trajectory. Then, in 20KBytes image size, the acoustic sensors do not need to capture the object signal with t_s < 0.3 since the *network synchronization problem* blocks the on-time transmission of the sensor data for the visual compensation. If the visual sensors generate 40KBytes image, only t_s =0.4 supports the expected SCR value. In case of 60KBytes, all the simulated sampling times do not support the stable object tracking.



(a) image size: 20KBytes

(b) image size: 40KBytes

(c) image size: 60KBytes

Fig. 17. SCR results achieved in string scenario.

For the tree scenario, we plot only SCR result for 20KBytes image size in **Fig. 18** since the other cases support no visual compensation. This is because the complex network confi guration leads to the large end-to-end delivery delay of the visual image to a server. Even in

20KBytes, only acoustic sensor 4 and 5 which are near the server support the expected SCR = 0.6.

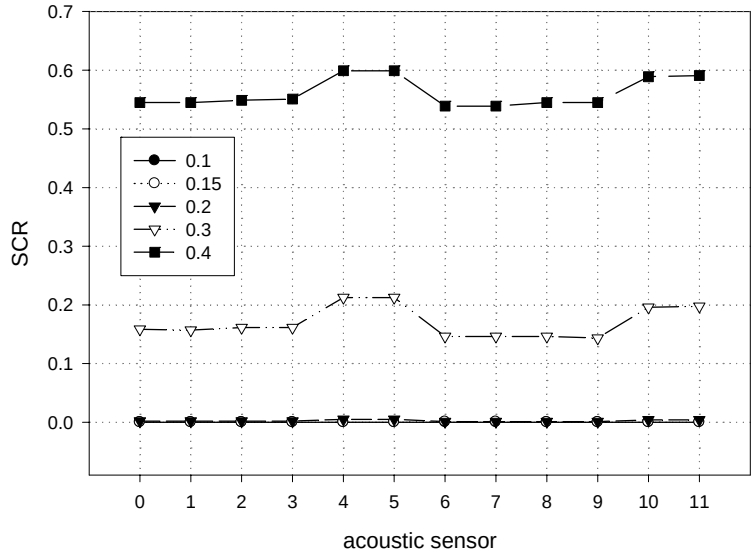


Fig. 18. SCR results achieved in tree scenario when image size is 20Kbytes.

6. Conclusions

The research works in this paper is based on our developed tracking system in which acoustic sensors mainly track target object with Particle Filter (PF) algorithm and the assisting functions to eliminate the estimation error inherent in PF are done by visual sensors performing localization algorithm. Based on the developed tracking system, this paper has identified the *network synchronization problem* caused by unbalanced delivery delay of sensor data between multiple sensor types. For example, visually captured image size is larger than the acoustic sensor data size, so that visual images arrive at a processing server later than acoustic sensor data. In this situation, the server performs the visual compensation only when the sampling times of the sampled data satisfy the successful condition for visual compensation. In order to efficiently deliver the acoustic sensor data to a server, we have proposed a time-based aggregation algorithm.

For the possible solution for the non-synchronization problem in a network, we separate the network queues to differently serve the sensor traffic and non-sensor traffic. The traffic differentiation model is achieved by Weight Round Robin (WRR) where the weights are allocated based on our proposed Delay-based Weight Allocation (DWA) algorithm. We have shown that the differentiation model sufficiently mitigates the unbalance in end-to-end delivery delay of sensor data and supports high level of visual compensation assistance. Finally, we have investigated the behavior of the tracking system in terms of acoustic sampling interval and visual image size.

References

[1] J. Lee, S. Oh, S. Hong, "Enhancing Particle Filtering Performance in Tracking through Visual Information Association," submitted to *IEEE Transactions on Signal Processing*, http://msdl.ee.sunysb.edu/~skjung/papers/association.pdf.
[2] C. H. Knapp and G. C Carter, "The Generalized Correlation Method of Estimation of Time Delay," *IEEE Trans. on Acoustic, Speech, and Signal Processing*, vol. ASSP-24, no. 4, pp. 320-327, 1976.
[3] J. H. Dibiase, H. F. Silverman, and M. S. Brandstein, "Robust Localization in Reverberant Rooms," *Microphone Arrays: Signal Processing Techiniques and Applications*, pp. 157-180, 2001.
[4] N. Strobel, T. Meier, and R. Rabenstein, "Speaker Localization using Steered Fltered-and-sum Beamforemrs," *Proc. of Erlangen Workshop on Vision, Modeling, and Visualization*, pp. 195-202, Erlangen, Germany, 1999.
[5] J. Vermaak and A. Blake, "Nonlinear Filtering for Speaker Tracking in Noisy and Reverberant Environments," *IEEE Int. Conf. Acoust., Speech, Signal Processing* (*ICASSP-01*), pp. 3021-3024, Salt Lake City, UT, May 2001.
[6] D. B. Ward and R. C. Williamson, "Particle Filter Beamforming for Acoustic Source Localization in a Reverberant Environment," *Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing* (*ICASSP-02*), vol. II, pp.1777-1780, Orlando, FL, USA, May 2002.
[7] C. Jue, J. P. Le Cadre, and P. Perez, "Sequential Monte Carlo Methods for Multiple Target Tracking and Data Fusion," *IEEE Trans. Signal Processing*, vol.50, pp. 309-325, February 2002.
[8] Jinseok Lee, Jaechan Lim, Sangjin Hong, Peom Park, "Tracking an Object in 3-D Space using Particle Filtering based on Sensor Array," Proc. of *IEEE International Conference on Computer and Information Technology* (*CIT*), pp. 242, September 2007.
[9] P. M. Djuric´and J. H. Kotecha and J. Zhang and Y. Huang and T. Ghirmai and M. F. Bugallo and J. Miguez, "Particle Filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19-38, September 2003.
[10] A. Doucet, N. de Freitas, and N. Gordon, Eds., "Sequential Monte Carlo Methods in Practice," *New York: Springer Verlag*, 2001.
[11] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "A Novel Approach to Nonlinear and Non-Gaussian Bayesian State Estimation," *IEEE Proceedings-F: Radara, Sonar and Navigation*, vol. 140, pp. 107-113, 1993.
[12] J. Carpenter, P. Clifford, and P. Fearnhead, "An Improved Particle Filter for Non-linear Problems," *IEE Proceedings-F: Radara, Sonar and Navigation*, vol. 146, pp. 2-7,1999.
[13] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A Tutorial on Particle Filters for Online Non-linear/Non-gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174 -188, February 2001.
[14] R. Okada, Y. Shirai and J. Miura, "Object Tracking Based on Optical Flow and Depth," *IEEE/SICE/RSJ International Conference*, pp. 565-571, December 1996.
[15] S. Khan and M. Shah, "Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1355-1360, October 2003.
[16] A. Bakhtari, M. D. Naish, M. Eskandari, E. A. Croft and B. Benhabib, "Active-Vision based Multisensor Surveillance -An Implementation," *IEEE Trans. on Systems, Man and Cybernetic -Part C : Application and Riviews*, vol. 36, no. 5, pp. 668-680, September 2006.
[17] D. N. Zotkin, R. Duraiswami and L. S. Davis, "Joint Audio-Visual Tracking Using Particle Filters," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 11, pp. 1154-1164, January 2002.
[18] M. S. Arulampalam, B. Ristic, N. Gordon and T. Mansell, "Bearings only Tracking of Manoeuvring Targets Using Particle Filters," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 2351-2365, 2004.
[19] Y. Boers and J. N. Driessen, "Interacting Multiple Model Particle Filter," *Proc. of the IEE Radar Sonar Navigation*, vol. 150, No. 5, 2003.
[20] A. S. chhetri, D. Morrell and A. P. Suppappla, "Scheduling Multiple Sensors using Particle Filters in Target Tracking," *Proc. of the IEE Statistical signal Processing*, pp. 549-552, Tempe, USA,

September 2003.
[21] J. Lee, S. Hong, P. Park, W. D. Cho, "Object Tracking Based on RFID Coverage and Visual Compensation in Wireless Sensor Network," *Proc. of the IEEE International Symposium on Circuits and System*, pp. 1597-1600, New Orleans, USA, May 2007.
[22] M. Stanacevic, G. Cauwenberghs, "Micropower Gradient Flow acoustic Localizer," *IEEE Transaction on Circuits and Systems I*, vol. 52, pp. 2148-2157, 2005.
[23] S. Hong, J. Lee, A. Athalye, and P. Djuric, "Design Methodology for Domain-Specific Parameterizable Particle Filter Realizations," *IEEE Transactions on Circuits and Systems I*, vol. 54, pp. 1987-2000, 2007.
[24] J. Lee, S. Jung, Y. Kyong, X. Deng, S. Hong, and W-D. Cho, "Data Traffic Analysis in Wireless Fusion Network with Multiple Sensors," *Proc. of IEEE International Midwest Symposium on Circuits and Systems* (*MWSCAS*), Montreal, Canada, August 2007.
[25] 4XEM PTZ Pan/Tilt/Zoom IP Network Camera, http://www.4xem.com/products/wired/IPCAMWPTZ/index.html.
[26] Network Simulator-2, http://www.isi.edu/nsnam/ns.

Sangkil Jung received the BS in Statistics from Chungnam National University, South Korea in 1999, the MS degree in Information and Communications Engineering from Gwangju Institute of Science and Technology, South Korea in 2001, and the PhD degree in Electrical and Computer Engineering from State University of New York at Stony Brook, USA in 2007. His research areas cover the algorithm design and performance evaluation based on software and hardware integration. The target applications are the voice and sensor with compression and aggregation on wireless mesh networks, and the multi-modal object tracking underlying a multi-hop wireless network. He has also tackled the security problems appearing in wireless sensor/mesh and broadband wireless access networks. He is now with Samsung Electronics.

Jinseok Lee received dual B.S. degrees in Electrical and Computer Engineering from Stony Brook University, NY, USA and Ajou University, Korea in 2005. He has received the Ph.D degree in the Department of Electrical and Computer Engineering from State University of New York at Stony Brook, USA in 2009. His research interests include multi-modal signal processing, sensor node architecture optimization, complex signal processing algorithm design, and various estimation problems for wireless sensor networks.

Sangjin Hong received the B.S and M.S degrees in EECS from the University of California, Berkeley. He received his Ph.D in EECS from the University of Michigan, Ann Arbor. He is currently with the department of Electrical and Computer Engineering at State University of New York, Stony Brook. Before joining SUNY, he has worked at Ford Aerospace Corp. Computer Systems Division as a systems engineer. He also worked at Samsung Electronics in Korea as a technical consultant. His current research interests are in the areas of low power VLSI design of multimedia wireless communications and digital signal processing systems, reconfigurable SoC design and optimization, VLSI signal processing, and low-complexity digital circuits. Prof. Hong served on numerous Technical Program Committees for IEEE conferences. Prof. Hong is a Senior Member of IEEE.