

On the Impact of Wireless Multimedia Network for Multi-Modal Activity Recognition

Akika Yamashita
Ochanomizu University
Department of Computer Sciences
Email: akika@ogl.is.ocha.ac.jp

Eng Keong Lua
MONASH University
Faculty of Information Technology
Email: lua.eng.keong@monash.edu

Masato Oguchi
Ochanomizu University
Department of Computer Sciences
Email: oguchi@computer.org

Abstract—Due to the rapid development of sensor devices, many types of multimedia sensor data can be collected and analyzed to develop useful multimedia applications such as human activity recognition. Our objective of this research work is to study and analyze the impact of the wireless multimedia network quality on human activity recognition accuracy in an end-to-end networked multimedia system environment - Ocha House. We utilized the new improved Generalized Time Warping (GTW) algorithm and its variants for temporally aligning multi-modal sequences from multiple subjects performing similar human activities. In Ocha House, we captured the human activity data with camera sensors which are transmitted via wireless network for server processing and analysis. We evaluated the human activity recognition accuracy using GTW and its variants by taking into consideration of the wireless network data packet loss phenomenon due to common wireless network protocol behaviors as a result of network flow control and congestion management. We discovered that GTW and its variants are sensitive to wireless network packet loss and observed interesting characteristics resulting in deteriorated human activity recognition accuracy - when wireless network packet loss rate is nominal, the human activity recognition accuracy decreases dramatically.

I. INTRODUCTION

In recent years, with the rapid development of small and high technological sensor devices, many types of multimedia sensor data can be collected and converted to develop useful networked multimedia applications such as human activity recognition. There are many past and existing research works focused solely on computer vision and image processing algorithms in terms of their human activity recognition accuracy. However, these research works lack the understanding of end-to-end networked system characteristics affecting the captured multimedia sensor data quality in real time transmission that will impact the accuracy of human activity recognition. Similarly, many past and current works focus on the investigation of real time wireless LANs (WLAN) to improve the wireless communication quality without discussing in details its effects in multimedia application performance. Thus, our objective of this research work is to study and analyze the impact of the WLAN communication quality on the human activity recognition accuracy in the end-to-end networked multimedia system environment — Ocha House. Ocha House is a Japanese home setup to conduct cyber-physical system experiments using several multimedia camera sensors to monitor and capture human motions and activity data in an end-to-end wireless multimedia network environment. That is, the collected human motions and activity data in Ocha House are transmitted real

time over the WLAN to a server for processing and analysis on human activity recognition application.

We utilize the new improved Generalized Time Warping (GTW) algorithm [1] and its variants (pDTW — Procrustes dynamic time warping), pDDTW — procrustes derivative dynamic time warping, and pIMW — procrustes iterative motion warping) proposed by Feng et al. at Carnegie Mellon University [2], for temporally aligning multi-modal sequences from multiple human subjects performing similar human motions and activities. Human activity data are captured and transmitted via WLAN communication for server processing and analysis using GTW and its variants in the Ocha House. We evaluated the accuracy of the human activity recognition using GTW and its variants by taking into consideration of the natural WLAN data packet loss phenomenon due to wireless network flow control and congestion protocol management. We conduct experiments using three types of captured multimedia data transmitted over the WLAN for processing and analysis at the server: (1) extremely long and short sequences with large size of image matrix, (2) data are smoothed, moderately aligned to almost the same length of sequences with small size of image matrix, (3) all backgrounds of small image matrix data are removed. We discovered that GTW is sensitive to WLAN packet loss resulting in deteriorated human activity recognition accuracy — when WLAN packet loss rate is high, the GTW algorithm accuracy decreases dramatically.

Outline. We discuss the related work of this research in Section 2. We describe the Ocha House environment and its end-to-end human activity recognition system in Section 3. In Section 4, We explain the wireless network packet loss phenomenon as the natural result of network control and congestion management. We describe the GTW algorithm and its variants that are used for multi-modal sequences alignment of multiple subjects for human motion activity recognition in Section 5 In Section 6, we explain our experiments and discuss the experimental results. Finally, we conclude this paper in Section 7.

II. RELATED WORK

First, we discuss the research about WLAN communication quality. A WLAN defines multiple transmission rate to keep the bit and/or frame error rates to be proper values, and appropriate transmission rates are selected depending on the bit error state. The multi-rate control works as follows: when a fine radio environment can be obtained, a higher transmission rate is selected, if a higher transmission rate cannot be kept

due to a poor radio environment, a lower transmission rate is selected. With such a mechanism, the multi-rate control can accommodate multiple terminals in different conditions. 11, 5.5, 2 and 1 Mbps are defined for IEEE802.11b and 54, 48, 36, 24, 18, 12, 9 and 6 Mbps are defined for IEEE802.11a/g as standard values in multi-rate control. Like [3] and [4], efficient rate adoption method and technique have been proposed, and [5] have evaluated the performance of multi-rate throughput with real-terminal.

Secondly, we discuss the research about aligning human motions from sensory data in the context of computer graphics and computer vision. In particular, we focused on the technique of Time warping. The technique is to align several sequences. For example, it aligns the sequence of video or frames in which different people are doing the same actions but the timing is not the same. As shown in [6], CTW (Canonical Time Warping) combines DTW (Dynamic Time Warping) with canonical correlation analysis to align data of different dimensionality temporally. Existing methods using CTW and DTW have following limitations: (1) computational complexity is quadratic in space and time, (2) only solved the problem of aligning two sequences, therefore it is unclear how to extend it to the alignment of multiple sequences, (3) rely on dynamic programming to find the optimal path, however it is unclear how to constrain the temporal warping adaptively. To overcome these limitations, GTW that allows an efficient and flexible alignment among two or more multi-dimensional time series of different modalities have been proposed [1]. The deterioration of the captured data quality through WLAN transmission affecting the accuracy of the GTW system has not been studied before. Hence, this is the first work to investigate the impact of the WLAN in GTW system.

Our research is different from others because we have considered end-to-end system. When the human action recognition system is used, to run all the system at local environment is not efficient and it is better to separate the data storage terminal and data processing and analyzing terminal. Then the captured and collected data in the storage should be sent to the terminal for data processing, and there should be packet loss depending on the quality of WLAN communication environment. Thus only talking about the WLAN communication quality or image processing model is not enough. In this paper, we evaluated how packet loss rate affects the accuracy of pDTW (procrustes dynamic time warping), pDDTW (procrustes derivative dynamic time warping), and pIMW (procrustes iterative motion warping) and GTW (generalized time warping) (details in Section 5).

For any Human Activity Recognition System, it requires the integration of the sensor devices to the network such as WLAN, for the transmission of captured data for analysis. Some researchers attach sensor devices such as gesture-based movement acceleration and medical measurement sensor devices directly to users and send the captured data via wireless network to the server for ANALYSIS (computer vision and image processing). In this work [7], Francesca De Simone et al. evaluated the video quality at the premises of two academic institutions. The video data include the packet loss data when transmitting via IP networks and groups of people are asked to rate the quality. In our research, we evaluated the video quality with GTW algorithm and its variants, which is the

reliable quantitative evaluation. The research works that are related to Human Recognition System in the smart house are also performed. Simon Moncrieff et al. in [8] have developed "anxiety framework" to determine hazards such as abnormal activity inside the smart house. Lian Wang et al. in [9] have used wearable sensors and wireless networks to recognize multi-user activities inside the smart house.

In these research, it exists the same problems of WLAN packet loss due to the network transmission but such research neglected the persistent network impact on the system quality. As such, our research work is the first that presents new findings and important observations that cannot be ignored in a total system view for end-to-end system quality.

III. OCHA HOUSE — HUMAN ACTIVITY RECOGNITION

To evaluate the performance of human activity recognition system, we collected the real data in Ocha House (Fig. 1). Ocha House is an experimental smart house constructed in March 2009. The purpose of the Ocha house is to implement and evaluate ubiquitous computing applications and cyber-physical systems research. We have setup cameras and other sensors like motion acceleration sensor and sound sensor to capture and collect "real" human actions inside the House. That are a natural data of people acting spontaneously. Particularly, we discuss the image data captured with camera in this paper. The camera is monitoring inside the house and start recording automatically if there is some actions and moving object. We collected a lot of kinds of human actions inside the house, for example, people open the door, sit on the chair, wipe the desk, open the door of shelf, etc. The human activity analytic system can detect what action people are doing inside the house with image processing. In this paper, we applied GTW and its variants as image processing methods, and align and/or classify different people with the same actions. By analyzing these data, we can detect anomalies in human activities. We evaluated the GTW and its variants accuracy with naturally captured data.

As shown in the Fig. 2, we have installed four cameras to record the video data from four different angles. The data captured with the cameras will be sent to storage immediately through WLAN. The reason why we use WLAN is that we separated the system into three parts, data capturing part of cameras, data storage part, and data analytic part. Camera and storage are tied with Access Point (AP). Data sending protocol is UDP (User Datagram Protocol) as we put emphasis on higher data rate than image data quality even if there will be packet loss. In UDP, due to the influence of interference and noise around, packet loss will happen and throughput will decrease as a result. We assume the following situation when packet loss happens. The bandwidth of each terminal decreases and the quality of communication will deteriorate in the following situations: (1) several terminals are sharing the same AP and (2) several APs are near using closer channel. We used the wireless network packet loss data assuming situation (1) in our research. We discuss the wireless network packet loss data phenomenon caused by network flow control and congestion management in following section IV.



Fig. 1. Ocha House

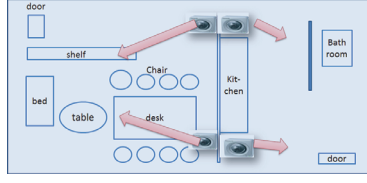


Fig. 2. Inside Ocha House

IV. NETWORK PACKET LOSS PHENOMENON

IEEE (Institute of Electrical and Electronics Engineer) 802.11 is a WLAN standard and IEEE 802.11g used in our experiments can communicate about 54Mbps in the 2.4GHz band. The target protocol of IEEE802.11 standard is Medium Access Control (MAC) as data link layer and physical layer. In wireless network, frame transmission is performed in form of broadcast and several terminals can share the bandwidth. Thus, if several terminals send the frames at the same time, the signal overlaps, frames cannot be sent correctly. When the frames conflict, the frames will be lost and retransmission happens. As the result, the throughput will decrease. To avoid frame collision, CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) is used. This is autonomous distributed control to avoid frame collision. CSMA can avoid frame collision as far as possible by controlling the timing of each terminal sending frames and detecting the usage of Wireless channel before each terminal sends frame. As shown in Fig. 3, each terminal can only send frame when a certain idle time is detected. The interval of sending frame is called DIFS (DCF Inter-Frame Space). Frame cannot be sent immediately, terminals have to wait during DIFS time. If there is no signal sensed within DIFS time, terminals judge that there is no signal and start to send their frames. However, sometimes, several terminals detect the channel is idle and send the frames at the same time, and frame collision happens. To avoid this, Backoff control is defined. Backoff is random time based on Contention Window (CW). Terminals have to wait during Backoff time after DIFS idle time. As the backoff time is randomly set, frame collision will decrease than only sensing carriers. If frame collision happens, the range of CW will be double, and it can reduce the probability of re-collision. Thus, if many terminals are sharing one WLAN bandwidth, frame collision will happen. In our experiment, seven times frame collision leads to one time packet loss.

In our paper, "noise" means the noise of radio wave (electric wave) that is delivered from other electromagnetic devices, existing at the space of the experiment. When we use WLAN to transmit packets, not all packets reach the receiver side by several causes. There are two main causes of the packet loss in WLAN: noise of radio wave and congestion on the transmitted route. We used UDP for the transmission of video data frames. Since UDP is an algorithm whose transmitted data is affected when packet loss happens, different from TCP, the throughput must be affected by the packet loss because throughput is the number of received packets in a unit time when they are transmitted through WLAN.

As the lost packets are just abandoned and will not reach to the receiver, throughput should decrease in such a case. Flow control in WLAN is transmission rate control. Therefore, the number of video data frames received at the receiver side

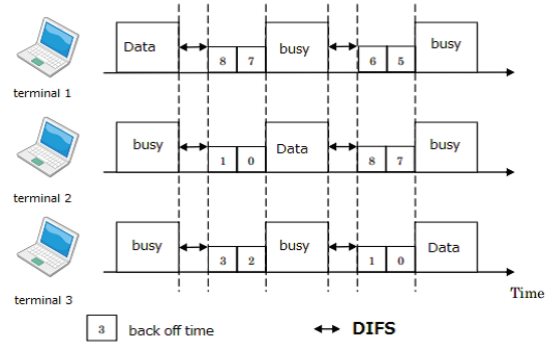


Fig. 3. Backoff control in WLAN

should be affected by the flow control in WLAN. In WLAN, congestion typically occurs when many senders deliver their packets almost simultaneously. As the noise is not strong in the condition of this experiment, the main reason of losses should be due to congestion.

V. GTW AND VARIANTS

The Generalized Time Warping (GTW) algorithm and its variants for temporally alignment of multi-modal sequences from multiple human subjects performing similar human motion activities are proposed in Feng et al. [1]. We will explain in detail in this section.

Recently, accurate alignment of human motions of different style or speed have been challenged in image processing research area. In the paper [1], Feng et al. have challenged to align multi-modal alignment of time series of people performing almost the same activities from different sensors. They proposed GTW which is the extension of DTW (Dynamic Time Warping) which is an old method to align human behavior [1] among two subjects only. Before GTW, many kinds of time warping algorithms have been proposed, for example, CTW (Canonical Time warping) [6] and DMW (Dynamic Manifold Warping) [10]. CTW can temporally align data of different dimensionality, for example, motion capture, video, and CTW combines DTW with canonical correlation analysis. DMW is extension of CTW, which can incorporate more complex spatial transformations using manifold learning. However, these kinds of time warping algorithm have limitations because they rely on DTW. The limitations are as follows: (1) Computational complexity is high, (2) Only two sequences can be aligned, not multiple sequences, (3) They rely on dynamic programming to find the optimal path. To overcome these limitations, GTW has been proposed, which can efficiently and flexibly align two or more multi-dimensional time series of different modalities, and has linear complexity because it uses a Gauss-Newton algorithm to optimizes the time warping function efficiently. These approaches differ from existing approaches based on dynamic programming. See details about Objective function in [1].

In the experiments, Feng et al. utilized three types of input video data to evaluate the accuracy of GTW and compared GTW with other variant methods such as pDTW (Procrustes dynamic time warping), pDDTW (procrustes derivative dynamic time warping), and pIMW (procrustes iterative motion warping) for the temporally aligning of multi-modal sequences

from multiple subjects performing similar human activities. pDTW [11] is used for shape for alignment and they extended pDTW to align multiple sequences. pDDTW is a technique to do invariant translation of DTW, using derivatives of the original features [12]. pIMW [13] is extension of IMW which time warping and spatial transformation are alternated to align two sequences. Input data in each experiment that were conducted by Feng et al. are as follows:

- Experiment1: time series alignment with known ground truth
- Experiment2: several video sequences alignment of different people doing similar actions
- Experiment3: three different sensor alignment of people doing similar actions

In Experiment1, synthetically generated 3-D spatio-temporal signals is used as input data. As a result, GTW performed the best. pDTW fails since the sequences have been distorted in space. pDDTW also fails since the feature derivatives do not capture the structure of the sequence well. pIMW is more prone to over-fitting. In Experiment2, GTW and its variants have been applied to align video sequences, the silhouette with background subtraction was extracted. Each sequence is two cycles of human walking sequences. As a result, pDTW nor pDDTW could align the video sequences since they lack the ability to solve for correspondence among signals of different nature. pIMW was well for top three components in space. However, it overfits all the dimensions and a biased time warping path was obtained. Only GTW could warp the sequences accurately in both space and time. In Experiment3, even if input data are different types captured from different sensor, and people in one of three sequence performing a little different action from others, GTW could solve the temporal correspondence among three sequences.

VI. EXPERIMENTS AND RESULTS

A. Experimental System of Collecting Data

In this section, we will describe how we collected the various multimedia sensor data over the wireless network in Ocha House. For quantitative evaluation, we first carried out the experiments with real terminals. Fig. 4 shows the experiment environment of transferring captured data with camera. A certain size of image frame data will be sent from sender side with Ethernet Converter (EC). We also fixed the transmission rate for quantitative evaluation. In IEEE802.11g, eight kinds of multi-rate values are defined (54, 48, 36, 24, 18, 12, 9, and 6Mbps). In our experiment, we used AP "MZK-MF300N [14]" as AP made by Planex company, which can fix the transmission rate. We used 54, 36, 18, and 6 Mbps, and auto transmission rate as the representative value.

As for the receiver side, several Android terminals are communicating with AP as background terminals to interfere with the main stream, video data transmission. We used Android terminal Nexus S [15] (NS) as background terminals, the number of background terminal NS is 0 - 5. The transmission rate of NS is fixed to 54Mbps and they are communicating with AP using Iperf [16] which is a software to send UDP packets to another terminal. The main terminal transmitting video data and 5 NS compete for bandwidth. As there are several background terminals communicating with AP, the packet sent from main terminal (video stream) will conflict with the packet sent from background terminals, and more back

ground terminals there are, less packets from main terminal can be reached to the receiver. When the transmission rate of main terminal is lower, the throughput should be lower as main terminal has less chance to send packet. We have analyzed in detail what happens at main terminal when its transmitted packets conflict with those from background terminals.

First, we have estimated the ability of EC used in the experiment. The EC is connected to PC at the sender side with wired LAN, and the transmission throughput of PC at the sender side is 100 Mbps. Thus the packets that sent out from sender PC are converted to WLAN packets, and sent out from EC with five pattern of transmission rate, 54, 36, 18, 6 Mbps and auto. This means some packets might be lost at EC as the wired LAN speed is much higher than that of WLAN. The buffer overflow on EC with various transmission rate is shown in Fig. 6. This graph shows the packet loss rate at EC, which is calculated from the used sequence number captured with AirPCap [17]. In other words, the rate of lost sequence number after EC. When the transmission rate is 6 Mbps, about 6% of packets are lost at EC, and the loss rate decreases as the transmission rate increases. As shown in the graph, the packet loss rate at EC is not so significant.

We evaluated how much quality of video data can reach the receiver. The result is shown in Fig. 5. This graph shows the result of each transmission rate. Horizontal axis is the number of background terminals (NS) interfering with the main communication, vertical axis is the packet loss rate of main terminal which is transmitting the main video data. For example, when the transmission rate is 36 Mbps, the main video data will suffer 20% loss with one background terminal (NS), 40% loss with two NS, 60% loss with three NS, about 80% loss with five NS. According to the graph, when the transmission rate is lower than 18Mbps, not all packets reach the receiver even when no background terminal exists. The transmission rate is too low to carry all the packets yielded at the sender PC in this case. When the transmission rate is more than 36Mbps, almost all packets can reach the receiver if no background terminal exists. However, as the number of background terminal increases, the packet loss rate increases dramatically, which means a conflict occurs between the transmitted packets from main terminal and those from background terminals. The examples of packet loss frame data are shown in Fig. 7. In each frame, people are doing the action, "opening the door" and three sequences are selected as the input data. As the packet loss rate increases, the pixels will drop because the packet could not reach to the receiver. We used these packet loss data to input to GTW and its variants and found that GTW is good for aligning multiple sequences, but sensitive to packet loss data.

B. Experimental Result Discussion

We discuss the result of our experiments in this section. As for the experiment method, we applied the project source codes used in [2] which were developed by Feng et al. Input data are captured frames in the Ocha House. We applied pDTW, pDDTW, pIMW, GTW algorithm respectively. For each experiment, we selected three sequences of different people doing the same action. The results are shown in Fig. 8, 9, and 10. For each result, five patterns of packet loss rate, 0%, 20%, 40%, 60%, and 80% are given. The graphs are

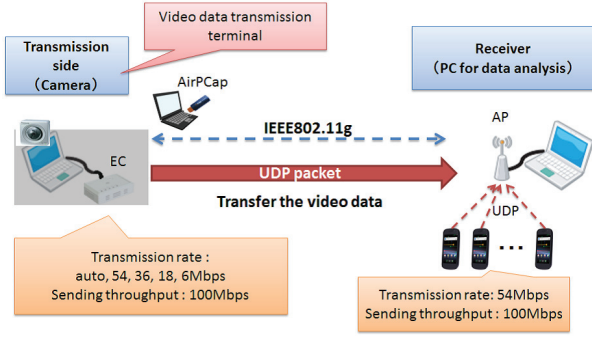


Fig. 4. Experiment environment of data transfer

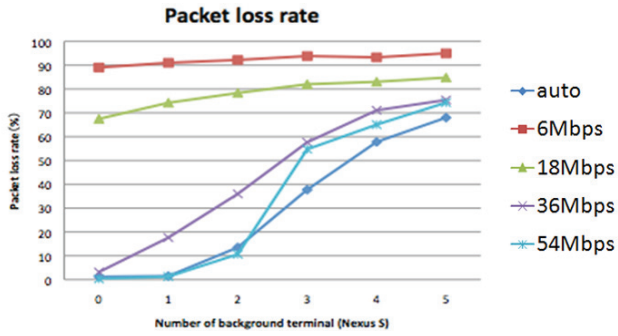


Fig. 5. Result of data transmission interference experiment

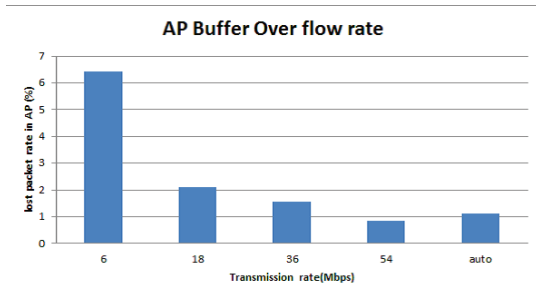


Fig. 6. Buffer overflow on EC

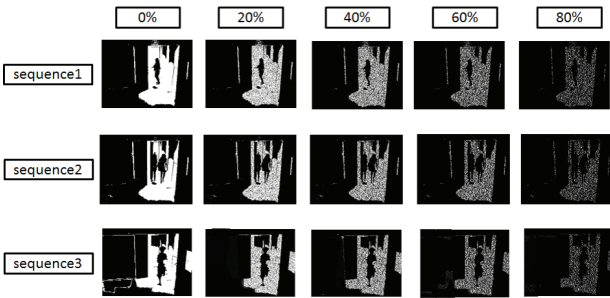


Fig. 7. Examples of packet loss frames

error rate of alignment. In the experiments, we have applied three types of captured input video data to each algorithm: (1) extremely long and short sequences with large size of image matrix, (2) data are smoothed, moderately aligned to almost the same length of sequences with small size of image matrix, (3)

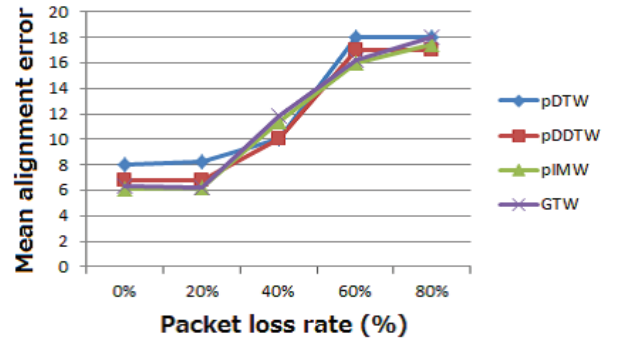


Fig. 8. Big matrix, extremely long and short sequences

all backgrounds of small image matrix data are removed. We will compare the GTW accuracy with other algorithms, pDTW, pDDTW, pIMW. We found that GTW cannot solve the multi-modal temporal alignment problem efficiently in the situations whereby the transmitted data are distorted due to packet loss phenomenon caused by degraded quality in wireless network transmission. Sometimes, pDDTW and pIMW can align the sequences better than GTW.

1) *Big Image Matrix - Extremely Long and Short Sequences*: Big matrix means that many pixels are selected per each frame. Long sequence is composed of many frames that people take longer period of time to do one action, for example, about 50 frames. Short sequences is composed of a few frames that people do one action quickly, for example, about 5 frames. For this experiment, three sequences X_1 , X_2 , and X_3 , which is three videos of people opening the door. The size and length of each matrix X_i is as follows. X_1 : 400×50 , X_2 : 400×18 , X_3 : 400×9 . Note that space 400 of matrix is big enough, and length of three sequences are varied as 50, 18, and 9. As the Fig. 8 shows, the alignment error rate is the lowest when the packet loss rate of input frame is 0%, (which is the perfect data frame without packet loss), but the alignment error rises dramatically as the packet loss rate increases and the error rate when the packet loss rate is 80% is about three times as that of perfect data. GTW alignment error rate is lower than other methods such as pDTW, pDDTW, and pIMW when packet loss rate is low, but when the packet loss rate is 80%, GTW alignment error rate is almost the same as other methods. Especially when the packet loss is 40%, pDDTW error rate is the lowest. Thus, GTW cannot tolerate packet loss data and even be the highest alignment error rate in this experiment.

2) *Small Image Matrix — Smoothed with Almost Same Length of Sequences*: In response to the result of Fig. 8, we reduced the size of matrix from 400 to 12, because the matrix space was too large. A few pixels representative one frame in small matrix. We also aligned the length of sequences to be almost the same and smoothed each frame with a function defined in OpenCV (Open Computer Vision)library [18]. The size and length of each matrix X_i in this experiment is as follows. X_1 : 12×50 , X_2 : 12×48 , X_3 : 12×50 . Note that the space of matrix is small enough (12) and the length is long enough to align in this experiment. As shown in the Fig. 9, GTW and other time warping alignment error rate are lower when the input frame has no packet loss (0%), but when the packet loss rate is 80%, GTW error rate is the worst than any

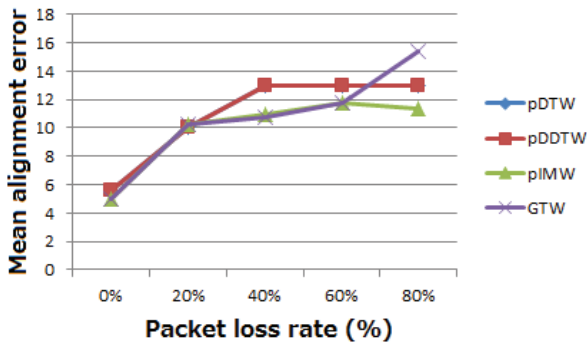


Fig. 9. Small matrix, smooth frames, almost the same length of sequences

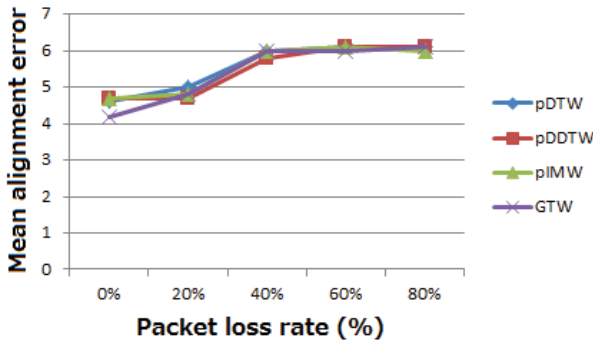


Fig. 10. Without background

other time warping algorithm. pIMW could align the sequences the best when the packet loss rate is 80%.

3) *Small Image Matrix — Without Background*: Lastly, we removed the background of all frames to align the sequences better. Each frame only has the information of people doing motion activity without background in this experiment. The size and length of each matrix are the same as the experiment in Fig. 9. The result is shown in Fig. 10. Even though the whole alignment error rate became lower and alignment error rate didn't rise so much as the packet loss rate increases, alignment error rate of GTW was the lowest only when the packet loss rate is 0% (which is the perfect data without packet loss). When the packet loss rate is higher than 20%, GTW alignment error rate become the same as its variants (pDTW, pDDTW, and pIMW).

VII. CONCLUSION

In our Ocha House, we have collected the video data frames of multiple human subjects with various similar human activities which are transmitted real time over the WLAN network to a processing and analysis server running GTW and its variants (pDTW, pDDTW, pIMW) for temporally aligning multi-modal sequences from multiple human subjects performing similar human activities. The captured video data frames are affected by the natural WLAN packet loss phenomenon due to the wireless network flow control and congestion protocol management in an end-to-end wireless networked environment. We discover that the alignment error rate will rise dramatically and the accuracy of human activity recognition deteriorated as the wireless network packet loss rate increases. When the wireless network packet loss rate increases to 80%, the

alignment error rate of GTW is almost the same as its other time warping variants — in some cases, pDDTW and pIMW can align the sequences better than GTW. Thus, GTW is very sensitive to the quality of the captured video data frames and cannot solve the multi-modal temporal alignment problem efficiently in such situations.

We suggest looking into the problems and weaknesses of GTW characteristics in network packet loss environment as the intuition to help GTW tolerate network packet loss. The weaknesses of GTW in the network packet loss environment are as follows: (1) It cannot align small spaces and shorter length matrices. (2) When aligning bigger matrices, some pixels must be selected as representative values. So, GTW should be able to align longer and bigger matrix features better with appropriate selection representative value. We believe that adjustment of such kind of features to improve captured data quality can help GTW to tolerate network data loss environment.

REFERENCES

- [1] Feng Zhou and Fernando De la Torre, "Generalized time warping for multi-modal alignment of human motion," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [2] Carnegie Mellon University, "Human sensing lab," <http://www.cs.cmu.edu/~ftorres/>.
- [3] P. A. K. Acharya, A. Sharma, E. M. Belding, K. C. Almeroh, and K. Papagiannaki, "rate adaption in congested wireless networks through real-time measurements," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 1535–1550, Nov. 2010.
- [4] Zhongming Zheng, "Constrained energy-aware ap placement with rate adaption in wlan mesh networks," in *IEEE Global Telecommunications Conference (GLOBECOM2011)*, 2011.
- [5] Saeo Iwaki, Tutomu Murase, and Masato Oguchi, "Throughput analysis and measurement on real terminal in multi-rate wireless lan," in *ACM International Conference on Ubiquitous Information Management and Communication (ICUIMC2011)*, 2011.
- [6] Feng Zhou and Fernando De la Torre, "Canonical time warping for alignment of human behavior," in *Neural Information Processing Systems (NIPS)*, 2009.
- [7] Francesca De Simone, Matteo Naccari, Marco Tagliasacchi, Frederic Dufaux, Stefano Tubaro, and Touradj Ebrahimi, "Subjective quality assessment of h.264/avc video streaming with packet losses," *EURASIP Journal on Image and Video Processing* 2011, 2011.
- [8] Simon Moncrieff, Svetha Venkatesh, Geoff West, and Stewart Greenhill, "Multi-modal emotive computing in a smart house environment," *Pervasive and Mobile Computing*, vol. 3, pp. 74–94, 2007.
- [9] Lian Wang, Tao Gu, Xianping Tao, Hanhua Chen, and Jian Lu, "Recognizing multi-user activities using wearable sensors in a smart home," *Pervasive and Mobile Computing*, vol. 7, pp. 287–298, 2011.
- [10] D. Gong and G. G. Medioni, "Dynamic manifold warping for view invariant action recognition," in *International Conference on Computer Vision (ICCV2011)*, 2011.
- [11] I. L. Dryden and K. V. Mardia, "Statistical shape analysis," 1998, Wiley.
- [12] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *SIAM International Conference on DATA MINING (SDM2001)*, 2001.
- [13] K. Pulli E. Hsu and J. Popovic, "Style translation for human motion," *ACM Trans. Graph.*, vol. 24(3), pp. 1082–1089, 2005.
- [14] "Mzk-mf300n," <http://planex.co.jp/product/router/mzk-mf300n/>.
- [15] "Nexus s," <http://www.google.com/nexus/s/features.html>.
- [16] "Iperf," <http://sourceforge.net/projects/iperf/>.
- [17] "Aircap," <http://www.riverbed.com/products-solutions/products/network-performance-management/wireshark-enhancement-products/Wireless-Traffic-Packet-Capture.html>.
- [18] "Opencv library," <http://opencv.org/>.