Activity Recognition for Dogs Using Off-the-Shelf Accelerometer

Tatsuya Kiyohara, Ryohei Orihara, Yuichi Sei, Yasuyuki Tahara and Akihiko Ohsuga

Graduate School of Information Systems, University of Electro-Communications, Chofu-city, Tokyo, Japan

Keywords: Activity Recognition, Accelerometer, Time Series Data Mining, Sensor Data Mining, Acceleration Sensor, Dynamic Time Warping(DTW), DTW-D.

Abstract: Dogs are one of the most popular pets in the world, and more than 10 million dogs are bred annually in Japan now (JPFA, 2013). Recently, primitive commercial services have been started that record dogs' activities and report them to their owners. Although it is expected that an owner would like to know the dog's activity in greater detail, a method proposed in a previous study has failed to recognize some of the key actions. The demand for their identification is highlighted in responses to our questionnaire. In this paper, we show a method to recognize the actions of the dog by attaching only one off-the-shelf acceleration sensor to the neck of the dog. We apply DTW-D which is the state-of-the-art time series data search technique for activity recognition. Application of DTW-D to activity recognition of an animal is unprecedented according to our knowledge, and thus is the main contribution of this study. As a result, we were able to recognize ten different activities with 65.8% classification F-measure.

1 INTRODUCTION

There are services for dog owners that record dog's activity in the form of life logs and report it to them. Examples of the services include the one provided by Whistle Lab's "Whistle" (Whistle Labs, 2013) and NTT docomo's "pet fit" (NTT DOCOMO, 2014). These commercial services recognize raw actions such as "walking", "running", "resting" and "sleeping". These services themselves are evidences of the demand to learn pets' behavior when the owners are away. However, the variety of actions recognizable by the current commercial services is limited and far from being satisfactory. In our analysis, which will be verified in Section 2 by analyzing the results of a questionnaire, there are three aspects of the demand for pet activity monitoring. The first aspect arises from the interest in short-term healthcare. The second aspect originates from the interest in long-term healthcare. The third and final aspect is related to problematic behavior of pets.

For example, vomit reporting is desired for a pet monitoring system because the action is directly related to internal health condition. The action should be detected from the aspect of short-term healthcare.

Eye-scratching is an action that can lead to a serious disease if repeated multiple times. If a pet monitoring system reports the number of times the action is occurred, early treatment by a veterinarian is possible and a serious condition can be avoided. Therefore the detection of the action is desirable from the aspects of long-term healthcare.

In Japan, approximately 70% of the dogs share the life space with human beings. In such circumstances, the pet may exhibit problematic behavior such as biting the furniture and entering the places where it should not, especially in the absence of owner. The owners need to know the problematic actions in order to take appropriate corrective measures, hence the third aspect of the demand for pet monitoring. An action related to this aspect is jumping. It is problematic because it could reflect pets' intention to touch things at higher place, which are kept there by the owners so that the pets could not play with them. Although there is a study on monitoring the actions of a dog, the accuracy of detection for those actions is not high.

In this article, we propose a method to monitor dog's behavior, which is especially effective in the recognition of those actions whose demand of detection is high, according to our analysis of the demands of the owners. The remainder of this paper is organized as follows. In Section 2, we will investigate and analyze a questionnaire to see whether there is a background to the kind of needs. In section 3, we will write about the work related to the activity recognition of the dog and the search technique of time series data. In Section 4, we will present the algorithm

Kiyohara T., Orihara R., Sei Y., Tahara Y. and Ohsuga A..
Activity Recognition for Dogs Using Off-the-Shelf Accelerometer.
DOI: 10.5220/0005212001000110
In *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART-2015), pages 100-110
ISBN: 978-989-758-074-1

for the calculation of Euclid distance, DTW distance and DTW-D distance. Section 5 is dedicated to the description of the experiments. We will describe the experimental environments, experimental procedures, experimental results and discussion. In Section 6, we will discuss the conclusions and recommendations for future works.

2 QUESTIONNAIRE SURVEY FOR NEEDS

2.1 Questionnaire Result

We performed a questionnaire survey with pet owners in order to investigate which actions of the pets should be recognized by a remote pet monitoring system. The questionnaire listed 22 typical actions of the pets and the owners were asked to tell if they were interested in knowing their occurrence when they were away. Furthermore, a free-format comment field was provided to collect the reasons why the owners were interested in knowing those actions. Figure 1 shows the questionnaire results and Table 1 shows the comments filled in the free-format field. The action that gathered the most interest from the owners is vomiting. In addition, the questionnaire result and the free-format description suggest that many owners are concerned with the health condition of their pets.

2.2 Questionnaire Analysis

Let us focus on behaviors in which more than 70% of the owners are interested. “Vomiting” and “shivering” directly reflect the health conditions of dogs, thus their monitoring is desirable from the aspect of short-term healthcare. “Coughing” can suggest respiratory diseases if its frequency is unusually high and so its monitoring is desirable from the aspect of long-term healthcare. “Scratching” can lead to a serious disease if done repeatedly and therefore its monitoring is also desirable from the aspect of long-term healthcare. “Barking”, “chewing”, “drinking”, “eating”, “urinating”, “defecating” and “jumping” are potentially problematic actions. “Barking” could make the neighbors complain. “Chewing” may indicate damage to the furniture. The problem with “drinking” and “eating” is that a pet might eat or drink something that the owner does not want it to. “Urinating” and “defecating” could mean a blunder. With “jumping”, a dog may try to take things at high places. As a result, the eleven behaviors in which more than 70% of owners are interested are related to the three aspects introduced in Section 1.

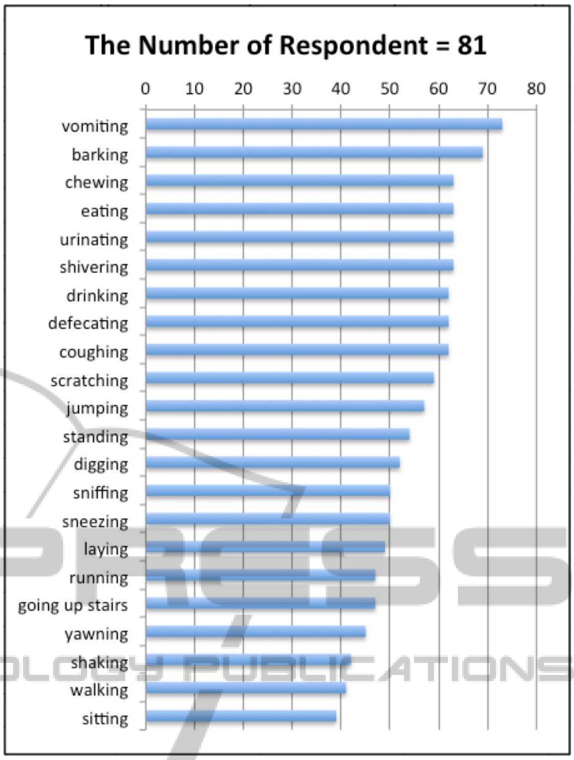

Figure 1: Results of Questionnaire.

3 RELATED WORK

3.1 Activity Recognition for Dogs

There is a study on activity recognition (Ladha et al., 2013). They use PCA-based feature extraction and empirical cumulative density function (ECDF) (Hammerla et al., 2013). They sample acceleration data at 30Hz. The acceleration data are divided into one-second frames and each frame is analyzed separately. A frame has 50% overlap with its predecessor and is created with sliding window procedure based on (Plötz et al., 2010). Each frame is labeled using the movie which is recorded by one annotator. The feature vector of each frame is trained and tested using 10-fold cross validation and is classified in each of the 16 actions and one rejection class using k-NN (k=1). They show the result as a confusion matrix. In their study, jumping in which 70% of the owners are interested is not recognizable. Furthermore, seven actions are with less than 50% recognition accuracies in their study, and recognition accuracy is less than 80% for 12 actions. Therefore, we must say that there is room for improvement in the recognition accuracy.

Table 1: Comments in Free-Format Field.

How much is my dog relaxed? What kind of facial expression does the dog have? When the owner is away, what kind of action does the dog often take?
I keep some cats. I am not worried about the state of my house when I am away, because the cats usually sleep. When I had a dog before, I was worried how the dog was doing. I think it depends on animal species.
It would be nice to talk to a dog at home via a mobile device, when the owner is away.
(Dog) Showing the stomach. Going around. Running around as energetically as possible. Shaking the tail buzzingly. Excited with the sound that promises food items even if they are invisible. (Cat) Making rumbling sound at the throat. Putting the face into a paper or plastic bag. Climbing the curtain when excited. Waiting at the door for a family member to come home. Grooming.
The reaction to the sound of phone calls and intercom during the absence of the owner.
I am concerned if the dog gets into trouble while I am away.
My dog silently vomits without having a cough. That makes it difficult for me to notice the vomiting instantly. I want to notice abnormality as early as possible.
Because my dog is elderly, I am very interested in knowing the behavior of the dog during the absence of my family. In addition, I am concerned if the dog does some action that leads to an illness.
I currently keep my cat in the room. When I go home, the room is so messy that I can imagine what the cat has been doing.
Because the dog spends the daytime alone everyday, I leash the dog. So the range that the dog can move within is narrow. Sometimes my dog can neither jump nor walk. But, I think if the dog spends time without doing any mischievous act, there would be no need of the leash... [in order to realize the situation] it would be nice if the whole of the dog's behavior could be recognized.
I want to see how the dog behaved during the earthquake.
I want to know the action of the dog when it thunders during the absence of me and my family. Because the dog comes to see me to the door when I go home, I want to know when the dog begins to move. Is it when I open the front door, I stop the bicycle, or I open the gate?

3.2 Time Series Data Mining

3.2.1 Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping

There is a problem of finding a subsequence that is similar to a query sequence in a large scale time series data. The problem is solved by calculating the distance between the sequences and the query using Dynamic Time Warping (DTW). However, the computational cost of DTW is high. As time series data, that is to be searched, becomes longer, the number of calculations of DTW increases linearly. As a result, the computational time for the search becomes enormous. The study proposed a method to solve the problem by eliminating unpromising candidates at early stages.

3.2.2 DTW-D: Time Series Semi-Supervised Learning from Single Example

Time series data with little up-and-downs tend to become close to any data in DTW distance. Because of this, a sequence with significant temporal change could be classified as data without one. In order to avoid a situation like this, Chen et al. proposed a distance measure called DTW-D.

4 ALGORITHMS FOR SIMILARITY CALUCULATION

Suppose there are two sequences $X = \langle x_i | i = 1, \ldots, N \rangle, Y = \langle y_j | j = 1, \ldots, M \rangle$. We would like to measure the distance between X and Y in order to measure the similarity of the sequences of X and Y in waveforms. The smaller the distance, the more similar X and Y. Some distances are commonly used.

4.1 Euclidean Distance

Euclidean Distance (ED) is the classic scale for measuring the similarity among the time series data. It is measuring the distance between the time series data of the same length. It is determined by summing up the distances between the data at the same time index.

$$ED = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}\ , \quad (N = M) \tag{1}$$

Figure 2 is a figure of alignment of the Euclid distance.

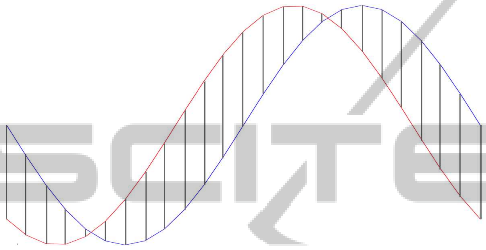

Figure 2: Alignment of the Euclid distance.

4.2 Classical DTW

A weak point of the ED is that it tends to be large when the time series data are out of phase. Dynamic Time Warping is a technique used to find distance more flexibly between time series data than ED. Figure 3 shows alignment of the data points that is used to calculate the DTW distance.

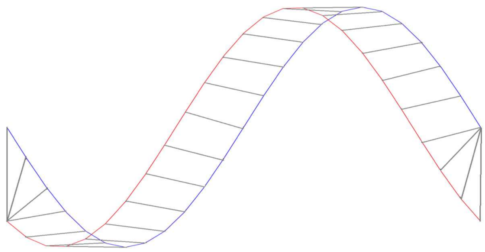

Figure 3: Alignment of the DTW distance.

At first, we calculate cost matrix $C \in \mathbb{R}^{N\times M}$ which is defined by the distance between each element using equation (2). An example of the cost matrix is shown in Figure 4.

$$C(i,j) := c(x_i,\ y_j) = abs(x_i - y_j),$$
$$(i = 1,\ldots,N,\ \ j = 1,\ldots,M) \tag{2}$$

Then, we calculate accumulated cost matrix AC using equation (3). An example of the accumulated cost matrix is shown in Figure 5.

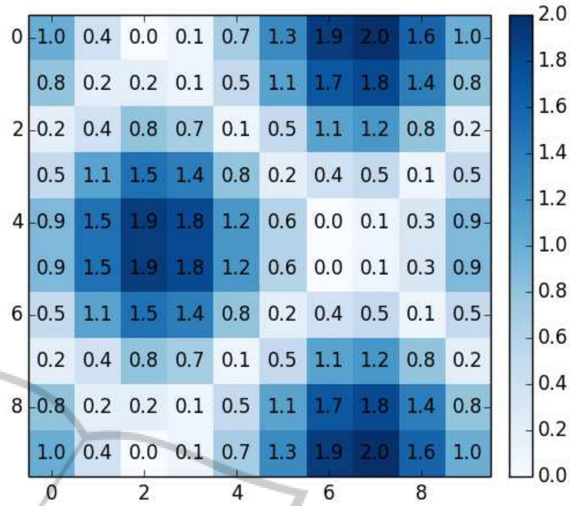

Figure 4: Cost Matrix.

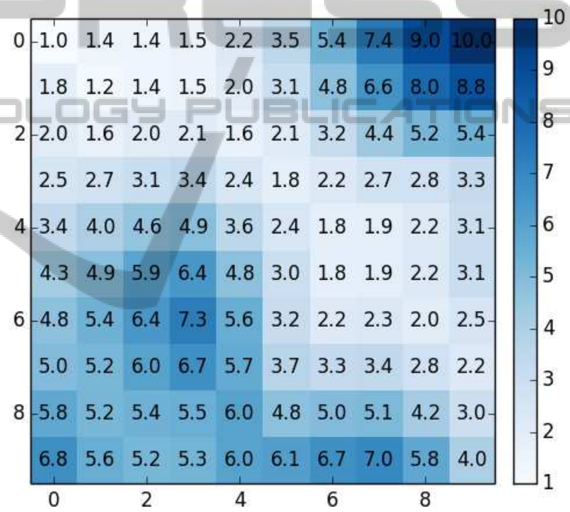

Figure 5: Accumulated Cost Matrix.

$$\begin{aligned}
AC(i,\ 1) &= \sum_{k=1}^{i} c(x_k,\ y_1) \\
AC(1,\ j) &= \sum_{l=1}^{j} c(x_1,\ y_l) \\
AC(i,\ j) &= C(i,\ j) + min\{AC(i-1,\ j-1), \\
&\quad AC(i-1,j), AC(i,j-1)\},\ (i,\ j \geq 2)
\end{aligned} \tag{3}$$

DTW distance is $AC(N,\ M)$. We have shown the above-mentioned algorithm in Algorithm.1.

4.3 DTW-D

Let us consider to calculate the distance between each sequence R, G, B in Figure 6. If we use ED measure, the alignment of data points will be similar to that in Figure 7. On the other hand, if we use DTW measure,

Algorithm 1: Calculate DTW Distance.

Input: sequence $X = \langle x_i | i = 1, \ldots, N \rangle$,
$Y = \langle y_j | j = 1, \ldots, M \rangle$

Output: DTW Distance $AC(N, M)$

/*Calculate Cost Matrix*/
for $i = 1$ to N **do**
  for $j = 1$ to M **do**
    $C(i, j) \Leftarrow abs(x_i - y_j)$
  end for
end for
/*Calculate Accumulated Cost Matrix*/
$AC(1, 1) \Leftarrow C(1, 1)$
for $i = 2$ to N **do**
  $AC(i, 1) \Leftarrow AC(i-1, 1) + C(i, 1)$
end for
for $j = 2$ to M **do**
  $AC(1, j) \Leftarrow AC(1, j-1) + C(1, j)$
end for
for $i = 2$ to N **do**
  for $j = 2$ to M **do**
    $AC(i, j) \Leftarrow C(i, j) + min\{AC(i-1, j-1), AC(i-1, j), AC(i, j-1)\}$
  end for
end for
return $AC(N, M)$

the alignment will be like that in Figure 8. In both cases, the counter-intuitive result that G is more similar to R than B is, will be derived, as shown in Figure 9.

In order to avoid a situation like this, Chen et al. (Chen et al., 2013) proposed a distance measure called DTW-D. DTW-D is calculated by equation (4) where ε is a small positive constant placed in order to avoid the division by zero. As shown in Figure 10, based on DTW-D, B is more similar to R than G is.

$$DTW\text{–}D(x,y) = \frac{DTW(x,y)}{ED(x,y) + \varepsilon} \tag{4}$$

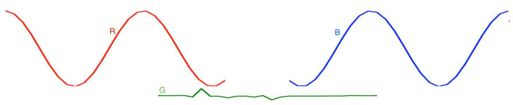

Figure 6: Three sequences to calculate distance.

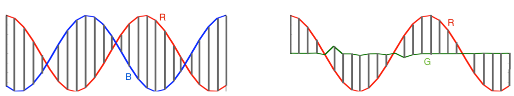

Figure 7: alignment of the ED between R and B, R and G.

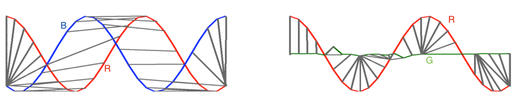

Figure 8: alignment of the DTW distance between R and B, R and G.

ED	R	G	B
R	0	3.4	6.7
G		0	3.5
B			0

DTW	R	G	B
R	0	3.3	3.4
G		0	3.2
B			0

Figure 9: ED and DTW Distance in the three Sequence [R, G, B].

DTW/ED	R	G	B
R	0	3.3/3.4	3.4/6.7
G		0	3.2/3.5
B			0

DTW−D=DTW/ED	R	G	B
R	0	0.97	0.51
G		0	0.91
B			0

Figure 10: DTW-D distance is the distance that DTW distance divided by ED.

5 EXPERIMENTAL PROTOCOLS

5.1 Experimental Environments

The experimental subjects and environments shown in Table 2. The conditions of the experiment are as follows. The acceleration data was collected at sampling frequency 25Hz. The video was recorded in order to put the ground truth label. We prepared the acceleration sensor shown in Figure 11 on the left side. The sensor was attached to the neck of the dog as shown in the Figure 11 on the right side. The acceleration sensor which we used for the experiment is AX3 Watch of Axivity (Axivity Ltd., 2011). The sensor is equipped with 3 axes MEMS which works as an accelerometer. Sampling frequency can be selected from several predetermined values provided by the tool between 12.5Hz and 800Hz. It has a mounted NAND flash memory of 512MB to store the data. The maximum recording time is 14 days at 100Hz, and 30 days at 12.5Hz. Measurement range of the acceleration is ±16g. This sensor has IP68-rated dust- and water-proof capability that is standardized by International Electrotechnical Commission 60529. Acceleration data is transferred to the PC from the sensor by using USB. We have shown the definition of the activity of the dog in Table 3. Figure 12 shows appearance and dimension of the subjects.

5.2 Experimental Procedures

5.2.1 Procedure of Our Approach

Using ELAN (MPI for Psycholinguistics, 2013), which is a tool for video annotation, every sample of the acceleration data is labeled. The label consists of one of 10 activities shown in Table 3. "Unspecified"

Table 2: Information of the experimental subjects and environments.

Name	Yuzu	Oreo
Breed	Pembroke Welsh Corgi	Toy Poodle
Sexuality	Female	Male
Age	4-year	8-month
Weight	10.7kg	2.7kg
Body Length	60cm	43cm
Body Height	30cm	30cm
Environment	Room (Wooden Flooring)	Outdoors (Paved Road or Ground)

Table 3: Definition of Behavior (the number of DTW-D frames & PCA-based frames shown in parentheses).

walking: (2781 & 216 frames) Walking. Movement of the left and right limb is alternating and not aligned.
eating: (5223 & 401 frames) Put food in the mouth, and swallow.
sitting: (4140 & 319 frames) Sitting quietly with buttocks on the floor or ground.
laying: (5746 & 442 frames) Lying down and put his head against a fixed object such as a floor or ground.
sniffing: (370 & 29 frames) Sniffing the smell of the floor or ground.
running: (961 & 74 frames) Running. Movement of the left and right limb is almost aligned.
jumping: (660 & 50 frames) Foot of all is away from such as a floor or ground.
drinking: (1068 & 82 frames) Drinking such as water from the dish on the floor or ground.
shaking: (670 & 50 frames) Shaking itself to shake off the water.
scratching: (47 & 5 frames) Scratching eyes by foreleg. Scratching the front side from the chest by hindleg.

label is given to the behavior that cannot be judged as one of the 10 activities.

Let us call a subsequence of 25 samples a frame. A new frame is created by sliding the 25-sample window forward by one sample. As a result, adjoining two adjoining frames shares 24 samples of each other, that is 96% of the frames. When the same ground truth label appears in more than 20 samples in a frame, that is 80% of the frame, the whole frame is given the ground truth label. This is because, average F-measure became maximum at 80% in 52%$\sim$100%.

Figure 11: The appearance of the sensor and how the dog wears it.

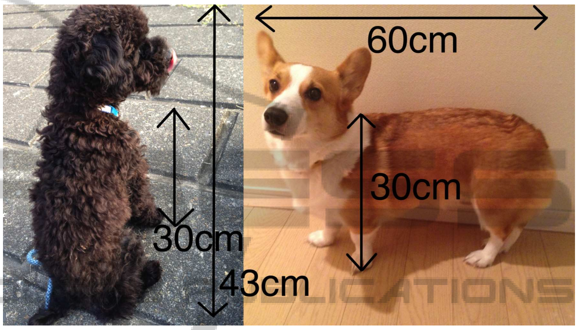

Figure 12: Size of dogs. (left) Oreo, (right) Yuzu.

Otherwise, the frame is labeled "mixed" and used for the test and training frames, but that frame is not listed in a result. We choose one frame from all frames and set it aside as a test frame. Remaining frames is assumed to be training frames. The test and training frames are chosen so that there are no shared samples. The distance between the test and training frames is calculated using each of the Euclidean, DTW, and DTW-D methods. We infer a label of a test frame from the label of the training frame nearest to the test frame. In other words, we used the nearest neighbor method. This is because, in a preliminary study comparing performance of k-nearest neighbor methods for the data, average F-measure became maximum at $k = 1$ in $k = 1,3,5$. Recognition accuracy is calculated through cross validation. Overview of the experimental setting is shown in Figure 13.

5.2.2 Procedure of Existing Approach

The existing PCA-based approach was also applied to our data set. Each parameters were chosen to be as close as possible to the existing study. A new frame is created by sliding the 25-sample window forward by 13 samples. As a result, adjoining two adjoining frames shares 12 samples of each other, that is 48% of the frames. When the same ground truth label appears in more than 19 samples in a frame, that is 76% of the frame, the whole frame is given the ground truth la-

Table 4: Precision, Recall and F-measure at ED, DTW Distance and DTW-D Distance of 2 dogs.

	Precision			Recall			F−measure		
	ED	DTW	DTW−D	ED	DTW	DTW−D	ED	DTW	DTW−D
unspecified	0.6153	0.6260	0.6248	0.5840	0.5986	0.5892	0.5992	0.6120	0.6065
walking	0.5364	0.5218	0.5257	0.3632	0.4311	0.5850	0.4331	0.4721	0.5538
eating	0.8000	0.7968	0.8307	0.8279	0.8273	0.8269	0.8137	0.8118	0.8288
sitting	0.5295	0.5723	0.6658	0.7877	0.7717	0.6833	0.6333	0.6572	0.6745
laying	0.8644	0.8915	0.9517	0.9574	0.9619	0.9495	0.9085	0.9253	0.9506
sniffing	0.5619	0.5873	0.7774	0.6135	0.6270	0.5568	0.5866	0.6065	0.6488
running	0.7469	0.6821	0.5834	0.1904	0.1988	0.5057	0.3035	0.3078	0.5418
jumping	0.8215	0.8038	0.5906	0.3987	0.5152	0.7212	0.5369	0.6279	0.6494
drinking	0.9614	0.9575	0.8281	0.9794	0.9916	0.9925	0.9703	0.9742	0.9029
shaking	0.7008	0.7129	0.6820	0.6328	0.6522	0.7075	0.6651	0.6812	0.6945
scratching	0.1316	0.1091	0.2439	0.1064	0.1277	0.4255	0.1176	0.1176	0.3101

Table 5: Precision, Recall and F-measure at ED, DTW Distance and DTW-D Distance of Yuzu.

	Precision			Recall			F−measure		
	ED	DTW	DTW−D	ED	DTW	DTW−D	ED	DTW	DTW−D
unspecified	0.5930	0.6107	0.5947	0.5797	0.6002	0.5816	0.5863	0.6054	0.5880
walking	0.6554	0.6263	0.5776	0.3681	0.4375	0.6085	0.4714	0.5152	0.5926
eating	0.8901	0.8847	0.8772	0.8660	0.8660	0.8763	0.8779	0.8752	0.8768
sitting	0.5815	0.6182	0.6952	0.8006	0.7872	0.7062	0.6737	0.6925	0.7007
laying	0.8777	0.9033	0.9564	0.9581	0.9629	0.9502	0.9161	0.9322	0.9533
sniffing	0.6381	0.6520	0.7606	0.6457	0.6417	0.5630	0.6419	0.6468	0.6471
jumping	0.8558	0.8738	0.6802	0.4596	0.6046	0.7710	0.5981	0.7147	0.7228
shaking	0.0576	0.1361	0.1322	0.0370	0.0926	0.1481	0.0451	0.1102	0.1397
scratching	0.2381	0.1698	0.2473	0.2128	0.1915	0.4894	0.2247	0.1800	0.3286

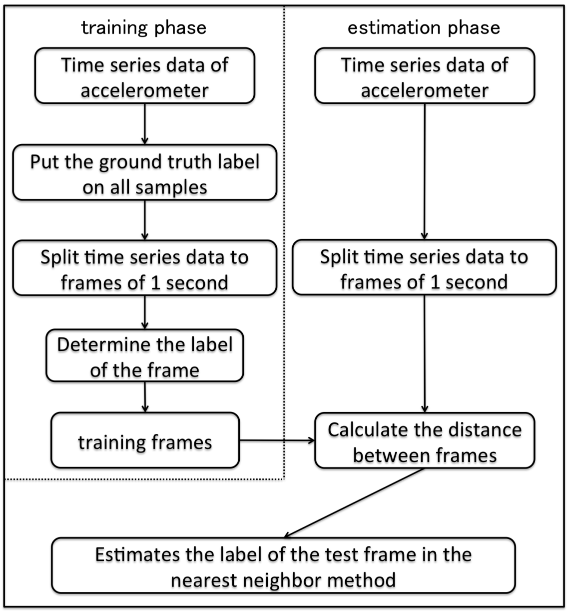

Figure 13: Overview of the experimental setting.

bel. We choose one frame from all frames and set it aside as a test frame. Remaining frames is assumued to be training frames. In some cases part of the test frame and training frames have 48% overlap. Each frame is normalized by inverse ECDF. We projected them into first 25 principal components in order to reduce the dimension of the feature of the frames. The distance between the feature vectors of the test frame and training frames are calculated. The label of the test frame is estimated using nearest neigbor method. Recognition accuracy is calculated through leave-one-out cross validation.

5.3 Experimental Results and Discussion

5.3.1 Analyses of the Result of Our Approach

Table 4 and Figure 14 show accuracies and confusion matrix of our approach obtained through the experiments using data of the two dogs. Table 5 and Figure 15 show accuracies and confusion matrix obtained through the experiments using data of Yuzu alone. Table 6 and Figure 16 show accuracies and confusion matrix obtained through the experiments using data of Oreo alone. In Table 4, 5 and 6, a cell is marked red if the corresponding distance measure gives the best result among the three measures.

According to Figure 4, the DTW-D has yielded high F-measures as compared to the DTW and ED. Using DTW-D, subtle differences in the actions of the dog are recognized more precisely than using the

Table 6: Precision, Recall and F-measure at ED, DTW Distance and DTW-D Distance of Oreo.

	Precision			Recall			F-measure		
	ED	DTW	DTW−D	ED	DTW	DTW−D	ED	DTW	DTW−D
unspecified	0.6602	0.6928	0.7510	0.7271	0.7259	0.6691	0.6920	0.7090	0.7077
walking	0.1707	0.2705	0.3312	0.0808	0.2154	0.4038	0.1097	0.2398	0.3640
eating	0.8142	0.8120	0.8541	0.8648	0.8641	0.8598	0.8387	0.8373	0.8569
sitting	0.1899	0.1628	0.2000	0.3226	0.2258	0.0968	0.2390	0.1892	0.1304
sniffing	0.7079	0.8214	0.8171	0.5431	0.5948	0.5776	0.6146	0.6900	0.6768
running	0.8462	0.8263	0.6319	0.2747	0.3465	0.7180	0.4148	0.4883	0.6722
jumping	0.3333	0.5455	0.4912	0.0851	0.1277	0.5957	0.1356	0.2069	0.5385
drinking	0.9614	0.9575	0.8255	0.9794	0.9916	0.9925	0.9703	0.9742	0.9014
shaking	0.8194	0.8635	0.8956	0.9295	0.9471	0.9824	0.8710	0.9034	0.9370

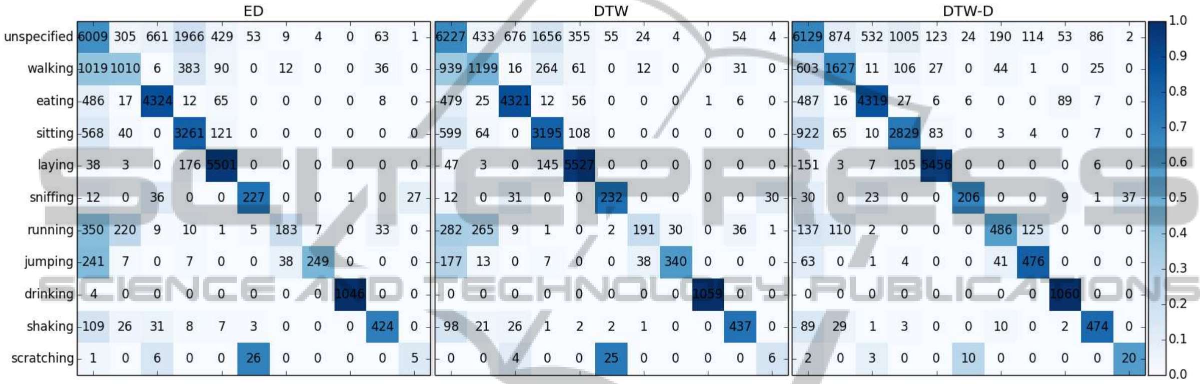

Figure 14: Confusion Matrix of Euclidean Distance, DTW Distance and DTW-D Distance of 2 dogs.

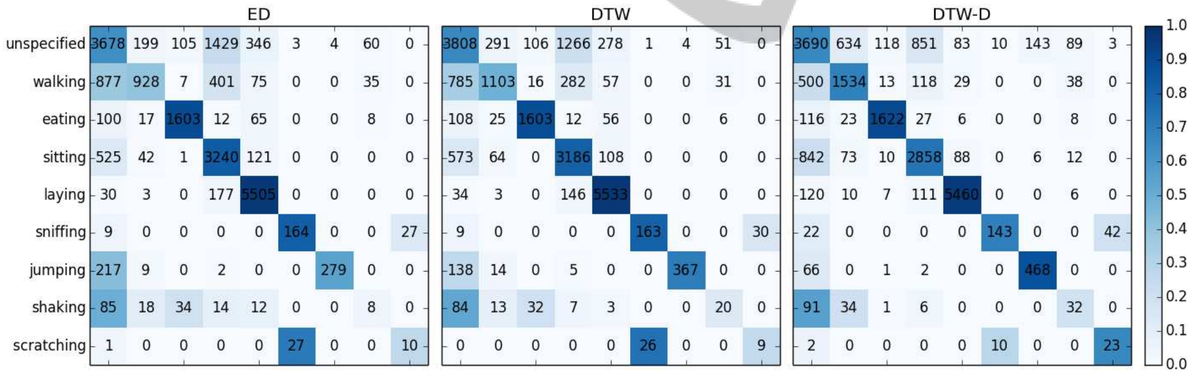

Figure 15: Confusion Matrix of Euclidean Distance, DTW Distance and DTW-D Distance of Yuzu.

DTW and ED. With DTW-D, it is expected that the accuracy of activity recognition will be stable even when the number of actions to be recognized increases. If an action appears only in a part of a frame, such as jumping, it is difficult for a statistical method to detect the difference in the feature value from other actions.

It can be explained that a few samples with significant feature can be obscured by many ordinary samples in a statistical method. On the other hand, the methods that calculate the similarity of waveforms, such as DTW, are able to detect the difference resulting in superior accuracy.

In “drinking”, F-measure of DTW-D is lower than F-measure of ED and DTW. The reason is because a part of “eating” frames are estimated to be “drinking” frames as shown in Figure 14. “Scratching” has resulted an extremely low F-measure. There are only 47 frames of “scratching” in the data. The recall cannot be good simply because of the shortage of the data. The resulted few true positives are further overwhelmed by vast amount of false positives, yielding the poor precision. This can explain the remarkably poor F-measure.

The F-measures from the data of the two dogs roughly falls between those of Yuzu-alone and Oreo-

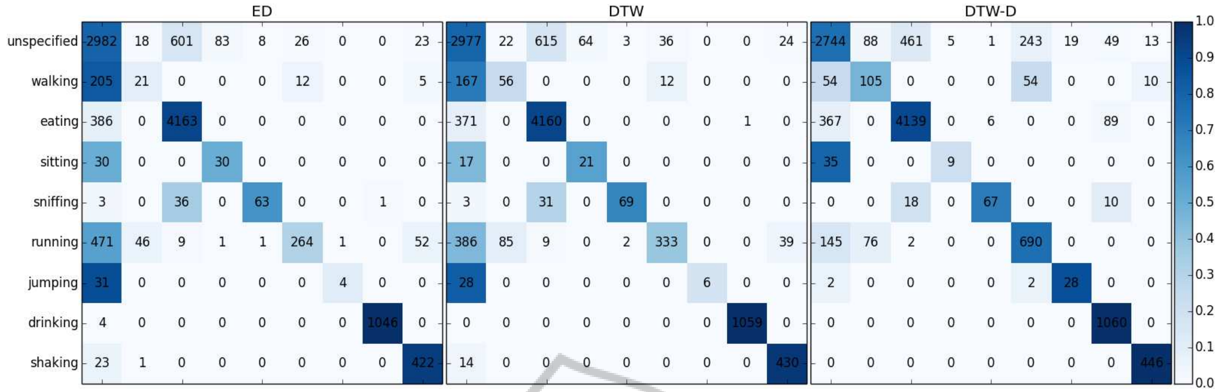

Figure 16: Confusion Matrix of Euclidean Distance, DTW Distance and DTW-D Distance of Oreo.

alone. However, there are some behaviors that has resulted very poor F-measures in the one-dog data, such as "shaking" for Yuzu and "sitting" for Oreo. Improving them is a challenge for the future.

5.3.2 Comparison with the Existing Approach

Table 7 and Figure 17 show accuracies and confusion matrix of the existing approach obtained through the experiments using data of the two dogs. Comparing this with Table 4 and Figure 14, it can be said that our approach has resulted higher F-measures than the existing approach. This could be explained by a theory that the amount of data might be too small to perform the statistical feature extraction. We also think that valuable information of the data could have been lost by the interpolation used in the existing approach.

6 CONCLUSIONS AND FUTURE WORKS

6.1 Conclusions

As seen in the emergence of commercial services that recognize *simple* behaviors of the dogs and to record them as life log, the desire to record the behavior of the dog has been increasing. However, the activity recognition ability of the services is limited and the need to record more detailed actions will arise in the future. We have investigated the needs and analyzed what kind of actions of the dog should be recognized by a pet monitoring system. As a result, we have found that there are three aspects of the demand for pet activity monitoring, namely, short- and long-term healthcare and problematic behavior. The action that the owners wanted to know the most was "vomiting". Furthermore, we have found that there are approximately 70% of the owners who would like to monitor "jumping", whose recognition accuracy was low in the existing study (Ladha et al., 2013). We observed that the reason behind low recognition accuracy of "jumping" was that they used statistical technique in the study. In addition, in our data set, our approach has higher F-measure than the existing approach. Therefore, we focused on the waveform of time-series data. We applied DTW-D, which is a

Table 7: Precision, Recall and F-measure at the existing study.

PCA–based	Precision	Recall	F–measure
unspecified	0.2952	0.3346	0.3137
walking	0.1232	0.1574	0.1382
eating	0.3384	0.2793	0.3060
sitting	0.1740	0.2351	0.2000
laying	0.3100	0.1614	0.2123
sniffing	0.0244	0.0345	0.0286
running	0.0208	0.0270	0.0235
jumping	0.0423	0.0600	0.0496
drinking	0.1429	0.0244	0.0417
shaking	0.0000	0.0000	0.0000
scratching	0.0000	0.0000	0.0000

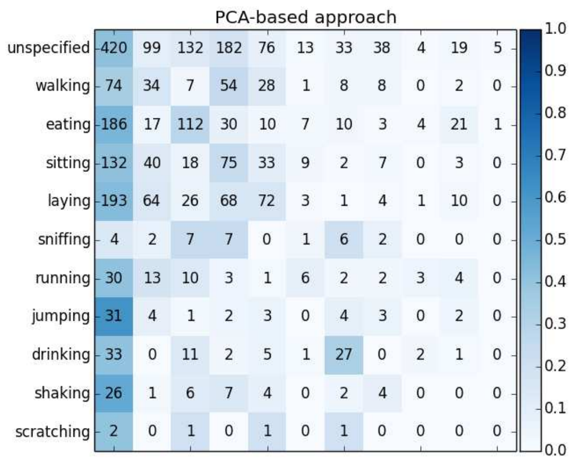

Figure 17: Confusion Matrix of the existing study.

method to measure the similarity of the waveform, to activity recognition in dogs for the first time. As a result, recognition accuracy for “jumping” is particularly improved as compared to the previous study.

It can be said that it is difficult for a statistical method to differentiate an action which appears only in a part of a frame, such as jumping. On the other hand, methods that calculate the similarity of waveforms such as DTW perform well for the actions and result better recognition accuracy. “Vomiting” which is the most desired action to be monitored, is also a brief action. We are optimistic to detect it better with DTW-D.

6.2 Future Works

6.2.1 Measurement of Heart Rate and Respiratory Rate at Rest

By measuring the respiratory rate and heart rate at rest, it is possible to detect the heart or lungs diseases at early stages. It also makes it possible for dogs to receive the appropriate treatment by a veterinarian. For human beings, there is a study by Poh et al. (Poh et al., 2011). However, because this study measures the transition of the reflection of light in the skin, application of this method to dogs with lots of hair is difficult. Therefore, we think that the measurement of heart rate by acceleration sensor is effective.

6.2.2 Pet Location Monitoring in a Room

Whether the behavior becomes problematic or not depends on the place where the pet is kept. If the detailed position of the dog in the room was available, it would further enhance the usefulness of the activity recognition. The research of Paasovaara et al. (Paasovaara et al., 2011) could be a hint. Their study proposed the concept of human-dog interaction with social media. They planned to use a RFID device for indoor position detection as one of human-dog interactions.

6.2.3 Improvement of Recognition Accuracy

There are actions whose accuracy is low in this research and the existing research. We think that further improvement in accuracy becomes an issue. Many small sensors are available now. By the analysis of the behavior with low recognition accuracy, it can be decided what kind of sensor needs to be added. Currently, we are focusing on using sound. We want to improve the accuracy by adding microphone as a sensor in future.

6.2.4 Further Inspection of the Validity of Our Approach

We cannot say that our approach has been sufficiently validated by experiments shown in this paper, both in terms of the number of individual dogs and the variety of breeds. Ultimately we would like to have higher F-measures for any unknown dogs. However, as the first step, we will carry out an experiment using many dogs of the same breed and do cross validation between individuals to verify the robustness of the approach among the same breed.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Numbers 24300005, 26330081, 26870201.

In performing this study, We would like to thank everyone that has helped us questionnaire survey.

REFERENCES

Axivity Ltd., (2011), Axivity Ltd., viewed August 28th 2014, http://www.axivity.com.

Chen, Y., Hu, B., Keogh, E. &Batista, GEAPA (2013), ‘DTW-D: Time Series Semi-Supervised Learning from a Single Example’, in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining , Chicago, IL, USA, pp. 383-391.

Hammerla, N. Y., Kirkham, R., Andras, P. & Ploötz, T. (2013). ‘On Preserving Statistical Characteristics of Accelerometry Data using their Empirical Cumulative Distribution’, in Proceedings of the 2013 International Symposium on Wearable Computers, Zurich, Switzerland, pp. 65-68.

JPFA, (2013), Japan Pet Food Association, viewed August 28th http://www.petfood.or.jp, in Japanese.

Ladha, C., Hammerla, N., Hughs, E., Olivier, P. & Plötz, T. (2013), ‘Dog’s Life: Wearable Activity Recognition for Dogs’, in Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, Zurich, Switzerland, pp. 415-418.

MPI for Psycholinguistics, (2013), ELAN, ver. 4.6.2, Max Planck Institute for Psycholinguistics, Wundtlaan, Nijmegen, Nederland.

NTT DOCOMO, (2014), NTT DOCOMO, INC., viewed August 28th 2014, http://www.docomopet.com, in Japanese.

Paasovaara, S., Paldanius, M., Saarinen, P., Hakkila, J. & Vaananen-Vainio-Mattila, K. (2011), ‘The Secret Life of My Dog Design and Evaluation of Paw Tracker Concept’, in Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services , Bonn, Germany, pp. 231-240.

Plötz, T., Moynihan, P., Pham, C. & Olivier, P. (2010), 'Activity Recognition and Healthier Food Preparation'. In Activity Recognition in Pervasive Intelligent Environments, eds L CHEN, CD NUGENT, J BISWAS & J HOEY, Atlantis Press, Paris, pp. 313-327.

Poh, M. Z., McDuff, D. & Picard, R. (2011), 'A Medical for Non-contact Health Monitoring', in Special Interest Group on Computer Graphics and Interactive Techniques Conference, Vancouver, BC, Canada, Article No. 2.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J. & Keogh, E. (2012), 'Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping', in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, China, pp. 262-270.

Whistle Labs, (2013), Whistle Labs, Inc., viewed August 28th 2014, http://www.whistle.com