An Ensemble Classifier Based on Three-way Decisions for Social Touch Classification

Gangqiang Zhang[1], Qun liu[1], Yubin Shi[1], Hongying Meng[2]

[1]Chongqing Key Laboratory of Computational Intelligence,

Chongqing University of Posts and Telecommunications, Chongqing, 400065, PR China, Chongqing, China

[2]Department of Electronic and Computer Engineering, Brunel University London, UK

Abstract. Social touch is an important form of social interaction. In Human Robot Interaction (HRI), touch can provide additional information to other modalities, such as audio, visual. One of the application is the robot therapy that has great social significance. In this paper, an ensemble classifier based on three-way decisions is proposed to recognize touch gestures. Firstly, features are extracted from on six perspectives and four classifiers are constructed on different scales with different pre-processing methods. . Then an ensemble classifier is used to combine the four classifiers to classify the gestures. The proposed method is tested on the public Corpus of Social Touch (Cost) dataset. The experiments results not only verify the validity of our method but also show the better accuracy of our ensemble classifier.

Keywords: touch gesture recognition, preprocessing, ensemble classifier, three-way decisions

1 Introduction

Touch behavior is one of the important non-verbal forms of social interaction, which can describe the intensity emotions communicated by other modalities [1]. Touch is able to affect the emotions, attitude and social behavior in the communication between humans [2]. As a novel subject, it has drawn growing attention. Humans can understand the meaning of social touch such as emotions. Robot and other interfaces also need to understand social touch. In the social human-robot interaction, touch gesture can be used together with audio-visual cues to improve affect recognition performance [3]. So far, many researches have been carried out to design suitable devices for capturing and classifying social touch to reach the social intelligent interaction. Envisioned applications for these interfaces are like: robot therapy, remote communication and interactive stuffed animal [4-10].

The contributions of this paper is to explore an ensemble classifier based on three-way decisions for the classification of social touch gestures. We use Corpus of Social Touch (CoST) introduced by Jung et al. [5] as our experimental dataset. In order to achieve a higher recognition accuracy, two kinds of preprocessing methods are proposed. The first one depends on the analysis of the procedure of data collection and the other one is inspired by banalization of gray-scale image. Firstly, four base classifiers which are trained on different scale dataset by different preprocessing methods. Then based on three-way decisions, a new ensemble classifier is proposed to recognize touch gestures.

The remainder of the paper is organized as following. Section 2 gives an overview of the related work. In Section 3, methods of data description and processing, features extraction and classification are described. The ensemble classifier based on three-way decisions is illustrated in Section 4. Finally, the conclusion is given in Section 5.

2 Related Work

To spark the further study of social touch, the Social Touch Gesture Challenge was organized in 2015 [11], which focused on the recognition of touch gesture with social meaning performed by hand on a pressure-sensitive surface. Two datasets were given, CoST [5] and Human-Animal Affective Robot Touch (HAART) [12]. The result of challenge was summarized in the 2015 ACM International Conference on Multi-model Interaction (ICMI). In the challenge, Hughes et al. [13] used deep neural networks with hidden Markov models (DNN-HMMS), geometric moments and gesture level features to identify the two datasets, they got 56% accuracy of CoST, 71% accuracy of HAART. Balli Altuglu et al. [14] used image features, Hurst exponent, Hjorth parameters and autoregressive model coefficients as features, they got accuracy from 26% to 95% of CoST and around 60% to 70% of HAART. Gaus et al. [15] used the random forests classifier, and got the accuracy 59% and 67% separately. Ta et al. [16] proposed 273 features and used random forests classifier, got the accuracy for CoST 61.34%, and HRRAT 70.91%. The above 4 papers received top 4 in this challenge.

Three-way decisions is a kind of decision-making model that conforms to human cognition, and it believes that people can make decisions immediately in the process of actual decision-making if they have full confidence in accidence in acceptance or rejection. For those which they cannot make immediate decisions, people tend to postpone the judgment [17]. The essential ideas of three-way decisions are commonly used in everyday life [18] and widely applied in many fields and disciplines, including medical decision-making [19], [20], [21], social judgment theory [22], hypothesis testing in statistics [23], peering review process [24], and management sciences[25], [26].

3 Preprocessing Dataset

3.1 The CoST dataset

grab	hit	pat	slap	massage
tap	tickle	poke	rub	poke
press	scratch	squeeze	stroke	

Figure 1. The 14 touch gestures in CoST

Figure 2. CoST experimental device [REF]

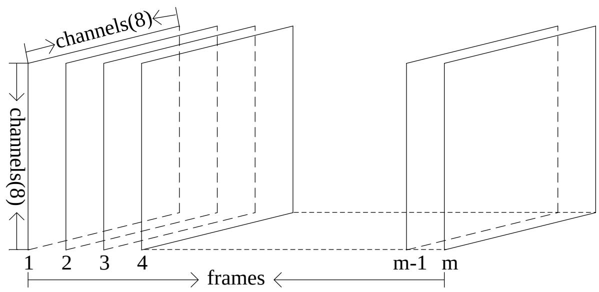

Figure 3. An example of a gesture which consists of m frames and 8*8 grid pressure sensor per frame

The CoST dataset was introduced in [5]. Fourteen different gestures shown in Figure 1are collected through the experiment device shown in Figure 2 [REF]. During the gathering process, the simulated arm is covered with 8*8 grid pressure sensor, and the pressure values of the 64 channels are ranged from 0 to 1023. Before performing gestures, gesture sample video was shown on the computer screen. Then participants pressed the start button and performed, and the end button is pressed if they finished. There are 31 participants' gestures collected in CoST dataset and every gesture was performed 6 times in 3 variations. The dataset provided by the social touch Gesture challenge 2015 only contains two variations, normal and gentle. In this challenge, the organizer divided the dataset into training set (3524 gesture data) and test set (1769 gesture data) randomly. Figurre 3 shows an example of a gesture which consists of m frames and 8*8 pressure matrix per frame.

3.2 Preprocessing methods

As we all know removing the noise from the datasets is beneficial on receiving high quality data. In this section two different preprocessing methods proposed in our work, named "cutout" and "removing backgroud".

"Cutout". After performing a gesture, the participants need to press a key to see next gesture's instruction shown on the computer monitor. So many additional invalid frames may be contained during segmentation between keystrokes and the next gesture. We used the threshold of the maximum pressure sequence of each frame of the gesture as the reference, then truncated the signals bellowing the threshold on the earlier part and later part of each frame. For example, to the complete signal for gesture "hit" shown in Figure 4, it can be seen that the effective signals are only between the red lines, so it is feasible to truncate the extra signals below the reference so as to eliminate the noisy data.

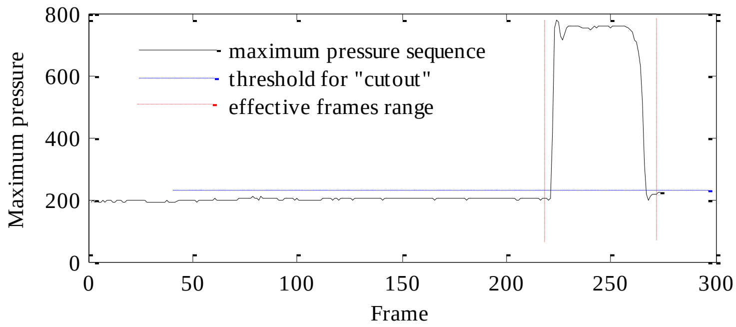

Figure 4. "Cutout" example of the gesture "hit".

In order to find the suitable sequence thresholds for "cutout", we tried the different ratio for the mean, median and maximum of the maximum pressure sequence. For instance, the ratio range for the mean and median are from 10% to 150%, the max range is from 10% to 80%, the step size is 5%. After using random forest classifier to perform the recognition, It was found that the highest recognition accuracy could be obtained using 100%-mean of the maximum pressure sequence as the threshold.

"removing background". Some gestures are always performed too fast or too gently, so the signals captured by the sensor may be very weak and difficult to be classified accurately. As shown in Figure 4, it shows some pressure frames of the gesture "tap". The dark area represents the pressure sensitive area, and the deeper color indicates the greater pressure. In order to enhance the valid part of frames as much as possible, we used the binarization method of gray-scale image for preprocessing. Each pressure value of the channel will be preserved if it is greater than the threshold, otherwise will be set to 0. Similarly, to find the suitable threshold for "removing background", we tried Otsu[?] value, the mean value and 50% maximum for all 64 sensors as the thresholds respectively. Using the random forest classifier to do many experiments, it was found the Otsu value could perform the best as the threshold. Figure 5 shows the comparison for some frames of the gesture "tap" before and after "removing background".

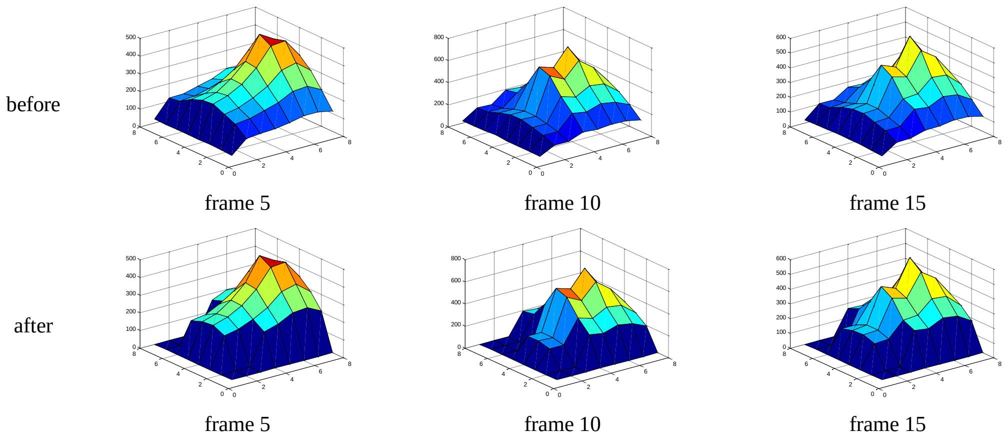

Figure 5. Frames of gestures “tap” before and after “removing background”

3.3 Feature extraction

We extract features from six different perspectives, some of them are referenced the work of other researchers, some are defined by ourselves through analyzing the characteristics of dataset. The total number of features is 331. They are illustrated as follows:

Basic features. This part of the features are defined in [5], including duration, mean pressure, maximum pressure, mean pressure per column, mean pressure per row, pressure variability and displacement. Details can be found in [5]. There are total 24 features.

Histogram-based features. Histogram-based feature extraction is a common method used in image processing field. The whole pressure range (0-1023) is discretized by bins and each pressure for frames of the gestures is put into the bins depending on its value. The number of each bins is used as the features. Siewart et al. evaluated the amount of bins from 2 to 32, and find 8 bins performed best [27].

Sequence features. We use statistical methods to extract the sequence features [16]. We compute statistical metrics of the mean, maximum and sum for the pressure matrix of each frames separately as the features, including maximum, mean, median, mode, range, midrange ($(\max + \min)/2$), variance, standard deviation, coefficient of variation and peak count. Peak count is the number of cross points for the intersection of sequence curve and a certain threshold line, we use the 50% maximum, mean, midrange and median as the thresholds and plot 4 different threshold lines respectively. Hence, there are total 39 features.

Gradient-based features. In order to extract the difference between the channels, we calculate the absolute difference between each single channel and its neighbor channel including horizontal, vertical and diagonal for each frame. Then 210 values of one frame can be obtained and regarded as the gradient of this frame. Firstly, the mean and maximum of these 210 gradient values for each frame are calculated to consist of two sequences, then the statistical metrics like the above are computed on these two sequences as the features. So there are 26 features totally.

Contact area features. It is necessary to extract the contact area for different gestures because of their characteristics. Contact area can be described as the number of channels whose pressures are greater than a certain threshold. We take Otsu value, mean and 50% maximum of each frame as the thresholds to consist of three sequences. Then statistical metrics like the above are computed on these three sequences as the features. In addition, we also take the contact area of the frame with the maximum summed pressure to be the features. There are a total of 42 features.

Channel-based features. Since there are 64 pressure sensors used on simulated skin to capture the variation of gestures, it is important to extract features based on every channels so as to get more precise and complete information. Hence, we captured the mean value of pressure, the variation of mean pressure and the percentage which is the number of pressure points greater than the Otsu threshold of each frame and the total number of frames for a channel. The number of these features are 192.

3.4 Analysis of simple classification results

After using our preprocessing method, 4 different datasets can be gotten, they are the original data set, the dataset preprocessed by "cutout", the dataset preprocessed by "removing background" and the dataset preprocessed by both "cutout" and "removing background". We analysis and extract features from these datasets, then use random forest classifier to classify them by 10-fold cross validation separately. Table 2, 3 and 4 show the recall, precision and F-measure metrics for the above classification results.

From the Table 2, 3 and 4, we can see that the classifying results with "cutout" dataset has the highest mean recognition rate for all gestures. The dataset by "cutout" and "removing background" has higher mean recognition rate. However, the recognition accuracy for "original" dataset and "removing background" dataset are weaker. But from the classification results of the single gesture, "removing background" dataset is good at recognizing gesture "tap". "cutout" preprocess is helpful for recognizing gesture "grab", "hit", "pat", "press", "rub", "scratch", "squeeze" and "tickle", the preprocess of "cutout" and "removing background" can improve the recognition accuracy for classifying gesture "pinch", "stroke" and "tickle".

The above experimental results illustrate that different preprocessing methods are suitable for different gestures' classification. Obviously, every gesture has its own characteristic. For example, the gesture "poke" is a kind of quick gesture, if we preprocess the data by "cutout", the signal of this gesture will be shorter and shorter. Therefore, it is necessary to find an effective method to merge the advantages of different preprocessing methods to recognize different gestures effectively.

Table 1. The recall of 10-fold cross validation on train set.

Recall	original	cutout	removing background	cutout and removing background
grab	80.20%	82.90%	79.00%	**84.90%**
hit	70.60%	73.80%	65.50%	**74.20%**
massage	84.90%	**85.70%**	83.30%	81.70%
pat	53.80%	**54.60%**	51.00%	53.80%
pinch	73.80%	75.00%	74.60%	**79.40%**
poke	**79.80%**	76.60%	77.00%	76.20%
press	71.00%	**73.00%**	68.70%	72.20%
rub	50.00%	**51.20%**	46.40%	**51.20%**
scratch	53.20%	**56.70%**	50.00%	54.80%
slap	**66.70%**	65.90%	61.90%	60.70%
squeeze	50.80%	**55.60%**	47.20%	49.60%
stroke	74.90%	78.10%	75.30%	**83.70%**
tap	52.00%	52.80%	**53.20%**	50.80%
tickle	**71.70%**	67.70%	70.90%	68.50%
mean	66.70%	**67.80%**	64.60%	67.30%

Table 2. The precision of 10-fold cross validation on train set.

precision	original	cutout	removing background	cutout and removing background
grab	68.50%	**70.60%**	67.20%	69.00%

hit	**68.20%**	67.10%	65.00%	64.30%
massage	**83.90%**	82.40%	80.40%	82.00%
pat	59.50%	**62.60%**	55.40%	60.00%
pinch	70.20%	**72.10%**	68.40%	71.70%
poke	**72.60%**	71.00%	70.30%	71.60%
press	74.60%	78.00%	71.20%	**78.10%**
rub	64.30%	**68.60%**	65.40%	67.50%
scratch	**58.30%**	56.70%	54.10%	57.50%
slap	65.40%	65.40%	62.70%	**66.50%**
squeeze	62.10%	**65.40%**	59.80%	**64.40%**
stroke	70.40%	70.00%	68.00%	**71.70%**
tap	53.70%	**53.80%**	**53.80%**	51.60%
tickle	59.00%	**63.90%**	58.90%	63.20%
mean	66.50%	**67.70%**	64.30%	67.10%

Table 3. The F-measure of 10-fold cross validation on train set.

F-measure	original	cutout	removing background	cutout and removing background
grab	73.90%	**76.30%**	72.60%	76.20%
hit	69.40%	**70.30%**	65.20%	68.90%
massage	**84.40%**	84.00%	81.80%	81.80%
pat	56.50%	**58.30%**	53.10%	56.70%
pinch	72.00%	73.50%	71.30%	**75.30%**
poke	**76.00%**	73.70%	73.50%	73.80%
press	72.80%	**75.40%**	69.90%	75.10%
rub	56.30%	**58.60%**	54.30%	58.20%
scratch	55.60%	**56.70%**	52.00%	56.10%
slap	**66.00%**	65.60%	62.30%	63.50%
squeeze	55.90%	**60.10%**	52.80%	56.10%
stroke	72.60%	73.80%	71.50%	**77.20%**
tap	52.80%	53.30%	**53.50%**	51.20%
tickle	64.70%	**65.80%**	64.40%	**65.80%**
mean	66.30%	**67.50%**	64.10%	66.80%

4 An Ensemble Classifier Based on Three-way Decisions

In order to combine the advantages of the above 4 preprocessing methods together, an ensemble classifier based on three-way decisions is proposed in this section. Firstly, we will introduce the theory of ensemble classifier. Then we will describe the three-way decisions theory and discuss the method of calculating threshold in our algorithm based on three-way decisions. The algorithm will be described in detail. Finally the experimental results and analysis will be given in Section 4.5 and 4.6.

4.1 Theory of ensemble classifier

The principle of ensemble classifier is to use several base classifiers to produce their own classification results, then use the vote mechanism to select the best prediction. There are many different ensemble classifiers that are commonly used, such as simple vote [28] and Bayes vote [29]. Simple vote is one of simple combination way according to a certain vote rules. However, simple vote cannot work well because different base classifiers have different accuracy rate. It is necessary to add a weight for those classifiers with better recognition rate, this is principle of Bayes vote. Let $G = \{g_1, g_2, ..., g_{14}\}$

denotes the 14 different gestures, $C=\{c_1,c_2,c_3,c_4\}$ denotes the 4 base classifiers. The weight adopted in our algorithm is defined below:

$$w_{ij}=conf_{ij}\times p_{ij} \tag{1}$$

Where w_{ij} is the weight of base classifier C_i voting to the gesture g_j, $conf_{ij}$ denotes the accuracy rate which classifier C_i can make the right decision for the gesture g_j. In our algorithm, we take F-measure of classifier C_i for the gesture g_j as $conf_{ij}$. p_{ij} denotes the prediction probability that the gesture g_j is predicted correctly by classifier C_i.

4.2 Three-way decisions theory and m-category classification transformation

In order to take advantage of better base classifiers and reduce the influence of worse base classifiers, three-way decisions is introduced. In [30], *U* is supposed as a finite nonempty set and *C* is a finite set of criteria. Three-way decisions is to divide, based on the set of criteria *C*, *U* into three pair-wise disjoint regions, *POS*, *NEG*, and *BND*, called the positive, negative, and boundary regions, respectively. The positive and negative regions can be used to induce rules of acceptance and rejection; whenever it is impossible to make an acceptance or a rejection decision, the third non-commitment decision is made. In [31], the authors reveal the model of three-way decisions as Fig. 6.

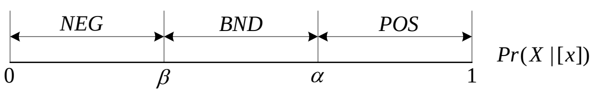

Figure 6. The model of three-way decisions [31?].

$Pr(X\mid[x])$ is the conditional probability of classifying, its definition is as following:

$$\begin{aligned}
&POS_{(\alpha,\beta)}(X)=\{x\in U\mid Pr(X\mid[x])\geq\alpha\},\\
&BND_{(\alpha,\beta)}(X)=\{x\in U\mid \beta<Pr(X\mid[x])<\alpha\},\\
&NEG_{(\alpha,\beta)}(X)=\{x\in U\mid Pr(X\mid[x])\leq\beta\}.
\end{aligned} \tag{2}$$

However, the three-way decisions can be just used to solve the two-category classification problems. In this paper there are *m* (*m*=14) gestures to be classified. So it is necessary to transform *m*-category classification problem to *m* two-category classification problems at first. Referring the method of literature [32] , *obj1, obj2, obj3 are supposed* to belong to one of 3 categories *C1, C2, C3*. The transformation procedure of a 3 two-category classification is shown in Table 4.

Table 4. The transformation procedure with three-way decisions

test data	*C1(*$C1,\neg C1$*)*	*C2(*$C2,\neg C2$*)*	*C3(*$C3,\neg C3$*)*	decision results
obj1	*BND*	*NEG*	*POS*	*C1(BND), C2(NEG), C3(POS)*
obj2	*BND*	*NEG*	*BND*	*C1(BND), C2(NEG), C3(BND)*
obj3	*POS*	*NEG*	*NEG*	*C1(POS), C2(NEG), C3(NEG)*

4.3 Calculating the threshold of three-way decisions

Reasonable thresholds α and β play key roles in solving the decision problems. Yao and Wong [33] introduced the Bayes decision procedure into rough set model, proposed the decision theoretic rough set that gave the theoretical basis of threshold calculation. Jia et al. [34] proposed an adaptive learning parameters algorithm based on the decision theoretic rough set. Their algorithm summarizes the 6 decision risks in three-way decisions, and build the target model of the minimum total risk to search the optimal thresholds α and β .

In our model, we need transform *m* category classification to *m* two-category classification. The target model is not suitable for our problem because it is just used for two category classification. Therefore, the statistical method to calculate the threshold of three-way decisions is put forward.

Firstly, we use Random Forest (RF) in Weka to calculate the prediction probability on train dataset by leave-one-out cross validation [34]. In order to introduce the methods about calculating the thresholds α and β , we give an example on gesture "slap" and list its prediction probability distribution on Table 5.

Table 5. Prediction probability distribution of 4 base classifiers on gesture "slap".

slap		*Q1*	*median*	*Q3*	*mean*	*mode*
original	*TP*	43.00%	54.00%	71.00%	56.00%	52.00%
	FP	33.00%	42.00%	56.00%	44.00%	41.00%
	TN	13.00%	20.00%	32.00%	21.00%	25.00%
	FN	0.00%	0.00%	2.00%	2.00%	0.00%
cutout	*TP*	46.00%	61.00%	74.00%	60.00%	74.00%
	FP	33.00%	41.00%	51.00%	42.00%	31.00%
	TN	12.00%	19.00%	26.00%	18.00%	27.00%
	FN	0.00%	0.00%	1.00%	2.00%	0.00%
removing background	*TP*	43.00%	56.00%	72.00%	57.00%	43.00%
	FP	32.00%	44.00%	55.00%	43.00%	30.00%
	TN	15.00%	22.00%	30.00%	21.00%	22.00%
	FN	0.00%	0.00%	2.00%	2.00%	0.00%
cutout and removing background	*TP*	47.00%	63.00%	80.00%	62.00%	62.00%
	FP	37.00%	42.00%	51.00%	44.00%	41.00%
	TN	12.00%	19.00%	29.00%	20.00%	13.00%
	FN	0.00%	0.00%	1.00%	2.00%	0.00%

Table 5 shows that every corresponding metrics value of *TP* are the largest, but the corresponding metrics value of *FN* are the smallest. It means that the confidence for making the correct decision is larger than the confidence of making the wrong decision. Therefore the threshold α should be selected from the prediction probability of *TP*, and β from *FN* part.

In order to minimize the decision risk and let the correct decisions of classification result be in the *POS* and *NEG* region, and non-commitment decision in *BND* region, the threshold α should cover the *TP* and distinguish from other 3 parts clearly. Similarly, the threshold β should cover *FN* as more as possible and distinguish from the other 3 parts to a large extent. We also observed that the statistical metrics of *FP* and *TN* are separated completely, just the metrics of *TP* and *FN* have some intersections. So the value of the threshold α can be only chosen from the *TP* and *FP*, the threshold β can be selected from the *FN* and *TN*.

Firstly, to the threshold α , regarding the mean and mode of *TP* and *FP* as the references, we can find that the first quartile (*Q1*) of *TP* is greater than the two above metrics of *FP*, but less than the mean and mode metrics of *TP*. So that the first quartile can be used as the threshold α .

Secondly, to the threshold β , we found that just the third quartile (*Q3*) and mean of *FN* can be the valid metrics since the number of negative samples in the experiment is much larger than the positive samples. From the experiment results, we also found that the mean of the *FN* is greater than *Q3*. To reduce the risk of decision making, the third quartile is chosen as β .

Similarly, we also analyzed the experiment results of the other 13 gestures, the conclusions are the same. Finally, we take the first quartile of *TP* as the threshold α and the third quartile of *FN* as the threshold β .

4.4 The model of the new ensemble classifier

In this section, based on three-way decisions, the model of a new ensemble classifier are described in detail that is combined with 4 base classifiers built on different preprocessing methods. Each base classifier uses the same feature set and RF algorithm as their classifiers. It is divided into two steps: (1) Decision of single base classifier: Transform the m-category classification to m two-category classification problems firstly; Then calculate three-way decisions' thresholds for every two-category classification according to the methods discussed in Section 4.3; Next make decisions for m two-category classification depending on the rules of three-way decisions; Finally select a reasonable result according to the procedure described in Table 4; (2) Final prediction decision: Through the step 1, Each prediction result has been divided depending on 3 regions *POS, BND, NEG*, Firstly, compute the weight for the decisions in 3 region *POS* by the Eq.(1), select a class in the region *POS* with the maximum weight as the final prediction result; If no results in the region *POS*, discuss the results in the region *BND* by the same way; Otherwise as well as the region NEG. These two steps are illustrated in Figure 6.

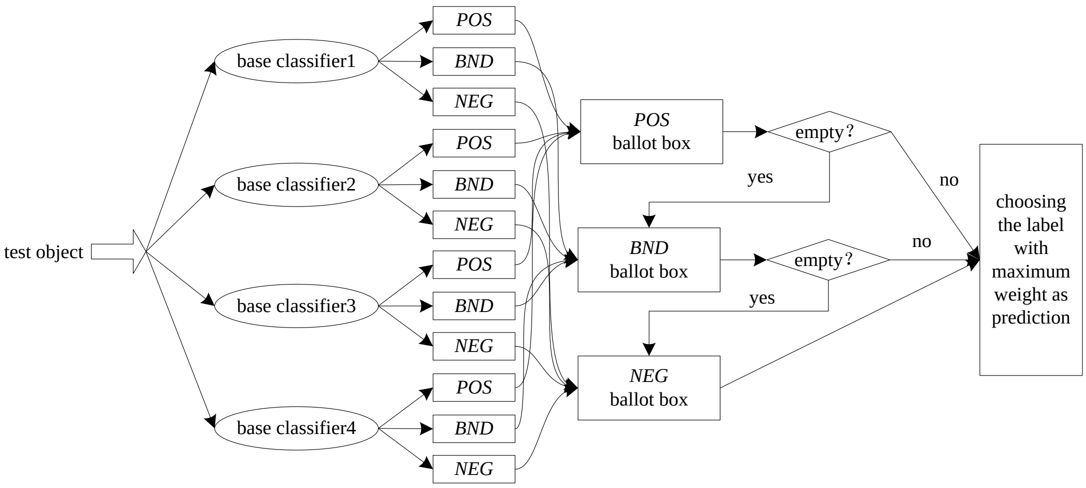

Figure 6. The procedure of vote based on three-way decisions.

Using the RF algorithm, 4 base classifiers are trained based on four datasets preprocessed by different methods. The algorithm in the training phase, the training set is preprocessed and extracted features to get the corresponding 4 datasets. Then 10-fold cross validation is used to get the F-measure of each classifier for each gesture as the *conf* in Eq.(1). Next three-way decisions threshold is calculated for each two-category classification through the method described in section 4.3. During the test phase, the last prediction result is computed inspired by [36] . The algorithm is shown in details as following:

Algorithm 1 A new ensemble classifier.
Input: train set S , test object obj , number of labels M .
Output: prediction result.
Training process:
1. Extract features on train set with different preprocessing to get the 4 data sets $\{S1, S2, S3, S4\}$.
2. Train the 4 base classifiers $\{c1, c2, c3, c4\}$.

```
3.    Get conf_ij .
4.    Calculate the three-way decisions threshold (α, β)_ij of each classifier for each gesture.
Test process:
1.    Preprocess the object obj and extract features to get {obj1, obj2, obj3, obj4} .
2.    Create ballot boxes POS, BND, NEG, Vote contains (label, weight).
3.    List<Vote> POS = new ArrayList<Vote>();
4.    List<Vote> BND = new ArrayList<Vote>();
5.    List<Vote> NEG = new ArrayList<Vote>();
6.    for i = 1 to 4 do    // traverse the 4 base classifiers.
7.        double[] dist = c[i]. distributionForInstance(obj[i]); // get the conditional probability of classifying.
8.        for j =1 to M do    // traverse m two-category classification.
9.            w = dist[j] * conf[i][j];    // calculate vote weight.
10.           According to the rules of three-way decisions to determin the results of base classifiers and then vote.
11.           if ( dist[j] >= alpha[i][j] ) do    // decision region is POS
12.               POS.add(j,w);
13.           end if
14.           if (dist[j] < alpha[i][j] && dist[j] > beta[i][j] ) do    // decision region is BND
15.               BND.add(j,w);
16.           end if
17.           if (dist[j] <= beta[i][j]) do      // decision region is NEG
18.               NEG.add(j,w);
19.           end if
20.       end for
21.   end for
22.   Count the vote in POS, BND, NEG 3 ballot boxes.
23.   If POS ballot box is not empty, then choose the label with maximum summation weight as the prediction.
24.   If POS ballot box is empty, then check the BND ballot box in the same way.
25.   If BND ballot box is empty, then choose the label with maximum summation weight in NEG ballot box as the prediction.
26.   Return the prediction gesture label.
```

To the complexity of the algorithm, its time complexity is analyzed based on the steps of the algorithm at first,: (1) during the preprocess and features extraction, it is $O(N \times R)$, N is the total frames (642431) of the 3524 gesture data, R is the pressure array length (64) of each frame; (2) During the process of training 4 base classifiers, it is $O(N_{tree} \times mtry \times d \times n)$, n is the gesture number(3524), $N_{tree}, mtry, d$ are the parameters of random forest algorism, they are number of decision trees in the forest, using features number when every node of decision tree splits ($mtry$ less than the total features number v =331), and maximum depth of decision tree respectively; (3) When calculating $conf_{ij}$, it is $O(N_{tree} \times mtry \times d \times n)$; (4) When calculating three-way decisions threshold $(\alpha, \beta)_{ij}$, it is $O(N_{tree} \times mtry \times d \times n^2)$; (5) At last, it is $O(np \times M)$ in the test process, np is 4 denoting the kinds of preprocessing method, M is the number of labels describing the 14 gestures. So, it is the calculation of $(\alpha, \beta)_{ij}$ which needs to take the most time, that is means the time complexity of our algorithm is $O(N_{tree} \times mtry \times d \times n)$ and linear. Moreover, its space complexity is $O(v \times n)$.

4.5 The analysis of the experimental results for our ensemble base classifier

Table 6, 7 and 8 show the comparison results between ensemble algorithm and each single base classifier on the metrics, such as recall, precision, and F-measure respectively. From the mean value of these 3 metrics, our ensemble algorithm receives all the best results. For the single gesture, the classification accuracy of gesture “hit”, “massage”, “pinch”, “poke”, “stroke” have been promoted on the recall; the classification accuracy of gesture “pat”, “rub”, “scratch”, “squeeze”, and “tap” have been improved on the precision; on the F-measure, the gesture of “pat”, “poke”, “squeeze”, “stroke”, and “tickle” have been promoted.

But there are still some gestures whose recognition rate are lower than the base classifiers. Because these gestures' data are confused due to the procedure of data collection and preprocess caused by their characteristics. However, most of gestures' recognition rate have been improved.

Table 6. The recall of evaluation on test set.

recall	original	cutout	removing background	cutout and removing background	Ensemble algorithm
hit	64.20%	56.70%	**65.00%**	56.70%	**65.00%**
massage	78.30%	**80.00%**	75.00%	**80.00%**	**80.00%**
pat	41.70%	**45.00%**	43.30%	44.20%	43.30%
pinch	65.00%	65.80%	66.70%	62.50%	**70.80%**
poke	76.70%	72.50%	81.70%	75.00%	**84.10%**
press	**77.50%**	76.70%	74.20%	72.50%	75.00%
rub	**43.30%**	40.00%	33.30%	38.30%	35.80%
scratch	**40.00%**	35.80%	35.80%	38.30%	36.60%
slap	56.70%	**68.30%**	55.80%	60.00%	63.30%
squeeze	38.70%	37.00%	38.70%	**39.50%**	38.60%
stroke	65.00%	66.70%	60.80%	67.50%	**70.00%**
tap	37.50%	38.30%	**40.80%**	26.70%	31.60%
tickle	**73.30%**	65.00%	63.30%	63.30%	72.50%
mean	59.67%	59.07%	57.99%	57.16%	**60.35%**

Table 7. The precision of evaluation on test set.

precision	original	cutout	removing background	cutout and removing background	Ensemble algorithm
grab	51.70%	**52.50%**	50.80%	50.00%	51.90%
hit	**55.00%**	50.00%	50.60%	45.60%	48.40%
massage	**77.00%**	69.10%	66.20%	59.30%	62.30%
pat	54.90%	58.70%	49.50%	54.70%	**61.90%**
pinch	**78.80%**	72.50%	72.10%	72.00%	76.50%
poke	80.70%	**87.00%**	76.60%	79.60%	81.40%
press	68.40%	**74.80%**	70.10%	72.90%	67.60%
rub	58.40%	50.50%	49.40%	53.00%	**62.30%**
scratch	62.30%	53.10%	47.80%	57.00%	**62.80%**
slap	46.30%	56.90%	**58.30%**	52.90%	53.90%
squeeze	58.20%	50.00%	58.20%	49.50%	**60.50%**
stroke	**61.40%**	59.70%	57.50%	58.70%	59.50%
tap	36.60%	40.00%	44.50%	33.70%	**45.70%**
tickle	56.80%	54.90%	57.10%	**62.30%**	57.60%
mean	60.46%	59.26%	57.76%	57.23%	**60.88%**

Table 8. The F-measure of evaluation on test set.

F-measure	original	cutout	removing background	cutout and removing background	Ensemble algorithm
grab	62.00%	**63.10%**	61.40%	60.30%	62.40%
hit	**59.20%**	53.10%	56.90%	50.60%	55.50%
massage	**77.70%**	74.10%	70.30%	68.10%	70.00%
pat	47.40%	**50.90%**	46.20%	48.40%	**50.90%**
pinch	71.20%	69.00%	69.30%	67.80%	**73.50%**
poke	78.60%	79.10%	79.00%	77.30%	**82.70%**
press	72.70%	**75.70%**	72.10%	72.30%	71.10%
rub	**49.80%**	44.70%	39.80%	43.30%	45.50%
scratch	**48.70%**	42.80%	41.00%	45.20%	46.30%
slap	50.90%	**62.10%**	57.00%	56.60%	58.20%

squeeze	46.50%	42.50%	46.50%	43.90%	**47.10%**
stroke	63.20%	63.00%	59.10%	62.80%	**64.30%**
tap	37.00%	39.10%	**42.60%**	30.30%	37.40%
tickle	64.00%	59.50%	60.10%	62.80%	**64.20%**
mean	59.21%	58.48%	57.24%	56.40%	**59.22%**

4.6 Comparison analysis of experimental results

In the Social Touch Gesture Challenge 2015, the highest recognition accuracy results are obtained by literature [13], [14], [15] [16]. In the literature [15], [16], they classified with two different algorithms respectively. In this section, we just compare the higher accuracy one in their works. The classification results are shown in Table 9, 10, and 11 on the metrics of recall, precision, and F-measure between other researches and ours.

From the Tables, it can be seen that Literature [16] gets the highest mean value on the above 3 metrics, but our results are close to it. Just comparing the single metrics, for the recall metric, the gesture “grab”, “hit”, “massage”, “poke”, “press” and “rub” have been improved the recognition accuracy by our algorithm; the gesture “pinch”, “rub”, “scratch”, and “squeeze” are higher than others on the precision metrics, as well as the recognition of gesture “grab”, “pinch”, “poke”, “rub”, “slap”, and “tickle” on the F-measure metric. Generally, the better accurate recognition rate has been obtained compared with the other researches.

Table 9. The recall comparison.

recall	literature[13]	literature[14]	literature[15]	literature[16]	Algorithm 1
grab	69.17%	69.17%	70.00%	66.67%	**78.33%**
hit	50.83%	62.50%	47.50%	61.67%	**65.00%**
massage	64.17%	65.83%	73.33%	75.83%	**80.00%**
pat	37.50%	42.50%	34.17%	**47.50%**	43.33%
pinch	55.83%	63.33%	**75.83%**	65.83%	70.83%
poke	47.50%	75.00%	**86.67%**	82.50%	84.16%
press	62.50%	66.67%	70.00%	73.33%	**75.00%**
rub	30.83%	35.00%	30.83%	34.17%	**35.83%**
scratch	18.33%	42.50%	**60.00%**	46.67%	36.66%
slap	54.17%	48.33%	**72.50%**	53.33%	63.33%
squeeze	19.33%	43.70%	43.70%	**48.74%**	38.65%
stroke	55.83%	64.17%	66.67%	**71.67%**	70.00%
tap	24.17%	34.17%	28.33%	**50.00%**	31.66%
tickle	70.00%	60.83%	61.67%	**73.33%**	72.50%
mean	47.15%	55.26%	58.66%	**60.80%**	60.38%

Table 10. The precision comparison.

precision	literature[13]	literature[14]	literature[15]	literature[16]	Algorithm 1
grab	46.11%	51.23%	**51.85%**	49.68%	48.98%
hit	39.35%	46.29%	44.53%	**52.11%**	50.00%
massage	65.25%	73.14%	77.19%	**83.48%**	59.25%
pat	47.36%	52.57%	61.19%	**62.63%**	61.44%
pinch	71.27%	65.51%	69.46%	75.23%	**77.27%**
poke	67.85%	64.28%	66.24%	**79.83%**	76.51%
press	46.58%	64.00%	**83.16%**	72.72%	72.80%
rub	36.63%	49.41%	62.71%	50.61%	**65.07%**

scratch	37.93%	52.04%	50.00%	60.21%	**66.66%**
slap	40.62%	**52.72%**	45.07%	49.61%	52.08%
squeeze	40.35%	52.52%	50.98%	54.20%	**57.57%**
stroke	44.96%	55.79%	**63.49%**	63.23%	59.31%
tap	33.72%	36.93%	47.22%	**48.38%**	45.67%
tickle	46.40%	57.03%	**60.16%**	56.41%	58.27%
mean	47.46%	55.25%	59.52%	**61.31%**	60.78%

Table 11. The F-measure comparison.

F-measure	literature[13]	literature[14]	literature[15]	literature[16]	Algorithm 1
grab	55.33%	58.86%	59.57%	56.93%	**62.45%**
hit	44.36%	53.19%	45.96%	**56.48%**	55.51%
massage	64.70%	69.29%	75.21%	**79.47%**	70.07%
pat	41.86%	47.00%	43.85%	**54.02%**	50.98%
pinch	62.61%	64.40%	72.50%	70.22%	**73.59%**
poke	55.88%	69.23%	75.09%	81.14%	**82.78%**
press	53.38%	65.30%	**76.01%**	73.02%	71.14%
rub	33.48%	40.97%	41.34%	40.79%	**45.50%**
scratch	24.71%	46.78%	**54.54%**	52.58%	46.31%
slap	46.42%	50.43%	55.59%	51.40%	**58.23%**
squeeze	26.13%	47.70%	47.05%	**51.32%**	47.17%
stroke	49.81%	59.68%	65.04%	**67.18%**	64.36%
tap	28.15%	35.49%	35.41%	**49.18%**	37.43%
tickle	55.81%	58.87%	60.90%	63.76%	**64.20%**
mean	45.90%	54.80%	57.72%	**60.54%**	59.27%

5 Conclusions

A new touch gesture classification method is proposed in this paper. . Firstly, two kinds of data preprocessing methods were proposed to extract features from six perspectives which are Basic features, Histogram-based features, Sequence features, Gradient-based features, Contact area features, Channel-based features. These two preprocessing methods, called “cutout” and “background removing”, are effective to eliminate the interference of noise data for some gestures. Then an ensemble algorithm is proposed to recognize touch gestures on CoST corpora based on three-way decisions. The four base classifiers of this ensemble classifier are the random forests algorithm built on different datasets through different preprocessing methods on CoST corpora. From the analysis of experiment results, the accuracy of touch gesture classifying is improved by our ensemble algorithm. In the Social Touch Gesture Challenge 2015, we know that the best performance is 60.8% from the references, our result is close to it. However, the recognition accuracy is still lower, especially for the gestures "hit", "pat", "rub", "scratch", "slap", "squeeze", and "tap". Although we proposed a statistical method to compute the thresholds of three way decisions and achieved the better results, it still has some drawbacks. In the future, we want to design the reasonable target model of the minimum total risk for the m category classification, then we can compute the better thresholds α and β to obtain the better classification results.

Acknowledgments: This work is supported by ChongQing postgraduate scientific research and innovation projects of China (CYS16161), partly funded by the National Nature Science Foundation of China (61379114) and the Natural Science Foundation of Chongqing (CSTC2014jcyjA40047) and the Doctoral startup project of CQUPT (A2014-20).

Reference

[1] Hertenstein M J, Verkamp J M, Kerestes A M, et al. The communicative functions of touch in humans, nonhuman primates, and rats: A review and synthesis of the empirical research[J]. Genetic, social, and general psychology monographs, 2006, 132(1): 5-94.

[2] Gallace A, Spence C. The science of interpersonal touch: an overview[J]. Neuroscience & Biobehavioral Reviews, 2010, 34(2): 246-259.

[3] Zeng Z, Pantic M, Roisman G I, et al. A survey of affect recognition methods: Audio, visual, and spontaneous expressions[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2009, 31(1): 39-58.

[4] Chang J, MacLean K, Yohanan S. Gesture recognition in the haptic creature[M]//Haptics: Generating and Perceiving Tangible Sensations. Springer Berlin Heidelberg, 2010: 385-391.

[5] Jung M M, Poppe R, Poel M, et al. Touching the void--introducing cost: Corpus of social touch[C]//Proceedings of the 16th International Conference on Multimodal Interaction. ACM, 2014: 120-127.

[6] Nakajima K, Itoh Y, Hayashi Y, et al. A Balloon-Shaped Interface Recognizing Social Touch Interactions[C]//Advances in Computer Entertainment: 10th International Conference, ACE 2013, Boekelo, The Netherlands, November 12-15, 2013. Proceedings. Springer, 2013, 8253: 182.

[7] Naya F, Yamato J, Shinozawa K. Recognizing human touching behaviors using a haptic interface for a pet-robot[C]//Systems, Man, and Cybernetics, 1999. IEEE SMC'99 Conference Proceedings, 1999, 2: 1030-1034.

[8] Tawil D S, Rye D, Velonaki M. Interpretation of the modality of touch on an artificial arm covered with an EIT-based sensitive skin[J]. The International Journal of Robotics Research, 2012, 31(13): 1627-1641.

[9] Silvera-Tawil D, Rye D, Velonaki M. Interpretation of social touch on an artificial arm covered with an EIT-based sensitive skin[J]. International Journal of Social Robotics, 2014, 6(4): 489-505.

[10] Stiehl W D, Breazeal C. Affective touch for robotic companions[M]//Affective Computing and Intelligent Interaction. Springer Berlin Heidelberg, 2005: 747-754.

[11] Jung M M, Cang X L, Poel M, et al. Touch Challenge '15: Recognizing Social Touch Gestures[C]// ACM on International Conference on Multimodal Interaction. ACM on Internation- al Conference on Multimodal Interaction, 2015:387-390.

[12] Yohanan S, MacLean K E. The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature[J]. International Journal of Social Robotics, 2012, 4(2): 163-180.

[13] Hughes D, Farrow N, Profita H, et al. Detecting and identifying tactile gestures using deep autoencoders, geometric moments and gesture level features[C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015: 415-422.

[14] Balli A. T, Altun K. Recognizing Touch Gestures for Social Human-Robot Interaction[C] //Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015: 407-413.

[15] Gaus Y F A, Olugbade T, Jan A, et al. Social touch gesture recognition using random forest and boosting on distinct feature sets[C]//Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015: 399-406.

[16] Ta V C, Johal W, Portaz M, et al. The Grenoble system for the social touch challenge at ICMI 2015[C] //Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015: 391-398.

[17] Liu D, et al. Three-way decisions and granular computing [M]. Science Press, 2013.

[18] Marinoff L. The middle way: Finding happiness in a world of extremes[M]. Sterling Publishing Company, 2007.

[19] Lurie J D, Sox H C. Principles of medical decision making[J]. Spine 24, 1999: 493-498.

[20] Pauker S G, Kassirer J P. The threshold approach to clinical decision making[J]. New England Journal of Medicine, 1980, 302(20): 1109-1117.

[21] Schechter C B. Sequential analysis in a Bayesian model of diastolic blood pressure measurement[J]. Medical Decision Making, 1988, 8(3): 191-196.

[22] Sherif M, Hovland C I. Social judgment: Assimilation and contrast effects in communication and attitude change[J]. Yale University Press, New Haven (1961).

[23] Wald A. Sequential tests of statistical hypotheses[M]//Breakthroughs in Statistics. Springer New York, 1992: 256-298.

[24] Weller A C. Editorial peer review: Its strengths and weaknesses[M]. Information Today, Inc., 2001.

[25] Woodward P W, Naylor J C. An application to Bayesian methods in SPC[J]. The Statistician, 1993: 461-469.

[26] Goudey R. Do statistical inferences allowing three alternative decisions give better feedback for environmentally precautionary decision-making?[J]. Journal of Environmental Management, 2007, 85(2): 338-344.

[27] van Wingerden S, Uebbing T J, Jung M M, et al. A neural network based approach to social touch classification[C]//Proceedings of the 2014 workshop on Emotion Representation and Modelling in Human-Computer-Interaction-Systems. ACM, 2014: 7-12.

[28] Bahler D, Navarro L. Methods for combining heterogeneous sets of classifiers[J]. Machine Learning, 2000, 39(3): 243-276.

[29] De Stefano C, Fontanella F, Scotto d F A. A novel naive bayes voting strategy for combining classifiers[C]// International Conference on Frontiers in Handwriting Recognition(ICFHR). Computer Society Washington, DC, USA: IEEE Press, 2012: 467-472.

[30] Yao Y. An outline of a theory of three-way decisions[C]//Rough Sets and Current Trends in Computing. Springer Berlin Heidelberg, 2012: 1-17.

[31] Liu D, Yao Y Y, Li T R. Three-way decision-theoretic rough sets[J]. Computer Science, 2011, 38(1): 246-250.

[32] Yao Y, Zhao Y. Attribute reduction in decision-theoretic rough set models[J]. Information sciences, 2008, 178(17): 3356-3373.

[33] Yao Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180(3): 341-353.

[34] Jia X Y, Li W W, Shang L, et al. An adaptive learning parameters algorithm in three-way decision-theoretic rough set model[J]. Dianzi Xuebao(Acta Electronica Sinica), 2011, 39(11): 2520-2525.

[35] Bouckaert R R, Frank E, Hall M, et al. WEKA manual for version 3-7-3[J]. University of Waikato, Hamilton, New Zealand, 2008.

[36] Yao Y.Y. Granular Computing and Sequential Three-Way Decisions[M]// Rough Sets and Knowledge Technology. Springer Berlin Heidelberg, 2013:16-27.