

FEATURE EXTRACTION AND CLASSIFICATION OF RAINBOW FISH BASED ON MORPHOMETRIC TRUSS NETWORK BY USING DISCRIMINANT ANALYSIS

IHSAN JATNIKA, SARIFUDDIN MADENDA, ERI PRASETYO WIBOWO, HUSTINAWATI

Gunadarma University, Indonesia

E-mail: {ihsan, sarif, eri, hustina}@staff.gunadarma.ac.id

ABSTRACT

Biodiversity is the strength of an ecosystem. Indonesia is a country that has a diversity of ornamental fish species both marine and freshwater ornamental fish. Maintaining its sustainability by identifying its organisms is important. Identification is important because all subsequent work sequences depend on the correct identification results. Identification is looking for and recognizing individual taxonomic characteristics and incorporating them into a taxon. The high diversity becomes its own obstacle in conducting a fast and precise identification. The computer aided of fish's identification and classification by using image processing technology has been widely carried out. The results of previous studies indicate the identification and classification of the fishes are narrowed only to the stage of distinguishing the characteristics of fish species at the family level. This research tries to identify and classify more deeply to be able to distinguish 3 species in one genus *Melanotaenia* (Rainbow fish).

This study extracts features based on morphometric network truss using distance measurement methods and produces features in the form of the ratio of segment length ratio toward standard length of the Rainbow fish. The learning method used for classification and introduction is the Linear Discriminant Analysis method.

The results showed that the distance measurement method can be used to obtain the characteristics of the Rainbow fish based on morphometric network truss, which is independent toward the fish's size and the magnification of objects in the image. Another result is that the Linear Discriminant Analysis method can produce an accuracy of 70.83% to 75.00% in the classification of Rainbow fish species.

Keywords: *Feature Extraction, Morphometric Truss, Linear Discriminant Analysis, Rainbow Fish*

1. INTRODUCTION

Fish biodiversity is the strength of an ecosystem, which can be characterized through inter-and intra-ecosystem diversity [1]. So that the identification of species that inhabit a waters becomes very important to assess the diversity of an ecosystem. The number of species inhabiting the earth have reached 28,000 species and included in 515 family. Of the 515 families, 9 of them have more than 400 species members, namely Cyprinidae, Gobiidae, Cichlidae, Characidae, Loricariidae, Balitoridae, Serranidae, Labridae, and Scorpaenidae, so that almost 33% exist species are only in nine families [2]. The high species diversity in a family would be an obstacle in quick and precise identification because some overlapping characters are found. Identification is very important because the whole sequence of work

depends on the results of the correct identification of the sample being studied

Fish identification can use morphometric and meristic method, as well as through fish descriptive characteristics. Morphometric is a measurement of fish body parts such as body length, head length, etc. Meristic is the calculation of fish body parts such as number of teeth, number of fin fingers, and others. Descriptive characteristic is observing fish characters such as tail fin shape, color patterns, coloring, and others. When identifying, an identification key is required. Fish characteristics that have been identified through morphometric, meristic, and descriptive measurements will guide a sample in an identification key sequence. From this identification key sequence, a taxonomic group of a sample--from its species to its class--can be known.

Fish identification by using morphometric is based on the fish shape. The shape of the fish is

reconstructed from several reference points that form a line segment called a truss network. The lines represent the skeletons or parts of fish shapes such as the head, body, and tail. The reference points taken are the points where the 2 bones meet. The 10 reference points that form the truss network are (1) the tip of the snout; (2) The tip of the upper head bone; (3) Base of dorsal fin; (4) The back end of the dorsal fin base; (5) the base of the upper tail fin; (6) base of lower tail fin; (7) The back end of the anal fin base; (8) Base of base of the fin; (9) Base of abdominal fin; and (10) The tip of the lower head bone. Discriminant analysis method is used to differentiate fish species based on measured segments of several fish samples [3].

Several research for automatic fish recognition has been developed with image processing methods and artificial intelligence techniques. Alsmadi developed a fish recognition method based on a combination of powerful feature selection techniques, image segmentation, and parameters of geometry, neural networks, and decision trees. This combination produces several features that are used as parameters for the fish recognition process, namely: size, shape, texture, color signature, and geometry parameters. The system is applied to more than 7 families. The system can categorize poisonous or non-poisonous fish and categorize non-poisonous fish into its family. This method successfully provides a high level of accuracy in the classification of 96.4% [4]. Alsmadi also developed a fish recognition method based on color texture measurement, using back-propagation classifier. The method used for feature extraction from color textures is the Gray Level Co-occurrence Matrix (GLCM). The study was conducted on 20 families and successfully extracted 6 features based on color texture measurements with an accuracy of testing reaching 84% [5].

Research on the fish introduction based on the size measurements and the fish shape by using the back-propagation classifier method, that is applied to 20 families and successfully extracted 18 parameters, used the angle and distance measurement method at several anchor point with an accuracy up to 86% [6]. Fish introduction based on the color signature, used the Back-Propagation Classifier method. Feature extraction produces 6 color patterns based on color histograms, and 9 features have been successfully extracted by using a color histogram based on the Gray Level Co-occurrence Matrix (GLCM). This study succeeded in classifying poisonous and non-poisonous fish with an accuracy rate of up to 84% for testing [7].

Pornpanomchai did fish introduction based on the shape and texture by comparing the two recognition techniques, Euclidean Distance Method (EDM) and Artificial Neural Network (ANN). The pre-processing stage includes size adjustment, gray image conversion, conversion to black and white, noise removal, edge detection using Sobel, and object segmentation. Feature extraction produces 8 characteristics that are used for identification, they are: width ratio, length ratio, area ratio, width and length ratio, boundary ratio, red average, green average, and blue average. This study was able to distinguish 30 species with a precision level of 81.67% and the processing time using EDM techniques was faster than ANN [8].

Alsmadi conducted a classification of fish based on extraction of significant feature combinations using Gabor filters, detection of reference location points, statistical measurements of texture and shape measurements. 4 features were extracted successfully using the Gabor filter, 24 features were extracted using angle and distance tools and 2 features were extracted using statistical measurements. The introduction of fish images uses a hybrid metaheuristic algorithm (genetic algorithm with iterated local search) with back propagation classifier (GAILS-BPC). Classify fish into hazardous, harmless, dangerous family predator and poisonous fish and classify harmless fish family species of ornamental fish and fish for consumption.

Based on the analysis result of the studies that have been described it can be concluded that previous studies of some fish morphological features that have been used as a basis for identification using fish descriptive characteristics such as color texture [5], [8], color patterns [7], and morphometrics such as size and shape [4], [8]. The method used for classification is artificial neural network, Euclidean Distance Method. Classification results generally succeed in achieving a high level of accuracy in the range of 81 - 97%. However, the classification results are generally only at the level of distinguishing species from several families, and no one has done the classification to distinguish species in one genus.

The high level of accuracy in previous studies may be due to significant differences in the morphological characteristics of the fish used in the study. Therefore, a study is needed to examine what morphological characters can be used as features to classify fish species that have similar characteristics. In addition, how the feature

extraction process can be carried out for this purpose.

This study tries to extract features based on truss network morphometrics and classify them using linear discriminant analysis (LDA). The aim is to find a feature set and methodology that is able to classify fish species that have similar characteristics. For this reason, this study took 3 species (Figure 1) of fish namely *Melanotaenia ajamaruensis* (ajamaru), *Melanotaenia parva* (kurumoi), and *Melanotaenia yapenensis* (yapen) which are members of 1 genus *Melanotaenia* (Rainbow Fish) as sample.

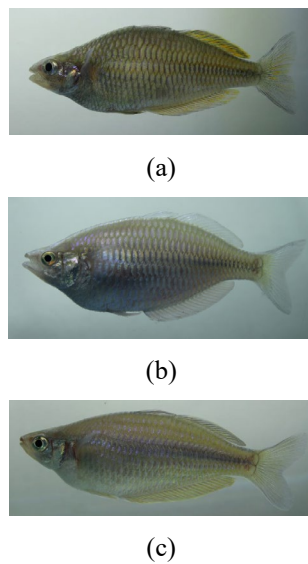


Figure 1. (a) *Melanotaenia ajamaruensis* (ajamaru) (b) *Melanotaenia parva* (kurumoi) (c) *Melanotaenia yapenensis* (yapen)

2. RESEARCH METHOD

Broadly speaking, the research methodology is presented in Figure 2 which consists of 10 stages.

The first stage is the image acquisition of rainbow fish species using a high-resolution camera (5184x3456). The second stage is detecting 10 reference points based on morphometric network truss, which is to get coordinates (x, y) from each reference point. The third step is extracting features to get features that can be used as characteristics of rainbow fish species by calculating the distance of truss line segments, calculating the reference length, and calculating the ratio length toward the reference length to produce a feature dataset in the learning process. The fourth stage is learning to get a Classification Model of the Model Species, using the Linear Discriminant

Analysis method, produces a Classification Model. The Fifth Stage is preprocessing to reduce the image resolution to 800x533 resolution. The sixth step is using the preprocessing results to re-detect the truss reference point. The seventh stage is to extract features and produce feature datasets. The eighth stage is used for the first classification test using the LDA classification model. The ninth step is to re-detect the reference points on different data groups. The tenth stage produces feature extraction back for the feature dataset second classification test by using the LDA classification model. Algorithm implementation using MATLAB 2017 software. We also use statistical software IBM SPSS to analyze the LDA model

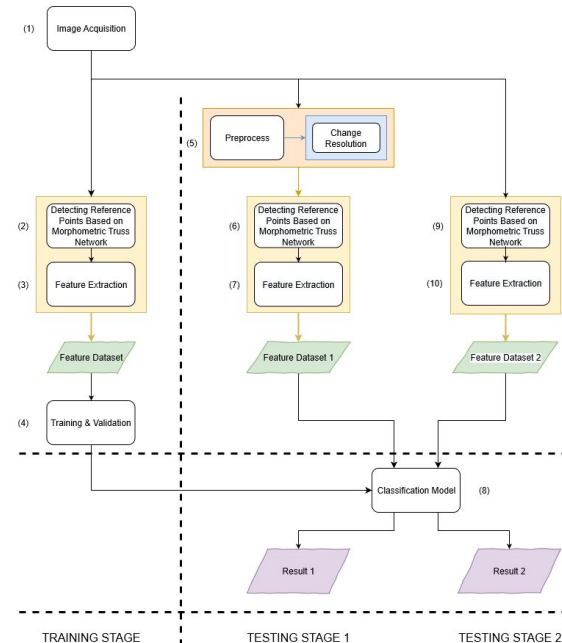


Figure 2. Research Stage

2.1 Reference Point Detection

The reference point based on morphometric truss network is as shown in Figure 3 namely, the tip point of the snout (T1), the tip of the upper head bone (T2), the base of the dorsal fin (T3), the back end of the base of the dorsal fin (T4), the base of the upper tail fin (T5), base of lower caudal fin (T6), back end of anal fin base (T7), base of anal fin base (T8), base of abdominal fin (T9) and tip of lower head bone (T10) of the image of rainbow fish species. Detection of reference points is done manually. The reference point is taken by determining 10 locations, and takes the coordinate values (x, y) from these points using the mouse-click function.

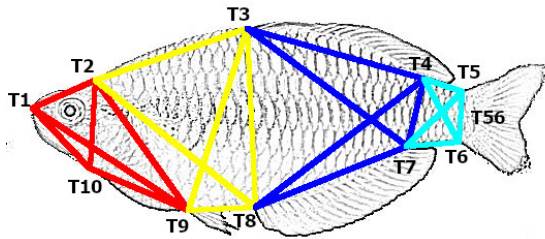


Figure 3. Truss morphometric network reference points for rainbow fish species

2.2 Feature Extraction

Feature extraction is carried out in 3 stages, namely (1) calculating the length of a segment formed from 10 reference points, (2) calculating normal length, and (3) calculating the ratio of segment length to normal length. The segment in question is as shown in table 1. Column Code Section shows the name of the segment formed from the points shown in the Point column. The Truss Characteristics column shows the parts of the fish represented by several segments. For example, sections D1 through D6 represent the shape of the fish head, and so on.

Table 1. Truss Segments (D1-D21)

Truss Characteristic	Code Section	Point
Head	D1	T1 – T2
	D2	T2 – T9
	D3	T9 – T10
	D4	T1 – T10
	D5	T2 – T10
	D6	T1 – T9
Front body	D7	T2 – T3
	D8	T3 – T8
	D9	T8 – T9
	D10	T3 – T9
	D11	T2 – T8
Rear body	D12	T3 – T4
	D13	T4 – T7
	D14	T7 – T8
	D15	T4 – T8
	D16	T3 – T7
Tail	D17	T4 – T5
	D18	T5 – T6
	D19	T6 – T7
	D20	T5 – T7
	D21	T4 – T6

Calculating the length of a segment is done by measuring the distance between 2 points. In algebraic geometry the measurement of the distance between 2 points, T1 (x1, y1) and T (x2, y2) for example, is done using the formula:

$$D = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (1)$$

$$= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

where:

D = distance between 2 points of T

x1 = coordinate value of x first point

x2 = coordinate value of x second point

y1 = coordinate value of y first point

y2 = coordinate value of y second point

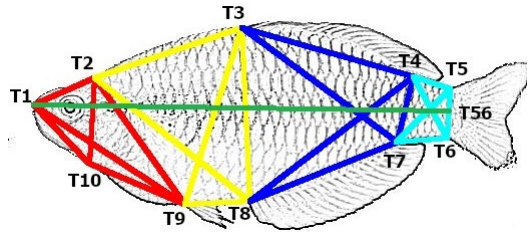


Figure 4. Normal Length (green line)

The next step is to determine the coordinates of the midpoint between point T5 and point T6. This point is named as T56. This point will be used to calculate the Normal Length. Normal Length is the distance between point T1 and point T56 as shown in Figure 4. This Normal Length will be used to get the measurement ratio.

The x coordinate value from point T56 is obtained by calculating the distance difference of x between points T5 and T6 divided by 2. Then, the result is added to the smallest x value between points T5 and T6 or subtracted from the biggest x value between points T5 and T6. Similarly, the y coordinate value from point T56 is obtained by calculating the distance difference of y between points T5 and T6 divided by 2. Then the results are added to the y value from point T5. After obtaining the x and y coordinates for point T56, the normal length (PN) can be calculated.

The last step of the Feature Extraction step is to calculate the Segment Length Ratio toward the Normal Length, which is the length of each segment of the morphometric network truss that has been obtained, proportionate to the Normal Length. This ratio is needed to get the independent features of the fish image toward the fish size and fish object enlargement in the fish image. The ratio is calculated by dividing the length of each section (D1 to D21) by Normal Length (PN). The ratio calculation is formulated as follow:

$$R_i = \frac{D_i}{PN} \quad (2)$$

as:

R_i = the ratio of i-th segment length toward the normal length

D_i = i-th segment length

PN = normal length

2.3 Classification

The classification process is carried out by using the Linear Discriminant Analysis (LDA) method. Discriminant analysis aims to classify an individual or observation into mutually exclusive/disjoint and exhaustive groups based on several independent variables and grouping new objects into the mutually exclusive groups. In addition, this discriminant analysis is also aims to (1) Determine whether there are real differences between the characteristics studied in distinguishing two or more groups; (2) Determine which independent variables make an important / meaningful contribution in differentiating the discriminant average values of two or more groups; (3) Group the data into two or more groups based on the data characteristics studied. The discriminant analysis model regarding linear combinations has the following form.

$$Y = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k \quad (3)$$

as:

Y = discriminant value (related variable)

λ_k = discriminant weighting value for the k-th variable

x_k = the k-th free variable

The steps for using the LDA method briefly can be seen in the following algorithm [9]:

1. Given a set of N samples $[x_i]_{i=1}^N$, each of which is represented as a row of length M .
2. Compute the mean of each class $\mu_i(1 \times M)$.
3. Compute the total mean of all data $\mu(1 \times M)$.
4. Calculate between-class matrix $S_B(M \times M)$.
5. Compute within-class matrix $S_W(M \times M)$.
6. The eigenvalues (λ) and eigenvectors (V) of W are then calculated.
7. Sorting eigenvectors in descending order according to their corresponding eigenvalues. The first k eigenvectors are then used as a lower dimensional space (V_k)
8. Project all original samples (X) onto the lower dimensional space of LDA.

3. RESULT AND ANALYSIS

The training data used in this study were 24, with the composition as stated in table 2.

Table 2: Types and Amount of Training Data Imagery

Fish Types	Number Of Imagery
Ajamaru	6
Kurumoi	8
Yapen	10
Total	24

The test data in this study uses the same image as the test data but is divided into 2 groups of datasets. The first dataset is the same dataset as the training data, but preprocessing has been carried out by reducing the image resolution to 800x533, and the feature extraction process is repeated. The second dataset is the same dataset as the training data, but the image is not preprocessed so it is still in 5124x3456 resolution.

The result example of training data feature extraction can be seen in table 3 at the end of this article. The data displayed are only 6 training data sets. A training data set contains 21 values of segment length ratio toward the normal length of 1 fish image. The Segment field shows the segment name. The Fish Data column contains the calculated ratio values by using the formula 2 above, for each type of fish from the training data.

3.1 Training

After conducting discriminant analysis to the data from feature extraction results using statistical software, the *Canonical Discriminant Function* coefficient (table 4) is obtained, and the Median functional values of each population (Functions at Group Centroids) is as follows (table 5).

Table 4. Canonical Discriminant Function coefficients

Variable	Function	
	1	2
R1	95.368	91.967
R2	-106.733	-20.651
R3	-44.064	-50.990
R4	272.717	89.961
R5	98.282	2.714
R6	-285.099	-70.985
R7	-76.355	-98.650
R8	307.645	-119.942

Variable	Function	
	1	2
R9	-17.790	-68.312
R10	-278.777	113.584
R11	-25.165	240.164
R12	0.000	0.000
R13	0.000	0.000
R14	-76.026	77.268
R15	0.000	0.000
R16	0.000	0.000
R17	0.000	0.000
R18	0.000	0.000
R19	0.000	0.000
R20	0.000	0.000
R21	0.000	0.000
(Constant)	119.289	-79.811

This Canonical Discriminant function is a linear function that includes coefficients for each segment of the truss network. This function will be used for fish species recognition.

From the table above, we can form the linear function as follows:

$$F1 = 95,368R1 - 106,733R2 - 44,064R3 + 272,717R4 + 98,282R5 - 285,099R6 - 76,355R7 + 307,645R8 - 17,790R9 - 278,777R10 - 25,165R11 - 76,026R14 + 119,289$$

$$F2 = -79,811 + 91,967R1 - 20,651R2 - 50,990R3 + 89,961R4 + 2,714R5 - 70,985R6 - 98,650R7 - 119,942R8 - 68,312R9 + 113,584R10 + 240,164R11 - 77,264R14$$

Table 5. Median functional values of each population

Population	Function	
	1	2
1.00	-0.361	3.760
2.00	6.784	-1.623
3.00	-3.782	-2.034

Table 5 shows the median value of each fish group. From these median (centroids), the cutting score can be calculated. Cutting score is the value that will be the limit to determine the group from the

calculation results of F1 function. The graph in Figure 5. shows the position of the median value and the position of the extracted image feature data for each fish species. The graph also shows that on the horizontal axis population 1 (Kurumoi) is quite close to population 3 (Yapen).

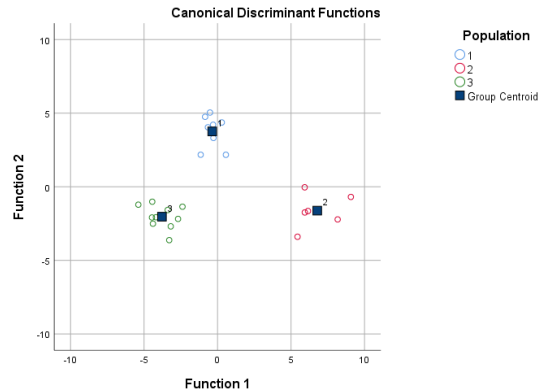


Figure 5. Plot Graph of Training Data

3.2 Testing

The first test was conducted on the first group test dataset, which is a dataset of the same image used in the learning phase. However, the image used has been pre-processed beforehand, resulting in a resolution of 800x533, which is smaller than the original size. The first test results contained 7 errors from 24 datasets. The second test was conducted on the second group dataset, which is a dataset of images with the same resolution as the training images. The second test result contained 6 errors from 24 datasets. The overall results of this testing phase can be seen in table 6.

Table 6: Testing Results

Test Dataset	Total Amount Of Test Dataset	Fish Classification Test Results	
		Error	Accuracy
Group 1	24	7	70,83%
Group 2	24	6	75,00%

4. DISCUSSION

The difference between ours and the previous studies is that we used fish species with similar characteristics, which are in the same genus in the taxonomic hierarchy. It is intended that the classification method used for objects that have similar characteristics is expected to work effectively on those that do not have similarities.

In previous studies such as [6], fish recognition was carried out based on shape measurements using

18 reference points to obtain 10 distance measurement features, and 5 angle measurement features. We used 10 reference points to extract 21 morphometric features in terms of the ratio of the distance between the morphometric segments to the normal length of the fish. In addition, the distance ratio maybe able to substitute the use of descriptors as in [3]. However, those 10 reference points are still obtained manually rather than automatically detected.

5. CONCLUSION

The experimental results show that the methodology we use has been able to classify 3 species of fish in the same genus that have similar characteristics. We have extracted twenty-one features from ten truss morphometric reference points. All these features are the result of measuring the size and shape of the fish in the form of the distance ratio between the truss morphometric network segments and the normal length of the fish image. We propose a truss morphometric network as a novel feature set that can be used by fish classifiers. The level of accuracy obtained using LDA method has only reached 75%.

FUTURE RESEARCH

From this study, it appears that the level of classification accuracy needs to be improved. Other classification methods such as neural networks need to be tried to increase the level of accuracy. The next research is to try using a fully automatic approach to detect the reference point and extract the rainbow fish species features.

REFERENCES:

- [1] Committee on Biological Diversity in Marine Systems, "Understanding marine biodiversity. a research agenda for the nation. 114 pp." National Academy Press, Washington, D.C., 1995.
- [2] Nelson. JS., *Fishes of the World* (4th ed), 4th ed. John Wiley & Sons-New Jersey, USA, 2006.
- [3] Strauss, Richard E., F. L. B. *The Truss: Body Form Reconstructions in Morphometrics Systematic Zoology*, Oxford University Press, Society of Systematic Biologists, Taylor Francis, Ltd., 1982, 31, 113-135
- [4] M. K. S. Alsmadi, K. B. Omar, S. A. Noah, and I. Almarashdah, "Fish recognition based on the combination between robust feature selection, image segmentation and geometrical parameter techniques using artificial neural network and decision tree," *CoRR*, vol. abs/0912.0986, 2009.
- [5] M. K. Alsmadi, K. B. Omar, S. A. Noah, and I. Almarashdeh, "Fish recognition based on robust features extraction from color texture measurements using back-propagation classifier," 2010.
- [6] M. K. Alsmadi, K. B. Omar, S. A. Noah, and I. Almarashdeh, "Fish recognition based on robust features extraction from size and shape measurements using neural network," *Journal of Computer Science*, pp. 1088–1094, 2010.
- [7] M. K. Alsmadi, K. B. Omar, and S. A. M. Noah, "Fish classification based on robust features extraction from color signature using back-propagation classifier," *Journal of Computer Science*, vol. 7, pp. 52–58, 2011.
- [8] C. Pornpanomchai, B. Lursthut, P. Leerasakultham, and W. Kitiyanan, "Shape- and texture-based fish image recognition system," *KASETSART JOURNAL: NATURAL SCIENCE*, vol. 047, no. 4, pp. 561 – 570, Jul-Aug 2013. [Online]. Available: http://kasetsartjournal.ku.ac.th/~kuj_files/~2013/~A1310041317311406.pdf
- [9] Tharwat, A.; Gaber, T.; Ibrahim, A. & Hassanien, A. E., "Linear discriminant analysis: A detailed tutorial", *Ai Communications*, vol 30, pp. 169-190, 2017.

Table 3. Feature Extraction Results

Segment	Fish Data					
	1 st Data (Kurumoi)	2 nd Data (Kurumoi)	9 th Data (Ajamaru)	10 th Data (Ajamaru)	15 th Data (Yapen)	16 th Data (Yapen)
R1	0.221909	0.230124	0.157635	0.17842	0.207459	0.21243
R2	0.193177	0.198682	0.145877	0.184242	0.17177	0.164071
R3	0.362769	0.35995	0.332213	0.329955	0.298557	0.299531
R4	0.221622	0.204543	0.233023	0.216158	0.189043	0.181698
R5	0.236186	0.24241	0.164427	0.194832	0.186773	0.189276
R6	0.432629	0.425687	0.387886	0.391516	0.388234	0.385178
R7	0.333441	0.328923	0.360764	0.318127	0.337752	0.329858
R8	0.403959	0.404868	0.399016	0.393316	0.341358	0.345568
R9	0.161266	0.165971	0.141204	0.131671	0.153777	0.14933
R10	0.433444	0.430255	0.423599	0.41492	0.379114	0.378142
R11	0.459563	0.462202	0.430372	0.412489	0.40263	0.400721
R12	0.393552	0.396435	0.413324	0.418315	0.377979	0.407065
R13	0.175207	0.176156	0.167871	0.170154	0.144902	0.143902
R14	0.373042	0.363051	0.405904	0.40722	0.420216	0.42794
R15	0.469769	0.46892	0.494071	0.498896	0.473102	0.490422
R16	0.462787	0.458005	0.474428	0.474463	0.440029	0.455794
R17	0.13282	0.132849	0.14372	0.147592	0.149941	0.129152
R18	0.116575	0.115475	0.109396	0.111121	0.100403	0.111878
R19	0.125592	0.132191	0.140319	0.135575	0.110385	0.118765
R20	0.185869	0.196313	0.19508	0.200603	0.170346	0.171928
R21	0.199262	0.19304	0.197509	0.193705	0.182548	0.182622