Motion analysis for broadcast tennis video considering mutual interaction of players

Naoto Maruyama, Kazuhiro Fukui
Graduate School of Systems and Information Engineering
University of Tsukuba, JAPAN
maruyama@cvlab.cs.tsukuba.ac.jp, kfukui@cs.tsukuba.ac.jp

Abstract

In this paper, we propose a new scheme of player recognition based on Cubic High-order Local Auto-Correlation (CHLAC) features. To achieve a high classification rate based on CHLAC features, various types of CHLAC features should be used, which are generated by controlling their parameters. However, some CHLAC features are unreliable for classification. To find the best CHLAC features, we apply the AdaBoost algorithm. Further, we add the information on the opposing player for classification enhancement. There are strong interactions between the actions of two competing tennis players, and thus it is effective to utilize their correlation for classification. Our approach of considering two player's interactions achieved a classification success rate of 96.09%, which is much more accurate than methods using only the target player's information, with which the classification success rate is 85.02%.

1 Introduction

Broadcast sports programs are often watched for more serious reasons than entertainment. For example, some people learn body movements from players in videos, and some people watch the video to analyze players' strategy. Given the variety of purposes for watching broadcast sports programs, there is a great demand for technology to obtain information corresponding to the viewers' desires. Accordingly, many methods and technologies for automatic sports video content analysis have been proposed[1, 2, 3]. In our research, we deal with broadcast tennis video, which is widely viewed. From the various possible factors, we focus upon the actions performed by players during play. The result can be used for various purposes such as event detection and strategy analysis.

The action of players in broadcast tennis video has been analyzed via various approaches. Miyamori et al. [4] recognized player action by using the silhouette of the player and the relationship between the position of the ball and player. In the research of Zhu et al. [5], a motion descriptor based on optical flow was used, and high performance was achieved in action recognition. However, these methods depend on the performance of object tracking algorithms, which are difficult to apply to broadcast tennis video because of the low resolution of the image frames and the complex movement of the target object. Further, more accurate action classification is required for detailed scene understanding.

In our method, CHLAC features[7] are used to describe the movement of a tennis player. Unlike existing works[4, 5], segmentation of the player is unnecessary before feature extraction, owing to the position invariance of CHLAC features. We obtain a variety of CHLAC features by changing their parameters, which produces an effect similar to using multiple image resolutions and multiple speeds of player action. By using many types of CHLAC features together, we can perform the classification considering both changes of video resolution and motion phase[8]. However, some CHLAC features do not contribute to the classification performance. To improve the performance, we need to select CHLAC features useful for classification. Accordingly, the AdaBoost algorithm[9] is applied to select useful CHLAC features efficiently.

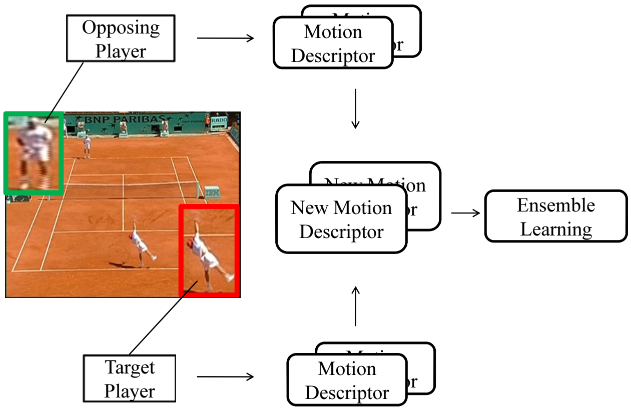


Figure 1. Basic idea of the proposed method.

For further improvement of the classification, we utilize information on the opposing player. There is very strong correlation between the behaviors of a target player and an opposing player in a singles tennis match, and thus this correlation can be exploited for classification enhancement. The whole process flow of the proposed method is shown in Fig. 1. CHLAC features are extracted from the opposing player in the same manner as from the target player. Then, new motion descriptors are obtained from all the possible combinations of connecting pairs of each player's CHLAC features. Finally, the useful new motion descriptors are selected by the Adaboost algorithm. In this case, Adaboost can find the best combination of several temporal-spatial parameters for integrating multiple CHLAC features, which is equivalent to the combination of the time and scale suitable to represent the information about the interaction between the two players.

The main contributions of this paper are as follows. 1) We propose a new framework of tennis player action recognition based on the selection of CHLAC features for classification by the Adaboost algorithm. 2) Information on opposing player is also utilized to consider the mutual interaction between two players for the improvement of classification.

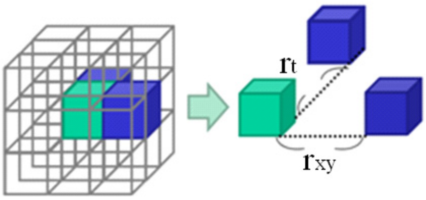


Figure 2. Parameters of temporal-spatial scales.

2 Concept of proposed framework

In this section, we propose a new framework of action recognition considering mutual interactions between two players. First, we explain CHLAC features as descriptors of player motion. Then, we describe the way to select useful CHLAC features with AdaBoost. Finally, we present an approach to integrating the information on both the target player and opponent for classification.

2.1 Motion descriptor using CHLAC features

There are many types of features to represent spatial-temporal information, such as the action of a tennis player. In this paper, we focus on Cubic High-order Local Auto Correlation (CHLAC) features to represent the action of a tennis player, since they are know to be among the most suitable features to represent human motion without high computational complexity. Further, CHLAC features are position invariant, with the value of the feature vector being unaffected by the position of the moving object. Due to this characteristic, segmentation of the player is not required. By changing the temporal parameter rt and the spatial parameter r, CHLAC features can represent various spatial-temporal scale motions as shown in Fig. 2.

2.2 Selection of valid CHLAC features with AdaBoost

Extremely large numbers of CHLAC features with various temporal-spatial scales can be easily generated by changing the parameters rt and r mentioned above. Among these features are many that are not effective for classification, and thus we must select CHLAC features with suitable parameters. For this purpose, we apply the AdaBoost algorithm [9] to select the effective CHLAC features. In this selection process, we regard each simple classifier, which classifies input data based on the Euclidean distance from the center of the class in a CHLAC feature vector space, as a weak classifier.

AdaBoost is an algorithm used to make a strong classifier by selecting the best weak classifiers and combining them with weights according to their performance [10].

$$Class(\mathbf{x}) = \arg\max_{k=1} H^k(\mathbf{x}). \tag{1}$$

$$H^k(\mathbf{x}) = \left(\frac{1}{|\sum_{t=1}^{T} a_t^k h_t^k(\mathbf{x})|} \sum_{t=1}^{T} a_t^k h_t^k(\mathbf{x}) \right), \tag{2}$$

where $k = (1, \ldots, C)$ is the category of player action and C is the number of category classes to be classified. Input vector $\mathbf{x}$ is a CHLAC feature vector derived from a set of sequential images of a player motion. $h_t^k(\mathbf{x})$ is the t-th selected weak classifier, which outputs -1 or 1. T is the total number of selected classifiers, and a_t^k is the weight of the t-th classifier selected by the AdaBoost algorithm. For details on how to obtain a_t^k, refer to [10]. $H^k(\mathbf{x})$ is the obtained strong classifier which outputs the similarity of the input vector $\mathbf{x}$ to category k. Finally, input player motion $\mathbf{x}$ is classified to the category in which the similarity $H^k(\mathbf{x})$ is the highest.

This process can also be considered the selection process of useful CHLAC feature vectors for classifying a player's motion from all of the possible CHLAC features.

2.3 Consideration of mutual interaction between two players' actions

We can consider the information on the opposing player by adding CHLAC features extracted from that player. How to integrate such the additional CHLAC features is a problem to be solved. In this case, we can exploit the proposed framework of selection and integration by using Adaboost without any modification. The solution is to add only the additional CHLAC features into the set of CHLAC features from the target player. Such adaptable integration of additional information is another advantage of our framework.

3 Framework of proposed method

In this section, we explain the process for action classification in the proposed method. Our framework of action recognition consists of two phases: the training phase and the test phase, which are shown in the flow diagram in Fig. 3. The same preprocessing is used in both phases.

3.1 Training phase

(1)Preprocessing Our method starts with preprocessing of input images as shown in Fig. 5. First, we calculate subtraction images from input images to extract the movement of a player. Next, we binarize the subtraction images for noise elimination. Threshold values for binarization are calculated by using Otsu's method[11]. Then, we divide the binarized image into two images to extract CHLAC features from each player. Because CHLAC features are position invariant, segmentation of a player is unnecessary before extracting CHLAC features.

(2)Feature extraction After preprocessing, CHLAC features are extracted from a set of 20 frames of binarized subtraction images, which are selected based a focus frame, as shown in Fig.4. This extraction operation is conducted by the following equation.

$$\hat{\mathbf{x}}(\mathbf{a}_1, \mathbf{a}_2) = \sum_r f(\mathbf{r}) f(\mathbf{r} + \mathbf{S}\mathbf{a}_1) f(\mathbf{r} + \mathbf{S}\mathbf{a}_2) \ , \tag{3}$$

$$\mathbf{S} = \begin{pmatrix} r_{xy} & 0 & 0 \\ 0 & r_{xy} & 0 \\ 0 & 0 & r_t \end{pmatrix}, \tag{4}$$

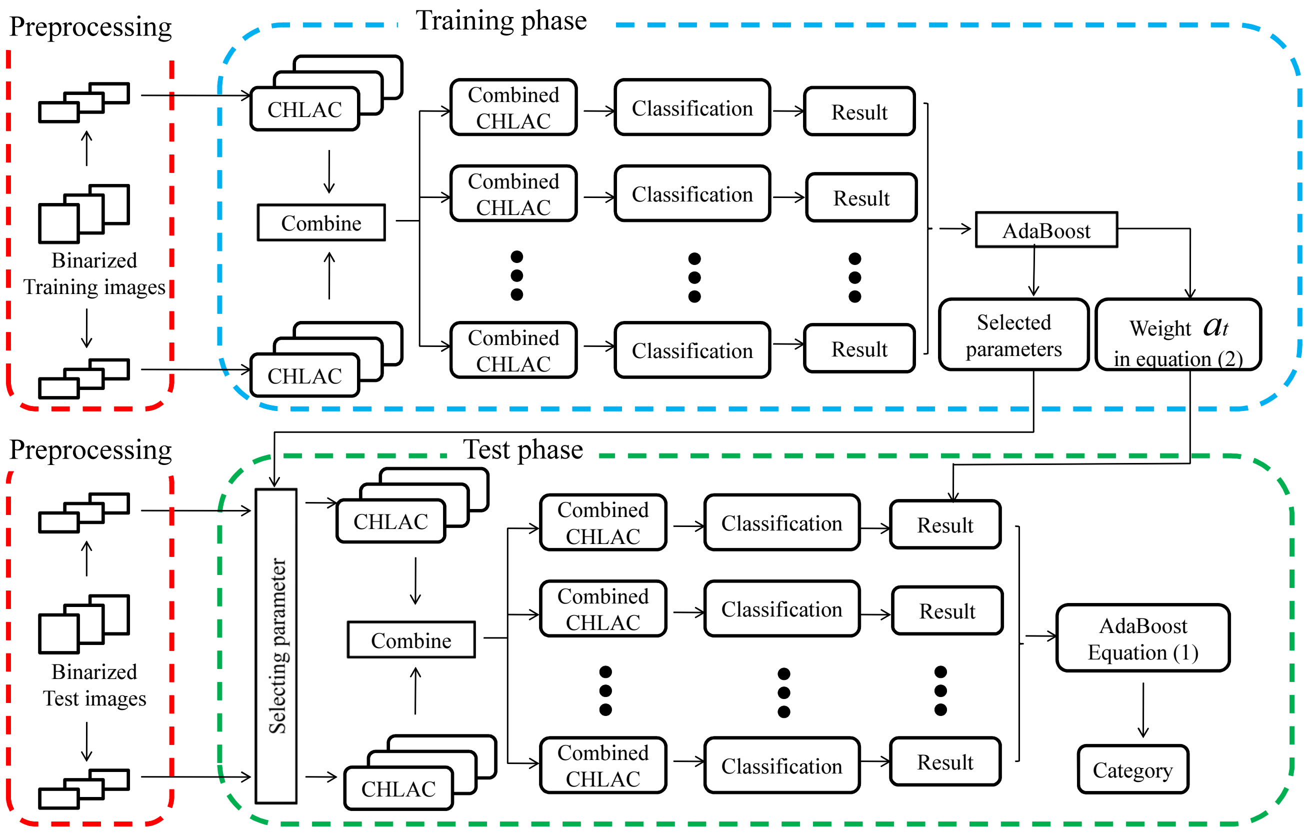


Figure 3. Flow diagram of player action classification

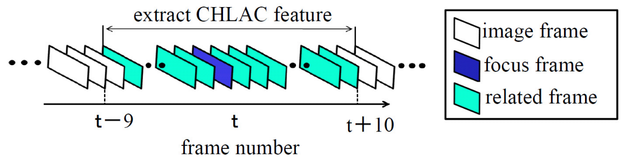


Figure 4. The manner to extract CHLAC features from an image sequence.

where $\mathbf{r}$ is a reference point vector in three-dimensional spatiotemporal space, $f(\mathbf{r})$ is the image intensity of the reference point $\mathbf{r}$, and $\mathbf{a}_1$ and $\mathbf{a}_2$ are displacement vectors from the reference point $\mathbf{r}$. Parameters of r_{xy} and r_t represent the scale of displacement vectors. r_{xy} corresponds to the scale factors of the direction of the x- and y-axes and r_t corresponds to the scale factor of the time-axis direction.

(3)Combining CHLAC features To consider the interaction between each player's action, we use CHLAC features of both players together to obtain new motion descriptors. This is accomplished by connecting all the possible pairs of each player's CHLAC features. The dimension of the CHLAC feature vector is 251, and thus the combined CHLAC features are obtained as 502-dimensional vectors.

(4)Classification To boost the classification ability of the combined CHLAC features, we apply Multiple Discriminant Analysis (MDA) to them.

In the weak classifier, the combined CHLAC features are classified based on the Euclidean distance from the center of each class. The class center vector is calculated from the projected combined CHLAC features onto the discriminate space. We execute this classification process for all the combined CHLAC features.

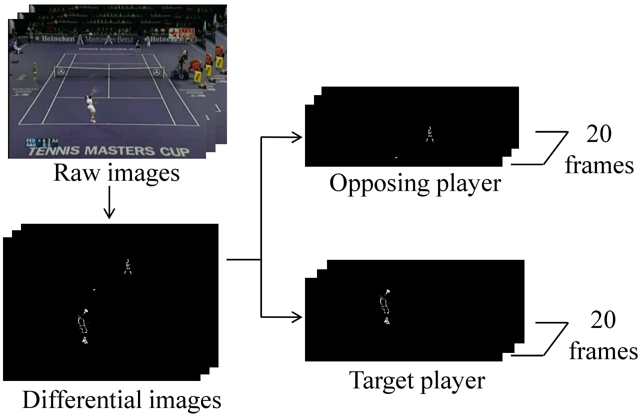


Figure 5. Preprocessing before feature extraction

(5)Selecting useful CHLAC features Useful combined CHLAC features and their weights of a_t^k in Eq. (4) are selected by using the AdaBoost algorithm.

3.2 Test phase

In the test phase, the useful CHLAC features selected in the training phase are extracted from preprocessed images of the target player and the opposing player, separately. After that, two CHLAC features of both the players are connected according to the list of

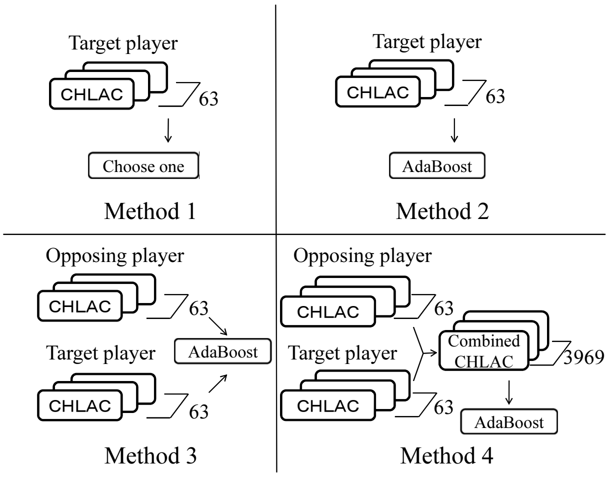


Figure 6. Four methods used in experiment.

the proper combination obtained in the training phase.

Next, all the combined CHLAC features are classified in the same manner in the training phase, and then a final result is obtained by using the classification results of all the result of each the combined CHLAC features by the majority rule considering the weight of each classifier.

4 Experiments

4.1 Experimental conditions

We conducted an experiment to evaluate the performance of the proposed method. Service, forehand, backhand and wait are selected as classification category. The player in the front side is regarded as the target of action classification. Data for the experiment was collected from broadcast video of 10 tennis matches. The video contains various players and different resolutions. We collected data on right-handed male players. The number of examples of each motion is shown in Table 1. The data were captured at 30 frames per second (fps) from the Internet. The resolution of the videos was not uniform; therefore, we resized each video such that the size of the players was almost the same. We used 10-fold cross-validation to evaluate the performance of the proposed method. In each test, 1 of the 10 matches was used as test data and the other 9 matches were combined to form the training data. This process was repeated 10 times, with each of the matches used exactly once as the test data. From each player, 63 CHLAC features were extracted by varying r_{xy} from 1 to 9 and r_t from 1 to 7. These numbers were decided based on a preliminary experiment. All the processing was executed on an Intel Xeon 3 GHz quad core processor.

4.2 Results and discussion

We compared the accuracy of the four methods shown in Fig. 6. Method 1 uses only 1 CHLAC feature, namely, the best performing of the 63, extracted from the target player. In Method 2, useful CHLAC features are selected from the 63 of them by using AdaBoost. Method 3 uses CHLAC features extracted from both player for classification without connecting the feature vectors of each player. In this case, CHLAC features selected from among 126 of them by Adaboost are used for the target player's classification. In method 4, that is, the proposed method, 3969 combined CHLAC features are obtained by changing the pair of CHLAC features of the target and the opposing players to be connected. The final useful CHLAC features are selected by using Adaboost from all the combined CHLAC features.

Table 1. Number of examples of each motion.

Service	Forehand	Backhand	Wait	Total
224	496	268	368	1356

Table 2. Performance of each method.

Method	Accuracy
Method 1	64.74%
Method 2	85.02%
Method 3	86.13%
Method 4(proposed method)	**96.09%**

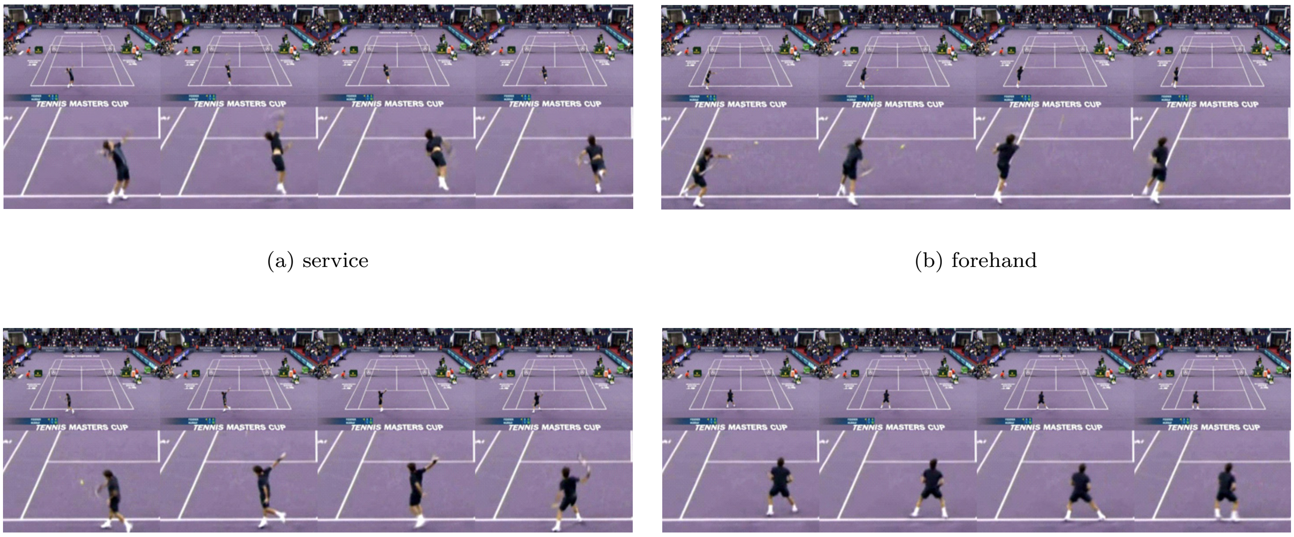
The accuracy of each method is listed in Table 2. The proposed method (method 4) achieved the highest classification success rate of 96.09%. ¿From this result, we can see that using various CHLAC features in combination is effective for player action classification. Furthermore, using information of the opposing player leads to further improvement of the classification accuracy.

The performance of method 3 is improved slightly compared with method 2. This means that the information on the opposing player was useful in predicting the target player's action. However, the information on the correlation between the two players might be insufficient since each player's CHLAC features was used individually. Additionally, in method 3, only a few CHLAC feature of the opposing player were selected by Adaboost, and thus information on the opposing player's action was scarcely considered for classification. On the other hand, in the proposed method, the information on the correlation between the two players was much more extensively considered as compared with method 3, owing to using the forcibly combined CHLAC features.

Fig. 7 shows the examples of the set of sequential images of each action, which was successfully classified by the proposed method. The lower images in each figure show the enlarged images of the neighborhood regions of the target player. From the enlarge images, we can see that each action is complicated and the differences among "forehand", "backhand" and "wait" are small.

5 Conclusions and future work

In this paper, we proposed a new scheme of recognizing the action of a player in broadcast tennis video. The proposed method classified the motion of a player using effective High-order Local Auto-Correlation (CHLAC) features selected by the Adaboost algorithm. In addition, we added information on the opposing player to enhance the classification.



(a) service

(b) forehand

(c) backhand

(d) wait

Figure 7. The examples of the set of sequential images of each action, which was successfully classified: The lower images in each figure show the enlarged images of the neighborhood regions of the target player.

Evaluation experiments demonstrated that our proposed method can achieve much higher performance than a simple method using only information on the target player for action classification. In the future work, we plan to apply the results of this research to more detailed analysis of tennis scenes.

Acknowledgments

This work was supported by KAKENHI (22300195).

References

[1] Lamberto Ballan, Marco Bertini, Alberto Del Bimbo, Giuseppe Serra,"Action Categorization in Soccer Videos Using String Kernels", Proc. 7th International Workshop on Content-Based Multimedia Indexing, pp.13–18, 2009.

[2] Y. Rui, A. Gupta, and A. Acero,"Automatically extracting highlights for TV baseball programs", Proc. ACM Multimedia, pp. 105–115, 2000.

[3] D. D. Saur, Y.-P. Tan, S. R. Kulkarni, and P. J. Ramadge,"Automated analysis and annotation of basketball video", Proc. Symp. Electronic Imaging: Science and Technology: Storage and Retrieval for Image and Video Databases, vol. 3022, pp. 176–187, 1997.

[4] H. Miyamori, "Improving accuracy in behavior identification for content-based retrieval by using audio and video information", Proc. International Conference on Pattern Recognition (ICPR), vol.2, pp.826-830, 2002.

[5] G. Zhu, C. Xu, Q. Huang, W. Gao and L. Xing, "Player Action Recognition in Broadcast Tennis Video with Applications to Semantic Analysis of Sports Game", Proc. 14th annual ACM international conference on Multimedia, pp.431–440, 2006.

[6] N. Otsu, and T. Kurita, "A New Scheme for Practical Flexible and Intelligent Vision Systems", Proc. IAPR Workshop on Computer Vision, pp.431–435, 1988.

[7] T. Kobayashi, N. Otsu, "Action and Simultaneous Multiple-Person identification Using Cubic Higher-Order Local Auto-Correlation", Proc. International Conference on Pattern Recognition (ICPR), pp. 741-744, 2004.

[8] Y. Horita, S. Ito, K. Kaneda, T. Nanri, Y. Shimohata, K. Taura, M. Otake, T. Sato and N. Otsu, "High Precision Gait Recognition Using a Large-Scale PC Cluster", Proc. IFIP International Conference on Network and Parallel Computing (NPC 2006), pp.50–56, 2006.

[9] Y. Freund and R.E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", In Computational Learning Theory: Eurocolt '95, pp.23–37. Springer-Verlag, 1995.

[10] R.E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions", Machine Learning, vol.37, no3, pp.297–336, 1999.

[11] Otsu. N, "A threshold selection method from gray-level histograms", IEEE Trans. Systems, Man, and Cybernetics, 9(1), pp.62–66, 1979.