Journal of Computer and Communications, 2019, 7, 95-104
http://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Research on Gesture Recognition Based on Improved GBMR Segmentation and Multiple Feature Fusion

Xianfei Zhu, Weizhong Yan, Dongzhi Chen, Cuicui Gao

R & D Center, Shanghai Aerospace Electronics Co., Ltd., Shanghai, China
Email: anchorandy@163.com

How to cite this paper: Zhu, X.F., Yan, W.Z., Chen, D.Z. and Gao, C.C. (2019) Research on Gesture Recognition Based on Improved GBMR Segmentation and Multiple Feature Fusion. *Journal of Computer and Communications*, **7**, 95-104.
https://doi.org/10.4236/jcc.2019.77010

Received: May 26, 2019
Accepted: July 7, 2019
Published: July 10, 2019

Abstract

Aiming at addressing the problem of interactive gesture recognition between lunar robot and astronaut, a novel gesture detection and recognition algorithm is proposed. In gesture detection stage, based on saliency detection via Graph-Based Manifold Ranking (GBMR) algorithm, the depth information of foreground is added to the calculation of superpixel. By increasing the weight of connectivity domains in graph theory model, the foreground boundary is highlighted and the impact of background is weakened. In gesture recognition stage, Pyramid Histogram of Oriented Gradient (PHOG) feature and Gabor amplitude also phase feature of image samples are extracted. To highlight the Gabor amplitude feature, we propose a novel feature calculation by fusing feature in different directions at the same scale. Because of the strong classification capability and not-easy-to-fit advantage of Adaboosting, this paper applies it as the classifier to realize gesture recognition. Experimental results show that the improved gesture detection algorithm can maintain the robustness to influences of complex environment. Based on multi-feature fusion, the error rate of gesture recognition remains at about 4.2%, and the recognition rate is around 95.8%.

Keywords

GBMR, Depth Information, PHOG Feature, Gabor Feature Fusion, Adaboosting

1. Introduction

In the field of pattern recognition and human-computer interaction, gesture recognition has become one of the research hotspots. For example, when the lunar robot performs the space task, based on the accurate gesture semantics in-

terpretation, the robot can complete the corresponding motion control.

Many researchers focus on the salient object detection problem for image [1]. However, the complexity of the lunar environment makes the astronaut's gesture detection and gesture recognition challenging. Due to the number restriction of samples in lunar environment, we choose to test the algorithm in the earth environment firstly.

The initial gesture recognition process mainly use the machinery and typical equipment, such as data gloves, to obtain manual space information. At present, compared with wearable devices, gesture recognition based on computer vision can adapt to the freedom of human action, so based on this, the researchers have put forward a lot of gesture detection and recognition algorithms. As mentioned by T.H. Kim [2], segmenting a single image into multiple coherent groups remains a challenging task in the field of computer vision. In this paper, based on gesture detection and recognition part, we classify the algorithms into several categories respectively.

In the gesture detection stage, the state-of-the art algorithm can be divided into two parts: the algorithm based on motion information and the algorithm based on appearance feature extraction. The typical examples in the first one can be listed as the Background Subtraction (BS) method and the Optical Flow (OF) method. BS method needs to obtain the image background in advance, and it is based on the assumption that the color of foreground and background is obviously different, so it is susceptible to the influence of external conditions such as illumination change. OF method doesn't have to obtain the background image in advance, but it also requires a relatively constant illumination condition. Methods in the second one are modeled under different color spaces, also susceptible to external conditions and not robust to complex background.

In the recognition stage, we also divide the relative algorithms into two parts: the algorithm based on template matching and the algorithm based on artificial neural network. The prior one is to compare target and the template through learning a large number of samples, and the category judgment is carried out by the similarity measure. The second one is also built on the premise of a large number of learning samples. The complexity of the network structure and numbers of parameters make it a great challenge for the practical application.

With the increasing popularity of depth camera, because of its color-data and depth-data simultaneous acquisition capability, more and more visual tasks begin to use deep acquisition equipment. The fusion of depth-data and color-data can contribute to the feature extraction of samples. Based on these image acquisition equipment, by adding depth data, this paper realizes the detection of gesture in complex environment.

In this paper, a novel gesture detection and recognition algorithm is proposed. In gesture detection stage, applying saliency detection via Graph-Based Manifold Ranking (GBMR) algorithm, the depth information of foreground is added to the calculation of superpixel. By increasing the weight of connectivity domains in graph theory model, the foreground boundary is highlighted and the impact

of background is weakened. In gesture recognition stage, Pyramid Histogram of Oriented Gradient (PHOG) feature and Gabor amplitude also phase feature of image samples are extracted. To highlight the Gabor amplitude feature, we propose a novel feature calculation by fusing feature in different directions at the same scale. Because of strong classification capability and not-easy-to-fit advantage of Adaboosting, this paper applies it as the classifier to realize gesture recognition. The structure flow of algorithm is shown in **Figure 1**.

2. Gesture Detection in Complex Background

2.1. Saliency Detection Based on Manifold Ranking

The Saliency Detection via Graph-Based Manifold Ranking (GBMR) was proposed by Chuan Yang [3]. By constructing a regular graph and establishing query of background seed point based on the boundary, the saliency detection diagram can be built through applying the idea of manifold ranking. The flow of algorithm is shown in **Figure 2** [3].

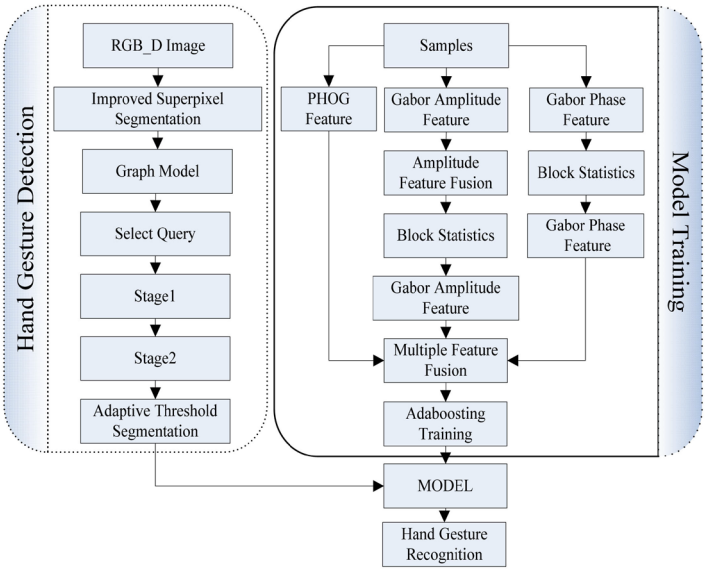


Figure 1. Structure flow of the algorithm.

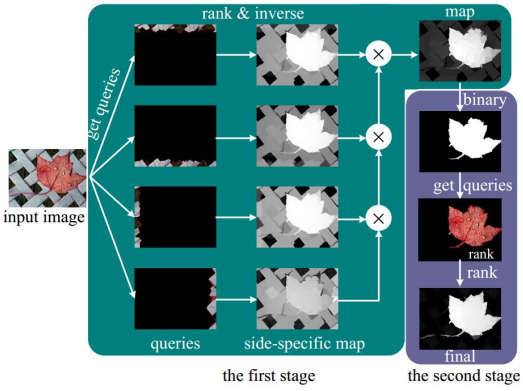


Figure 2. Flow of GBMR algorithm.

The GBMR algorithm process can be expressed as follows:

After Superpixel (SP) segmentation process of input image, the K regular graph of single layer image is constructed to establish the relationship between the SP blocks, and MR algorithm is applied to calculate the ranking score between the query point and then on-query point.

Given a dataset $\mathrm{X}=\{x_1,x_2,\cdots,x_n\}\in R^{m\times n}$, the algorithm uses vectors to record the tagging of data. When $y_i=1$, its corresponding x_i can be seen as a query point; when $y_i=0$, its corresponding x_i can be seen as equal-marked data.

Define the ranking function: $f:\mathrm{X}\to R^m$ which is used to output the corresponding rank score $f=(f_1,f_2,\cdots,f_n)^T$ for x_i.

1) Construct a graph model $\mathrm{G}=(\mathrm{V},\mathrm{E})$ based on dataset X, where V is a point set and E is an edge set.

2) Calculation of E-based association matrix $W=[\omega_{ij}]_{n\times n}$, where

$$\omega_{ij}=\exp(-\frac{dist^2(x_i,x_j)}{\sigma^2}).$$

3) Calculation of the degree matrix of graph $D=\mathrm{diag}(d_{11},d_{12},\cdots,d_{nn})$.

4) The manifold ranking function is $f^*=(D-\alpha W)^{-1}y$.

2.2. Improved GBMR Algorithm

In the case of SP segmentation of input image, if RGB color information is considered only, the gesture segmentation result is not effective when the background is complicated. The incomplete segmentation of target or the segmentation with parts of background will adversely affect subsequent graph theory modeling and ranking algorithm process. In this paper, we consider adding depth information to the SP segmentation so that the target boundary can be highlighted. In the calculation of boundary weights in graph theory model, depth information is also added to weaken the influences of background.

2.2.1. SP Segmentation with Depth Information

In this paper, in the process of implementing SLIC superpixel segmentation [4], we consider adding depth information that can be called D_SP. When the similarity measurement of pixel level is carried out, the formula is updated as:

$$d_c=\sqrt{(l_i-l_j)^2+(a_i-a_j)^2+(b_i-b_j)^2} \tag{1}$$

$$d_s=\sqrt{(D_i-D_j)^2} \tag{2}$$

$$d_{xy}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{3}$$

In the formula, d_c represents (l,a,b) distance measure in the CIELAB color space of pixels *i* and *j*. d_s represents the distance measure between depth pixel values of depth image. d_{xy} represents the spatial coordinates of different pixels in the image. As a result, the final distance metric calculation formula can be obtained as follows:

$$Dist = \sqrt{(\alpha * d_c)^2 + (\beta * d_s)^2 + (\gamma * d_{xy})^2} \tag{4}$$

In the formula, the parameter α, β, γ represent the balance weight of d_c, d_s, d_{xy} respectively. Based on the above distance measurement method, the boundary of target in the SP segmentation stage is more clear, which can contribute to the subsequent mapping model and ranking score algorithm. The result is shown in **Figure 3**.

2.2.2. Improved Graph Model

In order to make edge weights of nodes greater in the graph model, this paper considers updating the weight calculation as:

$$\omega_{ij} = \exp(-\frac{dist^2(x_i, x_j)}{\sigma_1^2} - \lambda * \frac{dist^2(D_i, D_j)}{\sigma_2^2}) \tag{5}$$

where λ and σ are the balance coefficients. The $dist^2(x_i, x_j)$ and $dist^2(D_i, D_j)$ of each sub-block are measured by χ^2 distance.

2.3. Experimental Results

The experimental hardware environment in this paper is Intel Core i3 processor, the main frequency 3.60 GHz. We select the ChaLearn Kinect dataset [5] to validate the algorithm, and the results are as follows **Figure 4**.

(a) (b) (c)

Figure 3. D_SP segmentation result. (a) RGB Image; (b) Depth Image; (c) D_SP Segmentation.

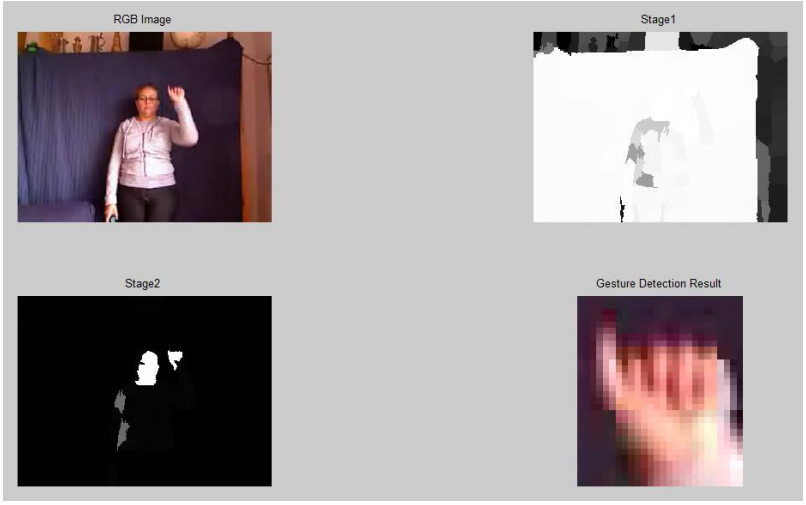


Figure 4. Gesture detection results.

3. Gesture Recognition of Multi-Feature Fusion

3.1. PHOG Feature

The PHOG feature idea was initially proposed by Anna Boschd [6], calculating the HOG feature at different scales and stitching them eventually. It includes spatial multiscale information, which contains more abundant feature information than HOG feature. The process of feature extraction algorithm is as follows:

1) After segmenting the Region of Interest (ROI) in the input images and RGB image grayscale processing, canny operator is applied to obtain the edge information of image.

2) Image layering. The first layer focused on the entire input image, labeled as Level = 0. The second layer will be image 2 * 2 separation, labeled as Level = 1. The third layer will image 4 * 4 separation, labeled as Level = 2.

3) Calculate the gradient direction by pixel at each layer. Divide π or 2π angle into several parts, and obtain the statistical histogram to generate the one-dimensional vector (HOG feature).

4) Combine the HOG feature levels of each layer to obtain the PHOG features of the entire image.

In this paper, the example result of PHOG features extraction is shown in **Figure 5**.

3.2. Gabor Feature

The Gabor feature origins from Gabor transform, which is extended from one-dimensional Gabor filter to two-dimensional image feature extraction.
The two-dimensional Gabor kernel function is defined as:

$$\phi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2}\exp\left(-\frac{\|k_{u,v}\|^2\|z\|^2}{2\sigma^2}\right)\cdot\left(\exp(ik_{u,v}z)-\exp\left(-\frac{\sigma^2}{2}\right)\right) \quad (6)$$

In the formula, u and v represent the direction and scale of the Gabor nucleus respectively, $z=(x,y)$ represents the coordinates of a given point in the image, $k_{u,v}$ represents the central frequency of the filter. $k_{u,v}$ on one certain direction and scale can be calculated as:

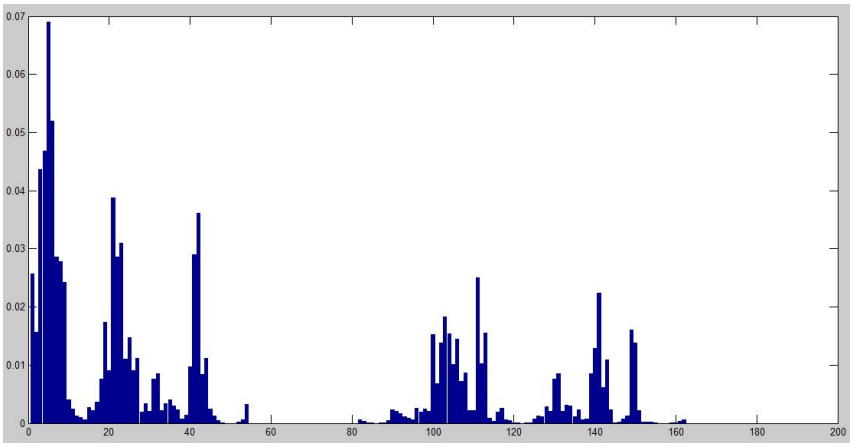


Figure 5. PHOG histogram result.

$$k_{u,v} = k_v e^{i\phi_u} \tag{7}$$

In the formula, $k_v = k_{max} / f^v$, $v \in \{0,1,2,3,4\}$; $\phi_u = \frac{\pi u}{8}$, $u \in \{0,1,2,\cdots,7\}$.

If the grayscale value of the input graph is $I(z)$, the Gabor feature of image is the convolutional result of $I(z)$ and Gabor kernel function. The result can be expressed as:

$$Q_{u,v}(z) = I(z) \otimes \phi_{u,v}(z) \tag{8}$$

In the formula, $Q_{u,v}(z)$ represents the feature description of the image $I(z)$ in the u direction and v scale.

As shown in **Figure 6**, the graph (a) is a Gabor nucleus with different scales in different directions, and the graph (b) is a Gabor feature result of different scales in different directions for the input image.

3.2.1. Gabor Amplitude Feature Fusion

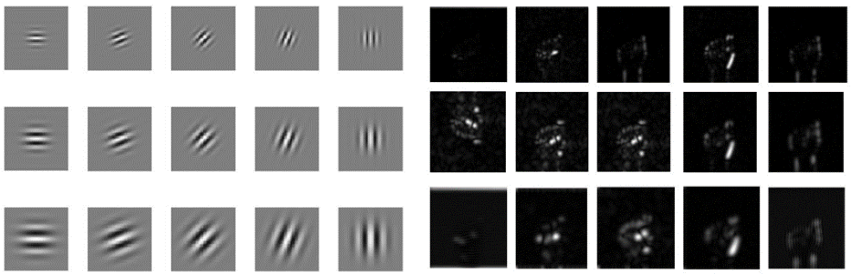
When extracting the feature data of input image, grayscale processing is firstly implemented. In this paper, the Gabor conversion of image is carried out by 8 directions and 5 scale Gabor filters. If 40 image features are cascade directly, the dimension of feature data will be expanded 40 times. And some data for the description of image will not have much impact on the whole, which results in the redundancy. In this paper, by fusing the feature of Gabor in different directions on the same scale, the dimension is lowered while retaining the valid ones.

Take the maximum value of Gabor feature in different directions on the same scale, that is,

$$Q'_v(z) = \max(Q_{u,v}(z)) \; u \in \{0,1,2,\cdots,7\} \tag{9}$$

In the formula, $Q'_v(z)$ represents the eigenvalues of each v scale with different directional feature fusion, and $z = (x, y)$ represents the coordinates of a given point in image. The fused feature result is shown in **Figure 7**.

In order to further reduce the Gabor feature dimension, the Gabor fusion diagrams of each scale are divided into non-overlapping sub-diagram. The mean and standard deviation of each sub-graph are calculated and recorded as (m, γ). Each sub-graph (m, γ) are cascade, which constitutes the eigenvectors of the fusion graph. Assuming that the fusion graph of each scale is divided into



(a) (b)

Figure 6. Gabor feature extraction result. (a) Gabor nucleus with different scales in different directions; (b) Gabor feature result of different scales in different directions.

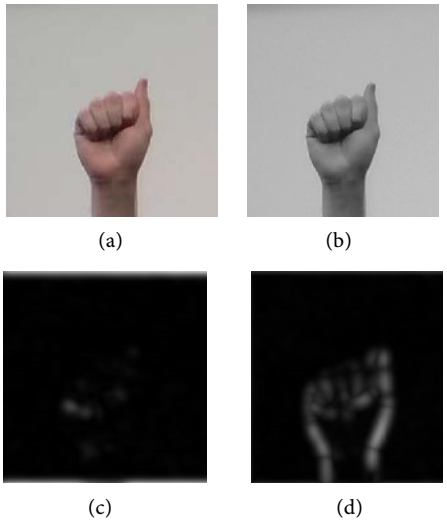


Figure 7. Comparison of Gabor feature result. (a) RGB Image; (b) Grayscale Image; (c) Feature Result before fusion; (d) Feature Result after fusion.

n sub-graph and Gabor scale parameter is set as v, the final Gabor feature dimension of image to be measured will be $2\times n\times v$.

3.2.2. Gabor Phase Feature

When the Gabor filter is carried out, the amplitude and phase feature are output. The real part and the imaginary part of the Gabor filter coefficient are obtained by the Quadrant Binary encoding (QBC) for each pixel point. The phase feature is described by Local XOR Pattern (LXP) operator. It is improved by the Local Binary Pattern (LBP) [7]. The main idea is that each pixel point is "XOR" operation with its adjacent pixels. After the binary sequence is output in a certain direction, the phase value of the corresponding pixel point will be output by weighted operation.

3.3. Multi-Feature Fusion

After the PHOG feature of image samples and the fused Gabor amplitude feature and phase feature are obtained, the data can be merged as the final feature vector output of input image.

3.4. Adaboosting Classifier

Adaboosting classifier is an adaptive enhanced high-precision classifier, which puts the weak classification algorithm as the base classification algorithm in the BOOSTING [8] framework, training the sample set to produce base classifiers. After multiple rounds of iteration, base classifiers are weighted to obtain the strong classifier. It maintains a set of probability distributions on the training set, adjusting the distribution of training set by the error rate of base classifiers. The base classifier of next round cycle can give a higher weight to judge hard examples. Considering its not-easy-to-fit advantage, this paper selects Adaboosting as classifier for category classification.

3.5. Experimental Results

In the experiments, gesture recognition samples are selected from the American Sign Language (ASL) database, with 12210 samples of 6 categories. The number of samples for each category is shown in **Figure 8**. After obtaining the vector of multi-feature fusion of samples, the Adaboosting classifier is applied for training. The number of weak classifiers is set to 30, and 300 pictures in each category are selected as the test set to obtain the recognition error of the model. As shown in **Figure 9**, the error of cross-validation results remains below 0.042.

The comparison of different algorithms is shown in **Figure 10**.

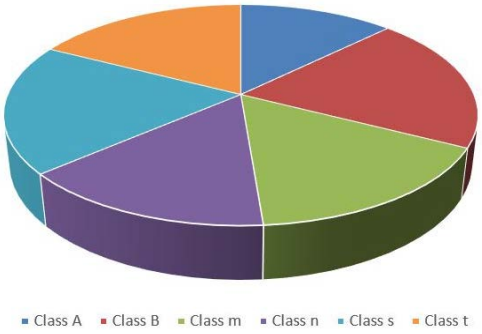


Figure 8. The number of samples for each category.

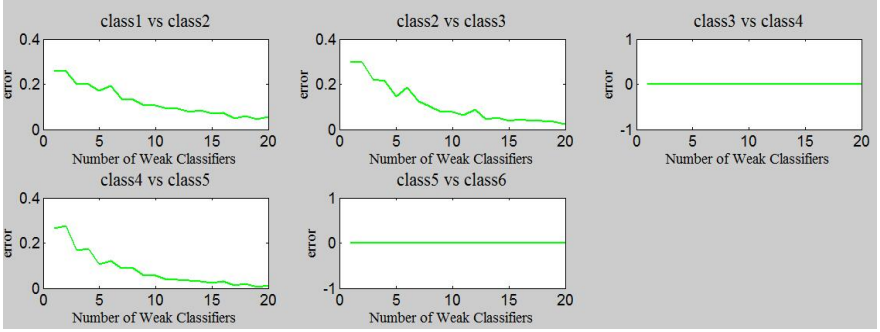


Figure 9. Result of cross-validation.

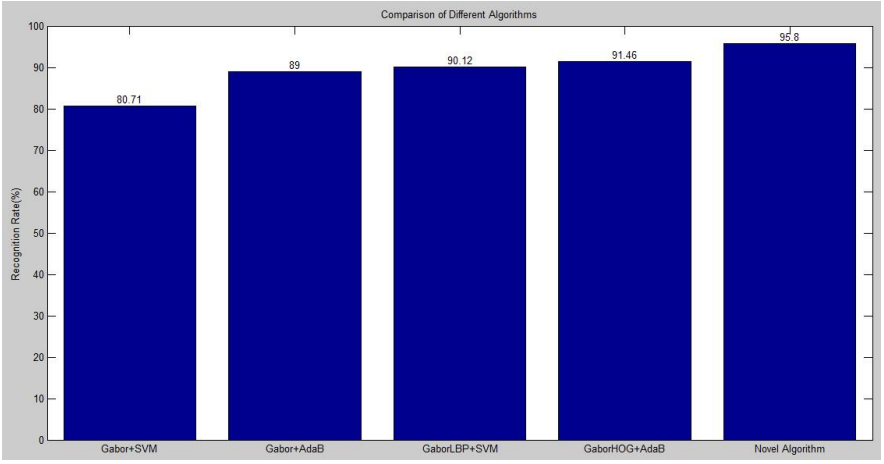


Figure 10. Comparison of different algorithms.

4. Conclusions

In this paper, an improved gesture detection and recognition algorithm is proposed, the contributions can be expressed as follows:

1) In gesture detection stage, the GBMR algorithm is improved. The depth information is added to the boundary weight calculation of SP segmentation and graph theory model, highlighting the boundary of target region and weakening background impact.

2) In gesture recognition stage, the Gabor amplitude feature fusion is carried out in different directions on the same scale, highlighting texture information, and the dimension of Gabor amplitude feature are reduced by applying block statistics method.

3) In gesture recognition stage, multi-scale PHOG feature, Gabor amplitude feature fusion and phase feature are integrated, and Adaboosting classifier is applied to realize recognition.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X. and Shum, H. (2011) Learning to Detect Asalient Object. *IEEE PAMI*.

[2] Kim, T.H., Lee, K.M. and Lee, S.U. (2010) Learning Full Pairwise Affinities for Spectral Segmentation. *CVPR*.

[3] Yang, C., Zhang, L., Lu, H., et al. (2013) [IEEE 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)—Portland, OR, USA (2013.06.23-2013.06.28). 2013 *IEEE Conference on Computer Vision and Pattern Recognition—Saliency Detection via Graph-Based Manifold Ranking*, 2013, 3166-3173. https://doi.org/10.1109/cvpr.2013.407

[4] Achanta, R., Smith, K., Lucchi, A., Fua, P. and Susstrunk, S. (2010) SLIC Superpixels. Technical Report, EPFL, Tech. Rep. 149300.

[5] http://gesture.chalearn.org/

[6] Bosch, A., Zisserman, A. and Munoz, X. () Representing Shape with a Spatial Pyramid Kernel. *ACM International Conference on Image & Video Retrieval.* https://doi.org/10.1145/1282280.1282340

[7] Gabor, D. (1946) Theory of Communication. *Journal of Inst. Electronic Engineer*, **93**, 429-457.

[8] Freund, Y., Iyer, R., Schapire, R.E., *et al.* (2004) An Efficient Boosting Algorithm for Combining Preferences. *Journal of Machine Learning Research*, **4**, 170-178.