International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.5 (2016), pp.417-424
http://dx.doi.org/10.14257/ijmue.2016.11.5.39

A Novel Visual Saliency Detection Method Using Motion Segmentation

Man Hua and Yanling Li

School of Computer Science, Civil Aviation Flight University of China,
GuangHan, Sichuan, 618307, China
hua.man@163.com

Abstract

In this paper, we propose a novel visual saliency detection method using motion segmentation. We group the corner point trajectories using a two stage clustering algorithm. The most stable trajectories are pre-clustered using mean shift in the first stage. Then, we propose an unsupervised clustering method to cluster the trajectories and detect the number of motions automatic. At last, the motion saliency map is generated with the segmented spare feature points. Experimental results show that our proposed method is capable of achieving both good accurate and the stable performance.

Keywords*: Visual saliency, Motion segmentation, Trajectory clustering*

1. Introduction

The rapid development of digital video capture and editing technology has led to increased amounts of video data, creating the need for more effective techniques for video summary and retrieval. In recent years, the visual saliency detection attracted more and more researchers in the field of computer science and psychology of attention which can be applied to target recognition, video coding and event detection. For example, Itti *et al* [1] apply the detection of visual saliency in video coding. Through the visual mechanism simulation, computational resource is priority allocated to those regions which is easy to cause the viewer's attention, which can greatly improve the efficiency of the existing video analysis. In document [2], visual attention model is applied to image retrieval and obtained good results. These applications are based on the visual saliency detection. Therefore, how to detect the visual attention region more effectively is an important issue in the current study.

Nowadays, expressing visual saliency region by use of visual saliency map is the common calculation model. The visual saliency map using the value of each pixel represents the saliency size of corresponding point in original image, thus not only represent the significance of each position of saliency region but also obtain the range of saliency region. For example, Itti *et al* [3] proposed a method for detection of saliency map in a static image. This method combined the measurement results obtained by center-surround operator in different types and scales of visual space into a saliency map. Due to the outstanding performance in the detection results and computational speed, the method is widely concerned by the researchers. But in the past, the methods are based on static image. Therefore, how to detect the visual attention region of video more effectively is an important issue in the current research.

Effectively Motion segmentation is the key step of the visual saliency region detection. But the traditional motion segmentation algorithm which applied to the visual salient region detection of the past methods often neglected the characteristics of visual saliency

ISSN: 1975-0080 IJMUE
Copyright ⓒ 2016 SERSC

region detection. In the detection of visual saliency regions, the past research has shown that the motion characteristics are as follows:

1) The number of motion model of the video is unknown because of the probability of existing one or more moving objects.

2) As the objects may be rigid or non-rigid, the modelling of moving objects by using the parameter motion model is impossible.

3) Since the camera may be stationary or moving, the need for camera motion is adaptive. But when the camera in the jitter time or in the zoom in/out operation, it is difficult for extracting visual saliency area of the human eye. Consequently, in the motion segmentation is without considering the complicated camera motion.

In this paper, using a segmentation method based on trajectory clustering on account of the track is formed by the matching of feature points and the trajectory clustering can be better applied to rigid and non-rigid motion. In addition, different object's trajectory has big difference in length and direction, so it can be used to distinguish between multiple moving objects. Finally, using the trajectory clustering can also distinguish the global motion caused by camera moving and local motion of an object.

Aiming at the characteristics of motion in visual saliency region and the problems in existing methods, this paper proposed a novel segmentation method for visual saliency detection. The method using a hierarchical clustering method to cluster trajectories of feature points according to the characteristics of visual saliency detection. Firstly, classify trajectories in the time domain according to the length of them; secondly, using an unsupervised clustering algorithm for different types of motion segmentation and obtain the motion classification number automatically; finally, using the motion segmentation results, this paper proposed a method combining spatial and color sampling to generate the motion saliency map.

2. Motion Segmentation Based on Trajectory Clustering

This paper uses the Harris corner-point detection algorithm [4] to detect the feature points in video frames, and with the Pyramid-based feature detection and tracking method [5] to estimate the position of the feature points in the next frame. Assuming that a frame of image have N feature points, i feature point's descriptor in the time of t is: $\boldsymbol{p_i^t(x_i^t, y_i^t), i \in N}$, where $\boldsymbol{x_i^t, y_i^t}$ is the position of feature point and $\boldsymbol{x_i^{'}, y_i^{'}}$ is the position of this point in the next frame, $\boldsymbol{dx_i, dy_i}$ is motion vector of feature point , where $\boldsymbol{dx_i = x_i - x_i^{'}, dy_i = y_i - y_i^{'}}$, N is the number of feature points. First, remove the feature points which have too large motion vectors: the motion vector of feature points is greater than a threshold would be regarded as production of feature point matching error, and remove it. Second, obtain the M frames' trajectories of each feature point by using the algorithm based on path coherence be proposed in [6]. In this paper, $\boldsymbol{x_i} = \{\boldsymbol{p_i^1}, \boldsymbol{p_i^2} ... \boldsymbol{p_i^M}\}$ is used to represent trajectory of the i feature point.

In [7], the paper using the expectation maximization algorithm through iterative method to obtain trajectory classification, but in advance, the total number of motion model should be determined when using the EM method. In [8], trajectory also be used in motion segmentation, but the background of this method is static and the length of trajectory is fixed. Nevertheless, in the condition of camera motion will produce a large number of incomplete trajectory. Therefore, the method in [8] cannot be directly clustered. Studies show in [9], better result can be obtain if clustering of similar length trajectory. Consequently, trajectories are divided into two categories according to the duration of time, namely the complete trajectory and incomplete trajectory. The complete trajectory of feature points is visible in all frames. While the initial frames of incomplete trajectory and the end of it may not be the same. Thus, the similarity comparison of incomplete trajectory is extremely difficult.

Based on the complete trajectory, this paper presents a trajectory clustering algorithm according to the motion information of trajectory, and the segmentation information of feature point contains in trajectories can be obtained after the trajectory clustering results be got.

2.1. Trajectory Feature Extraction

As the hypothesis in [9], we suppose that the space displacement is similar between the start and end of the feature points' trajectory which belong to the same moving object. While an object moves, the assumption consider the displacement of feature points' trajectory are similar. As the hypothesis does not require the object or background motion meet the parameters motion model, it is suitable for our algorithm.

Based on it, we use the motion vector as feature of trajectory. First, we define a motion vector $\boldsymbol{v}_i(\Delta\boldsymbol{x}, \Delta\boldsymbol{y})$ of trajectory $\boldsymbol{x}_i$ as:

$$\begin{aligned} \Delta\boldsymbol{x} &= \sum_{\boldsymbol{k}=1}^{\boldsymbol{M}} |\, \boldsymbol{p}_i^{\boldsymbol{k}}(\boldsymbol{x}) - \boldsymbol{p}_i^{\boldsymbol{k}-1}(\boldsymbol{y})\,| \\ \Delta\boldsymbol{y} &= \sum_{\boldsymbol{k}=1}^{\boldsymbol{M}} |\, \boldsymbol{p}_i^{\boldsymbol{k}}(\boldsymbol{y}) - \boldsymbol{p}_i^{\boldsymbol{k}-1}(\boldsymbol{y})\,| \end{aligned} \tag{1}$$

As seen from formula (1), the motion vector of trajectory is composed of the summation of multi-frame motion vector. The advantage of this is that the motion vector of trajectory itself have statistical information and thus the stability is higher than the motion vector between two frames.

How to cluster the trajectory effectively and get the number of motion type after the feature trajectory been obtained is an unresolved issue nowadays. Influenced by noise, the work of Faisal *et al*. shown bad effect of clustering the motion vector directly. In order to solve this problem, we propose a mean shift algorithm to preprocess the motion vector, and then use an unsupervised clustering algorithm on the preprocessing results.

2.2. The Preprocessing Based on Mean Shift

The core idea of the mean shift algorithm is making the data closer to the high density point according to data points weighted by kernel function [10].

Firstly, we use non-parametric method of kernel density estimation to estimate the density of every motion vector:

$$\hat{\boldsymbol{f}}(\boldsymbol{v}) = \frac{1}{\boldsymbol{n}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{K}\left(\frac{|\,\boldsymbol{v} - \boldsymbol{v}_i\,|^2}{\boldsymbol{h}^2}\right) \tag{2}$$

Where $\boldsymbol{K}(\bullet)$ is kernel function, we adopt Gauss kernel function, and formula (2) is written as:

$$\boldsymbol{f}(\boldsymbol{v}) = \frac{1}{\boldsymbol{n}} \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \exp\left(-\frac{|\,\boldsymbol{v} - \boldsymbol{v}_i\,|^2}{\boldsymbol{h}^2}\right) \tag{3}$$

Where h is bandwidth and we set it as 2. The mean shift vector is defined as:

$$\boldsymbol{m}(\boldsymbol{v}) = \frac{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{v}_i \cdot \exp\left(-\frac{|\,\boldsymbol{v} - \boldsymbol{v}_i\,|^2}{\boldsymbol{h}^2}\right)}{\exp\left(-\frac{|\,\boldsymbol{v} - \boldsymbol{v}_i\,|^2}{\boldsymbol{h}^2}\right)} - \boldsymbol{v} \tag{4}$$

Mean shift is an iterative algorithm, exit it while the value of $\boldsymbol{m}(\boldsymbol{v})$ is less than a threshold. According to n original motion vectors $\boldsymbol{v}_i(\Delta\boldsymbol{x}, \Delta\boldsymbol{y})$, we estimated n motion vectors which have been clustered by mean shift clustering algorithm as: $\boldsymbol{u}_i(\Delta\boldsymbol{x}, \Delta\boldsymbol{y}), 1 \le \boldsymbol{i} \le \boldsymbol{n}$. In the implementation, we uses look-up table method to improve the calculation speed in order to avoid every re-calculation of kernel density function.

After mean shift processing, the vector with similar motion shift to its center value, expanding the gap between different types of data and reduce the difference of the same type of data. This will be conducive to the clustering of different motion. However, the mean shift algorithm cannot mark the feature points directly in classification. Hence we propose an unsupervised clustering algorithm to cluster the mean shift algorithm results in order to achieve the mark of feature points.

2.3. Unsupervised Clustering Algorithm of Motion Vector

The idea of unsupervised clustering algorithm is compare the similarity between a motion vector waiting for clustering and a certain one been clustered. If the similarity is high, then update the motion vector to be clustered into existing clustering and use the mean as the clustering center. If the motion vector clustering is not similar with all the motion vectors clustered, we need to create a new cluster. Algorithm steps are as follows:

The input of the algorithm: $\boldsymbol{u}_i, 1 \le \boldsymbol{i} \le \boldsymbol{n}$.

The output of the algorithm: the clustering results of $\boldsymbol{u}_i$.

Step 1: Initialization. Read $\boldsymbol{u}_1$, suppose the total number of clustering $\boldsymbol{MC}$ is 1, and clustering center is $\boldsymbol{u}_1$.

Step 2: input $\boldsymbol{u}_i, 1 < \boldsymbol{i} \le \boldsymbol{n}$ in a procedure until read all the $\boldsymbol{u}_1$. For every $\boldsymbol{u}_1$, obtain the most similar items with it in clustering by formula (5) :

$$\boldsymbol{j}^* = \arg\min_{\boldsymbol{j} < \boldsymbol{MC}} (|\boldsymbol{u}_i - \boldsymbol{MF}(\boldsymbol{j})|) \tag{5}$$

To determine whether the similarity of $\boldsymbol{u}_i$ and $\boldsymbol{MF}(\boldsymbol{j}^*)$ is greater than a threshold by formula (6) :

$$|\boldsymbol{u}_i - \boldsymbol{MF}(\boldsymbol{j}^*)| > \boldsymbol{\theta} \tag{6}$$

Where $\boldsymbol{\theta}$ is a threshold determined by the displacement of an object, it determines the total number of motion type. Therefore, if already obtained prior knowledge of the total number of motion type, we can determine the number of motion type by adjusting $\boldsymbol{\theta}$.

If formula (6) holds, it represent that similar trajectory with $\boldsymbol{x}_i$ was not found in the clustered trajectory, that means we need to create a new cluster, then into creating procedures , said go to step 3. If formula (6) does not hold, go to step 4.

Step 3: the creation method shown as formula (7) :

$$\begin{aligned} \boldsymbol{MC} &= \boldsymbol{MC} + 1 \\ \boldsymbol{MF}(\boldsymbol{MC}) &= \boldsymbol{u}_i \end{aligned} \tag{7}$$

The creation process is actually created a new cluster to record $\boldsymbol{u}_i$. Then go to step 2 after the creation of the new cluster.

Step 4: $\boldsymbol{u}_i$ found similar motion vectors in the existing clustering, namely $\boldsymbol{u}_i$ has high similarity with $\boldsymbol{MF}(\boldsymbol{j}^*)$. $\boldsymbol{u}_i$ will be marked as cluster belonging to $\boldsymbol{MF}(\boldsymbol{j}^*)$ and update the clustering center. The update method is shown as the following formula:

$$\boldsymbol{MF}(\boldsymbol{j}^*) = \frac{\boldsymbol{MF}(\boldsymbol{j}^*) \times \boldsymbol{C}(\boldsymbol{j}^*) + \boldsymbol{u}_i}{\boldsymbol{C}(\boldsymbol{j}^*) + 1} \qquad (8)$$

$$\boldsymbol{C}(\boldsymbol{j}^*) = \boldsymbol{C}(\boldsymbol{j}^*) + 1$$

Where $\boldsymbol{C}(\boldsymbol{j}^*)$ is the number of clustering contained in $\boldsymbol{MF}(\boldsymbol{j}^*)$. The adjustment process is actually using the average value of the displacement trajectory as the feature of this motion model. When the update has finished, go to step 2 and keep reading the next trajectory.

The total number $\boldsymbol{MC}$ of the model and the mean value of each cluster shift can be obtained at the end of the algorithm. As the trajectory with similar shift characteristics is classified as a class, we you can get the result of motion segmentation which represent by the trajectory feature points.

And now, the complete trajectory clustering was done. With the same way, the next step is incomplete trajectory clustering according to the similarity between incomplete and complete. First, this paper uses the similar longest common subsequence of [9] to compare the similarity between two trajectories. LCS is actually an edit distance for string matching application. It is obtained by the comparison of the similarity between incomplete and complete trajectory. Second, get the LCS distance of the incomplete trajectory and all the complete trajectory which have been clustered. At last, the minimum LCS distance was found as the clustering of incomplete trajectory.

Trajectory clustering is obtained by using the above method, as well as the feature point motion segmentation results.

3. Generation of Saliency Map

This paper proposed a motion saliency map generation method combine two attributes of color and space at feature point because of the sparse feature of feature point. Firstly, we obtain the main clustering color of each feature point, calculate the smallest rectangular area surrounded by the feature points of each cluster and get the color histogram $\boldsymbol{H}$ of it. Determine the maximum peak position $\boldsymbol{P}_{mbin}$ and keep the eight connected bin of $\boldsymbol{P}_{mbin}$ in $\boldsymbol{H}$, remove the other bins. Using $\boldsymbol{H}$ as color feature of the feature points set.

Secondly, we divide the current image frame into multiple plurality of 8X8 block size, extract the color histogram as color feature of this block. Suppose a feature point clustering contains NS feature points, then the similarity $\boldsymbol{S}_i$ between the I block and the feature point set of it is obtained by equation (9):

$$\boldsymbol{S}_i = \frac{1}{\boldsymbol{NS}} \sum_{\boldsymbol{j}=1}^{\boldsymbol{NS}} \exp\left[-\frac{\boldsymbol{dd}(\boldsymbol{i},\boldsymbol{j})^2}{2\boldsymbol{\alpha}^2} - \frac{\boldsymbol{di}(\boldsymbol{i},\boldsymbol{j})^2}{2\boldsymbol{\beta}^2} \right] \qquad (9)$$

Spatial distance and color distance of a block to a feature point clustering were calculated by equation (9), where $\boldsymbol{dd}(\boldsymbol{i},\boldsymbol{j})$ indicate the spatial distance from i block to j block, $\boldsymbol{di}(\boldsymbol{i},\boldsymbol{j})$ is histogram intersection from I block to j block, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is the adjustment coefficient of spatial and color distance and set to 20 and 10 respectively in this paper.

Suppose that a feature point set is composed of moving object, if a higher similarity between a block and the feature point set, a higher saliency the block will be, and therefore $\boldsymbol{S}_i$ can be seen as saliency of block i. After $\boldsymbol{S}_i$ regulated to $[0,255]$, Motion saliency map can be obtained.

4. Experimental Results

In order to verify the effectiveness and robustness of this algorithm, a variety of standard video were tested for experiment. The proposed motion segmentation algorithm is tested and compared with related work in the experiment.

Coastguard test videos are multiple moving objects scenes in camera motion conditions. The test sequence contains the global motion of camera and also contains the relative background of the vessel as well as local motion of undulation. In Figure 1, the first line is the coastguard video motion segmentation results. In the first 50 frames, there are two moving objects, through unsupervised clustering algorithm in this paper, the movement is divided into three categories. Where the first category of feature points is represented by a circle and the other two were treated with cross and square. As can be seen from Figure 3 (a) (b) the three different kinds of motion consist of moving objects and background are successful decomposition, leaving only one moving object after the 50 frames. By the pretreatment of mean shift and unsupervised clustering of this paper, feature points of moving objects and background are divided into two groups. As shown in Figure 3 (c) (d), moving objects are successful separated from background. As a conclusion from the experiment, the trajectory clustering algorithm proposed in this paper can adaptive determine the number of clusters under the scenes of motion camera.

The second line (walking man) of Figure 1 is the video clips of a pedestrian tracking by camera. As the character motion is non rigid, so the traditional motion parameter model couldn't work on Character motion modeling and the methods in [7,12] are invalid. The second line of Figure 1 is the result of the test videos of this paper's method. When θ is set to 5, the trajectory is divided into two categories. As can be seen from Figure 1 (e, f, g, h), the feature points belonging to the people were clustered into one group. This is because the method using trajectory clustering technology of this paper, by tracking feature points and generate the trajectory, the feature points segmentation is obtained. As the results shown in the experiment, this method has good adaptability to the rigid and non-rigid motion. However, while the following two frames have a covering will led to the failure of the feature point matching. But as a result of the pre classification according to the path length, this method still can be very good to different motion segmentation. From the test results of four consecutive frames of Figure 1, the robustness of the method can be achieved.

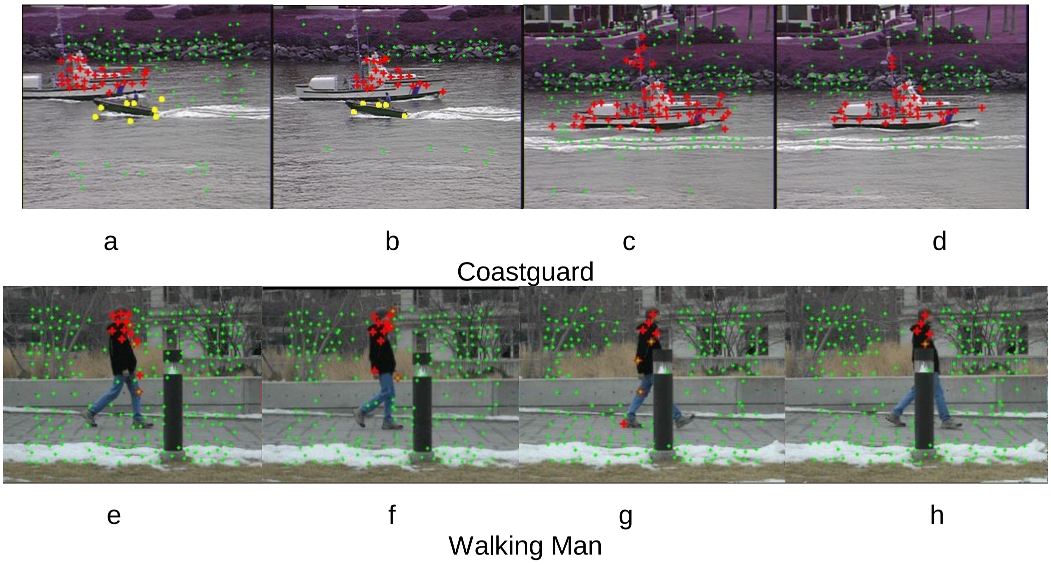


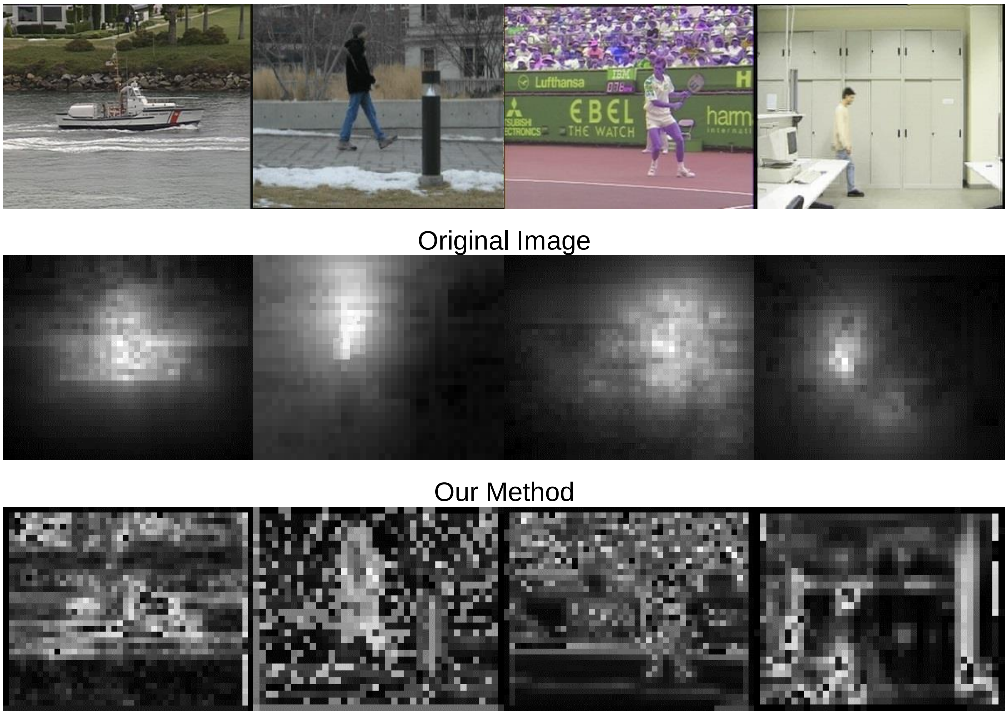
Figure 1. Experimental Results of Motion Segmentation

Copyright ⓒ 2016 SERSC

In order to validate the effect of visual saliency detection, a plurality of test video were tested and compared with the method in [11]. In Figure 2, the four columns show the experimental results respectively to the test videos of Coastguard, Waling man, Stefan and Laboratory. As can be seen from Figure 2, although the feature points are sparse feature, visual saliency region been generated is relatively complete due to the use of the motion saliency map generation method of the third section in this paper, and the results accord with the subjective judgment. As we using the motion segmentation to classify the global motion and local motion, the saliency map is less affected by global motion. Judging from the sequence of Coastguard, Waling man, Stefan, the effect of visual saliency region produced by our method is significantly better than the method of [11]. As for Laboratory, a video under static condition, Figure 4 illustrate that the visual saliency region is similar with the one in [11]. Our experimental results demonstrate that the visual saliency region were generated effectively whether the camera is moving or in static. Experimental results are satisfactory.

5. Conclusion

This paper proposed a novel visual saliency detection method using motion segmentation. Using the correspondence between the feature points of consecutive frames, feature points trajectory is obtained. Then, according to the trajectory information, feature points are classified through a special clustering algorithm. The new method obtain motion information according to the corresponding relationship of feature points, it contains less noise and does not depend on the specific motion model and use the motion information of consecutive frames . Also, it does not necessarily require any adjacent frames are continuous and stable enough. The experimental results demonstrate the effectiveness and robustness of the method in this paper.



Original Image

Our Method

Method in [4]

Coastguard Walking Man Stefan Laboratory

Figure 2. Experimental Results of Visual Saliency Detection

Acknowledgments

This work is supported by Joint Fund of the National Natural Science Foundation of China and the Civil Aviation Administration of China (No. U1233105).

References

[1] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention”, IEEE Transactions on Image Processing, vol. 13, no. 10, **(2004)**, pp. 1304-1318.

[2] F. Stentiford, “Attention-based similarity”, Pattern Recognition, vol. 40, no. 3, **(2007)**, pp. 771-783.

[3] L. Itti and C. Koch, “Feature combination strategies for saliency-based visual attention systems”, Journal of Electronic Imaging, vol. 10, no. 1, **(2001)**, pp. 161-169.

[4] C. G. Harris and M. J. Stephens, “A combined corrter and edge detector”, In: Proceedings Fourth Alvey Vision Conference, Manchester, UK, **(1988)**, pp. 147-l51.

[5] G. R. Bradski, and V. Pisarevsky, “Application in calibration, stereo, segmentation, tracking, gesture, face and object recognition”, IEEE Conference on computer vision and pattern recognition, SC, USA, **(2000)**, pp. 796-797.

[6] V. Salari and I. K. Sethi, “Feather point correspondence in the presence of occlusion”, IEEE Trans PAMI, vol. 12, no. 1, **(1990)**, pp. 87-91.

[7] S. J. Pundlik, “Real-Time Motion Segmentation of Sparse Feature Points at Any Speed”, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 38, no. 3, **(2008)**, pp. 731-742.

[8] X. Wang and K. Tieu, “Learning Semantic Scene Models by Trajectory Analysis”, in Proceedings of European Conference on Computer Vision (ECCV) Graz, Austria, **(2006)**, pp. 205-211.

[9] G. Antonini and J. P. Thiran, “Counting Pedestrians in Video Sequences Using Trajectory Clustering”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 8, **(2006)**, pp. 1008-1020.

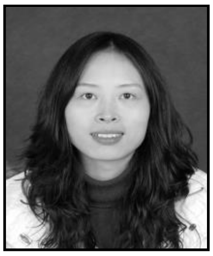
[10] D. Comaniciu and P. Meer, “Mean shift: a robust approach toward feature space analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, **(2002)**, pp. 603-619.

[11] Y. F. Ma and X. S. Hua, “A Generic Framework of User Attention Model and Its Application in Video Summarization”, IEEE Trans on Multimedia, vol. 10, no. 7, **(2005)**, pp. 907-919.

[12] O. L. Meur1 and D. Thoreau1, “A Spatio-temporal Model of The Selective Human Visual Attention”, IEEE International Conference on Image Processing(ICIP 2005), Genova, **(2005)**, pp. 1188-1191.

Authors

Man Hua, He received his B.S. (1999) in mechanical engineering from The Sichuan University and M.S. (2004) in computer science from The Southwest Jiaotong University, Chengdu, China. Since 2004 he has been with the Computer Science Department of The Civil Aviation Flight University of China. Currently, he holds the position of Associate professor. His research interests include image processing, video processing and information security.



Yanling Li, She received her M.S. (2008) in applied mathematics from The China University of Geosciences, Wuhan, China. Since 2008 she has been with the Computer Science Department of The Civil Aviation Flight University of China. Currently, she holds the position of lecturer. Her research interests include differential equation, neural network and image processing.

Copyright ⓒ 2016 SERSC