Acta Technica *62* No. 3A/2017, 263–274

The application of characteristic symbol detection and the scene recognition in football match video

Kuan Wang[1]

Abstract. In some football game videos, video data need to be segmented into different scenes to realize the retrieval of video data, thus analyzing the performance of different players in the field. How to effectively segment video data, therefore, becomes a significant application problem in the field of information retrieval. Video data is studied in the paper with camera pictures taken as the appearance of stereoscopic effect. A scene recognition algorithm similar to information pyramid is put forward at first so as to target at the moving scene of the player; the moving player in the camera can be seen as a circle symbol, whose detection algorithm is studied; meanwhile, in order to distinguish symbol pictures better, the algorithm of scene segmentation is researched from the minimum unit of the video. According to the situation when the accuracy and speed cannot be realized by using the current algorithm, a segmentation algorithm, based on the characteristic symbol detection and scene recognition algorithm, is figured out, whose accuracy improves gradually at a certain speed.

Key words. Scene recognition, characteristic symbol, scene segmentation.

1. Introduction

As a form of data, video [1] works better in helping people comprehend the information than text and carries much more information. A football match lasts 90 minutes, and the 90-minute video contains many massages for the coach, players and fans as well. People can see that the scenes and one scene switches to another very fast. The coach may focuses on the defense of midfield player, because the result of defense in midfield affects the pressure of defense line; players hope that they can see the moving ways of their counterparts so as to know about their attack modes; they care more about the wonderful moments of their favorite players.

When videos are uploaded to the Internet, they can be clipped and segmented in various ways and then they can be edited with titles. The way to manually handle with videos is inaccurate and time-consuming. Generally, the search engine defines

[1]Weinan Normal University, Shaanxi, 714099, China; E-mail: peterkuanwang@126.com

video in the way of text, which is limited. If video retrieval [2] is to be realized, the contents of the video itself should be taken into consideration.

2. Literature review

Scene recognition [3] is an important field of computer vision which is applied in many fields. It works in the data vectorization [4] of pictures in the scene, designing correspond data model and classifying different scenes. SIFT feature algorithm [5] has good recognition ability, but it has lower recognition rate in spatial distribution. Some researchers attempt to introduce certain information about spatial distribution to the semantic meaning, thus raising recognition rate. The literature points out that using spatial pyramid mode [6] to divide spatial area to extract SIFT features can greatly improve the scene recognition rate. Census transformation [7] is combined with special pyramid model in the literature, which works better than the former algorithm.

Pictures in video is stereoscopic [8], which is based on stereoscopic vision detection including camera calibration, picture data correction and picture fitting and three-dimensional positioning of objects. Characteristic symbol detection is crucial to three-dimensional positioning of objects.

This paper combines with spatial pyramid model and refers to the concept of similar information. A scene recognition algorithm is proposed and from the perspective of testing logo, players are taken as circular logo, to extract the edge information. And then a new symbol detection algorithm based on least squares fitting method is put forward. Combining scene recognition algorithm with symbol detection algorithm and from the perspective of scene segmentation, referring to the concept of entropy, an improved shot segmentation algorithm is proposed.

3. Research methods

3.1. Scene recognition algorithm similar to information pyramid

The traditional spatial pyramid mode uses the regular grid as the segmentation unit and extracts the features of correspond unit. But this method lacks in the description of semantic information of pictures. Based on spatial pyramid model, segmentation method with the similar information is adopted in the study so that the segmented grid units are of semantics.

The use of similar information can increase the computation of picture recognition, which normalizes small modules with several different colors to one color. This segmentation method combining similar information is based on entropy rate, which transforms the problem into another similar to clustering.

The function of the problem is expressed as

$$E(A) = F(A) + \beta G(A) \tag{1}$$

In the function, $F(A)$ is used to compute the entropy rate. The incoming picture is seen as an undirected graph. Panel points in the picture show the pixels, and edges of the picture indicate the connection of pixels. The function adopts the way of random walking to cover all points in the picture, so there are many pixel paths which can be used to compute the uncertain rate, namely, entropy rate. Clustering is analyzed based on several entropy rates and correspond pixel points. The bigger the entropy, the closer the pixel placer. Thus, the region of picture modules can be judged and divided. Expression $\beta G(A)$ is used to avoid the entropy rate tends to 0.

In using the method of segmenting similar information, information of similar modules of pictures can be clear, which lays a good foundation for the coming feature extraction. As is shown in Figs. 1 and 2, the way of segmenting similar information is applied and the picture is divided into several parts, and here there are 10 parts.

Fig. 1. Input image

Fig. 2. Similar image split

This kind of segmenting method ensures every pictures module has one or less objects when they are segmented in different resolving powers, which reduces the possibility that one picture module has more than one objects. This can improve the semantics of the picture module, which is shown in Fig. 3.

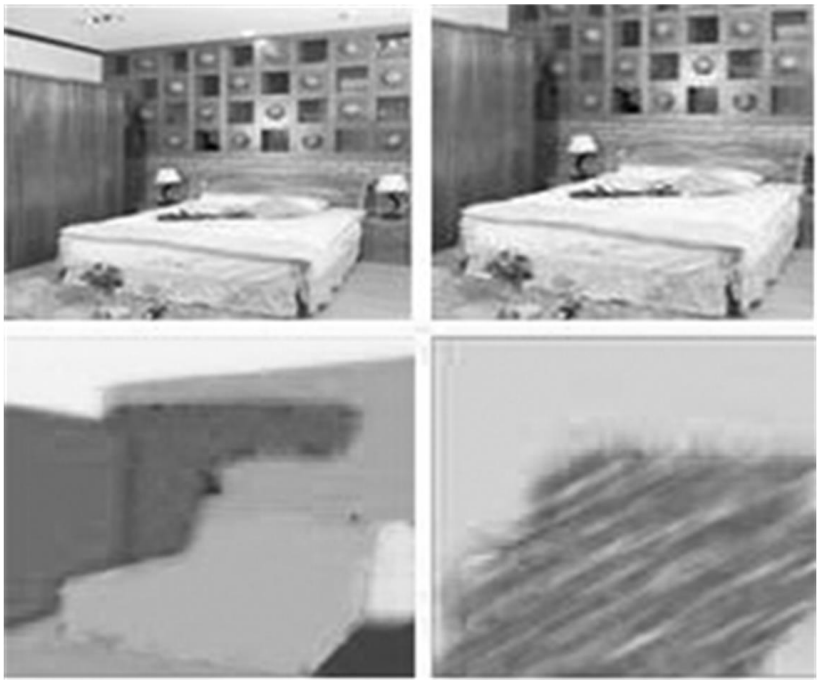

Fig. 3. Similar image split with different pixels

When the picture is segmented, features will be extracted by Census Transformation (CT), which is often used in connecting partial regions. It sets a founder with the size of 3×3. By comparing center point of the founder and the neighbor points, the transformation value can be known.

$$CT(q) = \underset{q' \in WT(q)}{\otimes} \text{MinMax}(U(q), U(q')), \tag{2}$$

$$\text{MinMax}(U(q), U(q')) = \begin{cases} 1, U(q) < U(q'), \\ 0, U(q) \geq U(q'). \end{cases} \tag{3}$$

In formulae (2) and (3), $CT(q)$ is the Census Transformation value corresponding to pixel q, $WT(q)$ is the transformation window corresponded to center point of the founder and $U(q)$ is the gray value of pixel q.

If the founder is $[(32,54,60),(45,54,70),(50,54,80)]^{\mathrm{T}}$, and has been transformed, there are 8 numbers which can be compared. According to (3), it can be transformed to $[(1,1,0),(1,1,0),(1,1,0)]^{\mathrm{T}}$, which corresponds to a binary number, whose range is [0, 255]. This is the CT value of the center point. Every change of pixel point through CT will reflect the transformation effect of the picture, which is shown in Fig. 4.

CT retains all features of the picture, and the picture is describe 256 degrees of vector, among which o and 255 is removed in that they go against the description. From the perspective of matrix theory, the picture structured by this method is of good sparsity and can help analyze main elements.

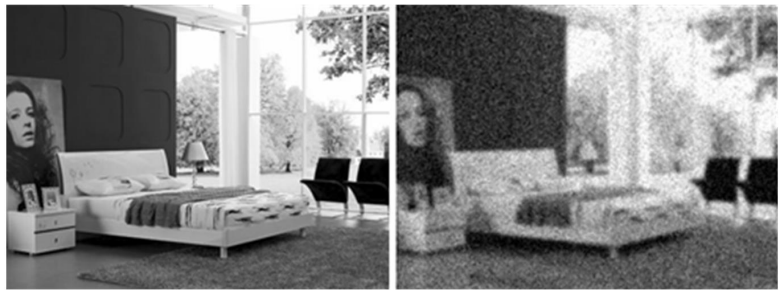

Fig. 4. Image transform with census

3.2. The algorithm of edge-least square fitting detection

In football match video, the characteristic symbols of players should be positioned if the moving track, the position and the defense in midfield of players are to be observed. Circles are common in characteristic symbol, which are easy to deal with and will not change with the pace of the geometry. In stereoscopic vision of the video, circle, the characteristic symbol, will be transformed into oval through affine. As the center of oval pictures does not mean the center of the three-dimensional circle picture, there is certain deviation between centers, and the center of the oval picture should be positioned.

The algorithms of least square oval fitting are:

3.2.1. Algebra fitting method. The method is to put discrete nodes to the equation to know the distances of deviation, and coefficient of the equation can be known in the method of least square. The method is easy to get results. If certain conditions are set, there will be less impossibility that the fitting result is oval. But it tends to be affected by abnormal nodes and isolated nodes, so the stability of it is not sure, which may lead to the difference of the least deviation function after transformation.

3.2.2. Geometry fitting method. This method defines that the deviation distance is the shortest distance from the point to the oval. It is different from algebra fitting method for there is no difference of the least deviation function, but it shows the exponential growth in computing.

3.2.3. Least median square method. The method requires that the median of residual sum of squares is the least. After extracting from the edge every time, 5 characteristic points are used to get corresponding parameters through linear transformation, this method features in excluding some characteristic points with great deviation, but the group numbers of correspondent oval parameters affect the performance of the algorithm.

In order to increase the precision of edge detection and make it the grade of subpixel, Zernike matrix is used here. It is an orthogonal complex which is of orthogonal invariant property.

The correspondent T-order polynomial of Zernike matrix is as the following:

$$W_{TF}(\beta,\,\alpha) = C_{TF}\mathrm{e}^{\mathrm{i}F\alpha}\,. \tag{4}$$

In the formula, T and F are integers and conform to the condition: $T \geq 0$ and $T-|F|$ is an even number. The parameter i refers to the imaginary unit. Polynomial formula CTF is defined as

$$C_{TF}(\beta,\,\alpha) = \sum_{d=0}^{\frac{T-|F|}{2}} \frac{(-1)^d(T-d)!\beta^{T-2d}}{d!(\frac{T+|F|}{2}-d)!(\frac{T-|F|}{2}-d)!}\,. \tag{5}$$

The T-order polynomial of Zernike is orthotropic in unit grid, namely, it conforms to the formula

$$\iint_{a^2+b^2\leq 1} W_{TF}(\beta,\alpha) \times W^*_{RV}(\beta,\,\alpha)\,\mathrm{d}a\,\mathrm{d}b = \frac{\pi}{T+1}\,. \tag{6}$$

Formula (6) is right when $T = R$ and $F = V$ and $W^*_{RV}f(x)$ means the complex conjugate.

Two-dimensional picture corresponding to Zernike matrix is defined by the formula (7):

$$Q_{TF} = \frac{T+1}{\pi}\iint_{a^2+b^2\leq 1} f(a,b) \times W^*_{TF}(\beta,\,\alpha)\,\mathrm{d}a\,\mathrm{d}b\,. \tag{7}$$

Considering that the picture is made up of scattered signals, Zernike matrix of two-dimensional picture $f(a,b)$ is

$$QTF = \sum_a \sum_b f(a,b) \times W^*_{TF}(\beta,\,\alpha),\ a^2+b^2 \leq 1\,. \tag{8}$$

Zernike matrix has less redundancy, strong ability of anti-interference, high precision and other merits. Good effects will appear if it is used in edge detection. The principle of detection is to compute the parameters of every pixel points in use of third-order Zernike matrix. According to the results, the pixel point will be judged whether it is the edge point. Parameters of the pixel point are: p, the gray coefficient of the background, $p+k$, the gray coefficient of foreground picture; om, the distance the foreground projects to the edge; θ, the included angle of project line and X axis. Symbol k is the span from the background gray value to the foreground.

At first, Canny operator is used to extract the correspondent pixel coordinate of the oval, and then Zernike matrix is used to further improve the precision of edge coordinate. According to the edge coordinate of sup-pixel, the parameter of the oval fitting is known and the correct oval center is also clear in the way of the least square fitting.

3.2.4. Scene segmentation algorithm In the algorithm of segmenting the scene, it is easy to mix the moving of camera or objects and the switching of scene. But the former is changing with the step of the position of edge, and the edge position

is relatively stable when pictures are switched. Based on the feature, "interframe information entropy is put forward.

$$E_p = -\sum_{t=0}^{W} |a_t^Y - a_t^W| \log |a_t^u - a_t^v| \,. \tag{9}$$

In the formula (9), u means any frame, v means a frame after u, t means the tth column of the histogram and $t \in [0, 255]$. 5 frames is an interval, which means that the difference of v and u is 5.

When adaptive threshold function is used to segment a group of moving videos with a camera or object, it will be mistaken as the switched candidate scene. Within a particular range, "interframe information entropy" is computed through Gaussian distribution formula. The related range can be changed by using the sliding of windows. Threshold R_s is shown in the following:

$$R_s = \eta + \beta\theta \,. \tag{10}$$

In the formula, η and θ refer to mean value and standard variance, in which β is the judging factor analyzed according to experiments. The parameter can be chosen properly to increase the rate of misjudging.

When the candidate scene is certain, further selection works should be carried out to remove the misjudging caused by the object or camera moving from the switching of scenes. Considering that there is overall change between frames in the moving video, but there change between frames in switching scene is partial. In accordance with the feature, the algorithm of scene edge detection is figured out, and the flow of it is as follows:

(1) Symbol A_i denotes the ith frame, and the interval is T_1, here its value is 2. At first, the pixel difference of two frames DA_i is figured out, shown in subfigure (b) of Fig. 5.

(2) The edge DA_i is detected by using Canny operator, and thus, WA_i (the corresponding moving effect) is shown in subfigure (c) of Fig. 5. The moving edge of switched scene is relatively stable and exhibits only a small change, but the edge change caused by the camera or object moving is greater.

(3) WA_i is simply expanded and segmented the effect picture of moving edge WA_i in the size of 6×6 window. The number of edge pixel in each unit is respectively collected. If one of them is over a pixel, it means the adjacent region of 6×6 is the edge. The edge expansion effect picture is shown in subfigure (d) of Fig. 5.

(4) After the edge picture of expansion is known, the edge information is gathered. Here the edge picture is divided into $Q \times H$ regions, and the resolving power of a single frame picture is $q \times h$. The number of unit region is $Q * H/(q * h)$. In the experiment, the unit region is set as 60×40, and the resolving power of a single frame is 1024×768. The information value of moving edge is figured out in the method of interframe information entropy:

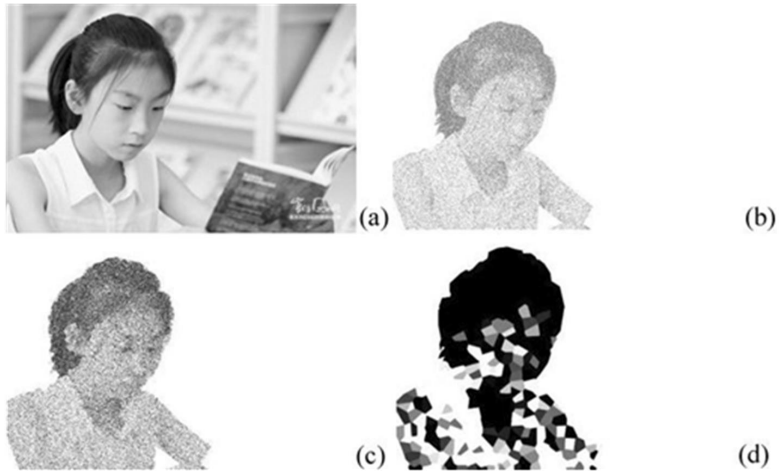

Fig. 5. Process of scene boundary detecting algorithm

$$E_p = -\sum_{v=1}^{Q\times H/(q\times h)} F_v^o/(q\times h)\log(F_v^o/(q\times h))\,. \tag{11}$$

In the above formula, F_v^o means the number of the edge pixel in unit region o, and $o \in [1, Q \times H/(q \times h)]$. Then the same computation is repeated in every frame so as to get the information difference of each frame using the formula

$$DA_v^F = |F_v - F_{v+1}|\,. \tag{12}$$

By doing this, a group number, namely, $DA_1^F, \cdots, DA_n^F$ misjudged in the video sequence is known, in which n means the number of the frame.

The whole process of segmenting scene develops from the group to the fine. The first step is based on interframe information entropy, which is to compute the candidate scene to be judged. As the similar performance of switching scenes and object or camera scene in general algorithms, it is impossible to select further. Therefore, in the second step, the distinguishing algorithm for switching scene and object or camera moving scene is designed based on the feature that the change appearing in switching scenes between frames is partial.

4. Experiment results and analyses

4.1. The experiment and analysis of scene recognition

In order to prove the algorithm put forward in the study is effective, different picture databases are compared and analyzed. Similar segmenting is adopted to segment the picture to two subregions with different grades: 128, 32, 8, 2. Owing to Census transformation, it is just based on the pixel intensity and lacks of multi-

dimensional information of pictures. Thus, some features of picture data are introduced. The feature dimension under the two grades is shown as: (38+5)×(8+2+128 +32)=7310. To extract characteristics, adopt the clustering algorithm to establish word dictionary for the characteristics, then get the word frequency of each scene, and the overall feature of the picture will be known, whose feature vector is expressed by 7310 dimensions.

10 kinds of scene are chosen here, and 300 pictures are used on average. The standard number of pixels of each picture is 1024×768 with the color omitted. In the data, the corresponding test set and training set are obtained based on Mahout, and the rate of them is loosely 1:4. The experiment is repeated 10 times to select the average recognition rate. Here the recognition rate is classified by using Bayes approach and sparse matrix is taken as the judging standard.

Table 1. Film scene labeled with alphabet

Film Scene	Alphabet
The Godfather	a
American Beauty	b
Raiders of the Lost Ark	c
The Silence of the Lambs	d
Paths of Glory	e
Toy Story 2	f
North by Northwest	g
The Sixth Sense	h
Crouching Tiger,Hidden Dragon	i
Homeless to Harvard	j

As what is shown in Table 1, it shows the scene of 10 movies. Symbols $[a, j]$ are used for better marking. For example, the movie American Beauty corresponds the letter b.

Table 2 shows that the scene recognition of each movie is high. From the main diagonal of the table, the average scene recognition can be computed as

$$(0.85 + 0.78 + 0.85 + 0.8 + 0.9 + 0.86 + 0.88 + 0.89 + 0.96 + 0.76)/10 = 0.853\,.$$

It can be seen that the average recognition rate put forward in the study is over 85 %. 8 out of 10 items have the recognition over 80 %, among which recognition rate of the movie Crouching Tiger, Hidden Dragon is up to 96 %, which shows the distinctive scene characteristics.

4.2. Simulation and analysis of characteristic symbol detection

Circle, the characteristic symbol, is studied. The design of detection algorithm is to solve the problem that project enter point of characteristics symbol is uncertain due to the deviation. Circle is adopted in the study, which will be projected to oval.

In the ideal state, the oval center is a point with the circle symbol projected. In practical detection, the oval centers which are fitted on the oval edge are not at one point.

Table 2. Scene recognition based on confusion matrix

	a	b	c	d	e	f	g	h	i	j
a	0.85	0.12						0.10		
b		0.78			0.13		0.02			
c			0.85	0.10		0.05				
d				0.8	0.13				0.07	
e					0.9			0.02		0.08
f						0.86	0.10		0.04	
g		0.12					0.88			
h			0.03		0.12	0.05		0.89	0.04	
i					0.04				0.96	
j				0.05			0.06			0.76

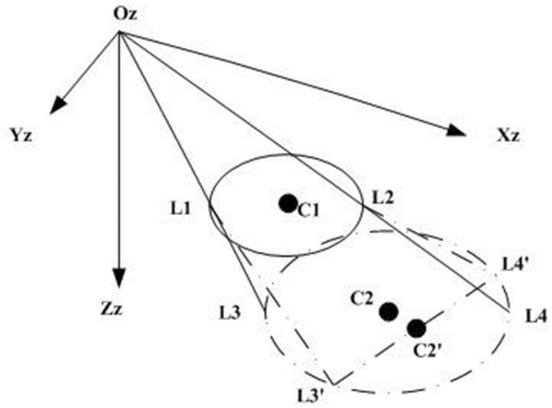

Fig. 6. Projection center deviation

As shown in Fig. 6, Oz-XzYzZz consists of the camera coordinate system. With C1 as the circle center, it establishes the photographed plane. The plane where the camera makes an image is the plane with C2 as the circle center. It is assumed that L1 and L2 are the diameters of an ellipse on a plane. OzL1 and OzL2 are extended to the space object plane, namely, they intersect with the plane whose center is C2 to L3' and L4'. In addition, from the figure, it is easy to find that, between the correspondent points, there holds $L3 \neq L3'$ and $L4 \neq L4'$.

L3' and L4' are connected and the center point C2' is selected. And then it will be clear that C2' and C2 are not the same points. It means that the projection point of the center on the space object plane is not in the circle point after it is transformed in the camera imaging plane. In the pretreatment of video data, it is usually divided into many pictures, so the deviation of the projection circle is common.

In order to reduce this kind of deviation as much as possible and better depict players in the field, it is important to compute the center of the projected oval. The method put forward in the study is to detect the edge and then extract the related edge coordinates of characteristic symbols, and get the edge coordinate of sub-pixel in using Zernike matrix. Hence, the detection precision can be improved. Through comparing the deviation of center of three-dimensional project imaging, the advantage and disadvantages of this method can be found.

Table 3. Coordinates of ellipse center with different algorithms

Algorithm	Center of grey	Ellipse fitting	Zernike fitting	Installed center
Ellipse 1	241.55, 209.58	242.15, 211.45	241.07, 208.70	240.65, 209.01
Ellipse 2	269.80, 271.05	269.96, 270.15	269.11, 269.75	269.00, 269.01
Ellipse 3	284.80, 286.05	284.96, 285.15	284.11, 284.75	284.00, 284.01
Ellipse 4	521.75, 234.30	521.39, 234.15	520.60, 233.75	520.80, 233.55
Ellipse 5	412.86, 270.59	412.34, 270.19	412.41, 269.50	412.13, 269.50

According to what is shown in Table 3, the root mean square of the center deviation, the pixel with maximum pixel, and the time for performance are computed using the following algorithms.

$$\text{Max}\Delta\text{Pix} = \max \sqrt{(X_t - XO_t)^2 + (Y_t - YO_t)^2}\,, \tag{13}$$

$$\text{RMSquare} = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left[(X_t - XO_t)^2 + (Y_t - YO_t)^2\right]}\,. \tag{14}$$

According to formulae (13) and (14), it can be concluded that the edge detection for the grade of sub-pixel is carried out by using Zernike matrix, and then the error of the ellipse center coordinates by fitting is relatively small, and thus, the precision is higher.

5. Conclusion

Focusing on football match video, a method of scene segmentation is put forward to better know the performance of players in the field. In order to carry out scenes segmentation, some pretreatment works should be done at first. The scenes images are carried out with edge detection and classification from the aspect of characteristic symbols and scene recognition. The algorithm of edge detecting of characteristic symbols solves the center deviation appearing in the circular projection; and the pyramid combined with similar information enhances the accuracy of scene recognition. As for the experiment, the picture pretreated is taken into consideration, and by combining with the scene segmentation algorithm, the segmentation experiments can be made in the condition of different scene pictures, which is the work in the

further study.

References

[1] Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, G. Toderici: *Beyond short snippets: Deep networks for video classification.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, Boston, MA, USA, IEEE Conference Publications (2015), 4694–4702.

[2] H. Yang, C. Meinel: *Content based lecture video retrieval using speech and video text information.* IEEE Transactions on Learning Technologies *7* (2014), No. 2, 142–154.

[3] B. Zhou, A. Lapedriza, J. X. Xiao, A. Torralba, A. Oliva: *Learning Deep features for scene recognition using places database.* Advances in Neural Information Processing Systems 27 (NIPS 2014) (2014), 487–495.

[4] V. Severo, H. A. S. Leitão, J. B. Lima, W. T. A. Lopes, F. Madeiro: *Modified firefly algorithm applied to image vector quantisation codebook design.* International Journal of Innovative Computing and Applications *7* (2016), No. 4, 202–213.

[5] B. Liao, H. Wang: *The optimization of SIFT feature matching algorithm on face recognition based on BP neural network.* Applied Mechanics and Materials *743* (2015), No.Chapter 4, 359–364.

[6] X. J. Peng, L. M. Wang, X. X. Wang, Y. Qiao: *Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice.* Computer Vision and Image Understanding *150* (2016), 109–125.

[7] C. Stentoumis, L. Grammatikopoulos, I. Kalisperakis, G. Karras, E. Petsa: *Stereo matching based on census transformation of image gradients.* SPIE Videometrics, Range Imaging, and Applications XIII *9528Q* (2015).

[8] M. T. El-Haddad, Y. K. Tao: *Automated stereo vision instrument tracking for intraoperative OCT guided anterior segment ophthalmic surgical maneuvers.* Biomedical Optics Express *6* (2015), No. 8, 3014–3031.

Received May 7, 2017