2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Video Attention Saliency Mapping using Pulse Coupled Neural Network and Optical Flow

Qiling Ni
Department of Electronic Engineering
Fudan University
Shanghai 200433 China
12210720035@fudan.edu.cn

Xiaodong Gu
Department of Electronic Engineering
Fudan University
Shanghai 200433 China
xdgu@fudan.edu.cn

***Abstract*—This paper proposes a biologically inspired video attention saliency detector by combining optical flow and topological properties. In this paper, we expound how to utilize feature informations of video and how to use topological property and optical flow based on attention detection for target tracking. Visual attention used in our model is a consequence of tuning of some saliency features such as color, shape and motion. The model (OFTPA-Optical Flow Topological Properties Attention) we proposed for motion attention saliency detection includes two stages. First stage focuses on extracting saliency features, including optical-flow velocity field, topological properties and Intensity, from video. The second integrates the traditional saliency features and an extra position prediction calculated from optical-flow field, to form bottom-up saliency maps which indicate where the object candidates are located. Spatiotemporal saliency maps are obtained from the phase spectrum of a video's hypercomplex Fourier transform. Experimental results show that the OFTPA model takes advantage over other models such as PQFT in complex background.**

Keywords—saliency map; optical flow; topological property; Unit-linking PCNN; video; attention

I. INTRODUCTION

A study predicts that nearly a million minutes of video content will cross the network each second in 2017, and it will take an individual about 2 years to watch. Video data generation rate is considerably larger than video data analysis rate. Video-based object detection has become one of the most challenging problems and has drawn increasing interest for its highly applications [13], such as video surveillance, video indexing and retrieval, communication, traffic control, machine intelligence, biological medical, etc. Therefore, the video attention detection has far-reaching significance and extensive application value. Attention detection can be divided into two basic categories: bottom-up attention and top-down attention [1]. Bottom-up combines multi-features such as color, motion and depth by weights to identify a saliency map. Top-down is an object selection with prior knowledge guide. It is about visual memory and learning.

In this paper, we are aimed at designing the simulation system of human vision by combining velocity, topological properties and position prediction. The paper introduces a Bottom-up model joint optical flow and topological properties for motion attention saliency detection. Optical flow theory which is widely used in moving object tracking is utilized first for motion attention saliency detection in this paper. We combine optical flow with topological properties attention which was proposed in literature [12]. However, literature [12] only applied frame difference for motion detection. The OFTPA model we proposed combines optical flow theory and position prediction for motion attention saliency detection. OFTPA model obtains saliency map by using frequency domain processing method which is proposed by PQFT model [11]. However, PQFT model detects moving object simply by frame difference and it has no topological properties for global object detection. Two main innovations presented by OFTPA model is optical flow and topological property of connectivity. The following sections detail how they contribute to mapping targets.

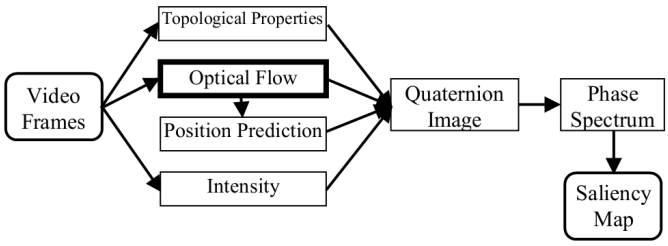

Fig. 1. Structure of OFTPA Model

Fig.1 illustrates the structure of OFTPA model. Firstly, we get three channels (topological properties, optical-flow velocity field and intensity) from video frames and one channel (position prediction) from optical flow field. Secondly, we obtain a quaternion image with four channels. Then, phase spectrum can be obtained by normalizing Fourier transform of quaternion image. Saliency map which indicate where the object candidates are located is obtained from phase spectrum by inverse Fourier transform.

Section II introduces definition and estimation of optical flow. In Section III, we present topological theory in psychological point of view. In Section IV, quaternion and its frequency-domain analysis are described. In Section V, the OFTPA model is presented and analyzed in detail. In Section VI, experimental results show that using OFTPA model, the spatiotemporal saliency maps can reflect the attention

This work was supported in part by National Natural Science Foundation of China under grant 61371148 and Shanghai National Natural Science Foundation under grant 12ZR1402500.

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

distribution more effectively. Conclusion and future work are given in Section VII.

II. OPTICAL FLOW

Optical flow is a widely used method for measurement of target velocity in video by computing difference of frame sequence. Watching video we are more likely to be focused on the moving targets. All things equal, the faster a target moves, the more attractive it looks, for the obvious reasons. Therefore, we believe that target's velocity is one of the factors affecting the attention detection performance. High-speed targets have an advantage in terms of human vision. Therefore, we introduce optical flow channel in order to detect object more effectively. In general, optical flow can be calculated from a sequence of images (i.e. video frame sequence) by relying on three fundamental assumptions [2]:

1) Object brightness invariance: The local changes in image intensity are caused only by the motion of a certain object with respect to the camera.

2) Spatial coherence: The motion is uniform over a small patch of pixels.

3) Temporal persistence: The image motion of a surface patch changes gradually over time.

In general, OF algorithms can be roughly classified into the following categories: "gradient" methods, "phase" methods, "region-based matching" methods and "feature-based" methods. Complex calculation is not required to calculate brightness gradients locally, gradient-based optical flow methods are suitable for real-time optical flow estimation for its ability of avoiding heavy computation. We choose the Lucas–Kanade algorithm [3], a well-known gradient-based algorithm for optical flow estimation in this paper.

$I(x,y,t)$ is the intensity of a pixel at location (x,y) and time t. According to fundamental assumption 1), object brightness does not change between consecutive frames, which leads to the equation,

$$I(x,y,t) = I(x+\delta x, y+\delta y, t+\delta t), \quad (1)$$

In equation (2), I_x and I_y are the derivatives of intensity in the x direction and y direction respectively, calculated at the given pixel location (x,y) and time t. (∂,β) is the velocity field called optical flow at point (x,y) at time t. Here, (∂,β) is assumed to be uniform over a small range L according to fundamental assumption 2), and the following simultaneous equations (3) are obtained from basic equation (2).

$$\begin{cases} S_{xx}\partial + S_{xy}\beta + S_{xt} = 0 \\ S_{xy}\partial + S_{yy}\beta + S_{yt} = 0 \end{cases} \quad (3)$$

In equations (3), $S_{xx} = \sum_{L} I_x I_x$, $S_{xy} = \sum_{L} I_x I_y$, $S_{xt} = \sum_{L} I_x I_t$, $S_{yy} = \sum_{L} I_y I_y$, and $S_{yt} = \sum_{L} I_y I_t$.

Resolving simultaneous equations (3), the velocity can be calculated by equations (4).

$$\begin{pmatrix} \partial \\ \beta \end{pmatrix} = \begin{pmatrix} \dfrac{S_{yy}S_{xt} - S_{xy}S_{yt}}{S_{xx}S_{yy} - S_{xy}S_{xy}} \\ \dfrac{S_{xx}S_{yt} - S_{xy}S_{xt}}{S_{xx}S_{yy} - S_{xy}S_{xy}} \end{pmatrix} \quad (4)$$

III. TOPOLOGICAL PROPERTIES

Chen Lin proposed topological perception theory in 1982 [4]. Topological properties are invariant in topological transformation, which is a one-to-one and continuous transformation. A topological transformation meets the following three conditions:

1) Not produce fracture;

2) Not be binding;

3) Two points can not be composed to one point, the points on graphs before and after transformation are one-one correspondence.

According to topological perception theory, we know that topological properties are more stable than other features and topological invariant perception is the beginning of human vision [4], [5]. Topological properties involve connectivity, and the number of holes, and inside/outside relations. Even if the local geometrical properties of an object change, its topological properties remain unchanged. The early process of visual cognition comes from the global perception which is decided by topological properties. The local geometrical properties are based on the global perception. Topological perception is prior to other perceptions of feature information in visual process. In visual process of human beings, topological properties are perceived first and topological invariance perception decides the separation of objects from backgrounds in visual process. There are enough reasons to presume that topological perception is fundamental to all vision systems. Therefore, we can utilize topological properties and topological perception to detect object in the image sequence. How to express topological properties of object in the image sequence? In this paper, we express one of the topological properties---connectivity by Unit-linking Pulse Coupled Neural Network (PCNN) hole-filter.

PCNN [7], a kind of neural network, developed from Eckhorn's linking field network, has been applied in image processing widely. In order to make PCNN simpler and more practical, a modified version, namely Unit-linking PCNN is presented in literature [8].

Fig.2 is the structure of unit-linking PCNN, quoted from "Image Shadow Removal Using Pulse Coupled Neural Network" [8].In Unit-linking PCNN hole-filter [9], neurons and pixels are one-to-one correspondents. The *j*th pixel's intensity I_j is the input parameter of its corresponding neuron, $F_j = I_j$. L_j gathers spiking information of neighborhood neurons shown in equations (5).

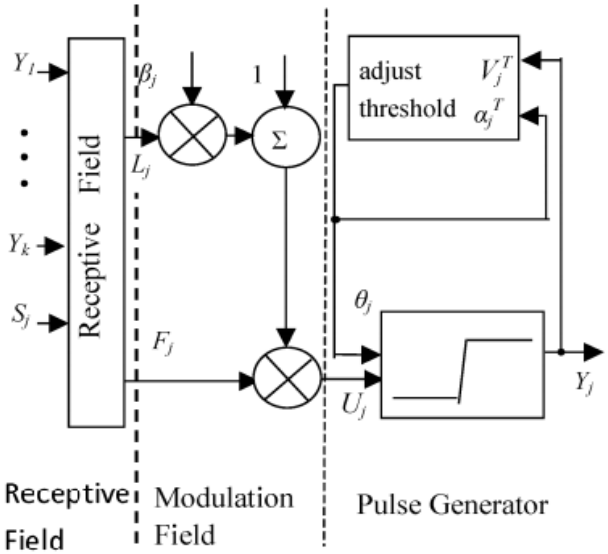

Fig. 2. Structure of Unit-linking PCNN

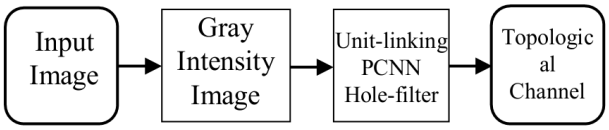

Fig. 3. Structure of Topological Channel

(a) (b)

Fig. 4. An example of Unit-linking PCNN hole-filter [12]

$$L_j = step(\sum_{k \in N(j)} Y_k) = \begin{cases} 1, & if \ \sum_{k \in N(j)} Y_k > 0 \\ 0, & else \end{cases} \quad (5)$$

Then, U_j shows the result of modulation,

$$U_j = F_j(1 + \beta_j L_j). \quad (6)$$

Every neuron in the last step judges whether sparking or not,

$$Y_j = step(U_j - \theta_j) = \begin{cases} 1, & if \ U_j > \theta_j \\ 0, & else \end{cases}. \quad (7)$$

In construction of topological channel, pixels in a gray image, calculated from input color image, connect to unit-linking PCNN neurons one-to-one. Neurons spark one by one and from outside to inside until obstacles are encountered. Fig.3 is the flow chat of topological channel construction.

There are a simple case, where holes of objects are filled, in Fig.4 to show Unit-linking PCNN hole-filter process. The image without holes represents the kind of topological properties---connectivity.

IV. QUATERNION

Quaternion has been applied in image processing [10], which has four dimensions, $q = a + bi + cj + dk$, and has the following properties:

$$jk = i^2 = j^2 = k^2 = -1$$

$$jk = i, kj = -i, ki = j, ik = -j, ij = k, ji = -k$$

$$| q | = \sqrt{a^2 + b^2 + c^2 + d^2}$$

$$\bar{q} = a - bi - cj - dk$$

Fourier transform and Fourier inversion of quaternion can be calculated by followings:

$$F[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu 2\pi((mv/M)+(nu/N))} f(n,m) \quad (8)$$

$$f(n,m) = \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{\mu 2\pi((mv/M)+(nu/N))} F[u,v] \quad (9)$$

Four dimensions of quaternion represent four features, when processing color images. In this paper, to build a quaternion image, we take four features of an image as four channels, optical flow, intensity, position prediction and a kind of topological properties---connectivity to a quaternion, and these features are processed in parallel [6]. The method can reduce operation time effectively and improve operation speed. Utilizing quaternion Fourier transforms, we get the phase spectrum of the quaternion image. A saliency map can be reconstructed from the phase spectrum.

V. OFTPA MODEL

The structure of OFTPA model we proposed is showed in Fig.1. Optical flow and topological properties can be obtained as introduced in section II and section III. Position prediction channel is calculated from optical flow. Euclidean norms of pixels' optical flow are chosen as the input parameters in the quaternion representation. Position prediction is a binary image. Pixel coordinate of position prediction is the product of optical flow (i.e. ∂ and β) and frame difference. Only when the pixel coordinate is prediction position based on optical flow of the last image, can the pixel's value be set to 1 in the current image. Then, we can get an image in which object prediction position is highlighted. What calls for special attention is that every channel's input is a 64*64 matrix, in which every element is normalized respectively.

The purpose of OFTPA model is to obtain the phase spectrum of quaternion images, and then reconstruct saliency maps from it. First, with quaternion Fourier transform of a quaternion image, we obtain its frequency domain expression in polar form; second, the amplitude spectrum is set to 1, which

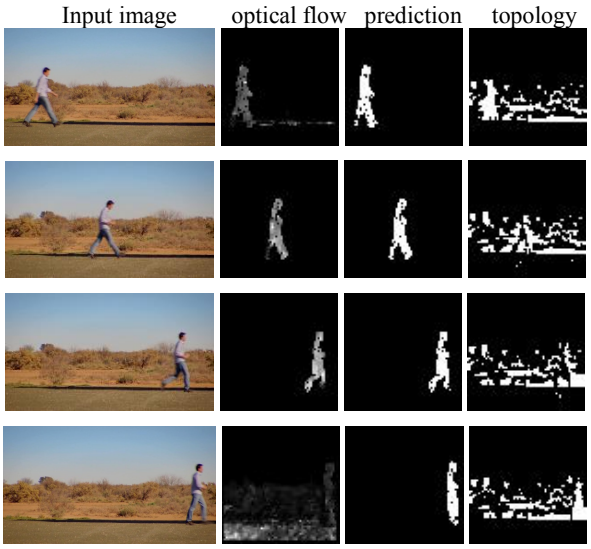

Fig. 5. Structure of Optical flow and Topological Channel

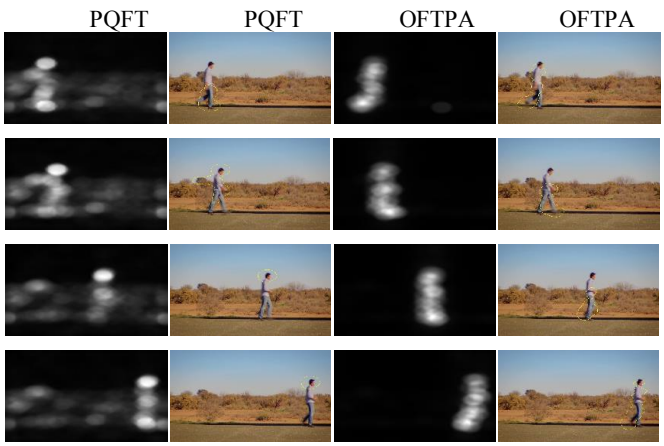

Fig. 6. Saliency maps of PQFT and OFTPA (α=0.5)

Table.1 Number of correct object in 92 frames

TPR	PQFT	OFTPA
0.75	2	27
0.5	20	43
0.3	38	71
0.2	40	83
0.1	82	91

only involves phase spectrum information $P(f)$. Doing inverse Fourier transform gets the spatial domain saliency maps which contain location information of object. In equations (3), G is a 2-dimension filter.

$$sM(x) = G(x) * |F^{-1}[\exp(P(f))]|^2 \quad (10)$$

VI. EXPERIMENT RESULTS

The criterion (11) is used in PQFT model [6] [11], where $O_i^{\max}$ is the biggest value in saliency map. $Mask_i$ is the area of *ith* target. The other criterion is TPR (the true positive rate), where A is attention detection result and G is ground truth.

$$Mask_i = \left\{(x,y) \middle| \alpha O_i^{\max} \le O(x,y) \le O_i^{\max}\right\} \quad (11)$$

$$TPR = \frac{area(A \cap G)}{area(G)} \quad (12)$$

Saliency maps of video frames can reflect the attention distribution. In this experiment, we choose 92 frames of one moving man from 169 image frames. Before operating processing, input images (720*1280) are compressed to 64*64 first, which improve processing speed and reduce processing time effectively. In addition, compression contributes to removing of image noise to improve the system robustness. Optical flow and prediction position are calculated under condition of one-frame-difference.

In Fig.5, there are respectively input images, optical flow, prediction position and topological channel from left to right. In images of optical flow velocity field, the target---moving man, is clearly highlight in most cases, meanwhile, background is inhibited effectively. Even though the current optical flow image is of limited effectiveness about highlighting object in a small number of cases, the position prediction image based on the last image can improve results. In images of topological properties, block objects are highlight, and the target is one of block objects. Topological image exclude some noise such as sky and road.

The first column and third column are saliency maps of PQFT and OFTPA model respectively in Fig.6. We find that our model pays more attention to the moving man, is more efficient than PQFT. OFTPA model has also demonstrated exceptional performance in background suppression. Attention area is highlighted with yellow circle in the second column and fourth column images. Yellow circle shows that OFTPA model we proposed can catch larger areas of object than PQFT. In addition, PQFT tends to catch part of object under simple background, which is poor at inhibiting noise of background. This limitation can be settled well in our model. OFTPA model has an advantage particularly with complex backgrounds. In Table.1, we can find that the number of correct object by OFTPA model is clearly bigger than PQFT in all TPRs. In conclusion, OFTPA model, which is more robust than PQFT model in complex background, processes more reliably and the accuracy of attention saliency detection is higher than PQFT model in all TPRs.

VII. CONCLUSION

In this paper, we propose a biologically inspired video attention saliency detection by using joint optical flow and topological properties. Our OFTPA model, combined optical flow theory with topological properties first for motion attention saliency detection, is effective experimentally. Comparing with current model such as PQFT model, OFTPA

model is more robust in the case of complex background, and proposes well under different criterions. We will make great efforts for attention selection applying in actual applications as target detection or target tracking in our further work. Our further research will use the learning ability of neural networks to improve the performance of motion attention selection to extract the salient target from complex background more efficiently.

REFERENCES

[1] J. M. Wolfe, "Guided Search 2.0: Arevised model of visual search," Psychonomic Bulletin & Review, vol. 1, No.2, pp.202-238, 1994.

[2] M Mammarella, and G Campa, "Comparing Optical Flow Algorithms Using 6-DOF Motion of Real-World Rigid Objects," IEEE Trans. System, Man, And Cybernetics, Vol. 42, No. 6, November 2012, pp.1752–1762.

[3] B. Lucas and T. Kanade, "An iterative image registration technique with applications in stereo vision," inProc. DARPA Image Understand. Workshop, 1981, pp. 121–130.

[4] L. Chen, "Topological Structure in Visual Perception," Science, vol. 218, pp. 699-700, Nov. 1982.

[5] L. Chen, "The Topological Approach to Perceptual Organization," Visual Cognition, vol. 12, pp. 553-637, Apr. 2005.

[6] C. L. Guo, Q Ma, and L. M. Zhang, "Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform," in 2008 CVPR, pp. 1-8.

[7] J. L. Johnson and D. Ritter, "Observation of periodic waves in a pulse-coupled neural network," Opt. Lett. , vol. 18, no. 15, pp. 1253–1255,1993.

[8] X. D. Gu, D. H. Yu, and L. M. Zhang, "Image shadow removal using pulse coupled neural network," IEEE Trans. Neural Networks, vol. 5, pp. 692-698, May 2005.

[9] X. D. Gu, D. H. Yu, and L. M. Zhang, "General Design Approach to Unit-linking PCNN for Image Processing," in 2005 Proc. IJCNN , pp. 1836-1841.

[10] Todd A. Ell and Stephen J. Sangwine, "Hypercomplex Fourier Transforms of Color Images," IEEE Trans. Image Processing, vol. 16, pp. 22-35, Jan. 2007.

[11] C. L. Guo and L.M. Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," IEEE Trans. Image Processing, vol 19, pp. 185-198, Jan. 2010.

[12] X. D. Gu, Y. Fang, Y. Y. Wang, "Attention selection using global topological properties based on pulse coupled neural network," Computer Vision and Image Understanding, 2013, vol. 117, pp. 1400-1411.

[13] Kim, W., Kim, C., "Spatiotemporal Saliency Detection Using Textural Contrast and Its Applications," IEEE Trans. Circuits and Systems for Video Technology, vol. 24, pp. 646 – 659, April 2014.