Optik 125 (2014) 6681–6686

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Moving target detection approach based on spatio-temporal salient perception

Gang Jin [a], Zhengzhou Li [b,c,*], Yuanshan Gu [b], Jialing Li [b], Dong Cao [a], Linyan Liu [a]

[a] China Aerodynamics Research and Development Center, Mianyang 621000, China
[b] College of Communication Engineering, Chongqing University, Chongqing 400030, China
[c] Key Laboratory of Beam Control, Chinese Academy of Sciences, Chengdu 610209, China

ARTICLE INFO

Article history:
Received 20 November 2013
Accepted 19 June 2014

Keywords:
Moving target detection
Spatial salient maps
Motion salient map
Spatio-temporal salient map

ABSTRACT

The differences in texture and motion between man-made object and natural scene are the key features for human biological visual system to detect moving object in scenery. The paper proposed a moving target detection approach based on spatio-temporal perception, which is a crucial function of the visual attention mechanism. The spatial feature including edge, orientation, texture and contrast of the image are extracted, and then the corresponding spatial salient map are constructed by fusing the features through difference of Gaussian (DOG) function, which can suppress the common and enhance the difference of local region. Then, the global motion, local motion and relative motion between continuous images are extracted by means of pyramid multi-resolution, and the moving salient map is constructed after the motion difference between moving target and background is confirmed. Finally, the spatio-temporal salient map is constructed by fusing the spatial salient map and the moving salient map through competition strategy, and the moving target could be detected by searching the maximum in the spatio-temporal salient map. Some experiments are included and the results show that the method can accurately detect the moving target in complex background.

1. Introduction

Fourier transform [1,2], optical flow [3–5] and feature matching [6,7] are the major methods to detect moving target in complex scene. Even some features are extracted; these methods still have poor performance of target detection and tracking due to low contrast between target and background, complex shape or changing pose of target. How to effectively detect, identify and track the target submerged in complex scene is always the challenge for the photoelectric detection system, and it is crucial to explore novel ideas to achieve a breakthrough [8,9].

Visual psychology studies have shown that the selective attention mechanism of human visual system has two sub-processes, namely pre-attentive mechanism and attention mechanism [10–12]. The pre-attentive mechanism is built by computing the salience of objects in the image sequence according to the bottom-up strategy. It belongs to the low-level cognitive processes. The attention mechanism focuses on the specific target by adjusting the selection criteria to meet the external requirements according to the top-down strategy. It belongs to the high-level cognitive processes [13,14]. The study about the selective attention mechanism mainly focuses on the pre-attentive mechanism, and gets many achievements including the computational models and the physical theory of the bottom-up strategy, but the study about the attention mechanism is few relatively due to the question how the external command participates in the calculation of attention. And so, the application in image processing study about the selective attention mechanism concentrates more on the bottom-up strategy [15–18].

The computational model of visual selective attention proposed by Itti is the representative for the bottom-up strategy, and it can achieve outstanding performance in detecting salient area for a static image [19–24]. Otherwise, it is hard to detect salient area or moving target in image sequence for the moving feature of target is rarely extracted and integrated into Itti model [18]. Actually, there are two paths in the simple cell of the visual cortex, namely, "what" path and "where" path [25]. The "what" path usually perceives the shape, color, texture and so on. The "where" path is sensitive to velocity and direction of moving target [26–28]. Therefore, it is necessary to reduce the difference between the Itti model and visual physiological system, and enhance the performance of moving target detection and tracking inspired by the visual psychology.

* Corresponding author at: College of Communication Engineering, Chongqing University, Chongqing 400030, China.
E-mail address: lizhengzhou@cqu.edu.cn (Z. Li).

http://dx.doi.org/10.1016/j.ijleo.2014.08.051
0030-4026/

It is necessary to extract the existing differences including the movement between the moving target and the background for identifying target from the background. A moving target detection approach based on the visual spatio-temporal salient perception is proposed in this paper. This approach references the Itti model and integrates with motion feature. The primary visual features of the image including color, intensity and orientation are extracted. Then the visual differences at variable scales are calculated by Center-Surround operator to form the spatial salient map. Subsequently, the salient regions in consecutive images are matched through the pyramid multi-resolution strategy, and the velocity of salient regions would be extracted, namely, global motion, local motion and relative motion. And the difference between moving target and background would be confirmed, and then the motion salient map is further built. Finally, the spatio-temporal map is constructed by competing and fusing the spatial salient map and motion salient map, and then the moving target could be enhanced and detected in spatio-temporal map. Theoretical analysis and experimental results show that this approach can quickly and accurately detect the target, and could improve the performance of target tracking.

2. The spatio-temporal salient perception model

The study about retina, optic nerve and visual cortex by nerve physiologist has shown that there are three paths in the visual attention system. Two paths jointly process static image, such as extract contour, edge and even fill the contour, and one path usually perceive the movement of the scene [28]. Therefore, motion perception is the important function of biological visual system. Neither of movement and changing shape of object has a great influence on object perception of human visual system.

In view of motion perception function of the visual nervous system, a method to detect moving target based on spatio-temporal salient map is proposed in this paper.

2.1. Spatial salient map

These spatial features such as edge, orientation, texture and contrast of the image are extracted, and the saliency of the pixel of image is calculated to form the spatial salient map using a local iterative approach. The model to calculate the spatial salient map is similar to the Itti model.

2.1.1. Visual features extraction

There is a lateral inhibitory effect during the primary information processing stage of visual system. This effect can extract the contrast of adjacent pixels by two different kinds of sensitive neurons, namely, central dark surrounding periphery light and central light surrounding dark periphery. The effect has the same function as the edge feature extraction. Given that the difference operator between the center and periphery is Θ, the edge extraction could be expressed by the following formula.

$$E(x, y; c, e) = \left| I(x, y, c) \Theta I(x, y, e) \right| \tag{1}$$

where $I(x, y, c)$ and $I(x, y, e)$ denote the image $I(x, y)$ with the scale of c and e, respectively, and x and y are image axis.

To simulate the function that these orientation-sensitive neurons of visual system can extract feature with variable orientation, the Gabor filters with four different orientations are introduced to filter image, and these four filtered edge images with orientation information are called the orientation feature maps

$$O_\theta(x, y) = G(x, y; \theta) * I(x, y) \tag{2}$$

where * denotes convolution, $G(x, y; \theta)$ is Gabor filter defined as

$$G(x, y; \theta) = \cos\left[2\pi(x\cos\theta + y\sin\theta)\right] \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \tag{3}$$

where σ_x^2 and σ_y^2 are the definition domain at x and y axis, θ is the direction. In this paper, θ is set as 0°, 45°, 90° and 135°, and these four edge image filtered by the Gabor filter with four different directions are denoted as $O_0(x, y)$, $O_{45}(x, y)$, $O_{90}(x, y)$ and $O_{135}(x, y)$. These four edge images could be added up linearly to get edge image $O(x, y)$.

Texture is the repeating patterns of local variations in image intensity. The gray level co-occurrence matrix GLCM of image $I(x, y)$ is calculated, and then second-order statistics of GLCM, such as correlation, entropy and contrast could be extracted to form the texture map $T(x, y)$ in this paper. For a gray-scale image with gray level L, the GLCM P_δ is defined as

$$P_\delta = \left\{ P_\delta(i, j) \,\middle|\, i, j \in [0, L-1] \right\} \tag{4}$$

where $P_\delta(i, j)$ is the element of GLCM, i and j are the gray value of two pixels with the distance δ, respectively.

The correlation coefficient is used to measure the similarity degree between the elements of GLCM at the direction of row and column, and it is denoted as

$$F_{correlation} = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} i \times j \times p_\delta(i, j) - u_x u_y}{\sigma_x^2 \sigma_y^2} \tag{5}$$

where u_x, u_y, δ_x and δ_y represent the means and the variances at the direction of row and column of GLCM, respectively.

The entropy describes the information amount possessed by image. If the scene is full of fine texture, the value of $p_\delta(i,j)$ is approximate equal, and the entropy will get the maximum value. And if the image has less texture, the value of $p_\delta(i, j)$ varies greatly, and the image entropy is the smaller. The entropy $F_{entropy}$ of image is defined as

$$F_{entropy} = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P_\delta(i, j) \log\ P_\delta(i, j) \tag{6}$$

The contrast can be understood as clarity of a scene, that is, the clarity of texture. Deeper the grooves of texture are, the greater its contrast ratio $F_{contrast}$ is, and the more clear visual image is. The contrast of image is defined as

$$F_{contrast} = \sum_{n=0}^{L-1} n^2 \left\{ \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} P_\delta^2(i, j) \right\} \tag{7}$$

where n is the absolute value of the difference between gray value i and j.

According to the definition of texture features above, the stronger the texture features subjectively is, the bigger the correlation coefficient $F_{correlation}$, the entropy $F_{entropy}$ and the contrast $F_{contrast}$ are. Without prior knowledge about the circumstances, these three features could be added up to form the texture feature as follows

$$F = F_{correlation} + F_{entropy} + F_{contrast} \tag{8}$$

The texture feature of the natural scenery usually is more obvious than the man-made objects, and so the value F of natural scenery is greater than that of man-made object.

2.1.2. Spatial feature map generation

After these spatial feature salient maps are extracted above, regularize every salient map and form the normalized edge feature map $E(x, y; c, e)$, orientation feature map $O(x, y)$ and texture feature