ATLANTIS PRESS
Advances in Computer Science Research (ACSR), volume 76
7th International Conference on Education, Management, Information and Mechanical Engineering (EMIM 2017)

Improved Spatio-temporal Context Target Tracking Algorithm

Yuanhang Shi[1, a], Junwei Zang[1, b] and Yuhuai Liu[1, 2, c]
[1]Institute of Information Engineering, Zhengzhou University, China
[2]Industrial Technology Research Institute, Zhengzhou University, China
[a]325871870@qq.com, [b]1506877508@qq.com

Keywords: Spatio-temporal context; SURF feature point; Target tracking; Obscure

Abstract. Visual tracking is one of the hot topics in the field of computer science. However, the effective use of context information can improve the robustness of video tracking due to factors such as illumination and rotation. The traditional spatio-temporal context algorithm does not detect the validity of the tracking result, so when the target is obscured for a long time, the tracking target is easily updated wrongly. Based on SURF feature point detection, an improved spatio-temporal context tracking algorithm is proposed. Using the SURF algorithm to extract the feature points from the initial model of the target as the evaluation criteria, the tracking target is evaluated and the target can be updated according to the standard when the target is blocked for a long time. Experiments show that the proposed algorithm can accurately update the target model when the target is obscured for a long time, and achieves reliable tracking.

Introduction

With the researchers on computer vision research, the development of computer vision has been developed by leaps and bounds. In the behavior-based interaction, video recognition widely used, making it a very popular topic nowadays. At present, video tracking can achieve ideal results in a controlled environment. However, robust, real-time target tracking still faces great challenges in uncontrolled environments, such as scale change, illumination change, local or global change.

In recent years, researchers have proposed a large number of visual tracking algorithm[1-5],Such as TLD [1], MIL [2], VTD [3]and other algorithms. In order to improve the robustness and accuracy of video tracking, many scholars put forward the use of temporal and spatial context correlation to improve the tracking effect. In references [6] some key points around the target are acquired. When the target tracking is lost, the lost target can be found by the key point. In references [7]through modeling the target time space, the SURF[8]feature points are used to build the support domain. The above algorithms are complex and inefficient in context. Zhang [9] proposed a Spatio-temporal context (STC) algorithm, which introduces the biological vision system, considering the surrounding context of the target, the position of the target in the next frame can be predicted by the relationship between the spatio-temporal relationship and the biological vision system, so that the tracking accuracy has been greatly improved and achieved very good results.

However, STC is also inadequate, it will not verify the validity of the surrounding context, that the surrounding background of the contribution of the target is the same, So when the target is blocked for a long time, very easy to follow the lost or drift. According to the above-mentioned shortcomings of the spatio-temporal context, this paper proposes a spatio-temporal context algorithm based on the feature points, and judges whether the spatio-temporal context algorithm is correct by the validity of the target feature points, When the target is obscured for a long time, after the target is lost, it can be retrieved according to the search feature points to improve the tracking accuracy. Experiments show that the target can be effectively retrieved after the target is lost.

Introduction to Spatio - Temporal Context Algorithm

Spatio-temporal context is the use of space on the target and the surrounding background of a region and the correlation between adjacent frames to determine the location of the next target

frame, and the maximum confidence position as the target location. The tracking problem is described as a confidence graph that computes the tracking target

$$c(x)= P(x|o) \tag{1}$$

Where $x \in R^2$ is the target position and o is the position in the scene. The tracing target center position is x*, The context feature set is define as

$$X^c = \{c(z) = ((I(z), z) \mid z \in \Omega_c(x^*)\},$$

Where $I(z)$ denotes the gray value at z, $\Omega_c(x^*)$ is the local context of the central location x *,The objective function can be converted from conditional probability

$$c(x) = \sum_{c(z)\in X^C} P(x, c(z) \mid o) = \sum_{c(z)\in X^C} P(x \mid c(z), o) P(c(z) \mid o) \tag{2}$$

It can be seen that the likelihood function can be divided into two parts, $P(x \mid c(z), o)$ is the spatial relationship between the target and the surrounding context, $P(c(z) \mid o)$ is the context a priori probability, for the context of the establishment of the model, where $P(x \mid c(z), o)$ is we need to learn it. The spatial context is defined as:

$$P(x \mid c(z), o) = h^{sc}(x - z) \tag{3}$$

$h^{sc}(x-z)$ is a function of the relative distance and direction of the target x and the local context position z, This function is a non-mirrored function that helps avoid interference with similar objects on the target. The prior function $P(c(z) \mid o)$ is defined as

$$P(c(z)|o) = I(z)\omega_\sigma(z - x^*) \tag{4}$$

Where $I(z)$ is the gray of point z and describes the appearance of this context z. ω Is a weighting function, z is closer to x, and the weight is larger. Defined as follows: $\omega_\sigma(z) = ae^{\sigma^2}$. a is a normalization constant and is a scale parameter. For the confidence graph function in Eq. (1), we can use a presence function to represent it

$$c(x) = b \times \exp(-|\frac{x - x^*}{\alpha}|^\beta) \tag{5}$$

Substituting (2) - (5) into (1) can be obtained

$$c(x) = be^{-\left|\frac{x-x^*}{\alpha}\right|^\beta} = \sum_{c(z)\in X^c} h^{sc}(x - z) I(z)\omega_\sigma(z - x^*) = h^{sc}(x) \otimes I(x)\omega_\sigma(x - x^*) \tag{6}$$

Where $\otimes$ denotes the convolution and the convolution operation is computationally large and can be accelerated by fast Fourier transform. After introducing the fast Fourier function,

$$h^{sc}(x) = F^{-1}\left(\frac{F\left(be^{-\left|\frac{x-x^*}{\alpha}\right|^\beta}\right)}{F(I(x)\omega_\sigma(x - x^*))}\right) \tag{7}$$

Tracking Process

For the first frame, we get the temporal context model of the target, and for the t frame, the position of the tracked object is x_t^*, then in the t + 1 frame, the space-time context model is

$$H_{t+1}^{sc}(x) = (1 - \rho)H_t^{sc}(x) + \rho h_t^{sc}(x) \tag{8}$$

Where ρ is a learning factor, The position of the target at t + 1 frame The new confidence graph is defined as follows:

$$x_{t+1}^* = \arg\max c_{t+1}(x) \tag{9}$$

Which $c_{t+1}(x)$ is expressed as:

$$c_{t+1}(x) = F^{-1}(F(H_{t+1}^{STC}(x)) \bullet F(I_{t+1}(x)\omega_{\sigma_i}(x - x_t^*)) \tag{10}$$

Substituting Eq. (8) into (10), we can get the target confidence graph extremum of the next frame.

Improved Spatio - Temporal Context Algorithm

SURF Feature Points Are Introduced. Since the temporal and spatial context uses the target position of the current frame to predict the image of the next frame without any detection of the target, the object is easily shifted and even lost following the occlusion. Based on the above analysis, in order to reduce the target is obscured the impact of tracking. In this paper, the introduction of accelerated robust features (SURF), the current frame of the target detection, combined with STC algorithm to achieve the target tracking.

Lowe[10]put forward SIFT(Scale Invariant Feature Transform) algorithm, SIFT feature points have good adaptability under the conditions of image rotation and scale transformation. However, SIFT algorithm has a large computational cost and a long time consuming. Therefore, Bay et al. Proposed an improved SIFT-based algorithm, SURF, which surpasses the SIFT algorithm in all respects, and the computational speed is three times that of SIFT. The SURF matrix uses the Hessian matrix to extract feature points. First, Define a Hessian Matrix at a Point

$$H(i, j, \sigma) = \begin{bmatrix} L_{xx}(i, j, \sigma) & L_{xy}(i, j, \sigma) \\ L_{xy}(i, j, \sigma) & L_{yy}(i, j, \sigma) \end{bmatrix} \tag{11}$$

Where $L_{xx}(i, j, s)$ is the result of convolution of the Gaussian second-order differential function $\frac{\partial^2 g(\sigma)}{\partial x^2}$ and point p, and when the local value of the Hessian matrix is maximum, the detected point is the point of interest.

SURF in the feature detection process does not need to SIFT as to directly create a pyramid image, so that the target template on the image filtering only need addition and subtraction operations, greatly reducing the operation speed. Euclidean distance is used to measure the similarity between feature points to find matching feature points.

Algorithmic Flow. In order to ensure reliable tracking of the target, the tracking result needs to be updated online. The traditional STC algorithm can not correct the drift after the target is obscured. This paper uses the SURF feature point to extract the search target to correct the initialization STC algorithm, Under the track can not be recovered after the loss of the problem.

In this paper, the target is selected manually, the target is initialized by the STC algorithm, the target feature points are extracted by the SURF algorithm, and the target is evaluated every 10 frames. The tracking target is matched with the number of the feature points of the template. If the matching result is valid, the tracking is continued. When the target is obscured or completely disappeared, the image is searched through the search mechanism. If the target is successfully searched, the target is updated and reinitialized, and the next frame is processed by the STC algorithm.

Experimental Results and Analysis

The experimental computer used for the Intel Core i3 2.83GHz, 4GB memory, the software environment for the Windows 7 operating system, Visual Studio 2010 and opencv2.4.11. T o verify the effectiveness of this algorithm, this algorithm and STC algorithm in the video to test the comparison, the video is the background of the laboratory, the character as the t arget resolution of 640*360.

Figure 1 and Figure 2 is the algorithm and STC algorithm in the video in the comparison, the video selected the upper body as the target, Figure 25, the 25th frame target in the fast-moving time, this algorithm Completed the tracking task, to avoid the tracking box drift. When the target moves to the semi-occluded state and attitude changes, this algorithm also achieved a robust tracking effect,

as shown in Figure 1, the video object will appear for a long time in the block, 319 frame when the target appears. When the video runs to 328 frames, the algorithm initially locates the target, but because of the inaccurate recognition of the deformed target, the target is repositioned and tracked at the 362nd frame. As shown in the figure, the algorithm can block the long- Reacquire the target location. Figure 2 shows the tracking effect of the STC algorithm in the video. It can be seen that the target tracking fails after the target is blocked

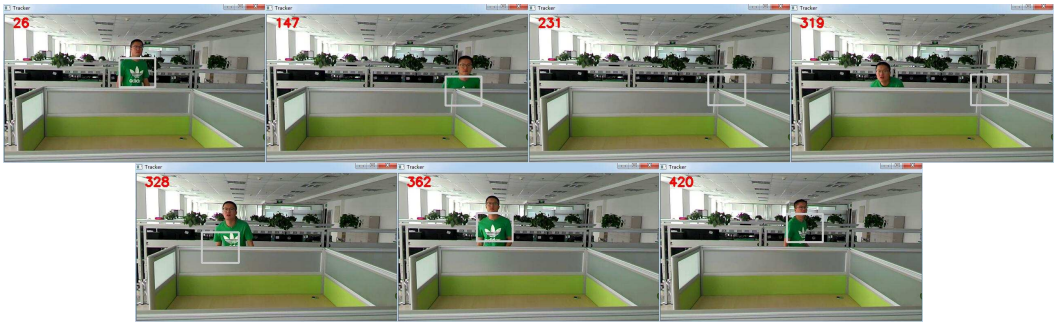

Fig.1 The algorithm for this video tracking results

Fig.2 STC tracking algorithm for video tracking results

Summary

In this paper, we propose an improved spatio - temporal context tracking algorithm based on detection and STC tracking, aiming at the problem that spatio - temporal context is easy to drift and sensitive to occlusion. Experiments show that the algorithm proposed in this paper can effectively solve the problem of long time occlusion and the drift and dropping of the target in the complex environment. However, the tracking algorithm has high requirements on real-time, so it can not meet the requirement of real-time when the target is lost. This is the subject that needs to be studied and studied in the next step.

References

[1] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409−1422

[2] Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning. IEEE Transactions On Pattern Analysis and Machine Intelligence, 2011, 33(8):1619−1632

[3] Kwon J, Lee K M. Visual tracking decomposition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 1269−1276

[4] Wu Yi, Lim J, Yang M. Online object tracking: a benchmark[C]//Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.

[5] Yilmaz A, Javed O, Shah M. Object tracking: a survey [J].ACM Computing Surveys, 2006, 38(4): 1-45.

[6] Grabner H, Matas J, Van Gool L, Cattin P. Tracking the invisible: learning where the object might be. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 1285−1292
[7] Wen L Y, Cai Z W, Zhen L, Dong Y, Li S Z. Online spatio-temporal structural context learning for visual tracking. In Proceedings of the 2012 European Conference on Computer Vision (ECCV). Florence, Italy: Springer, 2012. 716−729
[8] BAY H, TUVTELLARS T, GOOL L V. SURF: speeded up robust features [C]. Proceedings of the European Conference on Computer Vision, 2006:404-417.
[9] Zhang K H, Zhang L, Liu Q S, Zhang D, Yang M H. Fast visual tracking via dense spatio-temporal context learning. In: Proceedings of the 2014 European Conference on Computer Vision (ECCV). Czech Republic: Springer, 2014. 127−141
[10] Lowe D G. Distinctive image features from scale-invariant keypoints [J].Int.J.Comput. Vis., 2004,60(2):91-110