Target Tracking Algorithm Suitable for Fisheye Camera in Vehicle Environment

Jinghua Dong¹, Haiping Wei¹, Hongfei Yu¹* and Tingting Yang¹

¹School of Computer and Communication Engineering, Liaoning Shihua University, Fushun 113001, Liaoning, China

*1009930508@qq.com

Abstract. Moving target tracking is one of the core issues in the fields of automobile assisted driving and autonomous driving. This algorithm proposes a target tracking algorithm suitable for fisheye cameras in vehicle environment. A background point removal method based on median ordering is proposed to optimize the optical flow method to solve the problem of target scale change in fisheye camera videos. Multi-scale scaling processing for frames, tracking algorithm based on cyclic matrix structure (CSK), target tracking for multiple images after zooming, using fitting to determine the maximum corresponding value, reducing the error caused by target deformation on tracking results. The tracking test results of multiple fisheye camera videos in the vehicle environment show that the algorithm has high real-time tracking accuracy.

1. Introduction
Vision-based object detection is an important issue in the application scenarios of automobile assisted driving and autonomous driving. The fisheye camera has many advantages such as richer information acquisition, convenient and easy installation, low price, etc. So it has broad application prospects in target tracking. Compared with ordinary cameras, there are some difficulties in fisheye camera target tracking. Because the tracking target moves faster in the vehicle applications, the algorithm has higher real-time requirements. The size of the target object changes with the movement. When tracking the target, it is necessary to solve the problem of the scale change of the tracked target. Fish-eye camera causes deformation problems. The target tracking to solve the deformation problem.

According to the different modeling methods of target tracking methods, visual target tracking methods are divided into generative model method and discriminative model method [1]. The target area is modeled in the current frame, and the next frame is to find the most similar area to the model is the predicted position. The more famous ones are Kalman filter, particle filter, mean-shift, etc. Classic discriminative methods recommend Struck and TLD. In addition, there are two types of deterministic motion estimation methods and random motion estimation methods according to whether motion estimation is random [2,3]. The deterministic tracker includes a mean shift algorithm that uses a fixed area size iteration to measure maximum similarity. The CSK algorithm is a high-speed tracking algorithm proposed by Henriques et al. [4] in 2012. In the tracking algorithm evaluation[5] of CVPR2013, the algorithm has shown excellent performance in terms of real-time performance, anti-occlusion and background interference. The CSK tracker does not have the processing power of scales [6], it is impossible to accurately track the target that has undergone scale changes and deformation in the fisheye camera video.
Aiming at the above problems, this paper proposes a CSK algorithm combined with optical flow algorithm to track the target of fisheye camera video in vehicle environment. The optical flow algorithm is used to process the frames to calculate the target scale. And a background point removal method based on median ordering is proposed, so that as many feature points as possible are selected on the foreground, and the calculation error is reduced. Due to the characteristic that the target captured by the fisheye camera will deform during the movement, the frame is scaled at multiple scales to get five images. And each image is tracked using the CSK algorithm to get response value. After obtaining each group of data, they are fitted to determine the maximum corresponding value, and the target center position coordinate update of the current frame is performed. The scaling process of the picture can ensure that the entire matrix of the target is included in the fixed area size of the CSK algorithm, and the entire matrix is used for classifier training and iteration. At the level of real-time requirements, the efficient CSK algorithm is adopted to meet the real-time requirements of automotive applications.

2. Scale adaptive target tracking

2.1. Target scale calculation

In the current frame, the detected image area is \( p (m \times n) \). Cur is the feature points matrix selected in the \( p \) in the current frame. Pre is the feature points matrix in the same area in the previous frame. The number of feature points is \( I \). Suppose in the frame \( t-1(t>1) \), the positions of feature points a and b are \((x^a_{t-1}, y^a_{t-1})\) and \((x^b_{t-1}, y^b_{t-1})\). In the frame \( t(t>1) \), the positions of feature points a and b are \((x^a_t, y^a_t)\) and \((x^b_t, y^b_t)\). The distance between a and b points in frame \( t-1 \) is \( L_{t-1} = [(x^a_{t-1} - x^b_{t-1})^2 + (y^a_{t-1} - y^b_{t-1})]^ {1/2} \). The distance between a and b points in frame \( t \) is \( L_t = [(x^a_t - x^b_t)^2 + (y^a_t - y^b_t)]^{1/2} \). Use \( L_{t-1} \) and \( L_t \) to get the scale change \( R \):

\[
R = \frac{L_t}{L_{t-1}}
\]  

(1)

The matrix \( R_{all} \) stores the scale change \( R \) between any two feature points, and takes the median value of the matrix \( R_{all} \) as the scale change \( R \) generated by the moving target in the SZ region from frame \( t-1 \) to frame \( t \).

For I feature points, there is \( \text{CurTrans} = (1/R) * \text{Cur} \). Cur indicates the position of I feature points in the frame \( t \). CurTrans indicates the position of I feature points which are subjected to \((1/R)\) scale processing in frame \( t \). Find the full position offset \( T_{all} \) of I feature points:

\[
T_{all} = \text{Pre} - \text{CurTrans}
\]  

(2)

Since the algorithm is applied in the vehicle environment, the situation where the tracking target rotates is excluded. Therefore, the median value of the matrix \( T_{all} \) is selected as the position offset of the moving target from frame \( t-1 \) to frame \( t \) in the area \( p \), and is denoted as \( T \).

When the feature points are located on the background, it will cause greater interference to the result obtained. This paper proposes a median-ranked background point removal method to remove selected feature points on the background. The corresponding positional relationship of the I feature points in the frame \( t-1 \) and frame \( t \) is as follows:

\[
\begin{pmatrix}
X^t_{t-1} \\
Y^t_{t-1}
\end{pmatrix} = R
\begin{pmatrix}
X^t_t \\
Y^t_t
\end{pmatrix} + T
\]  

(3)

Where \( \begin{pmatrix}
X^t_t \\
Y^t_t
\end{pmatrix} \) indicates the positions of the feature points in frame \( t-1 \). The positions of the feature points in frame \( t-1 \) are obtained \( \begin{pmatrix}
X^t_{t-1} \\
Y^t_{t-1}
\end{pmatrix} \) according to the equation (3). Calculate the residual \( \bar{\beta} \) using actual positions \( \begin{pmatrix}
X^t_{t-1} \\
Y^t_{t-1}
\end{pmatrix} \) of the point in the matrix Pre:

\[
\bar{\beta} = \begin{pmatrix}
X^t_{t-1} \\
Y^t_{t-1}
\end{pmatrix} - \begin{pmatrix}
X^t_{t-1} \\
Y^t_{t-1}
\end{pmatrix}
\]  

(4)
For I feature points, use equation (4) to get I residuals $\hat{\beta}$ and rank, median is $\hat{\beta}_{\text{mid}}$. When the residual value of the feature points is greater than $\hat{\beta}_{\text{mid}}$, these feature points are removed from the Cur and the Pre. Use R and T and combine the target center of the previous frame $\text{Pos}_{t-1}$ to calculate the current frame target position $\text{Pos}_t$:

$$\text{Pos}_t = R \times (\text{Pos}_{t-1} - T)$$  \hspace{1cm} (5)

### 2.2. Target scale calculation

The scaling process is shown in the figure 1. The red frame is the target tracking frame, and the green frame is the sample area collected by the training classifier. Area p and p1 are the training sample areas of frame t-1 and frame t in the video. The size of p is $m \times n$, and the size of p1 is $M \times N$. The yellow dot indicates the target center position. The scale change of the moving target from the frame t-1 to the frame t is R. After scaling processing, the size of p1 is the same as p, and the area p1=$R \times p$ is transformed into: p1=$R \times p \times (1/R) = p$. Make the training area size of the sample samples consistent.

![Figure 1. Image scaling](image)

Due to the imaging characteristics of the fisheye camera, distortion will occur during the movement of the shooting target. As the target leans into the fisheye camera, the distortion will increase. The scale change $R$ obtained by the optical flow algorithm will be disturbed by the distortion of the target. In order to solve the interference caused by the distortion of the target on the scale change, a multiscale scaling method for the current frame is proposed in this experiment. The selection rules of the scale $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ are:

$$\begin{align*}
\alpha_1 &= 1 + (R - 1) \times 0.8 \\
\alpha_2 &= 1 + (R - 1) \times 0.9 \\
\alpha_3 &= 1 + (R - 1) \times 1.0 \\
\alpha_4 &= 1 + (R - 1) \times 1.1 \\
\alpha_5 &= 1 + (R - 1) \times 1.2
\end{align*}$$  \hspace{1cm} (6)

Where $R$ ($R > 1$) is the scale change obtained by the optical flow algorithm after the background point is removed. Scale the frame t according to five scales of $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ to generate five images $t_1, t_2, t_3, t_4, t_5$ about the frame t.

### 2.3. CSK target tracking

In order to meet the real-time requirements of tracking, most discriminative tracking algorithms use sparse sampling strategies for the background and targets, which restricts the tracking performance [4].

In the classifier training and update phase, CSK uses Ridge Regression to learn the classifier [7], samples the current frame, and finds the classifier parameter $\omega$ with the smallest loss function:

$$\min_{\omega} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| < \varphi(g_{m,n}) \omega > - s(m,n) \right|^2 + \lambda ||\omega||$$  \hspace{1cm} (7)
Where $G = \{g_{m,n}\}$ is the sample set required to train the classifier. $s_{(m,n)}$ is Gaussian function. $P$ is the sample collection area of size $M \times N$. $\varphi (p_{m,n})$ maps input $p_{m,n}$ into a high-dimensional feature space. $H$ is defined by $k(x,x') = \langle \varphi (x), \varphi (x') \rangle > \lambda$ is the regular term. The extended solution of the equation is $\omega = \sum \alpha p_{m,n}$. The solution is implicitly represented by the vector $\alpha$ , element coefficients are $\alpha_{m,n}$. Solving classifier parameter $\alpha$ using the K of the circular matrix:

$$\alpha = F^{-1} \frac{F(s)}{F(k)+\lambda}$$  \hfill (8)

Where $F$, $F^{-1}$ stand for Fourier transform and inverse Fourier transform. $K$ is a matrix with elements $k_{(m,n,p)}$.

In the fast detection phase, the response value of each position is calculated for the sampled $p$ area of five images, the size of region training classifier area $p$ is $m \times n$. The position with the maximum response value is the center of the current picture target. Collect samples in the area $p$ to train and update the classifier, and use the Gaussian kernel function to calculate the corresponding values based on the classifier parameters:

$$k = \exp \left( - \frac{1}{\sigma^2} \left( ||x||^2 + ||x'||^2 - 2F^{-1}(F(X)\otimes F(X')) \right) \right)$$  \hfill (9)

$$y = F^{-1}(F(k)\otimes F(\alpha))$$  \hfill (10)

Where $k$ is a matrix with elements $k_{(m,n,p)}$ and $y$ is the position of the current target. The response values of the five images are $y_1, y_2, y_3, y_4, y_5$. Use quadratic polynomial function to fit data series $(\alpha_1, y_1), (\alpha_2, y_2), (\alpha_3, y_3), (\alpha_4, y_4), (\alpha_5, y_5)$. Suppose the obtained fitting function is $P(\alpha) = b_0 + b_1 \alpha + b_2 \alpha^2$, and the mean square error between the fitting function and the data series is $Q(b_0, b_1, b_2) = \sum_{i=1}^{5} (P(\alpha_i) - y_i)^2 = \sum_{i=1}^{5} (b_0 + b_1 \alpha_i + b_2 \alpha_i^2 - y_i)^2$. The normal equation for the quadratic polynomial function fitting is

$$\begin{pmatrix}
\sum_{i=1}^{5} \alpha_i^2 & \sum_{i=1}^{5} \alpha_i & \sum_{i=1}^{5} \\
\sum_{i=1}^{5} \alpha_i & 5 & \sum_{i=1}^{5} \\
\sum_{i=1}^{5} \alpha_i^2 & \sum_{i=1}^{5} \alpha_i & 5
\end{pmatrix}
\begin{pmatrix}
b_0 \\
b_1 \\
b_2
\end{pmatrix} = 
\begin{pmatrix}
\sum_{i=1}^{5} y_i \\
\sum_{i=1}^{5} \alpha_i y_i \\
\sum_{i=1}^{5} \alpha_i^2 y_i
\end{pmatrix}$$  \hfill (11)

Get the fitted function $P(x)$ with the smallest mean square error. When the scaling ratio is $\alpha = -\frac{b_1}{2b_2}$, the response $y$ reaches the maximum. The rows and columns of this point in the sampling area are recorded as a matrix $Y$, which is the optimal target position $Pos$. The update scaling ratio is $R = -\frac{b_1}{2b_2}$. Update the latest target position $Pos$ with $Pos$ in the figure 2:

$$Pos = (1/R) \times Pos + Y.$$

**Figure 2. Target position update in frame t**

Scale the frame $t$ to the original image size, determine the final target center position $Pos_{tn}$ and the size of the tracking frame $q_t$ of the frame $t$ with the size of the tracking frame $q_{t-1}$ in the frame $t-1$:

$$Pos_{tn} = Pos' \times R$$  \hfill (12)

$$q_t = q_{t-1} \times R$$  \hfill (13)
3. Experiments

3.1. Evaluation criteria
In order to objectively compare the tracking performance of the proposed algorithm with the related algorithms mentioned, this paper evaluates the tracking algorithm in terms of tracking progress. The evaluation of tracking accuracy uses the accuracy curve [8]. The horizontal axis of the curve represents the precision error threshold, and the vertical axis represents the ratio of the number of video frames to the total number of video frames when the tracking progress of a single frame is higher than a certain threshold. The tracking accuracy of a single frame is the pixel difference between the tracking position of the single frame and the actual position. The accuracy curve reflects the overall situation of the algorithm's tracking accuracy during the entire video tracking process.

3.2. Result analysis
In order to evaluate the effectiveness of the algorithm and compare the performance of the algorithm, this paper selects the original CSK [4] and the optical flow algorithm as the comparison algorithm. Using the original code published by the literature author and the method proposed in this paper, the video sequence bus, car, car1, car2, car3 verify the algorithm's target tracking performance for fisheye camera video shooting in vehicle environment. These videos have multiple challenges, such as deformation, scale changes and fast motion.

![Figure 3. Tracking effect of 3 trackers on 4 videos](image)

![Figure 4. Accuracy curve of of 3 trackers on 4 videos sequences](image)

Figure 3 shows the tracking results of the proposed tracking method, CSK, and optical flow method in four video sequences. The tracking frames colors of the this paper, CSK and optical flow algorithm...
are red, yellow and green. For the 4 videos, due to the fixed size of the classifier trained by CSK and the interference of using the optical flow algorithm to track the background points, CSK and optical flow algorithms cannot achieve scale adaptive target tracking, or even lose targets. In the 4 videos, the targets have a large scale change due to the fast movement speed. Due to the fixed scale of the classifier, the CSK target tracking results are worse. The optical flow method has a large deviation in target tracking due to the deformation of the tracking target during the movement and some feature points on the background. Neither the CSK algorithm nor the optical flow method can recapture the tracking target. Both fail to track the target accurately or even lose target. The algorithm in this paper can accurately track the target even when the target has large scale changes and deformation problems.

Figure 4 shows the tracking accuracy curves of the three algorithms for the 4 videos sequences. In the accuracy curve, when the threshold is 10, the proportion of frames in the algorithm in this paper and the video sequence bus car1 and car2 is much higher than other algorithms.

4. Conclusion
Based on the CSK tracking method and the optical flow method, this paper proposes a target tracking method that combines the CSK algorithm with the optical flow algorithm for the scale problem in fisheye video tracking in the vehicle environment. The background point removal method and multi-scale scaling process are used to achieve the purpose of scale calculation and accurate tracking. At the same time, residuals and linear fitting are introduced to improve the accuracy of center positioning. Four vehicle fisheye video sequences were selected for verification and compared with CSK and optical flow algorithm. The experimental results show that the method in this paper is better than other methods in tracking the fisheye video in the vehicle environment, and it can achieve real-time tracking. At the same time, it is also robust to problems such as deformation, scale change and fast motion.

Acknowledgement
This work is supported by National Natural Science Foundation (NNSF) of China under Grant No.61702247; This work is supported by Key Project of the 13th Five-Year Plan for Economic Science in Liaoning Province under Grant JG18DA013; This work is supported by the Natural Science Foundation of Liaoning Province of China under Grant No.2019-ZD-0052;

References
[1] Ge Baoyi, Zuo Xianzhang, Hu Yongjiang. Review of Research on Visual Target Tracking Methods J. Journal of Image and Graphics, 2018,23(08):1091-1107.
[2] Liu Yating, Wang Kunfeng, Wang Feiyue. Visual target tracking based on tracklet correlation: present situation and Prospect J. Journal of Automation, 2017, 43 (11): 1869-1885.
[3] Li Mingjie, Liu Xiaofei, Zhang Fuquan, et al. Multi target detection and tracking algorithm based on particle filtering and background subtraction J. Application Research of Computers, 2018, 35(8): 2506-2509.
[4] HENRIQUES F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels C //European Conference on Computer Vision, Firenze, 2012:702-715.
[5] WU Y, LIM JW, YANG MH. Online object tracking: a benchmark C //IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013:1354-1362.
[6] Xu Yulong, Wang Jiabo, Li Yang, Li Hang, Zhang Gengning, Zhang Yafei. Scale-adaptive target tracking based on correlation filtering. J. Computer Application Research, 2016, 33 (11): 35133516 - 3520.
[7] Wang Yan, Wang Chunping, Fu Qiang, Xu Yan, Liu Yan. Scale-adaptive CSK tracking algorithm based on / 85 block J. Electro-optic and Control, 2017, 24 (02): 25-29.
[8] KAILATH T. The divergence and Bhattacharyya distance measures in signal selection J. IEEE Transactions on Communication,1967,15:52-60.