Research and Implementation of Object Tracking

Jie Tian*, Guanglong Wang, Weiwei Gao and Dan Fang
Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China

*Corresponding author e-mail: ysdmhj@163.com

Abstract. Aiming at the problems that the traditional MeanShift has a singleness feature, this paper introduces depth feature based on binocular vision, thus proposes a depth feature based real-time MeanShift tracking algorithm. Firstly, a similarity measurement method based on advanced quadratic-form distance is put forward. Secondly, the self-adaptive adjustment mechanism of feature weight coefficient and updating strategy of object model are improved based on the color and depth features. Finally, the algorithm is simulated by Matlab. Results show that the proposed algorithm performs well in tracking accuracy, robustness.

1. Introduction
Object tracking is an important topic in computer vision. It usually acquires objects such as pedestrians, vehicles, obstacles and markers by monocular cameras, providing position and size of the object in initial frame, and then estimates the state in subsequent frames. But features captured by monocular cameras are limited compared with binocular cameras. MeanShift has few parameters, strong robustness and fast speed. It is an important research topic in object tracking [1]. The real-time algorithms recommended by VOT 2015 include ASMS (Scale-adaptive Mean-shift) which is based on MeanShift framework [2], with an average frame rate of 125 FPS. However, MeanShift only considers color histogram, and it is susceptible to be disturbed when color of the object is close to the surrounding. The paper introduces depth by ZED, and proposes a real-time tracking method combining depth with MeanShift. A novel similarity measure, advanced quadratic-form distance (AQFD), is put forward instead of Bhattacharyya, which enhances the rationality and accuracy. Then the online self-adaptive adjustment mechanism of feature weights and updating strategy of object model are improved. Finally, the validity of the proposed method is verified by experiments.

2. Tracking Method Combining RGB-D with MeanShift

2.1. Advanced Quadratic-Form Distance (AQFD) based Similarity Measure
The conventional similarity measure Bhattacharyya assumes the subspaces of histogram are orthogonal, but the assumption is not applicable to depth histogram, thus we propose a novel similarity measure. Quadratic-Form Distance (QFD) based on Mahalanobis distance takes into account the relevancy between sub-intervals. Define as

\[ d_{aqfd}(p,q) = [(p - q)^T \Sigma^{-1} (p - q)]^{1/2} \]  

(1)

Where \( \Sigma \) represents covariance matrix of distances between subspaces. Let \( A = \Sigma^{-1} \),
\[ A = \begin{bmatrix}
1 - d_{11}/d_{\text{max}}, 1 - d_{12}/d_{\text{max}}, \ldots, 1 - d_{1m}/d_{\text{max}} \\
1 - d_{21}/d_{\text{max}}, 1 - d_{22}/d_{\text{max}}, \ldots, 1 - d_{2m}/d_{\text{max}} \\
\ldots \\
1 - d_{m1}/d_{\text{max}}, 1 - d_{m2}/d_{\text{max}}, \ldots, 1 - d_{mm}/d_{\text{max}}
\end{bmatrix} \tag{2} \]

Represents the relevancy between different subspaces, where \( \{d_{ij} \mid i = 1, \ldots, m; j = 1, \ldots, m\} \) denotes the distance between the \( i^{\text{th}} \) and \( j^{\text{th}} \) subspace, in other words, \( d_{ij} = |i - j| \), \( d_{\text{max}} = \max(d_{ij}) \). When \( A \) is a symmetric positive definite matrix, \( A = G^TG \). And Eq. (1) converts to

\[ d_{\text{w}}(p, q) = [(Gp - Gq)^T(Gp - Gq)]^{1/2} \tag{3} \]

Although QFD considers the relevancy between subspaces, the discrimination of different subspaces is not much. Thus we combine cosine distance with QFD to improve the discrimination. Cosine distance is defined as

\[ d_{\text{c}}(p, q) = \left(1 - \frac{p^Tq}{\|p\|\|q\|}\right)^{1/2} \tag{4} \]

Cosine distance defaults that subspaces are independent of each other. Combined with Eq. (3), Eq. (4) is optimized to

\[ d(p, q) = (1 - \frac{(p^T G^T G q)}{\|Gp\|\|Gq\|})^{1/2} \tag{5} \]

Let \( p' = \frac{Gp}{\|Gp\|}, \ q' = \frac{Gq}{\|Gq\|} \), then

\[ d(p, q) = (1 - \frac{(Gp)^T G q}{\|Gp\|\|Gq\|})^{1/2} = (1 - \frac{(p'^T q')^2}{2})^{1/2} \tag{6} \]

Which is called Advanced Quadratic-Form Distance (AQFD). From Eq. (2), we can see that the correlation coefficient between two subspaces whose distance is \( k \) with \( m \) subspaces is \( 1 - d_k / d_{\text{max}} = 1 - k/(m-1) \) which is linearly related to \( k \). We define \( G \) in Eq. (6) as \( G = \left[k_g \right]_{m \times m} \), where

\[ k_g = \exp(-|i - j|^2/(2\sigma^2)) \tag{7} \]

Where \( \sigma \) denotes attenuation factor. Generally, the range of \( \sigma \) is \( 5m\% \sim 10m\% \). The larger \( \sigma \) is, the slower the speed of attenuation is, and vice versa. From Eq. (7), we know that \( k_g = 1 \) and \( k_g \) decreases exponentially with increasing distance.

2.2. Weight Online Self-adaptive Adjustment Mechanism and Object Model Updating Strategy based on color and depth features

MeanShift based on multi-features often utilizes the joint histogram method \[3\], and the depth is employed as a new feature. However, the joint histogram generally has a high dimension, so the data is sparse. We establish object model based on color and depth histograms respectively, and the feature weights are self-adaptively adjusted to adapt to the real-time changes of the object. Probability density of the candidate model’s color feature \( p_{cu}(f) \) and depth feature \( p_{du}(f) \) are
\[
\begin{align*}
\rho_c(f) &= C_c \sum_{i=1}^{n_c} k \left( \frac{f - z_i}{h} \right) \delta \left[ b(z_i) - u_c \right] \\
\rho_d(f) &= C_d \sum_{i=1}^{n_d} k \left( \frac{f - z_i}{h} \right) \delta \left[ b(z_i) - u_d \right]
\end{align*}
\]

(8)

\(\rho_c(f)\) and \(\rho_d(f)\) represent AQFD of the candidate model’s color and depth features. \(\lambda_c\) and \(\lambda_d\) denote the corresponding feature weights. Define joint AQFD \(\rho(f)\) as

\[
\rho(f) = \lambda_c \rho_c(f) + \lambda_d \rho_d(f)
\]

(9)

Where \(\lambda_c + \lambda_d = 1\).

In order to adapt to the real-time changes of the object and surrounding, we adopts a self-adaptive adjustment mechanism of feature weights and online adjusts \(\lambda_c\) and \(\lambda_d\) to change the weights of color and depth features on the tracking effect. The idea is that according to the values of \(\rho_c\) and \(\rho_d\) in the previous frame to determine \(\lambda_c\) and \(\lambda_d\) of the current frame. Since the changes of object are continuous between the previous and current frames, it is reasonable to assume that the feature with a smaller AQFD in the previous frame obtains a higher weight in the current frame.

We denote the color and depth features of the \((t-1)\)th frame as \(\rho_c^{(t-1)}\) and \(\rho_d^{(t-1)}\) respectively. Let

\[
r^{(t)} = \ln(\rho_c^{(t-1)}/\rho_d^{(t-1)})
\]

(10)

And the corresponding function is shown in Figure 1.

\(r\) is to measure the size relation between \(\rho_c\) and \(\rho_d\), and the range is \((-\infty, +\infty)\). In consideration of \(\lambda_c, \lambda_d \in [0, 1]\), define \(\sigma\) (shown in Figure 2) as

\[
\sigma(x) = \frac{1}{1 + \exp(-x/b)}
\]

(11)

Where \(b\) controls the slope.

![Figure 1. Size relation between \(\rho_c\) and \(\rho_d\).](image)

![Figure 2. Function \(\sigma\) of different \(b\).](image)
\( \lambda_i = 1 - \lambda_i', \lambda_j = 1 - \lambda_j' \). And Eq. (12) demonstrates that the feature weights are only related to \( r \) and \( b \). In general, \( b = 1 \).

In practical applications, object which is influenced by factors such as light and deformation has gradual changes. Therefore, the establishment of a reasonable object model updating strategy is necessary for a long-term tracking [4]. It requires us to find a balance between adaptability and robustness. Classic updating strategy is

\[
q^{(t)} = \alpha p^{(t)} + (1-\alpha)q^{(t-1)}
\]  

(13)

Where \( p^{(t)} \) represents the tracking result of current frame, and \( q^{(t)} \) and \( q^{(t-1)} \) denotes the object model of current frame and previous frame respectively.

For multi-feature models, the complementarity between features is leveraged to improve object model updating strategy. After the \( (t-1) \)th frame converges, \( \rho_i^{(t-1)}, \rho_j^{(t-1)}, \lambda_i^{(t)} \) and \( \lambda_j^{(t)} \) can be obtained. AQFD reflects the matching degree between features and the object model in the current frame, and the weights reflect the credibility of features in the next frame.

We only update the feature with low matching degree and poor credibility, which can not only void model drifting due to wrong updating but also make the model have certain adaptability. Set the threshold as \( L, H, H \), and the proposed updating strategy is

\[
q^{(t)} = \begin{cases} 
\alpha p^{(t)} + (1-\alpha)q^{(t-1)} & \rho_j^{(t)} < \rho_i^{(t-1)} < \rho_j^{(t-1)}, \lambda_i^{(t)} < \lambda_j^{(t)} \leq \lambda_j^{(t-1)} \\
q^{(t-1)} & \text{others}
\end{cases}
\]  

(14)

Object model is updated only when the threshold of Eq. (14) is satisfied, and at most one feature model is updated in each frame.

3. Experiments

StereoLab put out 3D camera ZED in 2015. ZED is based on the principle of triangulation and employed for real-time acquisition of depth. We acquire an evaluation set by ZED and mark the groundtruth in each frame, which is convenient for quantitative comparison between methods. The evaluation set is processed by i5-7300HQ Processor, 8GB Memory and Matlab R2017b. The results are shown in Figure 3, and from top to bottom are results of classic MeanShift [5], our approach and Staple from CVPR 2016 [6].

![Figure 3. Results of MeanShift, our approach and Staple in evaluation set.](image-url)
From Figure 3, we know that the classic MeanShift tracks well before the object is occluded, but tracking task fails when the object is blocked by similar surrounding. Although the proposed method is based on MeanShift framework, the depth feature can well distinguish the object and the similar surrounding. Qualitatively, the proposed method and Staple can well achieve the interference experiment under similar surrounding. Comparing tracking results with the groundtruth, obtain pixel errors in each frame, and the result is shown in Figure 4. Calculate overlap rate \( r = \frac{A \cap B}{A + B - A \cap B} \), where \( A \) and \( B \) denote the area of tracking box and truth box respectively, and the result is shown in Figure 5.

![Figure 4. Pixel error of the proposed method and Staple.](image)

![Figure 5. Overlap rate of the proposed method and Staple.](image)

Figure 4 shows the average pixel error of our approach and Staple are 0.89 and 1.10 respectively. Figure 5 shows the average overlap rate of our approach and Staple are 0.92 and 0.90. The proposed method is superior to state-of-the-art Staple in the two evaluation indexes. Experiment demonstrates that our approach still has a high tracking accuracy under similar surrounding.

4. Conclusions

We introduce depth feature by ZED, and a real-time tracking method combining depth with MeanShift is proposed. We put forward a novel similarity measure AQFD. And then based on color and depth features, we implement an online self-adaptive adjustment of feature weights and improve the updating strategy of the object model. Experiment results demonstrate that our approach can track well under similar surrounding. The average pixel error and overlap rate of our approach are 0.89 and 0.92, which is superior to Staple, and the processing speed is about 30 FPS.
References

[1] GU Xingfang, MAO Yaobin, LI Qiujie, Survey on visual tracking algorithms based on mean shift, *Computer Science*, 39(12), 16-24, 2012. (in Chinese)

[2] Tomas Vojir, Jana Noskova, Jiri Matas, Robust Scale-Adaptive Mean-Shift for Tracking, *Pattern Recognition Letters*, 49(3), 250-258, 2014.

[3] NING J F, ZHANG L, Robust object tracking using joint color-texture histograms, *International Journal of Pattern Recognition and Artificial Intelligence*, 23(7), 1245-1263, 2009.

[4] COMANICIU D, MEER P, Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603-619, 2002.

[5] Comaniciu D, Real-time tracking of non-rigid objects using mean shift, *CVPR*, 142-149, 2000.

[6] Bertinetto L, Valmadre J, Golodetz S, et al., Staple: Complementary Learners for Real-Time Tracking, *Computer Vision and Pattern Recognition*, 1401-1409, 2016.