Target tracking using high-dimension data clustering

Shao Chunyan[1,2,3], Ding Qinghai[4], Luo Haibo[1,3,5], Li Yulian[6]

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China;
3. Key Laboratory of Opt-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China;
4. Space Star Technology Co., Ltd, Beijing 100086, China;
5. The Key Lab of Image Understanding and Computer Vision, Liaoning Province, Shenyang 110016, China;
6. Shenyang Metrology Testing Institution, Shenyang 110016, China)

Abstract: Inspired by the fact that a rigid body has consistent transformation for its individual part, a novel target tracking algorithm based on high-dimension data clustering is proposed. The proposed measure is proved to be available in object tracking mathematically. Thus, it is called the High-Dimension Data Clustering (HDDC) tracker. The frameworks of proposed method are as follows. First, Harris detector is utilized to extract the corners both in the template and the tracking region. Second, these feature points are grouped via their position information separately. Third, affine matrixes between the template and the tracking region are calculated among their respective feature groups. At last, high-dimension data clustering is carried out to measure these matrixes, and the feature points corresponding with the similar matrixes that are tracked targets. Extensive experimental results demonstrate that HDDC is efficient on measuring affine deformed objects and outperforms some state-of-the-art discriminative tracking methods.

Key words: high-dimension data clustering; affine deformed object; target tracking; rigid body

CLC number: TP391 **Document code:** A **DOI:** 10.3788/IRLA201645.0428002

采用高维数据聚类的目标跟踪

邵春艳[1,2,3],丁庆海[4],罗海波[1,3,5],李玉莲[6]

(1. 中国科学院沈阳自动化研究所,辽宁 沈阳 110016; 2. 中国科学院大学,北京 100049;
3. 中国科学院光电信息处理重点实验室,辽宁 沈阳 110016;
4. 航天恒星科技有限公司,北京 100086;
5. 辽宁省图像理解与视觉计算重点实验室,辽宁 沈阳 110016;
6. 沈阳计量测试院,辽宁 沈阳 110016)

摘 要: 根据刚体各部位具有变换一致性这一特性,提出一种采用高维数据聚类的目标跟踪方法。从数学理论方面证明提出的度量方法可以应用于目标跟踪,称其为高维数据聚类跟踪器(HDDC

收稿日期: 2015-08-13; **修订日期:** 2015-09-18

作者简介: 邵春艳(1987-),女,博士生,主要从事模式识别及图像处理方面的研究。Email:shaochunyan@sia.cn

导师简介: 罗海波(1967-),男,研究员,博士,主要从事图像处理与模式识别、成像跟踪、智能控制、并行信号处理器体系结构等方面的研究。Email:luohb@sia.cn

tracker)。该算法框架如下,首先,采用 Harris 检测器对模板与跟踪区域进行特征提取;然后利用这些特征的空间信息对所提取的特征进行编组;接着计算模板特征组与跟踪区域特征组间的仿射变换阵;最后,采用高维数据聚类对这些仿射变换阵进行度量,将那些相似仿射阵对应的跟踪区域作为跟踪目标。实验表明:HDDC tracker 能够有效地跟踪具有仿射形变的目标,并且性能优于先进跟踪算法。

关键词:高维数据聚类; 仿射形变; 目标跟踪; 刚体

0 Introductions

Two significant impacts on the efficiency of visual tracking algorithm are the target feature and the tracking algorithm. Many researchers have made great progress, and various tracking algorithms have been proposed these years[1-3]. The choice of the similarity measure is crucial in these tracking algorithms. Nowadays, due to the prediction of information theory, the information-theoretic based similarity measures are used widely[4-7]. Some information-theoretic similarity measures are commonly employed to similarity measure[1,8-10]. Targets are localized by iteratively finding the local minima of the distance measure functions. However, the spatial information of the targets in these algorithms is lost which limits its application on more general motion models. The tracker needs resort to using separate computational mechanisms to recover the scale and other information of the targets. Furthermore, these similarity measures have some disadvantages. First, the classical information-theoretic similarity measures such as the Bhattacharyya coefficient[1] or the Kullback-Leibler divergence[9] are not very discriminative[11], especially in high dimensions. Second, the sample-based classical similarity measures require a calculation that is quadratic in the number of samples, which makes it difficult to meet the real-time requirement in object tracking[9]. As a result, fast and effective measure methods for object tracking are worth of studying.

Yang et al. proposed a new simple symmetric similarity function between kernel density estimates of the template and target distributions in a joint spatial-feature space[8]. The similarity measure treats the location and other deformation in an integrated way and tracks the deformation incrementally. The proposed measure can be applied on general motion models, which is more robust and more discriminative compared with the information-theoretic similarity measures. Yang utilized the RGB feature and spatial position of the sample to a joint spatial-feature[8]. This joint feature could preserve the discriminative ability of the target. However, the joint spatial-feature is unstable when the illumination changes. Jose et al. defined a compound gain to measure the information gain between the pair-wise images[12]. They conducted several experiments on visual target distinctness with different information measures. Experimental results show that the compound gain is related to visual target distinctness as that perceived by human observers. Although they proved the method in several mathematical postulates, the results illustrate that it is not robust that the subjects are tested with prior knowledge of targets and the targets are not invariant in difficult surroundings[12].

The similarity measures above mentioned are all based on measuring the feature similarity between target template and target candidate. However, most features cannot preserve the spatial information of the target but some spatial feature. Hence, these algorithms are sensitive to the scale variation. Additionally, these similarity measures cannot measure globally, and they can enforce the local error accumulation which can be derived from the related formulas in literature[1,7,12-13].

Cootes and Sclaroff used matrix model that created from the Gaussian distances to measure[14-15]. The main drawback of this approach is that it does

not use the spatial relationship between the points in each set to constrain the search for the correspondences and the mapping[4]. Rangarajan derived a new non-quadratic distance measure that significantly outperformd the conventional quadratic assignment distance measure[4]. This non-quadratic distance measure is a function on the correspondence between the two feature sets which is time consuming.

A new measure for robust tracking based on high-dimension data clustering is proposed in this paper, so it is called the High-Dimension Data Clustering (HDDC) tracker. The measure proposed in this paper contains the invariant information between the template and the tracking frame which makes the algorithm recover the scale and other information of the target without other mechanisms. Moreover, the algorithm in this paper can preserve the spatial information of the target, which allows the application on more general motion models. Unlike the common similarity measure is focusing on measuring on the feature sets, the proposed measure is considering on the correspondence between the feature sets, which could accommodate feature variation and scale variation.

1 Mathematical proof

The proposed measure is based on comparing the transformation matrixes between pair-wise images. Targets will have different appearance such as shape, illumination when observers are in different locations. Finding the invariant model of the target in these images is very important for an observer. Obviously, the transformation of the target between two frames is geometrical transformation. In this research, the affine transformation model is just explored. According to the Pinhole Camera model[16], pixels in different frames have the following transformation when they are imaged from the same 3D point.

$$\begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{1}$$

Where (u_1,v_1) and (u,v) are the respective coordinates of the two pixels. Since our research is the affine transformation, the transformation matrix in Equation(1) is an affine matrix T which can be translated to

$$T=\begin{pmatrix} a_{00} & a_{01} & t_x \\ a_{10} & a_{11} & t_y \end{pmatrix} \tag{2}$$

For a rigid body, its space movement is consistent based on its definition in physics [17]. Hence, it can conclude that the transformation of feature points in a rigid body is consistent. Since most unmanned and manned vehicles can be accurately modeled as rigid bodies [18], the application of our approach could be extended to such vehicles.

Because the least square method has been well studied and provides several advantages [19], the least square method (LSM) is utilized to calculate the parameters in T in this paper (the LSM is detailed in literature[19]), which results the T to be

$$\begin{pmatrix} a_{00} & a_{01} & t_x \\ a_{10} & a_{11} & t_y \end{pmatrix} = \begin{bmatrix} x_1' & x_2' & \cdots & x_n' \\ y_1' & y_2' & \cdots & y_n' \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}^{+} \tag{3}$$

Where n is an integer with value of more than 3. Here the matrix with superscript "+" represents the Moore-Penrose matrix, and the points $\{(x_1',y_1'),(x_2',y_2'),\cdots,(x_n',y_n')\}$ are distributing on the same rigid body with point (u_1,v_1). Similarly, the points $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ and point (u,v) are in the same rigid body.

T can be denoted as a 6−D vector $\{a_{00},a_{01},a_{10},a_{11},t_x,t_y\}$ through converting the matrix into a column vector. Equation (3) illustrates different pair-wise feature groups of the same rigid body calculate a unique affine matrix, also a unique 6 −D vector. Similar with respect to 2 −D space, the vector represents a point in 6−D space. Inspired by this, the similarity of objects can be measured by clustering these vectors. Generally, due to the previous procedure of feature grouping, the transformation matrixes are approximate in a tolerant error range as shown in the latter experimental section. In order to testify the feasibility of the proposed measure, the

similarity measure is proved in mathematical aspect. Before the derivation, several prerequisites will be proposed.

Preconditions I_1 and I_2 are two images captured in different locations from the same sensor; the object in I_1 also appears in I_2 that is a rigid body; the feature points of the object in I_1 have one-to-one point mapping with the feature points of the object in I_2 (It is the ideal situation, and the pair-wise feature points sets are not one-to-one generally. In practical applications, at least 6 pair-wise points will make proposed approach available, as illustrated in Fig.1); the transformation between the two images is affine transformation; in tracking task, objects are the same when they have the same or partially approximate geometrical structure.

Proposition I object A from I_1 is the same with object B from I_2.

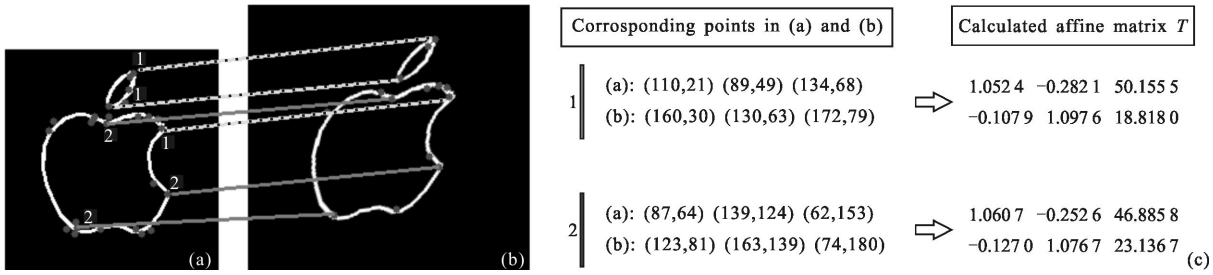


Fig.1 An illustration of proposed affine measure calculating. (a) an apple logo and its Harris corners; (b) the affine deformed result of (a) and its feature points. The solid line with group number 1 and the dotted line with group number 2 connect two different feature groups with different points shown beside. The corresponding affine matrixes are approximate; (c) the affine matrixes calculated between the corresponding points

Proposition II FG_A and FG_B are sets of the feature groups extracted from A and B respectively, the number of elements in FG_A and FG_B are N_A and N_B, where $N_A \leqslant N_B$. There are N_A' approximate affine matrixes mapping FG_A into FG_B.

The key insight of proposed approach is to take transformation consistency of rigid body for tackling the tracking problem. Hence, proposed approach is feasible when the above propositions satisfy $\bar{I} \Rightarrow \overline{II}$. In the following, the mathematical proof will be given.

Step 1: Extract feature points from A and B to get feature point sets $F_A=\{(Ax_1,Ay_1),(Ax_2,Ay_2),\cdots,(Ax_n,Ay_n)\}$ and $F_B=\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_n,By_n)\}$ respectively.

Step 2: $\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_i,By_i)\}$ are the feature set composed by the groups from FG_B which have the same transformation T with respect to FG_A. When $\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_i,By_i)\}$ are equal to $\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_{n2},By_{n2})\}$, we have

$$T\cdot\begin{bmatrix} Ax_1 & Ax_2 & \cdots & Ax_{n1} \\ Ay_1 & Ay_2 & \cdots & Ay_{n1} \\ 1 & 1 & \cdots & 1 \end{bmatrix}=\begin{bmatrix} Bx_1 & Bx_2 & \cdots & Bx_{n2} \\ By_1 & By_2 & \cdots & By_{n21} \end{bmatrix} \quad (4)$$

Which indicates that A and B have the same geometrical structure. On the basis of preconditions, A and B are the same, then $I \Rightarrow II$, the same as $\bar{I} \Rightarrow \overline{II}$.

When $\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_i,By_i)\}$ are the proper subset of $\{(Bx_1,By_1),(Bx_2,By_2),\cdots,(Bx_{n2},By_{n2})\}$, then

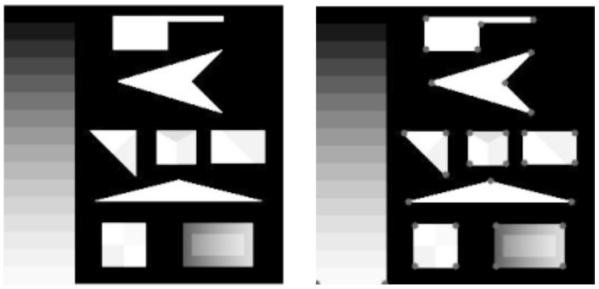
$$T\cdot\begin{bmatrix} Ax_1 & Ax_2 & \cdots & Ax_{n1} \\ Ay_1 & Ay_2 & : & Ay_{n1} \\ 1 & 1 & \cdots & 1 \end{bmatrix}=\begin{bmatrix} Bx_1 & Bx_2 & \cdots & Bx_i \\ By_1 & By_2 & \cdots & By_i \end{bmatrix} \quad (5)$$

Which proves that A and B are approximate partially, that is $I \Rightarrow II$, also $\bar{I} \Rightarrow \overline{II}$.

2 Description of the algorithm

2.1 Feasibility of the similarity measure

To verify the validity of the similarity measure, a test of the affine matrixes calculation to various types has been carried out using the test image designed in literature[20], which is shown in Fig.2(a). Fig.2(b) and Fig.2(c) are the Harris results of Fig.2(a) and its affine deformed image respectively, where the affine transformed matrix is set randomly.



(a) Test image (b) Harris corners of (a)

(c) Harris corners of (a) after affine deformed

Fig.2 Harris corner extracting results

32 corners and 33 corners are extracted in Fig.2(b) and Fig.2 (c) respectively. These corners are grouped manually via its space structures. As that three pair-wise points could work out one affine transformation matrix, groups with elements less than 3 are eliminated. Consequently, 8 pair-wise groups are obtained after the grouping in Fig.2 (b) and Fig.2(c) respectively. These groups provide 8 positive affine matrixes. In order to test the consistency of these affine matrixes, 8 negative affine matrixes are calculated by 8 pair-wise groups of different image structures. Following experiments will compare the 8 positive matrixes and 8 negative matrixes to validate the similarity measure.

Figure 3 analyzes the variation of each element in T among the groups in Fig.2. Experimental results illustrate that the six members of T have different spatial weights. Consequently, these members are separated into two parts as shown in the figure where members labeled with subscript s are members of positive T, while members with subscript d are negative. Figure 3 emphasizes the efficiency of the proposed measure. The affine transformation matrixes are similar when feature groups are in the same image structure. Each component of vector $\{a_{00},a_{10},a_{01},a_{11},t_x,t_y\}$ is approximate in a range of {1.052 0 –1.090 9, –0.1326– –0.1212, –0.263 2– –0.243 2, 1.023 0–1.112 3, 59.340 9–67.614 2, 27.676 4–41.554 5} respectively. Its standard deviation {0.012 7, 0.004 4, 0.007 8, 0.030 3, 2.551 9, 4.39} is much smaller comparing with {0.796 6, 0.352 6, 1.841 1, 1.017 2, 262.566 4, 81.289 1}, the standard deviation of affine matrixes calculated by different image structures.

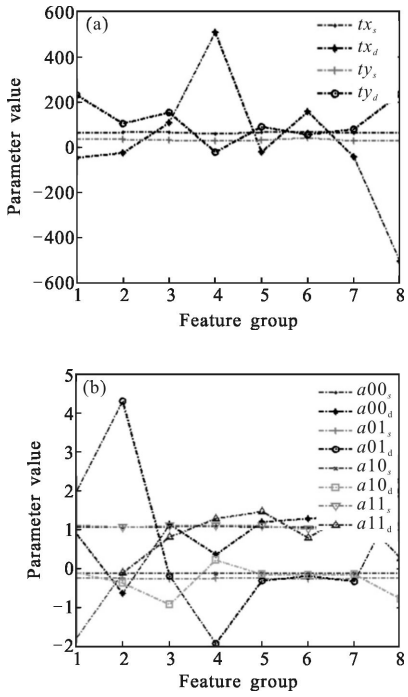


Fig.3 Variation of the members in T calculating from Fig.2

2.2 Framework of the proposed method

The proposed method contains three main processing procedures, feature grouping, calculation of affine matrixes and high-dimension data clustering.

2.2.1 Feature grouping

In the proposed algorithm, Harris corner detector is used as the feature extractor before feature grouping. This is because corner is a sort of robust feature to image variation and Harris corner detector is the best known corner detector for its simplicity and efficiency[21]. The saliency interest points can be obtained through tuning the parameters. Since T has

the form as equation (2), each three pair-wise points can work out one transformation matrix T. In the experiment, T is calculated by least square method (LSM) following equation (3). On one hand, the feature points in the template and the tracking frame are grouped into several groups by its spatial information. In order to meet the real time computational complexity, a linear grouping algorithm based on the spatial position is proposed, which has an execution time proportional of $O(n)$ to the number of the feature groups. For a $M\times N$ image, grouping on the feature point $P=\{(x_1,y_1),(x_2,y_2),\cdots,(x_s,y_s)\}$ set could be formulated as

$$G(P)=\{\underset{1\leqslant i\leqslant s}{p_i}\in\underset{1\leqslant j\leqslant n}{g_i}\{\frac{N}{n}\times(j-1)<x_i\leqslant\frac{N}{n}\times j,\frac{M}{n}\times(j-1)<y_i\leqslant\frac{M}{n}\times j\}\} \tag{6}$$

Then the transformation matrixes between the template feature groups and the tracking feature groups are calculated. Former proofs show that these affine matrixes could be transformed into several points in the 6-D space. Feature points contributing to the similar high-dimensional points are traced as the target. At last, similarity measure of two images is transformed into similarity measure of high-dimension data.

2.2.2 High-dimension data clustering based object tracking

After the calculation of affine matrixes, the similarity of these matrixes will be analyzed. For high-dimension data comparison, clustering is usually used [22]. Clustering refers to identifying groups or clusters in a data set. In contrast with the published clustering algorithms, density-based clustering is an approach where the clusters are considered to be higher density areas than the remainder of the data set. As a consequence, density-based clustering is used to group the 6-D spatial points in this paper. Points clustered into the same cluster are positive, and their corresponding affine matrixes connect to the feature points from the same object. Hartigan gives the general formalization of a density-based cluster as Equation (7)

$$\text{cluster}=\{x|p(x)>\lambda\} \tag{7}$$

Where $p(x)$ is the density at point x, and λ is the density threshold. Moreover, Hartigan defined the distance function *dis* linking between points. Because the density values are approximate in the experiments, *dis* is not considered in this paper. In the experiment, these points are normalized to align the distribution before calculating their distance. The coordinate of the point x in the 6-D space is $(a_{00},a_{10},a_{01},a_{11},t_x,t_y)$, the respective value of an affine matrix T. Previous experiment shows these 6 figures distribute with different scales, where $a_{00},a_{10},a_{01},a_{11}$ are much smaller than t_x and t_y. Hence, the 6 parameters are divided into two parts by their distribution and each part is normalized by the arctangent function. The linear density function $p(x)$ with the following form is considered.

$$p(T)=a\times\tan^{-1}(a_{00}+a_{10}+a_{01}+a_{11})+b\times\tan^{-1}(t_x+t_y) \tag{8}$$

Here a and b are the constant coefficients. In this paper, a and b are all set to 1. The final result of $p(T)$ is utilized for clustering. These matrixes with density value larger than a predefined value λ are clustered into the same cluster and their corresponding feature points are considered as the same. When all the density values are smaller than λ, tracking is stopped for losing target on the current frame. Figure 4 shows the frame work based on high-dimension data clustering.

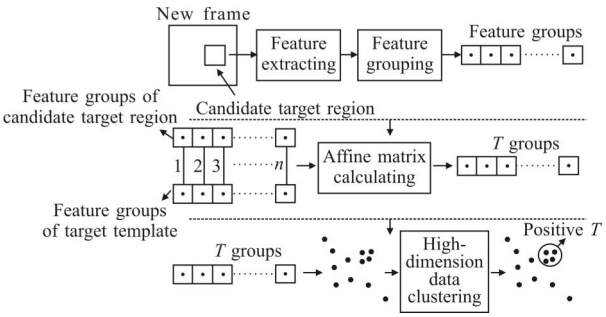


Fig.4 Tracking framework (Feature extracting in new frame is run surrounding the last target position. The target template groups are calculated from the template which is decided by the observer. The colorful lines with numbers in second part indicate the corresponding groups are one-to-one)

3 Experimental results and analysis

Additionally, in order to prove the efficiency and

performance of this approach, tracking on a series of flying plane images is carried out. Figure 5 (a1) and Figure5 (a2) present the 1st frame and 2nd frame of the tracking video. Figure 5 (b1) and Figure 5(b2) are the results of Harris corner detector on the template and the tracking region respectively, and Figure 5 (c1) and Figure 5 (c2) are their feature grouping results. The corners with the same number are grouped into the same one. Obviously, there are several groups with 3 or less members. In the following experiment, these groups are combined into one to calculate the affine matrix. The results of transformation matrixes of Figure 5 are listed in Table 1.

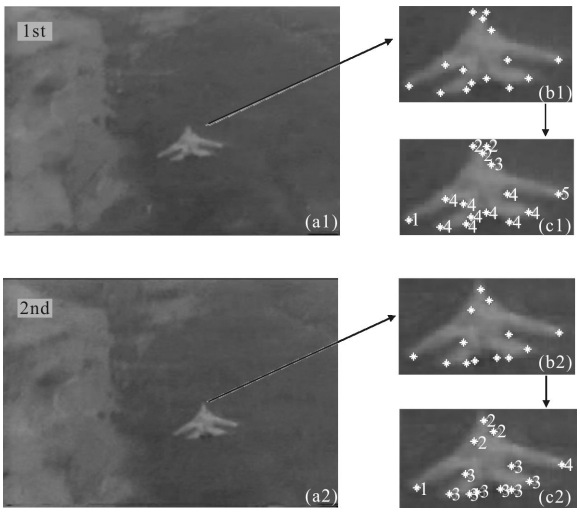


Fig.5 Feature grouping results

As Table 1 illustrates, G_1, G_3, G_5, G_7, G_9 are clustered together using previous clustering approach when the threshold λ is 1. Then the target is tracked by their corresponding feature sets. For the tracking frame with size of 352×240, the tracking costs nearly 148 ms for each frame. Hence, the proposed method is available for the real time application. The tracking result and results on several other frames were shown in Fig.6. Note that the high-dimension data clustering measure is effective on tracking.

Tab.1 Affine matrixes calculated from groups in Fig.5(c)

	a_{00}	a_{10}	a_{01}	a_{11}	t_x	t_y	$p(T)$
G_1	0.9166	−0.0210	−0.1821	0.8617	43.1904	26.0821	1.0059
G_2	−0.7569	0.0991	0.9572	−0.1880	206.2224	166.3845	0.6594
G_3	1.0086	0.0120	−0.1670	0.8571	23.6208	20.3643	1.0170
G_4	−0.9387	0.1235	−0.7244	0.0795	479.3164	124.0672	0.2352
G_5	0.9801	−0.0088	−0.1400	0.8438	25.7961	26.7949	1.0149
G_6	−1.0195	0.1429	−2.3581	0.4286	747.0816	66.7142	0.1339
G_7	0.9828	0.0001	−0.1220	0.8999	22.5372	16.5352	1.0208
G_8	−0.2157	−0.0212	−3.1637	0.5825	723.8520	73.0357	0.1334
G_9	0.9806	−0.0030	−0.1168	0.8842	22.6043	19.7936	1.0201

Proposed method is validated on different video sequences and compared with several state-of-the-art discriminative tracking methods, including OB(on-line boosting [23], the codes is available at the author′s webpage) and KMS [1](Kernel-based object tracking codes are integrated in the library in literature[24]). Two types of benchmark datasets are selected for the experiment. The first one is a traffic sequence from the Massachusetts Institute of Technology traffic data set [25]. The other one is a moving car sequence with scale variation which is available in the dataset in literature[24].

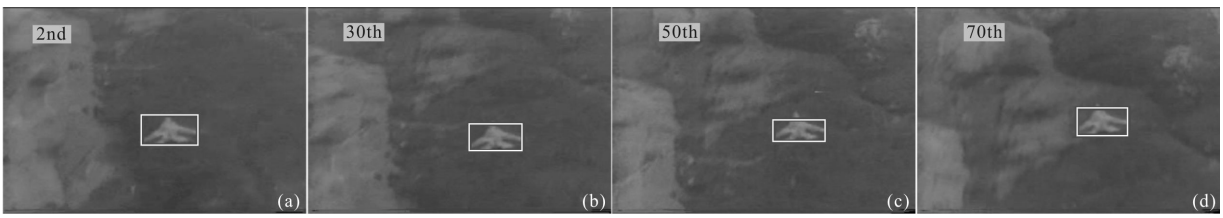
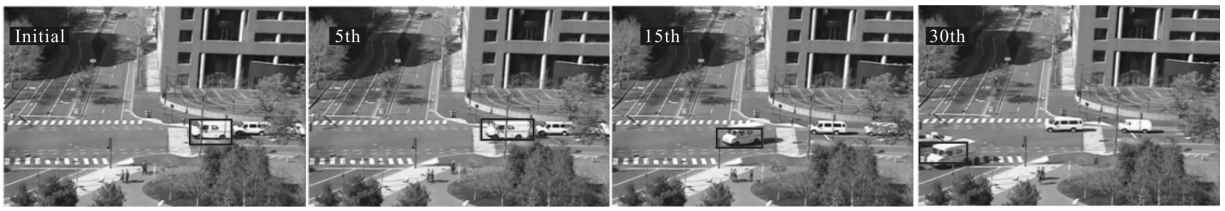


Fig.6 Tracking results

3.1 Tracking on short sequence

Figure 7 presents the comparison results on MIT traffic data set. The challenge of the sequence is complex backgrounds. OB (Online Boosting) is a typical classification tracker and insensitive to complex

(a) OB



(b) HDDC

Fig.7 Tracking performance comparison on sequence MIT traffic data set (High-dimension data clustering (HDDC) is able to provide accurate tracking while OB has drifting and inaccurate locating problems)

background. However, on short sequence, due to the few image samples, the tracking rectangle may sometimes drift, leading to inaccurate locating, as is shown in Figure 7. Relatively, proposed method locates the target more accurately. As the proposed method is based on the affine consistency of object, it can fit object transformation and model variation better, noting that even on the boundary in the last frame, HDDC still sticks to the right position.

3.2 Tracking on long time sequence

HDDC method is also validated as well as the two methods on a car sequence with scale variation. Figure 8 shows the tracking errors of these trackers. Comparison between the proposed method and the other two methods on the certain frames also can be obtained in Fig.8. The challenges in the car sequence include background changing, fast moving, zoom control, occlusion by trees, and the pan.

Proposed method provides more satisfactory results, whereas the tracking rectangles of other methods poorly adapt to the scale variation when the car moves near to the camera. OB tracker still drifts when back ground changes abruptly, and in frame 165 where the car has large occlusion, the tracker makes the wrong decision that the target is lost. KMS tracker seems to be robust to the occlusion. However, because it decides the tracking area without scale adaption, it lost the target since frame 249 where the car gets much bigger than that in the previous frames. From the curves presented in Fig.8, it can be concluded that OB and KMS may result in drifting when the background changes and scale of the target varies. The OB will lose target when occlusion happens and the KMS will lose target several times during the whole tracking process. With the time passage of tracking, proposed method has a little turbulence. This is because the pixels errors extend the range when the car moving in large scale, while

(a) HDDC

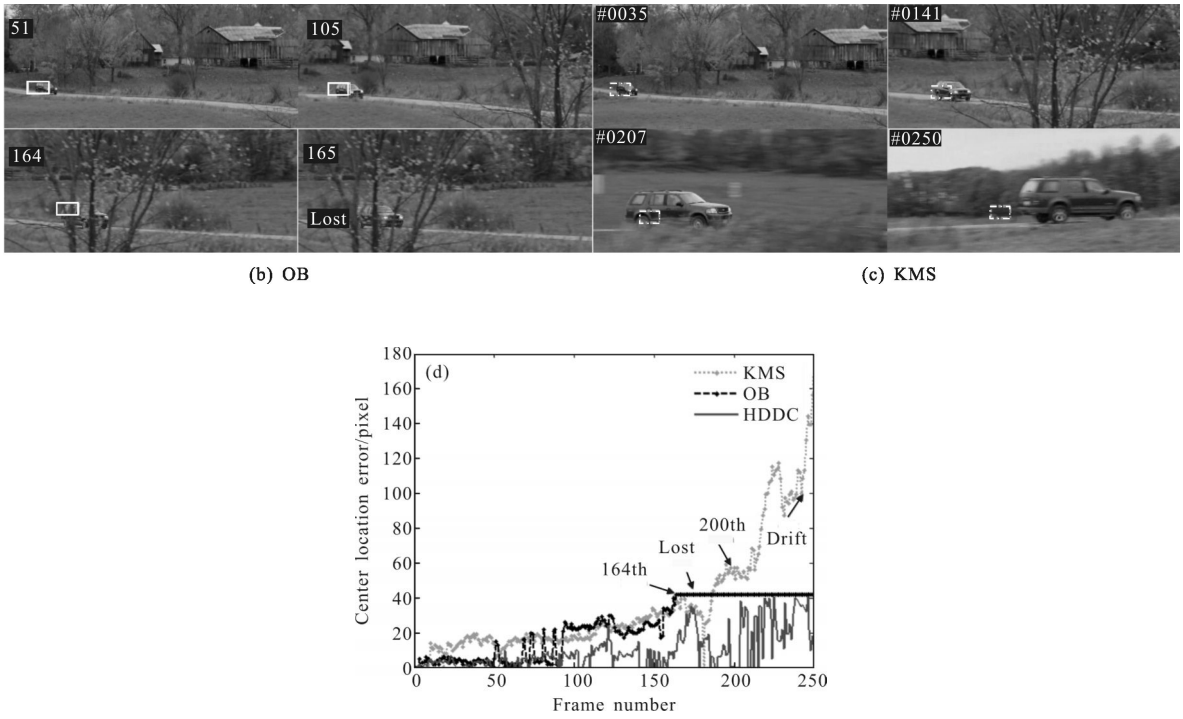


Fig.8 Tracking performance comparison on Car_Scale sequence (The tracking errors of different methods are also shown. HDDC=high-dimension data clustering tracker; KMS=kernel-based tracker; OB=online boosting tracker)

the trajectory seems to be steady. In a conclusion, compared with the other two methods, the approach provided steadier tracking with lower error.

4 Conclusions

In this paper a new similarity measure for rigid body tracking is proposed. It transformed the similarity of two images into the analysis of two affine matrixes. We demonstrate that the high-dimension data clustering based measure could be utilized as a similarity measure in object tracking, which avoids gathering limited negative samples for background description and positive samples for the target. Experimental results show that the proposed HDDC tracker is feasible and even outperforms some state-of-the-art tracking methods on providing accurate and stable tracking. Because there are still several limitations in our approach, a challenge for us in the future is trying to research more accurate feature grouping algorithm that can make the affine matrixes more approximate in high-dimensions.

References:

[1] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. ***IEEE Trans. Pattern Anal Mach Intell***, 2003, 25(5): 564-577.

[2] Cui X W, Wu Q Z, Jiang P, et al. Affine-invariant target tracking based on subspace representation [J]. ***Infrared and Laser Engineering***, 2015, 44(2): 769-774. (in Chinese)

[3] Yang Y F, Tian Y, Yang F, et al. Tracking of infrared small-target based on improved Mean-Shift algorithm [J]. ***Infrared and Laser Engineering***, 2014, 43(7): 2164-2169. (in Chinese)

[4] Rangarajan A, Chui H, Mjolsness E. A new distance measure for non-rigid image matching [C]//Energy Minimization Methods in CVPR, 1999: 237-252.

[5] Cazzanti L, Gupta M. Information-theoretic and set-theoretic similarity [C]//IEEE Int Conf on Symposium on Information Theory, 2006: 1836-1840.

[6] Cao W, Zhu M, He B G, et al. Overview of target tracking technology [J]. ***Chinese Optics***, 2014, 7 (3): 365-372. (in Chinese)

[7] Chen F, Wang Q, Wang S, et al. Object tracking via appearance modeling and sparse representation [J]. ***Image Vis Comput***, 2011, 29(11): 787-796.

[8] Yang C, Duraiswami R, Davis L. Efficient mean-shift tracking via a new similarity measure [C]//IEEE Conf on Computer Vision and Pattern Recognition, 2005: 176-183.

[9] Elgammal A, Duraiswami R, Davis L. Probabilistic tracking in joint feature-spatial spaces [C]//IEEE Conf on Computer Vision and Pattern Recognition, 2003: 781-788.

[10] Viola P, W M, W III. Alignment by maximization of mutual information [J]. ***Int J Comput Vision***, 1997, 24(2): 137-154.

[11] Hero A, Ma B, Michel O, et al. Applications of entropic spanning graphs [J]. ***IEEE Signal Proc Magazine***, 2002, 19 (5): 85-95.

[12] Garcia J A, Valdivia J F, Vidal X R F, et al. Information theoretic measure for visual target distinctness [J]. ***IEEE Trans Pattern Anal Mach Intell***, 2001, 23(4): 362-383.

[13] Yao Z, Lai Z, Liu W. A symmetric KL divergence based spatiogram similarity measure [C]//Int Conf on Image Processing, 2011: 193-196.

[14] Cootes T, Taylor C, Cooper D, et al. Active shape models: Their training and application [J]. ***Comp Vis Ima Unders***, 1995, 61(1): 38-59.

[15] Sclaroff S, Pentland A P. Model matching for correspondence and recognition [J]. ***IEEE Trans Patt Anal Mach Intell***, 1995, 17(6): 545-561.

[16] Wang K, Zhao L, Li R. Fisheye omnidirectional camera calibration-Pinhole or spherical model [C]//IEEE Conf on Robotics and Biomimetics, 2014: 873-877.

[17] Xiao Z Z, Qichoo C, Anand A, et al. Measurement of large deformation by digital image correlation method based on seed points [J]. ***Optics and Precision Engineering***, 2011, 19 (9): 2277-2282.

[18] Bras S, Izadi M, Silvestre C, et al. Nonlinear observer for 3D rigid body motion [C]//IEEE Conf on Decision and Control, 2013: 2588-2593.

[19] Zhao L R, Zhu W, Cao Y G, et al. Application of improved SURF algorithm to feature matching [J]. ***Optics and Precision Engineering***, 2013, 21 (12): 3263-3271. (in Chinese)

[20] Smith S M, Brady J M. SUSAN-a new approach to low level image processing [J]. ***Int J Comput Vision***, 1997, 23 (1): 45-78.

[21] Xu J J. Fast image registration method based on Harris and SIFT algorithm[J]. ***Chinese Optics***, 2015, 8(4): 574-579. (in Chinese)

[22] Kriegel H P, Kroger P, Sander J, et al. Density-based clustering [J]. ***Data Mining Knowl Discov***, 2011, 1 (3): 231-240.

[23] Grabner H, Bischof H. On-line boosting and vision [C]// IEEE Conf on Computer Vision and Pattern Recognition, 2006: 260-267.

[24] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C]//IEEE Conf on Computer Vision and Pattern Recognition, 2013: 2411-2418.

[25] Wang X, Ma X, Grimson E. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models[J]. ***IEEE Trans Pattern Anal Mach Intell***, 2009, 31(3): 539-555.