www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

NOVEL SVDD-BASED ALGORITHM FOR MOVING OBJECT DETECTING AND TRACKING UNDER DYNAMIC SCENES

Chunxiang Wang[1], Dongfang Xu[1] and Yongqing Wang[2]

[1] Basic Course Department, Henan Polytechnic, Zhengzhou 450046, China

[2] School of Computing, Zhengzhou University of Aeronautics, Zhengzhou 450015, China

Email: wyq-yongqing@163.com

Submitted: Jan. 14, 2016 ***Accepted: Mar. 21, 2016*** ***Published: June 1, 2016***

Abstract- Object detecting and tracking is an important technique used in diverse applications of machine vision, and has made great progress with the prevalence of artificial intelligence technology, among which the detecting and tracking moving object under dynamic scenes is more challenging for high requirements on real-time performance and reliability. Essentially analyzing, object detecting and tracking need to classify the objects and background into two different categories according to different features, where the detecting and tracking drift caused by noisy background can be effectively handled by robust maximum margin classifier, such as one-class SVM. But the time and space complexities of traditional one-class SVM methods tend to be high, which limits its wide applications to various fields. Inspired by the idea proposed by Support Vector Data Description (SVDD), in this paper we present a novel SVDD-based algorithm to efficiently deal with detecting and tracking moving object under dynamic scenes. The experimental results on synthetic, benchmark data and real-world videos demonstrate the competitive performances of the proposed method.

Index terms: Machine vision, moving object, object detecting and tracking, dynamic scene, support vector data description.

I. INTRODUCTION

With the prevalence of artificial intelligence technologies arising in our modern society, we “mankind” found that the jobs we could achieve good performance before can be easily accomplished by “machine” with even better performance.

Take the visual ability for example, whether or not we acknowledge it, the fact is, our visual ability is tend to be fragile, fatigue, aging and damageable, and which are on the contrary for machine vision, to a large extent.

Machine vision means utilizing computer to simulate the visual function of human eyes, identify the shape and movement of objects in the real world by extract information from images or image sequences.

Machine vision is a comprehensive technology, including digital processing technology, mechanical engineering technology, control technology, illumination technology, optical imaging technology, sensor technology, analog and digital video technology, computer software and hardware technology, man-machine interface technology, etc.

Machine vision emphasizes practicality, such as high capability on fault-tolerant and security, strong versatility and portability, high performances on real-time, fast speed and high precision. A typical industrial machine vision system includes light source, optical image system, image capturing system, image collection & digitalization module, image processing module, intelligent judgment decision module and mechanical control execution module (as shown in Fig. l).

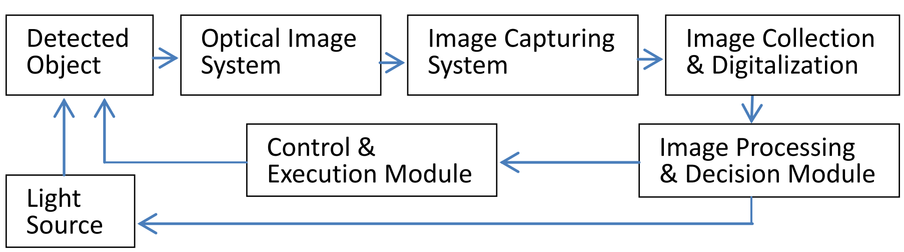


Figure 1. The flow diagram of typical industrial machine vision system

Compared to human being’s visual ability, machine vision has the merits of high efficiency, nondestruction, objective and accurate result, and fatigueless working, which is very helpful to achieve lots of successful applications in more and more fields, such as biological feature recognition, information retrieval, and so on. By utilizing of infrared-ray, ultraviolet-ray, X-ray, ultrasound and other advanced detecting techniques, machine vision has the significant advantages in detecting invisible objects, and in the circumstances under high dangerous scenes. We conclude the common applications of machine vision as below.

a. Machine vision in military and public security

The application of machine vision in security includes two scenarios, military security and public security.

a.i Military security

The military security scenario includes the cruise missile terrain recognition, the object detecting and locating (seen in Fig. 2), the radar terrain reconnaissance, the remote control aircraft guidance, the target identifying and guidance, the military alert system and automatic control of artillery, etc.

Figure 2. Machine vision in military object detecting and locating

a.ii Public security

The public security scenario includes the fingerprint automatic recognition, the iris feature automatic recognition, the synthesis of criminals' face, the automatic identification of handwriting, portrait, and seal, enhancing image quality to capture emergency in the monitoring system for closed circuit television, intelligent monitoring in traffic management system(seen in Fig. 3), etc.

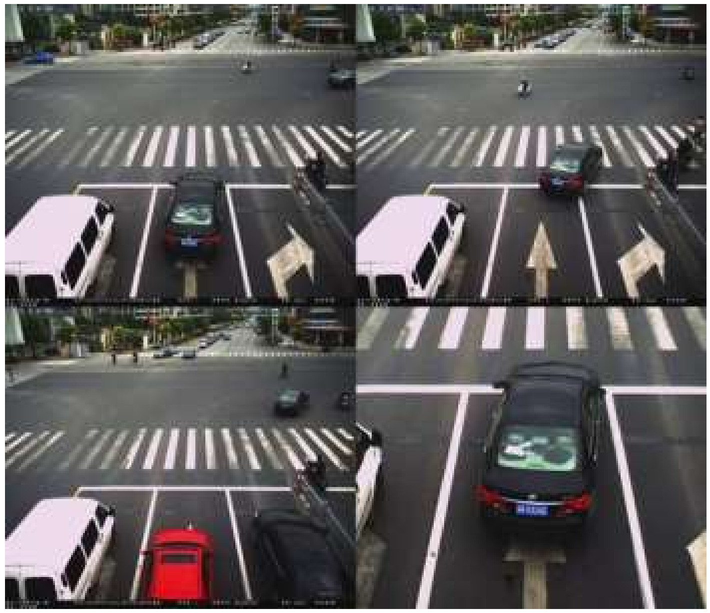


Figure 3. Machine vision in traffic monitoring

b. Machine vision in biomedicine

In the field of biomedicine, machine vision is used to assist doctors in medical images analysis (seen in Fig. 4), where digital image processing and information fusion technologies can be used for medical imaging data statistics and analysis in the scenes of X-ray perspective, nuclear magnetic resonance and CT images.

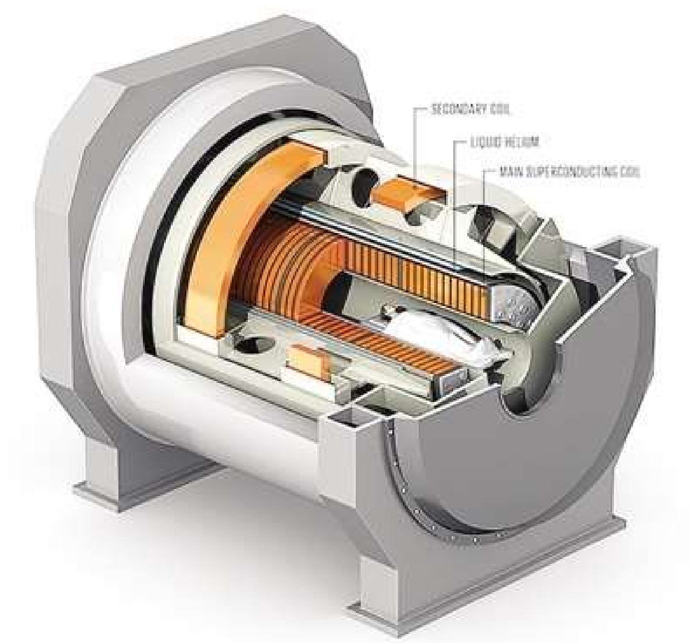


Figure 4. Machine vision in Magnetic Resonance Imaging

c. Machine vision in aviation and remote sensing

Machine vision can be used in diverse scenes of aviation and remote sensing (seen in Fig. 5), which can be listed but not limited to the following applications.

The applications include the reconnaissance, positioning and navigation in military scenes, the automatic cartography, satellite images and topographic map alignment, automatic surveying and mapping, the management of land and resources, the synthetic analysis and prediction to weather forecast, automatic environment and fire alarm monitoring, the detection and analysis of astronomy and space objects, transportation and air lines management, the collecting satellite remote sensing images, automatic identifying and classifying the ground targets according to the characteristics of image and graphics topography, etc.

Figure 5. Machine vision in remote sensing: Bird's Nest in Google Earth

d. Machine vision in industrial detection

Nowadays, machine vision has been successfully applied in the field of industrial detection (seen in Fig. 6), and greatly improves the quality and reliability of the product, which guarantees the speed of production.

For example, the quality detection in packaging and printing for products, quality detection for containers, drink filling, and bottle cap sealing in beverage industry, timber wood detection, semiconductor integrated packaging quality detection, coil quality detection, the industrial computed tomography of key mechanical parts, etc. At the customs, the application of X-ray and the machine vision technology can inspect the cargo without opening the package, which can greatly improve the speed of customs clearance, and saves a large amount of manpower and material resources. In the pharmaceutical production line, machine vision technology can be used to test the drug packaging, which can guarantee the package quality of drags.

Figure 6. Machine vision in industry detection

The applications of machine vision have many other topics, such as object detecting and tracking, which is a promising technique used in diverse applications of machine vision, and has made great progress with the prevalence of artificial intelligence technology. What we discussed in this paper is the detecting and tracking moving object under dynamic scenes, which is more challenging for high requirements on real-time performance and reliability.

Essentially analyzing, object detecting and tracking need to classify the objects and background into two different categories according to different features, where the detecting and tracking drift caused by noisy background can be effectively handled by robust maximum margin classifier, such as one-class SVM. But the time and space complexities of traditional one-class SVM methods tend to be high, which limits its wide applications to various fields. Inspired by the idea proposed by Support Vector Data Description (SVDD), in this paper we present a novel SVDD-based algorithm to efficiently deal with detecting and tracking moving object under dynamic scenes. The experimental results on synthetic, benchmark data and real-world videos demonstrate the competitive performances of the proposed method.

The rest of this paper is organized as follows. Section II provides a review on Object Detecting and Tracking. The method of Support Vector Data Description is introduced in Section III. Section IV proposes the novel SVDD-based algorithm in detail. Section V conducts the experimental results on synthetic, benchmark and real-world data set, and Section VI concludes this paper.

II. OBJECT DETECTING AND TRACKING

Moving object detecting and tracking under dynamic scenes has a very wide range of applications. It plays a key role in many fields such as smart video surveillance, traffic monitoring and some military occasions. Background modeling, moving objects detection and tracking is the focus of this paper. There are many experts and scholars have been studied in these fields for many years, and made great progress. However, moving object detecting and tracking still face many challenges in a complex scene with shadows interference and illumination changes.

a. Object detecting

Object detecting refers to a process of distinguishing a specific object from other objects, or objects of one type from other types. It includes not only the recognition of two very similar objects, but also the recognition of objects which belong to different types.

a.i The principle and procedure of object detecting

The basic principle of object detecting is the usage of object's characteristics information in radar echo, such as amplitude, phase, frequency spectrum and polarization, through the various multidimensional space transformation of mathematics to estimate the object's size, shape, weight and surface physical properties parameters, and finally accomplishing the identification in classifiers according to the identification functions, which determined by a large number of training samples.

The whole procedure of object detecting includes four steps as follows, whose logical relationship can be listed in Figure 7.

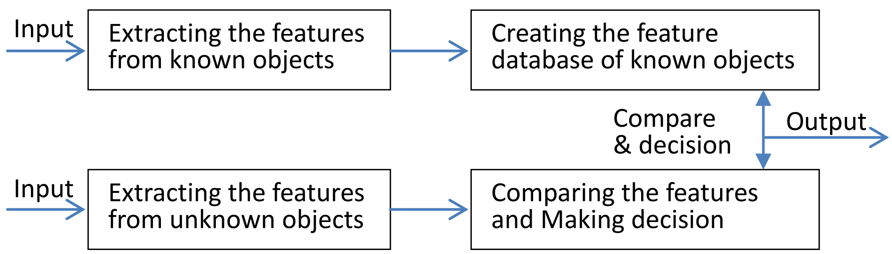


Figure 7. The flow diagram of object detecting

a.ii Representations of object shape

A. Yilmaz [1] summarized the object shape representations commonly employed for object detecting as follows.

1) Points

The object is represented by a point, that is, the centroid [2] or by a set of points [3]. In general, the point representation is suitable for recognition objects that occupy small regions in an image.

2) Primitive geometric shapes

Object shape is represented by a rectangle, ellipse [4], etc. Object motion for such representations is usually modeled by translation, affine, or projective transformation. Though primitive geometric shapes are more suitable for representing simple rigid objects, they are also used for recognition non-rigid objects.

3) Articulated shape models

Articulated objects are composed of body parts that are held together with joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. The relationships between the parts are governed by kinematic motion models, for example, joint angle, etc. In order to represent an articulated object, one can model the constituent parts using cylinders or ellipses.

4) Skeletal models

Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for detecting objects [5]. Skeleton representation can be used to model both articulated and rigid objects.

5) Object silhouette and contour

Contour representation defines the boundary of an object. The region inside the contour is called the silhouette of the object. Silhouette and contour representations are suitable for recognition complex non-rigid shapes [6].

a.iii Representations of object appearance

There are a number of ways to represent the appearance features of objects. Note that shape representations can also be combined with the appearance representations [7] for recognition. Some common appearance representations in the context of object detecting can be summed up as follows.

1) Probability densities of object appearance

The probability density estimates of the object appearance can either be parametric, such as Gaussian and a mixture of Gaussians [8], or nonparametric, such as Parzen windows [9] and histograms [4]. The probability densities of object appearance features (color, texture) can be computed from the image regions specified by the shape models (interior region of an ellipse or a contour).

2) Templates

Templates are formed using simple geometric shapes or silhouettes [10]. An advantage of a template is that it carries both spatial and appearance information. Templates, however, only encode the object appearance generated from a single view. Thus, they are only suitable for recognition objects whose poses do not vary considerably during the course of recognition.

3) Active appearance models

Active appearance models are generated by simultaneously modeling the object shape and appearance [11]. In general, the object shape is defined by a set of landmarks. Similar to the contour-based representation, the landmarks can reside on the object boundary or, alternatively, they can reside inside the object region. For each landmark, an appearance vector is stored which is in the form of color, texture, or gradient magnitude. Active appearance models require a training phase where both the shape and its associated appearance is learned from a set of samples using, for instance, the principal component analysis.

4) Multi-view appearance models

These models encode different views of an object. One approach to represent the different object views is to generate a subspace from the given views. Subspace approaches, for example, Principal Component Analysis (PCA) and Independent Component Analysis (ICA), have been used for both shape and appearance representation. Another approach to learn the different views of an object is by training a set of classifiers, for example, the support vector machines [12] or Bayesian networks [13]. One limitation of multi-view appearance models is that the appearances in all views are required ahead of time.

b. Object tracking

Object tracking can be simply defined as the estimation for trajectory in the image plane of an object moving around a scene.

b.i Mathematical description of object tracking

The object system can be regarded as a composition of dynamic system and observation system. The real state of the target system cannot be determined, so the target system can be identified as a black box, we can only through observations to estimate the state. And then, the object tracking problem can be transformed to the designing of state estimator, where the key point can be formulated as follows.

Given observations of $Z=\{z_i \mid i=1,...,k\}$, the goal is to solve the posterior probability density distribution function of object state x_k. Aiming at solving this state estimation problem, we suppose the state function of object to be

$$x_k = f_k(x_{k-1}, v_{k-1}) \qquad (1)$$

where f_k is the nonlinear function of state vector x_{k-1}, and $\{v_{k-1}, k \in N\}$ is the system noise. The corresponding object observing function is

$$z_k = h_k(x_k, n_k) \qquad (2)$$

where h_k is the nonlinear function of state observing process, and $\{n_k, k \in N\}$ is the observing noise. According to the Bayesian Theory and Chapman-Kolmogorov function, the estimation of state prior probability density at the k^{th} moment is

$$p(x_k \mid z_{1:k-1}) = \int p(x_k \mid x_{k-1}) p(x_{k-1} \mid z_{1:k-1}) dx_{k-1} \qquad (3)$$

The state posterior probability density at the k^{th} moment is

$$p(x_k \mid z_{1:k}) = \frac{p(z_k \mid x_k) p(x_k \mid z_{1:k-1})}{\int p(z_k \mid x_k) p(x_k \mid z_{1:k-1}) dx_k} \qquad (4)$$

where the value of equation (4) depends on the statistical properties of likelihood function $p(z_k \mid x_k)$ and observing noise n_k. Based on the criterion of minimum mean square error, the optimal estimation can be calculated as

$$\hat{x}_k = \int x_k p(x_k \mid z_{1:k}) dx_k \qquad (5)$$

The whole process of reasoning can be interpreted as two stages, prediction and update. Prediction: Substituting the state probability density at the $k-1^{th}$ moment $p(x_{k-1} \mid z_{1:k-1})$ into the move equation of object to get the prior probability density distribution of current state $p(x_k \mid z_{1:k-1})$, as shown in equation (3). Update: using the similarity measure function $p(x_k \mid z_k)$ of current observation data, and the prior probability distribution $p(x_k \mid z_{1:k-1})$ of state to determine the posterior probability density distribution of current state, as shown in equation (4). The tracking to the process of state's changes can be finally achieved by the whole procedure stated above.

b.ii Object tracking methods

Object tracking aims at generating the trajectories of moving objects over time by locating their positions in every frame of the video. Object tracker may also provide the complete region in the image that is occupied by the object at every time instant. According to the different strategies adopted in related methods, A. Yilmaz [1] divided the common object tracking into three categories.

1) Point tracking

Objects detected in consecutive frames are represented by points, and the association of the points is based on the previous object state which can include object position and motion. This approach requires an external mechanism to detect the objects in every frame.

Tracking can be formulated as the correspondence of detected objects represented by points across frames. Point correspondence is a complicated problem, especially in the presence of occlusions, misdetections, entries, and exits of objects.

Overall, point correspondence methods can be divided into two broad categories, namely, deterministic method [14-17] and statistical method [18-20]. The deterministic methods use qualitative motion heuristics [14] to constrain the correspondence problem. On the other hand, probabilistic methods explicitly take the object measurement and take uncertainties into account to establish correspondence.

2) Kernel tracking

Kernel refers to the object shape and appearance. For example, the kernel can be a rectangular template or an elliptical shape with an associated histogram. Objects are tracked by computing the motion of the kernel in consecutive frames. This motion is usually in the form of a parametric transformation such as translation, rotation, and affine.

Kernel tracking is typically performed by computing the motion of the object, which is represented by a primitive object region, from one frame to the next. The object motion is generally in the form of parametric motion (translation, conformal, affine, etc.) or the dense flow field computed in subsequent frames. These algorithms differ in terms of the appearance representation used, the number of objects tracked, and the method used to estimate the object motion. A. Yilmaz [1] divided these tracking methods into two subcategories based on the appearance representation used, namely, templates and density-based appearance models [21-25], and multi-view appearance models [26, 27].

3) Silhouette tracking

Tracking is performed by estimating the object region in each frame. Silhouette tracking methods use the information encoded inside the object region. This information can be in the form of appearance density and shape models which are usually in the form of edge maps. Given the object models, silhouettes are tracked by either shape matching or contour evolution. Both of these methods can essentially be considered as object segmentation applied in the temporal domain using the priors generated from the previous frames.

Objects may have complex shapes, for example, hands, head, and shoulders that cannot be well described by simple geometric shapes. Silhouette-based methods provide an accurate shape description for these objects. The goal of a silhouette-based object tracker is to find the object region in each frame by means of an object model generated using the previous frames. This model can be in the form of a color histogram, object edges or the object contour. A. Yilmaz [1] divided silhouette trackers into two categories, namely, shape matching and contour tracking. Shape matching approaches [28-32] search for the object silhouette in the current frame. Contour tracking approaches [33-37], on the other hand, evolve an initial contour to its new position in the current frame by either using the state space models or direct minimization of some energy functional.

b.iii Existing problems in object tracking

Object tracking has made significant progress during the last few decades, some robust trackers have been developed to track objects in real time in simple scenarios. However, the assumptions to validate tracking techniques tend to be unpractical in real world. For example, smoothness of motion, minimal amount of occlusion, illumination constancy, high contrast with respect to background, etc., are violated in many realistic scenarios and therefore limit a tracker's usefulness in applications. What we discussed in this paper is the difficulty in modeling and submerges the objects into the noisy background, which can be efficiently handled by the method we proposed.

III. SUPPORT VECTOR DATA DESCRIPTION

The tracking technique based on classification is a very efficient way to solve the object tracking problem. The classifier classifies the objects and background into two different classes, where the tracking drift caused by noisy background can be effectively handled by one-class SVM.

From the angle of mathematics, one-class SVM has the identical mathematical formulation in duality space, i.e., the Minimum Enclosing Ball (MEB) problem, under which relationship appears the Support Vector Data Description (SVDD) [38] method.

a. Mathematical formulation of MEB

Given a set of data points $S = \{x_i \mid i = 1,\ldots,m\}$, where $x_i \in R^d$, the Minimum Enclosing Ball (MEB) of S (denoted as MEB(S)) is defined as the smallest ball $B(c,R)$ that contains all the points in S, i.e., $B(c,R) = \{x_i \in R^d \mid \|x_i - c\| \le R\}$.

Let k be a kernel function with the associated feature map ϕ, i.e. $k(x_i,x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, where $\langle .,. \rangle$ denotes the inner product. Then the primal MEB problem in the kernel-induced feature space to find the MEB(S) with center c and radius R can be formulated as

$$\begin{aligned} &\min_{R,c} \ R^2 \\ &s.t. \ \ ||c - \phi(x_i)||^2 \le R^2, i = 1,\ldots,m. \end{aligned} \tag{6}$$

The corresponding dual is

$$\begin{aligned} &\max_{\alpha_i} \ \sum_{i=1}^{m} \alpha_i k(x_i,x_i) - \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i,x_j) \\ &s.t. \ \ \sum_{i=1}^{m} \alpha_i = 1, \ \alpha_i \ge 0, \ i = 1,\ldots,m. \end{aligned} \tag{7}$$

b. Mathematical formulation of SVDD

Given a training data sets $S = \{(x_i, y_i) \mid i = 1,\ldots,m\}$, where $x_i \in R^d$ and $y_i \in \{+1,-1\}$, the primal for the Binary SVM problem can be formulated as

$$\begin{aligned} &\min_{w,\rho,b,\xi_i} \ ||w||^2 + b^2 - 2\rho + C\sum_{i=1}^{m} \xi_i^2 \\ &s.t. \ \ y_i(w'\phi(x_i) + b) \ge \rho - \xi_i, i = 1,\ldots,m. \end{aligned} \tag{8}$$

The corresponding dual is

$$\begin{aligned} &\min_{\alpha_i} \ \sum_{i,j=1}^{m} \alpha_i \alpha_j (y_i y_j k(x_i,x_j) + y_i y_j + \frac{\delta_{ij}}{C}) \\ &s.t. \ \ \sum_{i=1}^{m} \alpha_i = 1, \ \alpha_i \ge 0, \ i = 1,\ldots,m, \end{aligned} \tag{9}$$

Where δ_{ij} is the Kronecker delta function, defined as following

$$\delta_{ij} = \begin{cases} 1, & if \ i = j, \\ 0, & if \ i \ne j. \end{cases} \tag{10}$$

We denote the pair (x_i, y_i) as z_i to simplify the notation. Introducing a modified feature map $\tilde{\phi}(z_i) = [y_i \phi'(x_i) \ \ y_i \ \ \frac{e'_i}{\sqrt{C}}]'$ and the associated kernel function $\tilde{k}(z_i,z_j) = y_i y_j k(x_i,x_j) + y_i y_j + \frac{\delta_{ij}}{C}$, then the dual of Binary SVM with form (9) can be rewritten as

$$\begin{aligned} \min_{\alpha_i} \quad & \sum_{i,j=1}^{m} \alpha_i \alpha_j \widetilde{k}(z_i, z_j) \\ s.t. \quad & \sum_{i=1}^{m} \alpha_i = 1,\ \alpha_i \geq 0,\ i = 1, \ldots, m. \end{aligned} \tag{11}$$

c. Mathematical formulation of CVM

Core Vector Machine (CVM) [39] is a promising technique for scaling up a Binary SVM to handle large data sets with the greedy-expansion strategy, where the kernels are required to be normalized to ensure the equivalence between the kernel-induced spaces of SVM and Minimum Enclosing Ball (MEB).

Considering only the situation where the kernel k satisfies $k(x,x) = \kappa$, a constant. This holds true for kernels like Gaussian, polynomial kernel with normalized inputs, and any normalized kernels [39]. Then the dual of the MEB problem (7) can be rewritten as

$$\begin{aligned} \min_{\alpha_i} \quad & \sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \\ s.t. \quad & \sum_{i=1}^{m} \alpha_i = 1,\ \alpha_i \geq 0,\ i = 1, \ldots, m. \end{aligned} \tag{12}$$

When the involved kernels fulfill the requirements mentioned above, any Quadratic Programming (QP) of form (12) can be identified as an MEB problem, and so the formulation of equation (11) does.

IV. NOVEL SVDD-BASED ALGORITHM

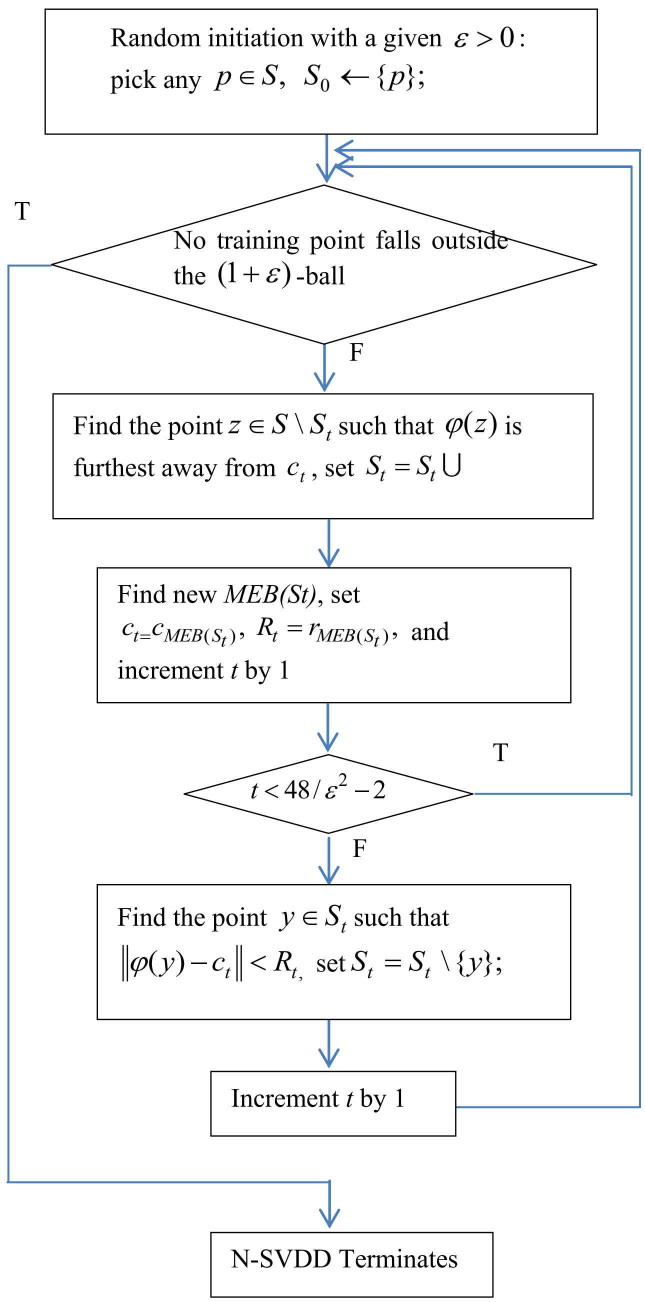
The quadratic optimization problem involved in both SVDD and CVM methods has high time and space complexities. Asharaf [40] claimed that, even though the decomposition or data-sampling techniques [41-46] can help to reduce the complexity of the optimization problem, they are still expensive for use in applications involving large data sets. A practical technique to overcome the problem to a large extent is introducing a fast MEB algorithm, which has the time and space complexities of $O(\frac{m}{\varepsilon^2} + \frac{1}{\varepsilon^4})$ and $O(\frac{1}{\varepsilon^2})$, respectively.

However, we found that the final core vectors obtained for formulating the final decision function is always more than necessary in implementations, which results in some redundancies in the process of storing and training. So, we propose the Novel SVDD-based algorithm by the strategy of inclusion-and-exclusion, which can reduce the redundancy as much as possible.

a. Novel SVDD-based algorithm

We formulate the proposed algorithm in Table 1 below.

Table 1. The flow chart of N-SVDD algorithm



b. The analysis on time and space complexities

We conclude the analysis on time and space complexities in Theorems 1 to 4 below.

Theorem 1. In the process of N-SVDD Algorithm, when the iteration satisfies $i \geq \frac{48}{\varepsilon^2} - 2$, if one point q falls into the interior of current MEB, i.e., $\|q - c_i\| < r_i$, it will fall into the interior of subsequence MEBs, i.e., $\|q - c_{i+j}\| < r_{i+j}$, $j \in Z^+$.

Theorem 2. N-SVDD Algorithm can achieve a $(1+\varepsilon)$-approximate MEB for training data set S within $O(\frac{1}{\varepsilon^2})$ iterations.

Theorem 3. In the iterations of N-SVDD Algorithm, there exists a subset $P \subset S$, whose points are at distance at most $(1+\varepsilon) r_{B(S)}$ from center $c_{B(S)}$, and the size of P is $O(\min\{\frac{1}{\varepsilon^2}, d\})$.

Theorem 4. The time and space complexities of N-SVDD Algorithm are $O(d^4 + d^2 m + \frac{d^3}{\varepsilon^2} + \frac{dm}{\varepsilon^2})$ and $O(\min\{\frac{1}{\varepsilon^4}, d^2\})$.

The detailed proofs of these theorems are omitted here for conciseness, the interested readers can refer to Wang [47-49].

V. EXPERIMENTAL RESULTS

We conduct the experiments on synthetic, benchmark and real-world data set to prove the validity and efficiency of the proposed N-SVDD algorithm. All the experiments were done on an AMD Sempron (tm) 2500+ 1.41GHz PC with 1GB RAM, and the software package we utilized is Matlab 7.0 toolbox.

a. Experiment on synthetic and benchmark data sets

Firstly, we implement the algorithms of Support Vector Machine (SVM), Core Vector Machine (CVM), Support Vector Clustering (SVC), and the proposed Novel SVDD-based (N-SVDD) algorithm on six synthetic data sets and three benchmark data sets.

The grid search method based on cross-validation is chosen to determine the values of the best model parameters in Gaussian kernel function $k(u,v)=exp(-p_1(u-v)'(u-v))+p_2$, such as $p_1=1$, $p_2=0$, penalty factor $C=100$.

a.i Data Sets in Experiments

1) Synthetic data sets: We generated six synthetic data sets, each of which is a 2 × 2 check-board data set, and the points follow the uniform distribution.

2) Benchmark data sets: The benchmark data sets we used are from the UCI machine learning repository (http://archive.ics.uci.edu/ml/), iris, glass and balance.

The details of data sets used in this section are listed in Table 2 below.

Table 2: Details of data sets

Data set	Sy. 1	Sy. 2	Sy. 3	Sy. 4	Sy. 5	Sy. 6	Iris	Glass	Balance
# Class	4	4	4	4	4	4	3	6	3
# Dim.	2	2	2	2	2	2	4	9	4
# Point	20	100	200	1000	2000	10000	150	214	625

a.ii Experimental results

We conduct the experiments with different ε on data sets mentioned before for all the algorithms to compare the performances. Training time and core vectors' number for all the algorithms, which vary with data size on the synthetic data under the best choice of ε are given in Figures 8-9. We can see that the proposed N-SVDD is of the smallest core vectors' number and the shortest training time, expect for the training time of SVM, which is of the lowest accuracy.

Generally speaking, we can conclude that in the compared algorithms, the proposed N-SVDD is of the shortest training time, the highest accuracy and the smallest core vectors' number.

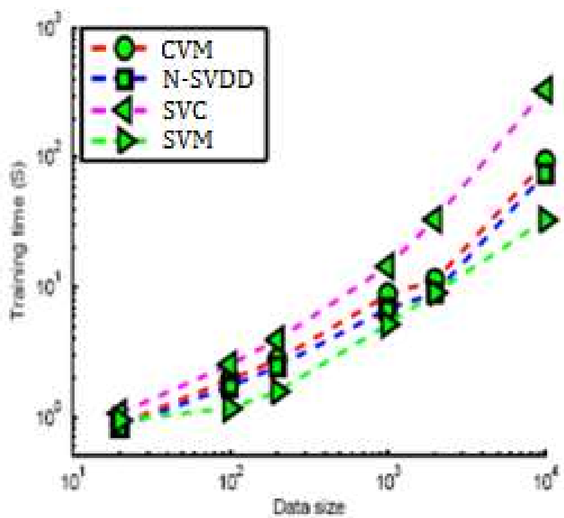


Figure 8. Training time vary with data size for different algorithms

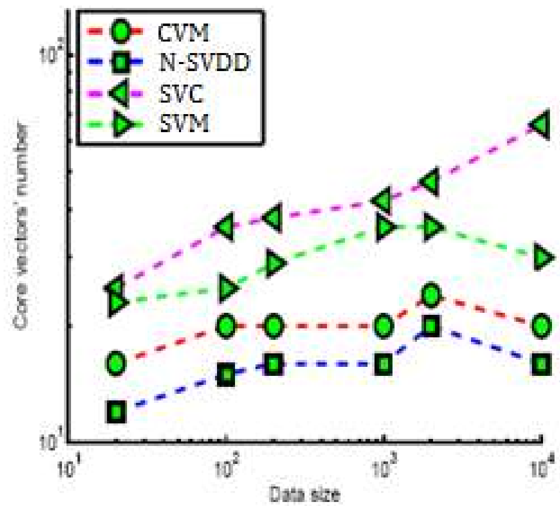


Figure 9. Core vectors' number vary with data size for different algorithms

b. Experiments on moving object detecting and tracking under dynamic scenes

We implemented the proposed N-SVDD algorithm to handle moving object detecting and tracking under dynamic scenes problem, where the tracking videos are downloaded randomly from internet by personal interest.

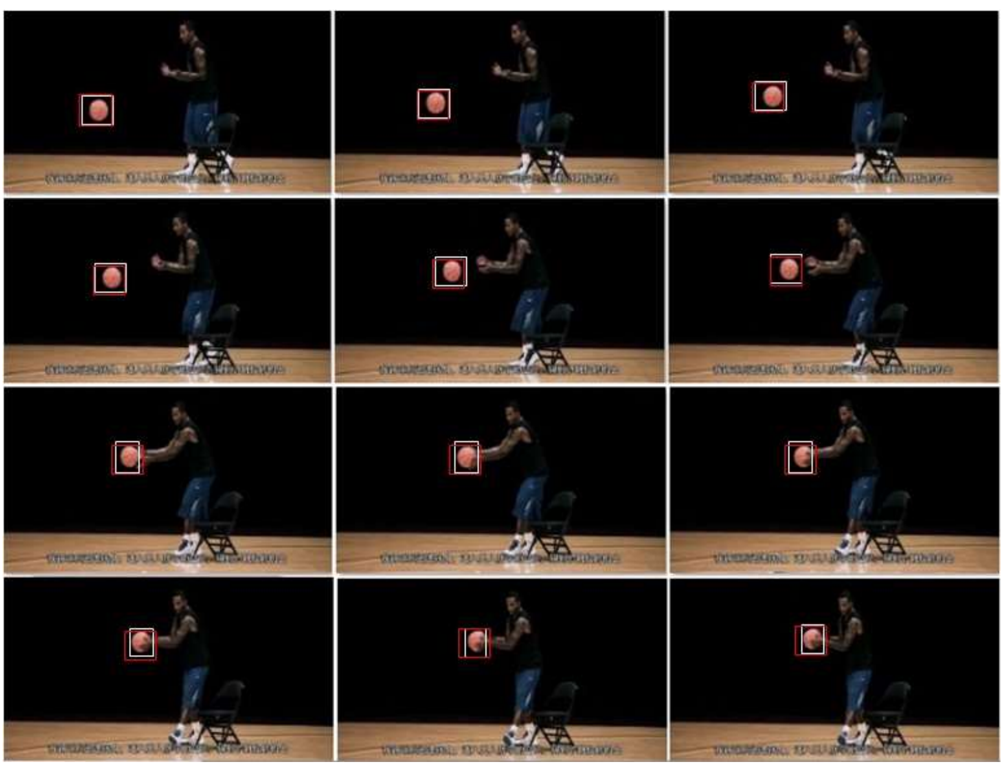


Figure 10. Detecting and tracking moving basketball under NBA star's dribbling in a contrastive background in playground (white squares stand for N-SVDD, red squares stand for Mean-Shift)

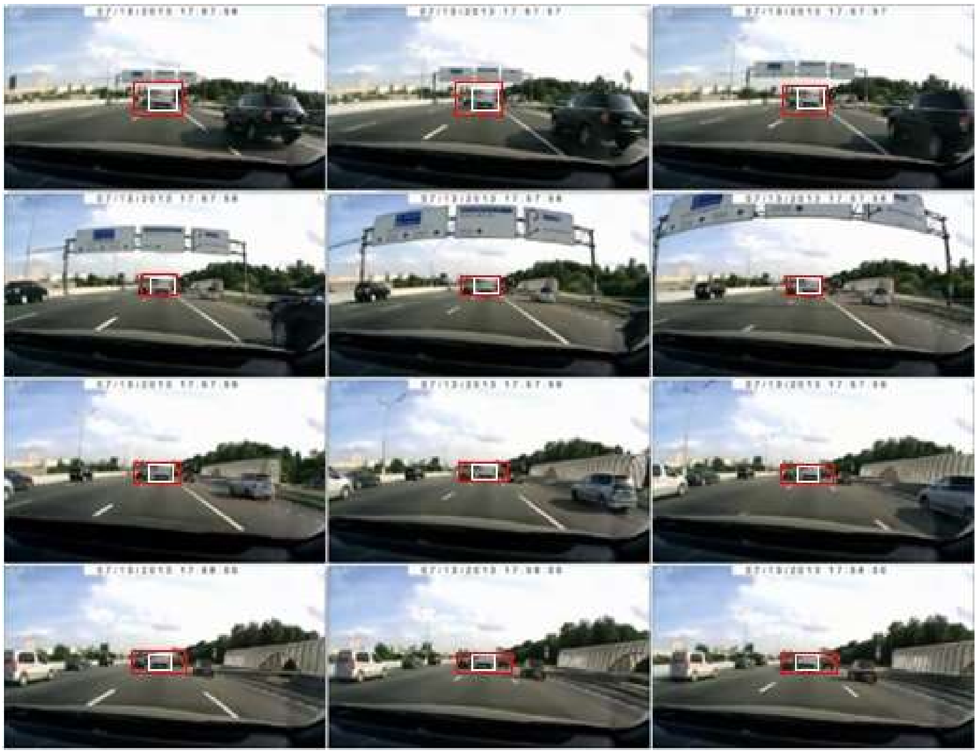


Figure 11. Detecting and tracking running truck in express way with dynamic noisy background caused by similar cars (white squares stand for N-SVDD, red squares stand for Mean-Shift)

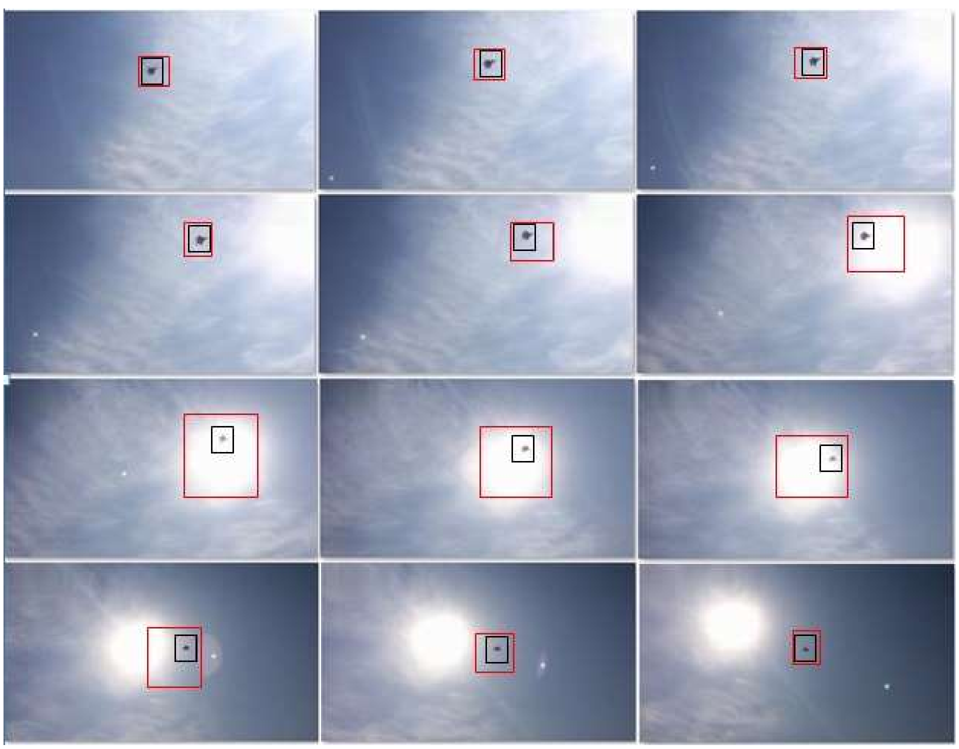


Figure 12. Detecting and tracking aircraft in clear sky with a high noisy background caused by strong sunlight (black squares stand for N-SVDD, red squares stand for Mean-Shift)

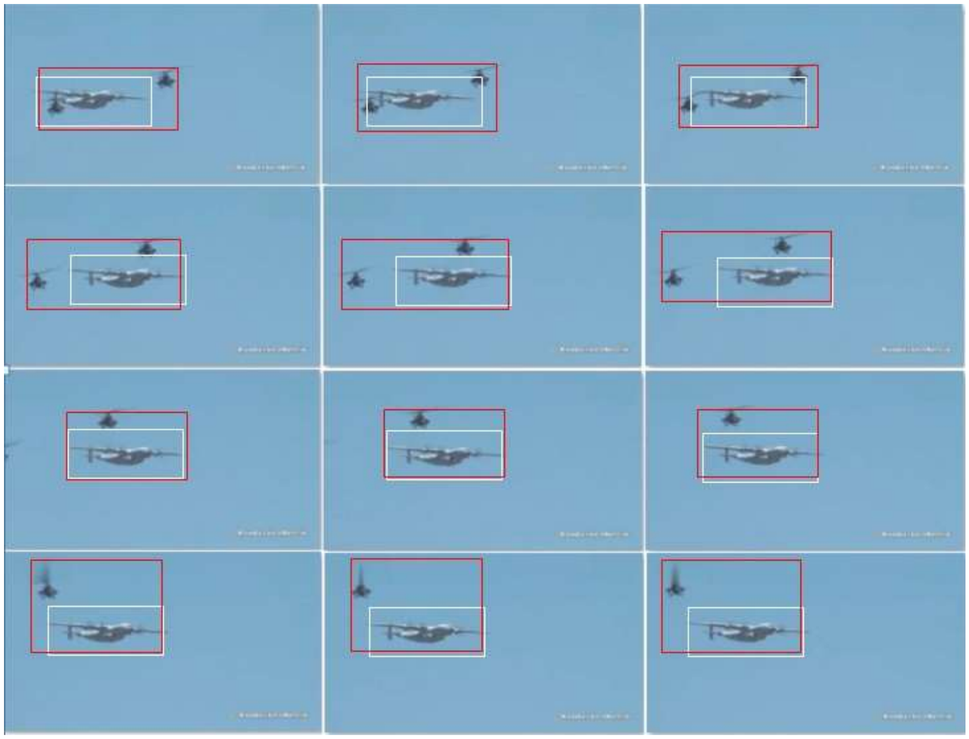


Figure 13. Detecting and tracking transport aircraft in a noisy background with similar helicopters (white squares stand for N-SVDD, red squares stand for Mean-Shift)

We detected and tracked a moving basketball under an NBA star's dribbling in a contrastive background in playground, a running truck in highway with dynamic noisy background caused by similar cars, an aircraft in clear sky with a high noisy background caused by strong sunlight, and a transport aircraft in a noisy background with several similar helicopters. We compare all the experiments implemented by N-SVDD and Mean-Shift.

Related results are given in Figures 10-13, whose detailed comparisons are analyzed as well.

We can see from Fig.10 that, in cases that moving object is distinguished from the background, and no much noise taking effect, the tracking results is pretty well, no matter what methods we take, N-SVDD or Mean-Shift, only tiny differences exist.

In Fig. 11 we can see that, when the tracked single moving object being approached too much by other similar objects, the tracking results by Mean-Shift are always mistaken by overdoing, while the tracking results by N-SVDD are more precise. The same conclusion can be drawn from Fig. 13, except for the background is changed from highway to sky.

The results presented by Fig. 12 can show that, when the background getting more noisy (such as the strong sunlight), the tracking results of Mean-Shift are tend to be less efficient than N-SVDD.

So, we can draw the conclusion that compared to classic Mean-Shift method, the N-SVDD method we proposed can maintain better tracking performance, especially when the backgrounds are noisier, such as the interferences caused by similar non-target objects, the changeable backgrounds, etc.

VI. CONCLUSIONS

Inspired by the idea proposed by Support Vector Data Description, we present a novel SVDD-based algorithm to handle moving object detecting and tracking under dynamic scenes efficiently. We prove theoretically that the proposed N-SVDD algorithm has time complexity of $O(m/\varepsilon^2 + 1/\varepsilon^3)$, which is linear in the number of training samples m for a fixed ε, and space complexity of $O(1/\varepsilon^2)$, which is independent of m for a fixed ε. The experimental results on synthetic and benchmark data, detecting and tracking results on moving objects with various dynamic backgrounds demonstrate the validity of the proposed algorithm.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 41001235), the Scientific & Technological Research Key Project of Henan Provincial Education Office (Grant No. 13A520404), the Scientific & Technological Project of Henan Provincial Science & Technology Office (Grant No. 132102210468), the Project of Science and Technology Bureau of Zhengzhou (Grant No. 20120435 and 20130713).

The authors are grateful to the anonymous referees for their valuable comments and suggestions to improve the presentation of this paper.

REFERENCES

[1] Alper Yilmaz, Omar Javed, and Mubarak Shah, "Object tracking: A survey", ACM Computing Surveys, Vol. 38, No. 4, 2006, pp. 1-45.

[2] C. VEENMAN, M. REINDERS, and E. BACKER, "Resolving motion correspondence for densely moving points", IEEE Trans. Patt. Analy. Mach. Intell., Vol. 23, No. 1, 2001, pp. 54-72.

[3] D. SERBY, S. KOLLER-MEIER, and L. V. GOOL, "Probabilistic object tracking using multiple features", In IEEE International Conference of Pattern Recognition, 2004, pp. 184-187.

[4] D. COMANICIU, V. RAMESH, and P. MEER, "Kernel-based object tracking", IEEE Trans. Patt. Analy. Mach. Intell., Vol. 25, 2003, pp. 564-575.

[5] A. ALI, and J. AGGARWAL, "Segmentation and recognition of continuous human activity", In IEEE Workshop on Detection and Recognition of Events in Video, 2001, pp. 28-35.

[6] A. YILMAZ, X. LI, and M. SHAH, "Contour based object tracking with occlusion handling in video acquired using mobile cameras", IEEE Trans. Patt. Analy. Mach. Intell., Vol. 26, No. 11, 2004, pp. 1531-1536.

[7] T. COOTES, and C. TAYLOR, "Robust real-time periodic motion detection, analysis, and applications", IEEE Trans. Patt. Analy. Mach. Intell., Vol. 23, No. 6, 2001, pp. 681-685.

[8] N. PARAGIOS, and R. DERICHE, "Geodesic active regions and level set methods for supervised texture segmentation", Int. J. Comput. Vision, Vol. 46, No. 3, 2002, pp. 223-247.

[9] A. ELGAMMAL, R. DURAISWAMI, and D. HARWOOD, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance", Proceedings of IEEE, Vol. 90, No. 7, 2002, pp. 1151-1163.

[10] P. FIEGUTH, and D. TERZOPOULOS, "Color-based tracking of heads and other mobile objects at video frame rates", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997, pp. 21-27.

[11] G. EDWARDS, C. TAYLOR, and T. COOTES, "Interpreting face images using active appearance models", In International Conference on Face and Gesture Recognition, 1998, pp. 300-305.

[12] S. AVIDAN, "Support vector tracking", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 184-191.

[13] S. PARK and J. K. AGGARWAL, "A hierarchical bayesian network for event recognition of human actions and interactions", Multimed. Syst., Vol. 10, No. 2, 2004, pp. 164-179.

[14] C. VEENMAN, M. REINDERS and E. BACKER, "Resolving motion correspondence for densely moving points", IEEE Trans. Patt. Analy. Mach. Intell, Vol. 23, No. 1, 2001, pp. 54-72.

[15] I. SETHI and R. JAIN, "Finding trajectories of feature points in a monocular image sequence", IEEE Trans. Patt. Analy. Mach. Intell, Vol. 9, No. 1, 1987, pp. 56-73.

[16] S. INTILLE, J. DAVIS, and A. BOBICK, "Real-time closed-world tracking", In IEEE

Conference on Computer Vision and Pattern Recognition (CVPR), 1997, pp. 697-703.

[17] K. SHAFIQUE and M. SHAH, "A non-iterative greedy algorithm for multi-frame point correspondence", In IEEE Intern. Conference on Comp. Vision (ICCV), 2003, pp. 110-115.

[18] M. ISARD and A. BLAKE, "Condensation-conditional density propagation for visual tracking", Int. J. Comput. Vision, Vol. 29, No. 1, 1998, pp. 5-28.

[19] S. ZHOU, R. CHELLAPA and B. MOGHADAM, "Adaptive visual tracking and recognition using particle filters", In Proceedings IEEE International Conference on Multimedia and Expo. (ICME), 2003, pp. 349-352.

[20] N. VASWANI, A. ROYCHOWDHURY and R. CHELLAPPA, "Activity recognition using the dynamics of the configuration of interacting objects", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003, pp. 633-640.

[21] S. BIRCHFIELD, "Elliptical head tracking using intensity gradients and color histograms", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1998, pp. 232-237.

[22] P. FIEGUTH and D. TERZOPOULOS, "Color-based tracking of heads and other mobile objects at video frame rates", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997, pp. 21-27.

[23] H. SCHWEITZER, J. W. BELL and F. WU, "Very fast template matching", In European Conference on Computer Vision (ECCV), 2002, pp. 358-372.

[24] D. COMANICIU, V. RAMESH and P. MEER, "Kernel-based object tracking", IEEE Trans. Patt. Analy. Mach. Intell, Vol. 25, 2003, pp. 564-575.

[25] A. JEPSON, D. FLEET and T. ELMARAGHI, "Robust online appearance models for visual tracking", IEEE Trans. Patt. Analy. Mach. Intell, Vol. 25, No. 10, 2003, pp. 1296-1311.

[26] M. BLACK and A. JEPSON, "Eigen-tracking: Robust matching and tracking of articulated objects using a view-based representation", Int. J. Comput. Vis., Vol. 26, No. 1, 1998, pp. 63-84.

[27] S. AVIDAN, "Support vector tracking", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 184-191.

[28] D. HUTTENLOCHER, J. NOH and W. RUCKLIDGE, "Tracking non-rigid objects in complex scenes", In IEEE Intern. Conference on Computer Vision (ICCV), 1993, pp. 93-101.

[29] J. KANG, I. COHEN and G. MEDIONI, "Continuous tracking within and across camera streams", In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003, pp. 267-272.

[30] J. KANG, I. COHEN and G. MEDIONI, “Object reacquisition using geometric invariant appearance model”, In Intern. Conference on Pattern Recognition (ICPR), 2004, pp. 759-762.

[31] N. K. Suryadevara and S. C. Mukhopadhyay, “Wireless Sensor Network Based Home Monitoring System for Wellness Determination of Elderly”, IEEE Sensors Journal, Vol. 12, No. 6, June 2012, pp. 1965-1972.

[32] N. K. Suryadevara and S. C. Mukhopadhyay, “Determining Wellness Through An Ambient Assisted Living Environment”, IEEE Intelligent Systems, May/June 2014, pp. 30-37.

[33] J. MACCORMICK and A. BLAKE, “Probabilistic exclusion and partitioned sampling for multiple object tracking”, Int. J. Comput. Vision, Vol. 39, No. 1, 2000, pp. 57-71.

[34] Y. CHEN, Y. RUI, and T. HUANG, “JPDAF based HMM for real-time contour tracking”, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001, pp. 543-550.

[35] A. MANSOURI, “Region tracking via level set PDES without motion computation”, IEEE Trans. Patt. Analy. Mach. Intell, Vol. 24, No, 7, 2002, pp. 947-961.

[36] D. CREMERS and C. SCHNORR, “Statistical shape knowledge in variational motion segmentation”, Image and Vision Computing, Vol. 21, 2003, pp. 77-86.

[37] A.YILMAZ, X. LI and M. SHAH, “Contour based object tracking with occlusion handling in video acquired using mobile cameras”, IEEE Trans. Patt. Analy. Mach. Intell., Vol. 26, No. 11, 2004, pp. 1531-1536.

[38] D. M. J. Tax and R. P. W. Duin, “Support Vector Domain Description”, Pattern Recognition Letters, Vol. 20, No. 14, 1999, pp. 1191-1199.

[39] I. W. Tsang, J. T. Kwok and P. M. Cheung, “Core vector machines: Fast SVM training on very large data sets”, Journal of Machine Learning Research, Vol. 6, 2005, pp. 363-392.

[40] S. Asharaf, M. N. Murty and S. K. Shevade, “Multiclass Core Vector Machine”, Proc. the 24th International Conference on Machine Learning (ICML 2007), 2007, pp. 41-48.

[41] Francesco Orabona, “Better Algorithms for Selective Sampling”, Proc. the 28th International Conference on Machine Learning (ICML 2011), 2011, pp. 433-440.

[42] Y. Liu, “Combining integrated sampling with SVM ensembles for learning from imbalanced datasets”, Information Processing and Management, Vol. 47, No. 4, 2011.

[43] M. Volpi, “Memory-Based Cluster Sampling for Remote Sensing Image Classification”, IEEE Transactions on Geoscience and Remote Sensing, Vol. 50, No. 8, 2012.

[44] Abhimanyu Das and David Kempe, "Submodular meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection", Proc. the 28th International Conference on Machine Learning (ICML 2011), 2011, pp. 1057-1064.

[45] N. K. Suryadevara, S. C. Mukhopadhyay, R. Wang, R. K. Rayudu, "Forecasting the behavior of an elderly using wireless sensors data in a smart home", Engineering Applications of Artificial Intelligence, Vol. 26, No. 10, 2013, pp. 2641-2652.

[46] G. Sen Gupta, S. C. Mukhopadhyay and M. Finnie, "Wi-Fi Based Control of a Robotic Arm with Remote Vision", Proceedings of 2009 IEEE I2MTC Conference, Singapore, May 5-7, 2009, pp. 557-562.

[47] Y. Q. Wang, Y. Li, and L. Chang, "Approximate Minimum Enclosing Ball Algorithm with Smaller Core Sets for Binary Support Vector Machine", Proc. the 2010 Chinese Control and Decision Conference (CCDC 2010), 2010, pp. 3404-3408.

[48] Yongqing Wang, "NEW INTELLIGENT CLASSIFICATION METHOD BASED ON IMPROVED MEB ALGORITHM", International Journal on Smart Sensing and Intelligent Systems, Vol. 7, No. 1, 2014, pp. 72-95.

[49] Yongqing Wang, "FACE RECOGNITION BASED ON IMPROVED SUPPORT VECTOR CLUSTERING", International Journal on Smart Sensing and Intelligent Systems, Vol. 7, No. 4, 2014, pp. 1807-1829.