Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 420286, 13 pages
http://dx.doi.org/10.1155/2013/420286

Research Article

Spatial Object Tracking Using an Enhanced Mean Shift Method Based on Perceptual Spatial-Space Generation Model

Pengcheng Han, Junping Du, and Ming Fang

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Junping Du; junpingdu@126.com

Received 10 January 2013; Accepted 20 March 2013

Academic Editor: Graziano Chesi

Object tracking is one of the fundamental problems in computer vision, but existing efficient methods may not be suitable for spatial object tracking. Therefore, it is necessary to propose a more intelligent mathematical model. In this paper, we present an intelligent modeling method using an enhanced mean shift method based on a perceptual spatial-space generation model. We use a series of basic and composite graphic operators to complete signal perceptual transformation. The Monte Carlo contour detection method could overcome the dimensions problem of existing local filters. We also propose the enhanced mean shift method with estimation of spatial shape parameters. This method could adaptively adjust tracking areas and eliminate spatial background interference. Extensive experiments on a variety of spatial video sequences with comparison to several state-of-the-art methods demonstrate that our method could achieve reliable and accurate spatial object tracking.

1. Introduction

Mathematical formalism is probably the most precise and logical language in science research. It is typical for researchers in pure natural sciences to attempt to describe observed phenomena using mathematical correlations. However, because of real-world scenarios, it is often very difficult to construct a perfect and permanent mathematical model for one specific issue in engineering fields [1]. During the last years, the effort concentrated in self-optimizing and self-adaption was leading to a new field between mathematics and applications, called intelligent modeling [2]. In this paper, we propose a new intelligent modeling method using the enhanced mean shift method based on a perceptual spatial-space generation model for spatial object tracking. Object tracking has been applied to many fields, such as video surveillance [3], robot recognition [4], and traffic control [5]. In spatial on-orbit docking, object tracking could be used to track spatial aircraft and assist with ground control. Because the spatial images are mainly generated from low-rate videos [6] or airborne spectral imagery [7], which are captured by the aircraft sensors [8, 9], their resolution and spatial-temporal coverage are not very ideal. In addition, because of differences in the sensor spectral bands, acquisition position, and contrast gradient setting, there are shifts in the relative position and scale zoom in multisource images with the same scene. All this will bring influence to spatial object tracking results.

Recently, multidimensional decomposition and multi-scale representation methods have been widely applied to image processing and computer vision. Mumford and Gidas proposed the stochastic model [10], in which truncation errors and noise interference could be isolated from the discrete domain. Witkin and Koenderink proposed the image scale-space model [11, 12], which cleared noise interference in fine scales, and analysis errors could decrease in coarse scales. Burt and Lindeberg proposed the coarse-to-fine model [13, 14], which could reduce useless gradients in gradient entropy calculation. In the object tracking field, the traditional method is rectangular block region tagging [15, 16]. In [17], Isard and Blake applied a local filter to object tracking field. Sun and Liu proposed using a combination of the local description and global representation in object tracking [18]. Recently, a graphics model based on a Bayesian neural network was also applied to continuous object tracking [19]. However, if we applied these methods to spatial object

tracking, spatial background clutter and moving object overlapping appeared in different scenes. The existing multiscale deviation will seriously reduce the spatial object tracking accuracy.

In this paper, we propose a spatial object tracking method using an enhanced mean shift method based on a perceptual spatial-space generation model. We detect the spatial object continuity and saliency between different scales in the perceptual spatial-space generation model. The enhanced mean shift considers the relevance between motion area and static background. It could achieve more robust object tracking. Our proposed method is shown in Figure 1. This paper is organized as follows. Section 2 describes the perceptual spatial-space generation model. Section 3 proposes an enhanced mean shift method. Section 4 shows experimental results. Section 5 is the conclusion.

2. Perceptual Spatial-Space Generation Model

Spatial images have a high amount of data structure, and their processing may not obey existing image processing model assumptions. In this paper, we propose a perceptual spatial-space generation model. It consists of two parts: a prototype pyramid, labeled $S[0,n] = (S_0, \ldots, S_n)$, and a set of perceptual transform rules, labeled $R[0,n-1] = (R_0, R_1, \ldots, R_{n-1})$. The set $I[0,n]$ is the Gaussian transform pyramid. Our goal is to set a common variable $p(I[0,n], S[0,n], R[0,n-1])$ and a maximized posterior probability $p(S[0,n], R[0,n-1] \mid I[0,n])$ that can be used to calculate the priority prototype pyramid and the perceptual transform rules.

2.1. Prototype Pyramid Generation. Generation model is a joint probability function with prototype S and image I, and Δ_R is a dictionary which includes image primitives, such as blobs, edges, crosses, and bars. It can be expressed as

$$p(I; S; \Delta_R) = p(I \mid S; \Delta_R)\, p(S). \tag{1}$$

The decomposition probability could be divided into primitives and texture.

Consider the following:

$$\begin{aligned} &p(I \mid S; \Delta_R) \\ &= \prod_{k=1}^{N_R} \exp\left(- \sum_{(u,v)\in R} \left(\frac{I(u,v) - R_k(u,v)^2}{2\sigma_o^2} \right) \right) \\ &\cdot \prod_{j=1}^{N_R} p\left(I_{R,j} \mid I_R; \sigma_j\right), \end{aligned} \tag{2}$$

where $S = \langle V, E \rangle$ is the properties graphics, V is the collection of primitives in S, (u, v) denoting the primitives in the dictionary Δ, and σ is the variance of the corresponding primitive features. The priority model $p(S)$ is an uneven Gibbs model, which defines the graphical attributes on S. It focuses on continuity properties in the perceptual model, such as smoothness, continuity, and typical functions.

Consider the following:

$$\begin{aligned} &p(S) \\ &= \exp\left\{ \varepsilon \sum_{d=0}^{4} \sum_{(u,v)\in R} R_d \left|\varphi(u,v)\right| + (1-\varepsilon) \sum_{(u,v)\in R} \varphi\left(B_u, B_v\right) \right\}, \end{aligned} \tag{3}$$

where R_d is the primitive mark in S and its connection degree is d. $\varphi(B_u, B_v)$ is potential association for two correlation functions. Because there is uncertainty in the inner perception posterior probability, the prototype pyramid may not appear continuous for each layer calculation. In order to ensure transition and consistency of single frame diagram, we define a set of graphical operation factors

$$R_k = \left(r_{k,1}, r_{k,2}, \ldots, r_{k,m(k)}\right), \quad r_{k,i} \in \sum_{\text{gram}}. \tag{4}$$

Graphical operation factors could synthesize detected graphic edges into pairs of characteristic bridges. Each bridge is related with the properties of probability function. The conversion from S_k to S_{k+1} is realized by a series of conversion rules R_k, and rule order could directly determine conversion efficiency. The generative rule graphic path from S_k to S_{k+1} could be expressed as

$$\begin{aligned} &p(R_k) = p(S_{k+1} \mid S_k) \\ &\quad = \prod_{i=1}^{m(k)} \left[p\left(r_{k,i}\right) + \frac{1}{m_k^i} \sum_{j=1}^{m_k^i} \left(w_k^{i,j} - \left| d_j - d_k^i \right| \right) \right], \\ &p(I[0,n], S[0,n], R[0,n-1]) \\ &\quad = \prod_{k=0}^{n} p\left(I_k \mid S_k; \Delta_{sk}\right) \cdot p(S_0) \cdot \prod_{k=0}^{n-1} \prod_{j=1}^{m(k)} p\left(r_{k,i}\right), \end{aligned} \tag{5}$$

where S is the optimal-calculated prototype. Under the condition that there will be no loss in perception model accuracy, we assume that S_{sm} begins to decay from S through the single operation factor. I_{sm} will gradually reduce the resolution, and S_{sm} will also have a related complexity. The posterior probability could be expressed as

$$\begin{aligned} P(sm) &= \log \frac{p\left(I_{sm} \mid S_{sm}\right)}{p\left(I_{sm} \mid S\right)} + \lambda_{sm} \log \frac{p\left(S_{sm}\right)}{p(S)} \\ &= \log \frac{p\left(S_{sm} \mid I_{sm}\right)}{p\left(S \mid I_{sm}\right)}. \end{aligned} \tag{6}$$

A layered reduced perception model will not completely adapt to the complex model S, so the first logarithmic ratio is often negative. The parameter λ is used to balance model fitness and complexity. If $\lambda = 1$, we could launch simplified $P(sm)$. λ_{sm} could be decided in the following range:

$$0 < \log \frac{p\left(I_{sm} \mid S_{sm}\right)}{p\left(I_{sm} \mid S\right)} + \lambda_{sm} \log \frac{p\left(S_{sm}\right)}{p(S)} < 1. \tag{7}$$

The transform between S_k and S_{k+1} is achieved using a group of greed detection. The accurate scale of graphical operation factor will make differences based on the subjective goal. We suppose that the graphical operation factor is between I and I_{sm}

$$a_1 = -\log \frac{p\left(I_{sm} \mid S_{sm}\right)}{p\left(I_{sm} \mid S\right)}, \qquad b_1 = \lambda_{sm} \log \frac{p\left(S_{sm}\right)}{p\left(S\right)},$$
$$a_2 = -\log \frac{p\left(I \mid S_{sm}\right)}{p\left(I \mid S\right)}, \qquad b_2 = \lambda_{sm} \log \frac{p\left(S_{sm}\right)}{p\left(S\right)}. \tag{8}$$

Based on formulas (6) and (7), we determined that the corresponding interval is

$$N\left(\left|r_a - r_k^{i,j}\right|; 0, \sigma_a^2\right) < \lambda_{sm} < N\left(\left|r_b - r_b^{i,j}\right|; 0, \sigma_b^2\right). \tag{9}$$

2.2. Perceptual Transform of Prototype Pyramid. In this section, our goal is to determine the optimum conversion path and deduce the hidden graphics prototype. Our method is scanning the prototype pyramid from top to bottom based on each primitive learning decision rule. Our method can be divided into three steps.

Step 1 (prototype pyramid independent calculation). We apply a pyramid algorithm in the bottom of image I and calculate the Gaussian pyramid S_0. Because each prototype layer is calculated using a MAP estimation [20], there is a certain loss in the continuity of the prototype pyramid. The specific formula is as follows:

$$\left(S\left[0, n\right], R\left[0, n-1\right]\right) = \arg\max \prod_{k=0}^{n} p\left(I_k \mid S_k; \Delta_k\right) \cdot p\left(S_0\right) \prod_{k=1}^{n} \prod_{j=1}^{m(k)} p\left(\lambda_{k,j}\right). \tag{10}$$

Step 2 (pattern matching from bottom to up). We match the image prototype attribute from S_k to S_{k+1} using an image registration algorithm from a previous study [21]. We use x, y, z, and e as a judgment function to process each node i and obtain the related image characteristics at each time. Specifically, the matching degree between the ith node at the kth scale and the jth node at the $(k+1)$th scale can be expressed as

$$\text{match}\left(i, j\right) = \frac{1}{Z} \exp\left\{-\frac{\left(x\left(i\right) - x\left(j\right)\right)^2}{2\sigma_x^2} - \frac{\left(y\left(i\right) - y\left(j\right)\right)^2}{2\sigma_y^2} - \frac{\left(z\left(i\right) - z\left(j\right)\right)^2}{2\sigma_z^2} - \frac{\left(e\left(i\right) - e\left(j\right)\right)^2}{2\sigma_e^2}\right\}, \tag{11}$$

where σ is the variance of related features. For similarity matching between S_k and S_{k+1}, this formulation allows the empty variable prototype to appear in S_k. We multiply S_k by the related variance σ and obtain a homologous S_{k+1} with subsidiary value. Pattern matching results will be used as the initial parameter for the Markov chain matching in the next step.

Consider the following:

$$P\left(S_k, S_{k+1}\right) = \left\{\left(x; y; z; e\right) \longleftarrow S^2 : \left(k'_{\min} < k < k'_{\min}\right) \times \left(\overline{\left(x; z\right) \in \text{match}\left[v_i\left(k\right), v_i\left(k+1\right)\right]}\right)\right\}. \tag{12}$$

Step 3 (Markov chain matching). Due to the uncertainty for the initial perception in the posterior probability model and the relative complexity for dynamic graphic structure mining, we use the reversibility of Markov chain matching to match a perceptual transform. The Markov chain includes 25 pairs of reversible jump. In each path of Markov chain matching, there is a reversible jump between X_1 and X_2 [22]. These reversible jumps are related to their corresponding grammars, and each pair of these rules is based on probability selection. We use this mechanism to optimize the perception conversion path, which could lead a cross-scale continuous perception prediction.

Consider the following:

$$\left(S\left[0, n\right], \text{Markov}\left[0, n-1\right]\right) = \arg / \max \prod_{k=0}^{n} p\left(Ik \mid Sk; \Delta R\right) \cdot \prod_{i}^{n} \prod_{j}^{m} g\left(r_{i,j}\right). \tag{13}$$

2.3. Object Contour Evolution. The problem of such matching strategy is that it does not contain any matching, which could split a long contour into short edges. We propose a new Monte Carlo contour detection method. This method mainly chooses the right standard in certain scale using spatial-space domain knowledge. Our proposed method uses the weight set $\{x_{i,t}, w_{i,t}\}$ to estimate the posterior probability density $p(x_{i,t}, w_{i,t})$. According to the resampling theory [23], it is feasible to calculate the sample specimen $Q(x, w)$ with appropriate weights from the normal density distribution X_t^i.

Consider the following:

$$w_t^{i,n} = \frac{p\left(z_t^i \mid x_t^i, z_t^{N(i)}, x_{0:t-1}^i, z_{1:t-1}^i, Z_{1:t-1}^{N(i)}\right)}{p\left(z_t^i, Z_{1:t-1}^{N(i)} \mid z_{1:t-1}^i\right) \cdot Q\left(x, w\right)}. \tag{14}$$

For the sequence sample, the important probability $Q(x, w)$ can be chosen using the following mode:

$$Q\left(x, w\right) = q\left(x_t^i \mid x_{0:t-1}^i, z_{1:t}^i, Z_{1:t}^{N(i)}\right) q\left(x_{0:t-1}^i \mid x_{1:t-1}^i, z_{1:t}^i, Z_{1:t-1}^{N(i)}\right). \tag{15}$$

The entire probability can be approximated with a simplified formulation

$$w_t^{i,n} = \frac{p\left(x_t^i \mid z_t^i\right) p\left(x_t^{i,n} \mid z_{t-1}^{i,n}\right)}{q\left(x_t^i \mid x_{0:t-1}^i, z_{1:t}^i, Z_{1:t}^{N(i)}\right)} \times \prod_{j \in N(i)} \left\{\sum_{t=1}^{N} p\left(z_t^i \mid x_t^{j,n}\right) p\left(x_t^{j,n} \mid z_t^{j,n}\right)\right\}. \tag{16}$$

We use only one part of the whole sample; $p(x_t^{i,n} \mid z_{t-1}^{i,n})$ simulates the interaction value between the two adjacent areas $x_t^{i,n}$ and $x_t^{j,l}$. The local probability $p(z_t^j \mid x_t^{j,l})$ is the weight of the interactive area. In this paper, we use a sequence of Monte Carlo simulations to estimate the interactive part $p(x_t^j \mid x_t^i)$ with its related estimation weight. The density importance function can be defined as follows:

$$\begin{aligned} &q\left(x_{0:t}^i \mid x_t^i, z_{1:t}^i, Z_{1:t}^{N(i)}, Z_{1:t}^{R(I)}\right) \\ &\quad = q\left(x_t^i \mid x_{0:t-1}^i, z_{1:t}^i, Z_{1:t}^{N(i)}, Z_{1:t}^{R(I)}\right) \\ &\qquad \times q\left(x_{0:t-1}^i \mid x_{1:t}^i, z_{1:t}^i, Z_{1:t-1}^{N(i)}, Z_{1:t-1}^{R(I)}\right). \end{aligned} \tag{17}$$

Based on the resampling theory, the sampling weight can be updated as follows:

$$\begin{aligned} w_t^{i,n} = w_{t-1}^{i,n} &\frac{p\left(x_t^i \mid z_t^i\right) p\left(x_t^{i,n} \mid z_{t-1}^{i,n}\right)}{q\left(x_t^i \mid x_{0:t-1}^i, z_{1:t}^i, Z_{1:t}^{N(i)}, Z_{1:t}^{R(I)}\right)} \\ &\times \prod_{j \in N(i)} \left\{ \sum_{t=1}^{N} p\left(z_t^i \mid x_t^{j,n}\right) p\left(x_t^{j,n} \mid z_t^{j,n}\right) \right\} \\ &\times \prod_{j \in N(i)} \left\{ \sum_{L=1}^{N^L} p\left(z_t^L \mid x_t^{k,L}\right) p\left(x_t^{k,L} \mid z_t^{L,n}\right) \right\}, \end{aligned} \tag{18}$$

where N is the sample retrieval in the ith part, L is the sample length of the jth part, and z_t^L is the sample set of the jth part. As we use the Markov property, the density function $p(z_t^k \mid x_t^{k,L})$ could do further approximation relied on the unit product of local observation similarity unit K.

Consider the following:

$$\begin{aligned} p\left(z_t^k \mid x_t^{k,L}\right) &= p\left(z_t^{N(k)}, z_t^{R(L)} \mid x_t^k, x_{0:t-1}^L\right) \\ &= \frac{p\left(z_t^k \mid x_t^k\right) \cdot p\left(x_t^k, Z_t^{N(k)}, Z_t^{R(L)} \mid x_{0:t-1}^k\right)}{p\left(z_t^k, Z_t^{N(k)} \mid Z_{1:t-1}^k, Z_{1:t-1}^{N(t)}, Z_{1:t-1}^{R(L)}\right)}. \end{aligned} \tag{19}$$

In our proposed framework, all parts of the detected object will be tracked at the same time. There is no need to calculate all unit probabilities $p(z_t^k \mid x_t^{k,L})$, so we can use the local possibility and directly estimate the weight of all units.

2.4. Markov Random Field Representation. We propose Markov random field representation for perception generation model. We define the pixel perception set X, in which any two arbitrary pixels are adjacent. The adjacent relation is an interactive relationship; in the an adjacent pixels system $E = \{\delta(x, y): (x, y) \in I\}$, if (x, y) is adjacent point of (w, v), then (w, v) is also adjacent with (x, y), and $p(I)$ is the Markov random field with respect to the perception set X

$$p\left(I_{x,y} \mid X(x, y)\right) = p\left(I_{x,y} \mid I(\delta \setminus X(x, y))\right). \tag{20}$$

We can define $I(\delta)$ as the pixels set that contains all pixels of X. The Markov property mentioned in formula (20) relies on the density distribution of the adjacent pixels. According to S. Geman and D. Geman [24], the Markov random field-related pixels in system X can be rewritten as the following Gibbs distribution:

$$p(I) = \frac{1}{2Z\left|\Sigma_C\right|^{1/2}} \exp\left\{ -\sum_X G_x(I(E)) \cdot \sum_C^{-1} \left(E_x^i - E_y^i\right) \right\}, \tag{21}$$

where $G_x(I(E))$ is the potential variable function defined in $I(\delta)$. Z is a conventional constant, which can maintain the $p(I)$ sum up to 1. For spatial object tracking, the perception latent variables that appear in pairs are difficult to describe accurately. If X has an amount of pixels, then G_x will be a multidimensional function. Here, we use the topological model to solve the above problems. Assuming that the spatial perceptual rules set is $\varphi(I)$, then $\varphi_{x,y}(I)$ could extract the local characteristics near pixels (x, y). A specific example is $\varphi_{x,y}(I) = \langle I, \beta_{x,y,s} \rangle$, $s = (I, \theta)$. We transform the image I with the Gibbs distribution

$$\begin{aligned} GD(I \mid \beta) &= \frac{1}{Z} \exp\left\{ GD\left(p\left(I^{(k)}\right)\right) - GD'(p(X)) \right\} \\ &= \frac{1}{Z} \exp\left\{ \sum_{s=1}^{S} \sum_{x,y} \beta_{x,y,s}\left(X_{x,y,s}(I)\right) \right. \\ &\qquad \left. - \sum_{x,y} \frac{p(I)}{p_{x,y,s} \cdot (I - p(I))} \right\}, \end{aligned} \tag{22}$$

where $\beta_{x,y,s}$ is the low-dimensional image characteristic function set, Z is the conventional constant that depends on β, and $X_{x,y,s} = (I, \beta_{x,y,s})$. Assuming that the normal distribution is $p(I)$, $p_{x,y,s}$ is the normal distribution of $\beta_{x,y,s}$ under $I - p(I)$. β could guarantee $f^*(I) = f(I \mid \beta)$. For each (x, y, s), the marginal distribution of $\varphi(x, y, s)$ is $p(x, y, s)$. With any given $f = f(I \mid \beta)$, $D(p \mid f^*) \le D(p \mid f)$, the specific calculation of the axiom is as follows:

$$\begin{aligned} &D(p \mid f) - D(p \mid f^*) \\ &= \log \frac{Z(\beta)}{Z(\beta^*)} \left\{ \sum_{s=1}^{S} \sum_{x,y} \beta_{x,y,s}^*\left(X_{x,y,k}(I)\right) - \beta_{x,y,s}\left(X_{x,y,k}(I)\right) \right\} \\ &= D(f^* \parallel f) \ge 0. \end{aligned} \tag{23}$$

In the formulation of $f(I \mid \beta)$, f^* can be seen as the best approximation of probability p, and it also can be marked as the "maximum likelihood." From the minimized $D(p \parallel f(I \mid \beta))$ to the maximized $E[\log f(I \mid \beta)]$, the marginal probability $\varphi(x, y, s)$ can be $p_{x,y,s}(\,)$ for any (x, y, s) under $p(I)$. So $H(f^*) - H(f) = D(p \parallel f^*)$.

Consider the following:

$$\begin{aligned} H(f^*) - H(f) &= E_q[\log q(I)] - E_{f^*}[\log f^*(I)] \\ &= E_q[\log q(I)] - E_q[\log f^*(I)] \\ &= D(q \,||\, f^*). \end{aligned} \tag{24}$$

In order to model the observed image, we assume that $\beta_{x,y,s}(\,) = \beta_s(\,)$, β does not depend on (x, y). We can continue to parameterize the β_s or normalize β in a low-dimensional scale. If we normalize β_s in $S = 1, \ldots, L$ and make $\beta_s(\varphi_{x,y,s}(I)) = \beta_{sl}$, we could rewrite the Gibbs distribution as follows:

$$\begin{aligned} GD(I \mid \beta) &= \frac{1}{Z} \exp\left\{ \sum_{s=1}^{S} \sum_{x,y} \sum_{l} \beta_{x,y,s} \left(X_{x,y,s}(I) \right) \right. \\ &\qquad \left. - \sum_{x,y} \sum_{l} \frac{p(I)}{p(x,y,s) \cdot (I - p(I))} \right\} \\ &= \frac{1}{Z} \exp\left\{ \sum_{s} \sum_{l} \beta_{sl} H_{kl}(I) - \sum_{x,y} \sum_{l} p'(I) \cdot H_k(I) \right\} \\ &= \frac{1}{Z} \exp\left\{ \sum_{s} \langle \beta_k, H_k(I) \rangle \right\}, \end{aligned} \tag{25}$$

where $H_k(I) = p_{x,y}\varphi_{x,y,s}(I)$ represents the quantity of the effective points $\varphi_{x,y,s}(I)$ which falls into the interval S and $H_k = (H_k, L)$ is the marginal matrix related to $\{\varphi_{x,y,s}, S\}$. If we want to find the maximum estimate coefficient β, we need to calculate the spatial-scale statistical coefficient $H_k(I)$

$$E_\beta[H_k(I)] = \left\{ \frac{1}{n} \sum_{k=1}^{n} H\left(p\left(I^{(k)} \right) \right) - H_k\left(I_{\text{observed}} \right) \right\}. \tag{26}$$

In other words, we need to match the spatial data and the related model. The most suitable model is determined by $H_k(I_{\text{observed}})$, and $E_\beta[H_k(I)]$ is an average parameter; value β is a natural parameter. In the perceptual model, there will be a global balance variable. We can make a minor adjustment

$$\begin{aligned} p\left(I_0 \mid I_{\rho 0} \right) &= \frac{1}{Z} \exp\left\{ \sum_{k} \sum_{x,y \in p_0} \beta_k \left(\Phi_{x,y,s} \right) \right\} \\ &\quad + \sum_{k} E_w \left[\log \frac{p(w \mid I_s)}{p(w \mid F)} \right], \end{aligned} \tag{27}$$

where $I_{\rho 0}$ is the approximation value of global observation estimated from observed spatial images. The local equilibrium parameter defined in any pixel will obey the specific distribution

$$\begin{aligned} &p(I) \\ &= \frac{1}{Z} \exp\left\{ \sum_{k=1}^{K} \left(C_{m,k} B_k + \varepsilon_m - \eta_j F_j(I) B_k \right) + \lambda \sum_{k=1}^{K} S\left(C_{m,k} \right) \right\}, \end{aligned} \tag{28}$$

where η_j is local approximation for pixel p_0. It is the only distribution that has similar effects with perceptual model.

3. Enhanced Mean Shift Method

3.1. Mean Shift. In this section, we propose an enhanced mean shift method. Our method uses the mathematical recursion method [25]. The spatial tracking object will be represented by a spatial histogram consisting of a weighted evaluation, in which the probability estimation function $p(y)$ and $q(x_0)$ are used to represent the potential motion probability in images $I(y)$ and $I(x_0)$. The histogram variables can be expressed as follows:

$$\begin{aligned} p(y) &= \frac{c}{|\Sigma|^2} \sum_{i} k\left(y_i \sum y_i^T \right) \delta\left[b_u(I(y_i)) - u \right] \\ &\quad + \frac{c}{2|\Sigma|^{1/2}} \sum w_{j,k} \left(y_j^T \sum^{-1} y_j \right), \\ q(x_0) &= \frac{c}{|\Sigma|^2} \sum_{j} k\left(x_j \sum x_j^T \right) \delta\left[b_u\left(I\left(x_j \right) \right) - u \right] \\ &\quad + \frac{c}{2|\Sigma|^{1/2}} \sum w_{j,k} \left(y_j^T \sum^{-1} y_j \right), \end{aligned} \tag{29}$$

where $y_i = (y_{Ti} - y)$, $x_j = (x_{Tj} - x)$, and Σ are the representatives of weight function, $b_u(I(y_i))$ and $b_u(I(y_j))$ are motion estimations for positions y_i and y_j, w_j is the box parameter, C is normalized constant, $u = 1, \ldots, m$, m contains all the binary pixels, and y and x are the center of related kernel function. The Bhattacharyya coefficient ρ will be used to detect similarity between tracking object area and the potential background.

Consider the following:

$$\begin{aligned} &\rho(p, q) \\ &= \frac{\sum_u \sqrt{p_y(x, y, \Sigma) q_x}}{\left\| L_x(i) \right\| \times \left\| L_y(j) \right\|} \min\left(\frac{\left\| L_y(j) \right\|}{\left\| L_x(i) \right\|}, \frac{\left\| L_x(i) \right\|}{\left\| L_y(j) \right\|} \right). \end{aligned} \tag{30}$$

We apply the first-order Taylor sequence extension, in which (x, y) are the coordinates of the center position in the previous frame, and then we obtain the following extended formulas:

$$\begin{aligned} \rho &= \sum_u \frac{1}{2} \sqrt{p_y q_x(x, y, \Sigma)} + \frac{c}{2|\Sigma|^2} \sum_j \omega_j k\left(y_j^T \sum x_i^T \right), \\ \omega_j &= \sum_u \sqrt{\frac{p_y}{q_x} \delta\left[b_u\left(I\left(y_j \right) \right) - u \right]}. \end{aligned} \tag{31}$$

The center of kernel function can be determined by the estimate of $\rho(x, y)$

$$\delta = \alpha_1 \sum_{i=1}^{n} \rho\left(x_p^i, x^i \right) + \alpha_2 \sum_{i=2}^{n} \left| \rho\left(x_p^i, x_p^l \right) - \rho\left(x^i, x^l \right) \right|. \tag{32}$$

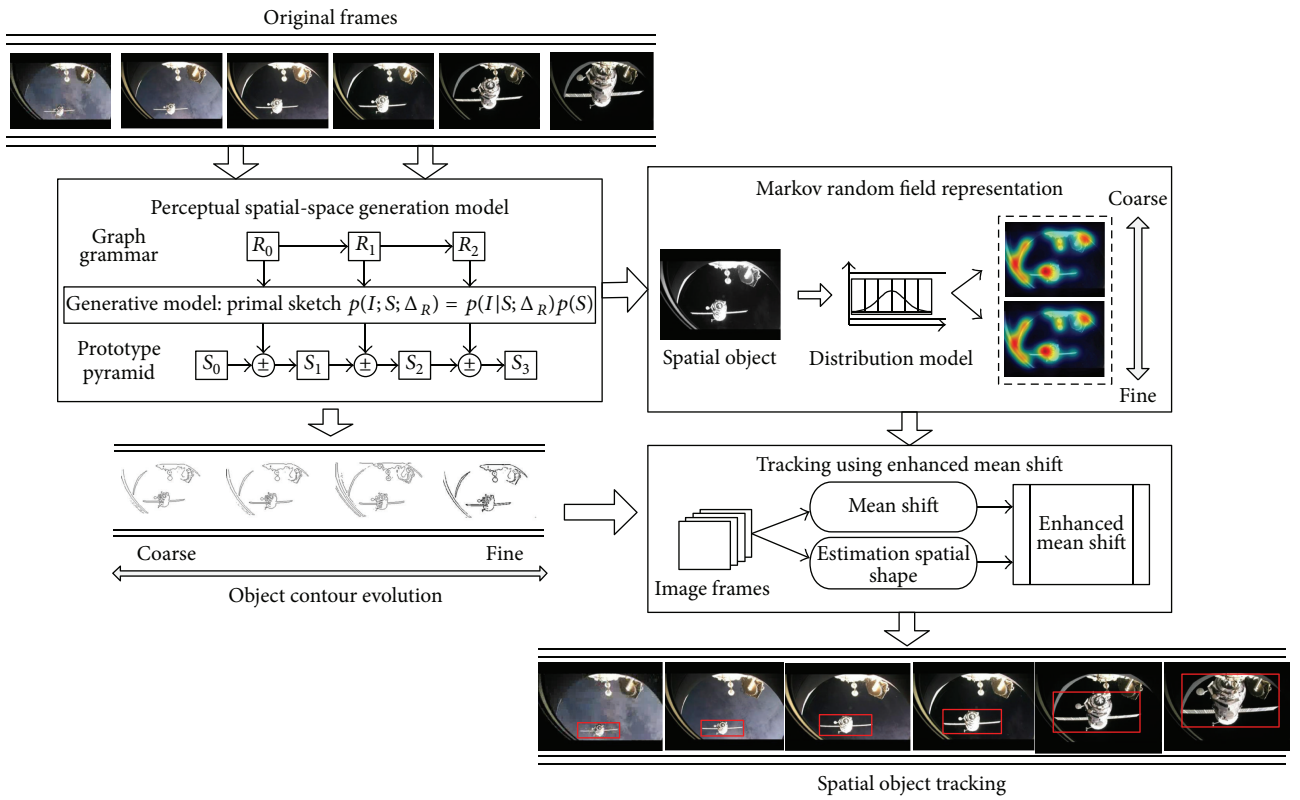


FIGURE 1: Enhanced mean shift method based on a perceptual spatial-space generation model.

In order to estimate the kernel function, the normalized bandwidth will be applied to a similarity judgment. The normalized bandwidth can be obtained by estimating $|\Sigma|\rho(x, y)$

$$\sum \rho(x, y) = \frac{2}{1-r} \frac{\sum_{j=1}^{n} \omega_j x_j \left(y_j^T \sum^{-1} x_i^T\right) g}{\sum_{j=1}^{n} \omega_j k \left(y_j^T \sum^{-1} x_i^T\right)} + \frac{1}{N_{\text{pre}}} \sum_{i=1}^{N_{\text{pre}}} \max_j \left\langle D_{\text{pre}}^i, D_{\text{cur}}^j \right\rangle, \tag{33}$$

where $y_j^T = (y_j - y)$ could accurately determine $\rho(x, y)$. Equations (32) and (33) will be calculated in an alternative iteration until the estimated parameters can cover all the variables.

3.2. Estimation of Spatial Shape Parameters. We also use the iterated function to determine boundary parameter $V_T^{(2)}$, which contains five fully adjustable affine box parameters. These parameters are the width, height, length, orientation, and center location. The orientation θ will be defined as the angle between the horizontal and width matrix. The box with width w and height h is the ellipse area between the long and short coordinate system. The relationship between θ, h, w and the bandwidth matrix Σ can be expressed as follows:

$$\Sigma = R^T(\theta) \begin{bmatrix} \left(\frac{h}{2}\right)^2 & 0 \\ 0 & \left(\frac{\omega}{2}\right)^2 \end{bmatrix} R(\theta), \quad \text{where } R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{34}$$

These parameters can be calculated using the octave decomposition method [26]; the specific formula is as follows:

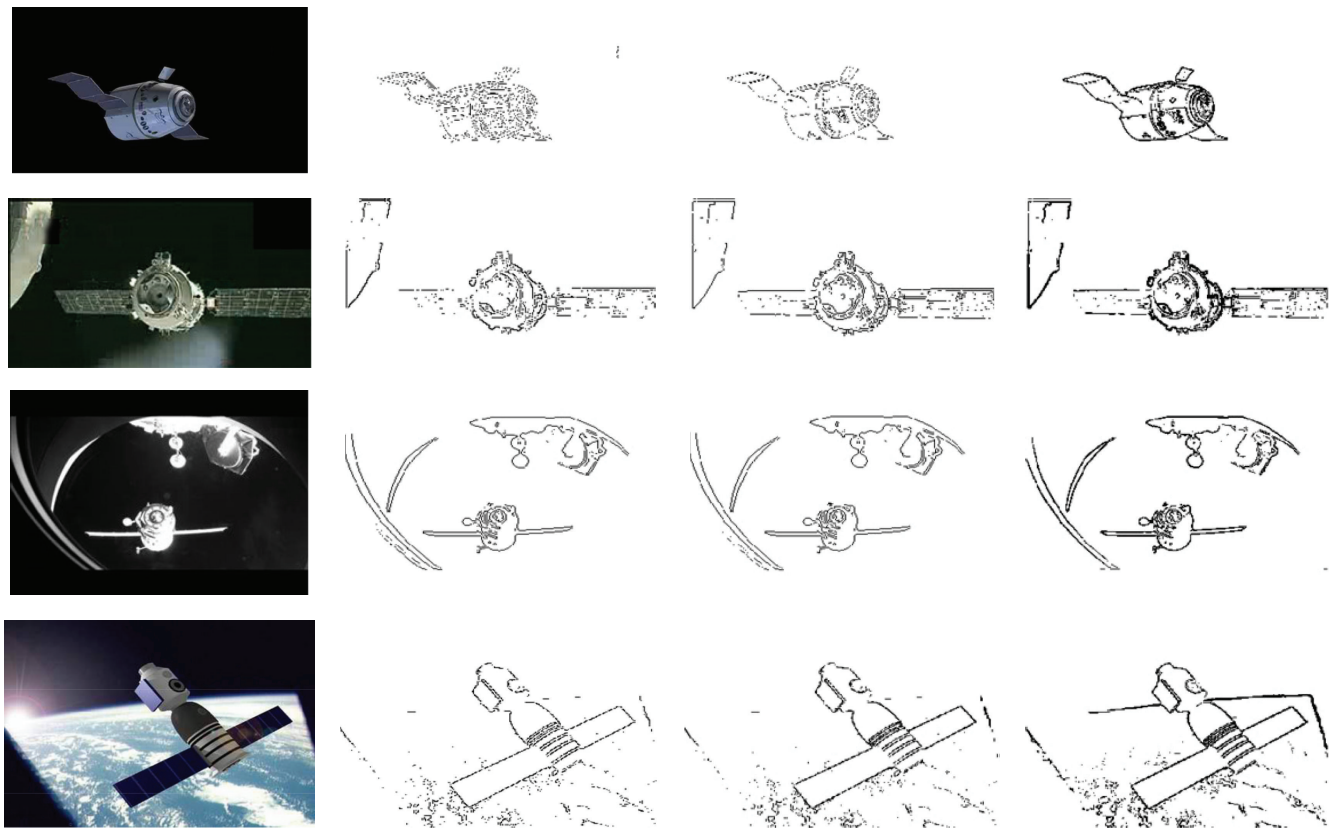
$$P\left(f_t^{\text{pre}} \mid f_t^{\text{cur}}\right) = \begin{cases} N\left(f_t^{\text{pre}} : f_t^{\text{cur}}, \text{diag}\left[\sigma_w^2, \sigma_h^2, \sigma_l^2, \sigma_o^2, \sigma_c^2\right]\right), & \text{if } f_t^{\text{cur}} \text{ is not null,} \\ P_{v1}, & \text{if } f_t^{\text{cur}} \text{ is null,} \end{cases} \tag{35}$$

where f_t^{pre} and f_t^{cur} are the octave decomposition components of the previous frame and the current frame.

3.3. Spatial Object Tracking Using Enhanced Mean Shift. In order to limit the possibility that background pixels appear in the tracking object, we use a relatively small elliptical area, in which the contract domain is determined by the factor K; in our experiments, $K = 0.7$. The elliptical area can be defined as follows:

$$EA = \sqrt{1 - \sum_{i=1}^{M} \sum_{u} p_u^{i,(j)}(y, \Sigma) q_u^{i,(j)}} + \frac{1}{2} \sum_{i=1}^{M} \frac{C_M}{|\Sigma_i|^{1/2}} \sum_{j=1}^{N} w_j^v k\left(y_j^T \sum_{v}^{-1} y_j\right). \tag{36}$$

In our proposed method, we should determine whether the previous frame motion area (Object A) will be used to guide the next frame object (Object B) tracking. If the number



(a) (b) (c) (d)

Figure 2: Contour evolution on spatial video sequences. (a) Original video frame and (b)–(d) contour evolution from a coarse to fine scale.

of continuous characteristic pixels and the Bhattacharyya coefficient for A are both higher, the initial rectangular tracking area for B will refer to A's settings, and the tracking area will be determined by the previous frame mean shift.

Consider the following:

$$V_t^{(2)} = \begin{cases} V_t^{(1)}, & \text{if } \text{dist}_t = \sum_{k=1}^{4} \left\| x_{t,k} - x_{t-1,k} \right\|^2 < T_1^{(2)}, \\ V_{t-1}^{(2)}, & \text{if } \text{dist}_t = \sum_{k=1}^{4} \left\| x_{t+1,k} - x_{t,k} \right\|^2 < T_2^{(2)}, \end{cases} \tag{37}$$

where $T_1^{(2)}$ and $T_2^{(2)}$ determine the threshold value. In order to solve drift and error propagation, we use enhanced mean shift for frame resampling. The object tracking resampling operation can be summarized as follows:

$$\rho \approx \sum_{i=1}^{M} \sum_{u} \frac{1}{2} \sqrt{p_u^i q_u^i \left\| x_{t,i}^{(2)} - x_{t-1,i}^{(2)} \right\|} + \sum_{i=1}^{M} \sum_{y_t \in R_t} \frac{R_{t-1}^{(\text{obj})}}{2|\Sigma_i|^{1/2}} \sqrt{w_t^i V_{t-1}^{(\text{obj})} \left(y_j^T \sum_{i}^{-1} y_j \right)},$$

$$\text{where } R_t \longleftarrow R_{t-1}^{(\text{obj})},\ V_t \longleftarrow V_{t-1}^{(\text{obj})},\ \rho_t \longleftarrow 0, \tag{38}$$

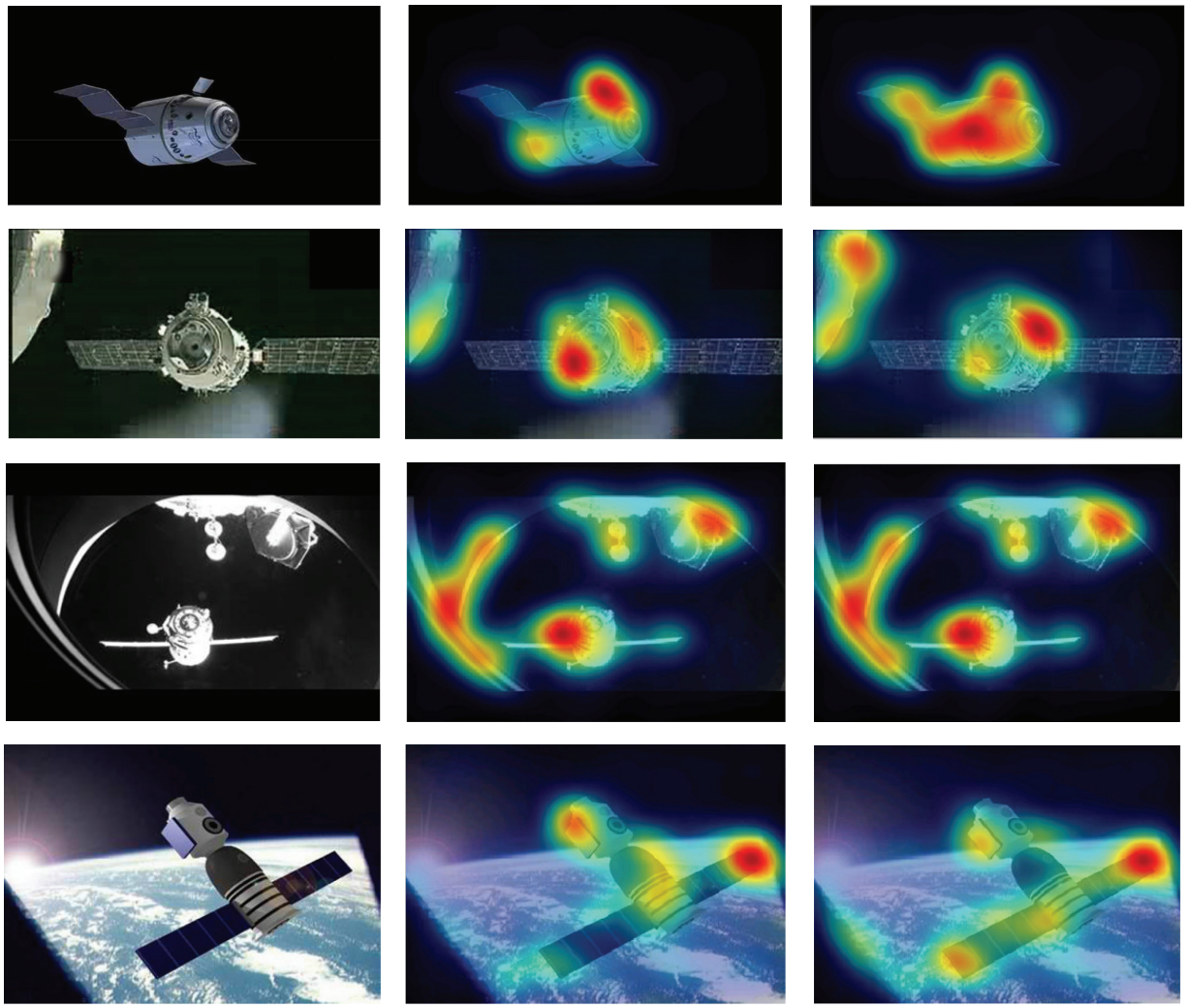
where $x_{t,I}^{(2)}$ and $x_{t-1,I}^{(2)}$ are the four-dimensional motion areas in frames t and $t-1$ and $T_3^{(2)}$ and $T_4^{(2)}$ are the threshold values determined by the graphic distance and the similarity shape.

4. Experimental Result

We conducted experiments on four different spatial video sequences, in which tracking objects are spatial satellites and aircrafts. We uniformed the sequence image size for the same spatial resolution; each frame is 320 × 256. The superiority of our proposed algorithm will be validated by an intuitive performance and objective evaluation.

4.1. Contour Evolution in Prototype Pyramid. Contour evolution experimental results are shown in Figure 2. Each prototype pyramid layer is calculated independently. The performances have shown that the contour evolution has a better continuity in the layer-by-layer prototype pyramid, and the approximate effects are derived from a perceptual spatial-space generation model and are closer to the human visual perception system.

4.2. Markov Random Field Representation. The Markov random field representation is shown in Figure 3. The significant representative region derivate from the perceptual spatial-space generation model contains significant feature information, which is closely related to the different color and



(a) (b) (c)

Figure 3: Markov random field representation in the perceptual spatial-space generation model. (a) Original frame and (b)–(c) initial and global smoothing representation.

texture distribution in the spatial area. The experimental results show that the region representation effect, which has been smoothed, could highlight the sensitivity of the motion area more than the initial representation.

4.3. Enhanced Mean Shift Object Tracking. We conducted experiments on the ten video sequences, in which the tracking objects include spatial satellite and aircraft, highway and park surveillance, and the human body. The enhanced mean shift conducted an iterative calculation 20 times on each video sequence. The normalized matrix bandwidth parameters are determined by a different experimental sample. In our experiments, it is 0.63 for video sequences 1, 2, and 3, 0.42 for video sequences 4 and 6, 0.56 for video sequences 5 and 8, 0.21 for video sequence 7, and 0.60 for video sequence 9 and 10.

4.3.1. Satellite-1, 2, and 3. Tracker-1: the particle filter will have a negative impact on the horizontal direction, and the motion estimation will appear noncontinuous. Tracker-2: the distance metric learning will affect motion area determination, and the rectangular window tracking will produce some deviation. Our proposed algorithm can better track the spatial objects, and satellites and aircraft can be completely contained in the rectangular window with a similar color distribution and background quiver. The edge deviation variance can be controlled well using spatial object detection, with no offset and blurring.

4.3.2. Automobile-1 and 2. In a nighttime environment, as the weak light and brightness, Tracker-1 and Tracker-2 obtain a vague tracking result, and the confusion area between background and the tracking object becomes larger. In the Automobile-2 sequence in particular, as the illumination from other automotive foreground lamps, Track-1 produces particularly serious deviation. Our proposed enhanced mean shift method could distinguish the tracking object from

Figure 4: Experimental results from the tracking, marked by the bounding box. The red (solid line) box is from our proposed method, the blue (dashed line) box is Tracker-1 from the particle filters in [27], and the green (dashed line) box is Tracker-2 from the online distance metric learning in [28]. The rows 1–10 contain images from the videos "Satellite-1," "Satellite-2," "Satellite-3," "Automobile-1," "Running," "Automobile-2," "Highway," "Walking," "Automobile-3" and "Walking-2," respectively.

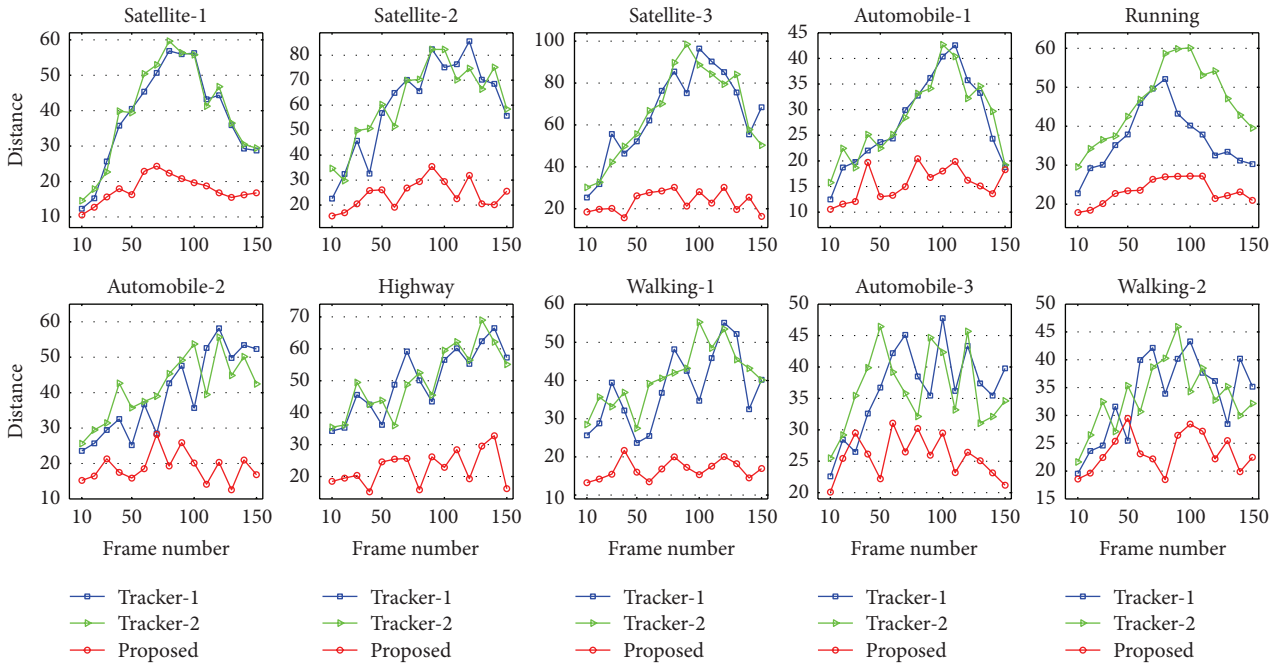


Figure 5: Euclidian distances between the tracked and artificially marked areas. Red curve: our proposed method; blue curve: *Tracker-1* in [27]; green curve: *Tracker-2* in [28].

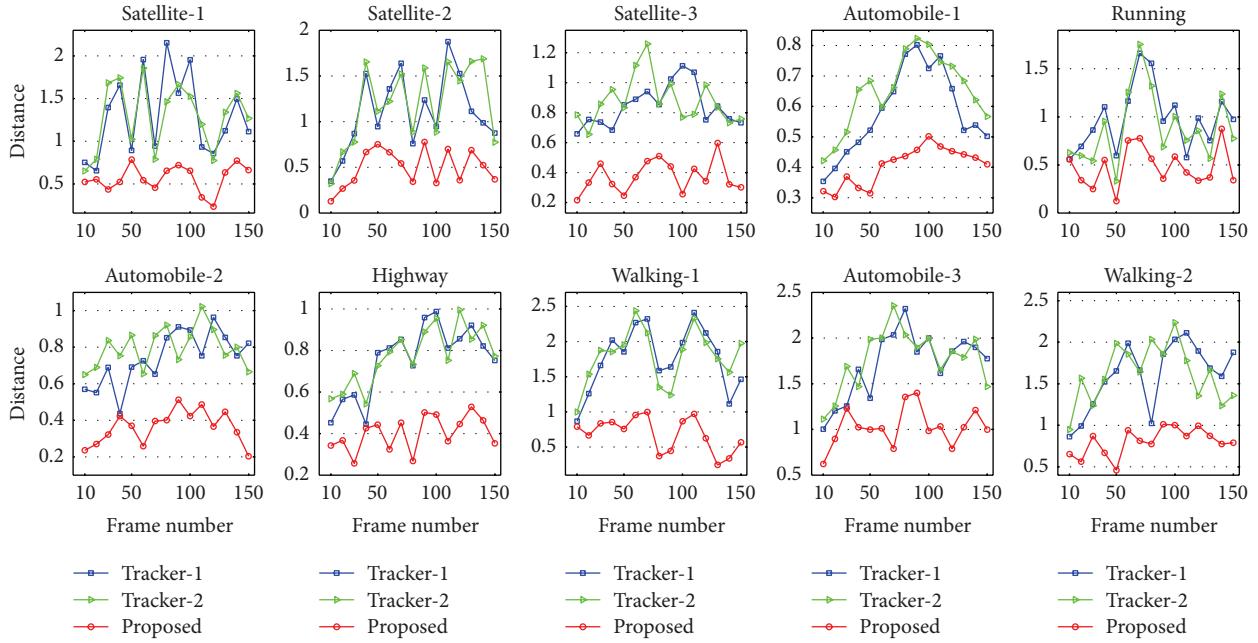


Figure 6: Bhattacharyya distances between the tracking method and the artificial markers. Red curve: our proposed method; blue curve: *Tracker-1* in [27]; green curve: *Tracker-2* in [28].

background confusion, and the tracking results do not appear to obviously deviate.

4.3.3. Highway. Tracker-1: this method will lose the tracking center in some frames. There are some cross-rectangular windows between the far and close vision sequences. Shape errors exist in the rectangle window estimations. Tracker-2: the tracking results also have a tracking deviation in the far vision. Our proposed method can maintain consistency between different visions and does not appear to have huge deviations.

4.3.4. Running, Walking. There are no obvious differences between the different methods. Except for some slow motion (walking body), Tracker-2 shows some partial deviations.

Table 1: Results of the averaged Euclidian distance over all frames in each video.

Video	Tracker-1	Tracker-2	Proposed
Satellite-1	38.41	39.60	17.82
Satellite-2	60.97	60.44	23.45
Satellite-3	65.41	56.37	23.32
Automobile-1	27.64	28.22	14.23
Running	24.53	25.76	16.44
Automobile-2	52.67	59.21	32.87
Highway	22.16	21.06	10.54
Walking	25.29	23.19	12.73
Automobile-3	36.53	36.47	25.71
Walking-2	33.46	33.43	23.42

4.3.5. Automobile-3, Walking-2. These cases are used to further test the robustness of our proposed method in the scene containing two or more moving objects. We could observe that Tracker-2 results in less accurate boxes probably due to its poor estimate of different moving object center position in the same scene. The performance of Tracker-1 is somewhat better; however, it also produces partial deviations. Our proposed method is more robust when dealing with scenarios of two or more moving objects.

4.4. Objective Evaluation. We use three objective evaluations for evaluating the object tracking performance (Figure 4).

4.4.1. Euclidian Distance. The Euclidian distance is the distance between rectangular windows obtained by tracking methods and the artificially marked. The specific calculation is as follows:

$$D = \frac{1}{4}\sum_{i=1}^{4}\sqrt{\left(x_{A,i} - x_{GT,i}\right)^2 + \left(y_{A,i} - y_{GT,i}\right)^2}, \quad (39)$$

where $(x_{A,i}, y_{A,i})$, $(x_{GT,i}, y_{GT,i})$, and $i = 1, 2, 3, 4$ are corner coordinates of the rectangular window calculated by tracking methods and artificially marked. Figure 5 shows the Euclidian distance between the tracked and artificially marked areas for our proposed method and the two other trackers on the videos.

The averaged Euclidian distance calculated on all videos is shown in Table 1. Compared to the distance values from Tracker-1 and Tracker-2, our proposed method clearly shows smaller and bounded Euclidian distances for the tested videos.

4.4.2. Mean Square Errors (MSEs). We have

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(x_i^A - x_i^{GT}\right)^2 + \left(y_i^A - y_i^{GT}\right)^2}, \quad (40)$$

where (x_i^A, y_i^A), (x_i^{GT}, y_i^{GT}) is the center of the tracking area obtained by methods and artificially marked, respectively. N is the total number of video sequence frames. The experimental results are shown in Table 2 and, it can be seen that our proposed method has the minimum MSE, which means it has the lowest tracking deviation. Our method has obvious advantages compared to other methods.

Table 2: Results of averaged MSE errors for the tracked box from our proposed method, *Tracker-1*, and *Tracker-2*.

Video	Tracker-1	Tracker-2	Proposed
Satellite-1	2.7865	6.2853	1.5866
Satellite-2	5.2368	13.4572	3.2914
Satellite-3	15.7326	12.3574	5.1233
Automobile-1	6.2357	7.8211	2.5671
Running	8.4596	7.7324	1.9635
Automobile-2	14.3687	18.5964	4.2158
Highway	4.2365	3.2151	1.1623
Walking	7.5642	9.2534	3.1486
Automobile-3	9.7749	10.1037	4.3357
Walking-2	8.9681	9.6614	4.0654

4.4.3. Bhattacharyya Distance. The Bhattacharyya distance is used to judge the deviation degree between the tracking area and the actual motion area. The specific calculation method is as follows: mean_A and mean_{GT} are the mean vectors with respect to the tracking area and are calculated by our method and artificially marked. The variables cov_A and cov_{GT} are covariance matrices with respect to the tracking area and are calculated by our method and artificially marked.

Consider the following:

$$\mathrm{BD} = \frac{1}{8}\left(\mathrm{mean}_A - \mathrm{mean}_{GT}\right)^T\left[\frac{\mathrm{cov}_A - \mathrm{cov}_{GT}}{2}\right]^{-1} \times \left(\mathrm{mean}_A - \mathrm{mean}_{GT}\right) + \frac{1}{2}\ln\frac{\left|\left(\mathrm{cov}_A + \mathrm{cov}_{GT}\right)/2\right|}{\left|\mathrm{cov}_A\right|^{1/2}\left|\mathrm{cov}_{GT}\right|^{1/2}}. \quad (41)$$

The Bhattacharyya distance between the tracked object area and the artificial marked region is shown in Figure 6. Among nine case studies, our proposed method has shown a marked improvement on the tracking accuracy as compared with the two existing trackers (Tracker-1 and Tracker-2). It is mainly due to our combination of perceptual spatial-space generation model and the enhanced mean shift. The averaged Bhattacharyya distance on different video is shown in Table 3. Our proposed method has the smallest average tracking deviation between different methods.

5. Conclusions

In this paper, we propose a new intelligent modeling method using the enhanced mean shift method based on a perceptual spatial-space model for spatial object tracking. The perceptual spatial-space model can obtain a continuous spatial object contour and highlight tracking object saliency. Enhanced mean shift method uses enhanced version mean shift, which focuses on the estimation of spatial shape parameters. The method could effectively cope with severe spatial

Table 3: Averaged Bhattacharyya distances for our proposed method, *Tracker-1*, and *Tracker-2*.

Video	Tracker-1	Tracker-2	Proposed
Satellite-1	1.2965	1.2891	0.5669
Satellite-2	1.1051	1.1902	0.4978
Satellite-3	1.2239	1.2536	0.5564
Automobile-1	0.8567	0.8944	0.4523
Running	0.9812	0.8845	0.4808
Automobile-2	1.8749	1.9861	0.7126
Highway	0.7746	0.7983	0.4533
Walking	1.7622	1.7906	0.6854
Automobile-3	1.7170	1.7699	1.0247
Walking-2	1.6012	1.6204	0.8040

interferences. The comparison between our method and other state-of-the-art methods demonstrates that our proposed method has a higher tracking accuracy and precision. In future research, we can incorporate more spatial object information, such as spatial textures and aircraft shapes, into our intelligent model to generate a more robust spatial object tracking method.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) (2012CB821206), the National Natural Science Foundation of China (no. 91024001 and no. 61070142), and the Beijing Natural Science Foundation (no. 4111002).

References

[1] W. D. Gray, "Simulated task environments: the role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research," *Cognitive Science Quarterly*, vol. 2, no. 2, pp. 205–227, 2002.

[2] D. Schröder, C. Hintz, and M. Rau, "Intelligent modeling, observation, and control for nonlinear systems," *IEEE/ASME Transactions on Mechatronics*, vol. 6, no. 2, pp. 122–131, 2001.

[3] S. C. Hsia, C. H. Hsiao, and C. Y. Huang, "Single-object-based segmentation and coding technique for video surveillance system," *Journal of Electronic Imaging*, vol. 18, no. 3, Article ID 033007, 10 pages, 2009.

[4] J. Lee, Y. Chee, and I. Kim, "Personal identification based on vector cardiogram derived from limb leads electrocardiogram," *Journal of Applied Mathematics*, vol. 2012, Article ID 904905, 12 pages, 2012.

[5] W. B. Wan, T. Fang, and S. G. Li, "Vehicle detection algorithm based on light pairing and tracking at nighttime," *Journal of Electronic Imaging*, vol. 20, no. 4, Article ID 043008, 11 pages, 2011.

[6] J. Xiao, H. Cheng, H. Sawhney, and F. Han, "Vehicle detection and tracking in wide field-of-view aerial video," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 679–684, June 2010.

[7] J. Kerekes, M. Muldowney, K. Strackerjan, L. Smith, and B. Leahy, "Vehicle tracking with multi-temporal hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, vol. 6233 of *Proceedings of SPIE*, April 2006.

[8] P. Yu and D. Z. Pan, "ELIAS: an accurate and extensible lithography aerial image simulator with improved numerical algorithms," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 2, pp. 276–279, 2009.

[9] D. G. Sim, R. H. Park, R. C. Kim, S. U. Lee, and I. C. Kim, "Integrated position estimation using aerial image sequences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 1–18, 2002.

[10] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quarterly of Applied Mathematics*, vol. 59, no. 1, pp. 85–111, 2001.

[11] A. P. Witkin, "Scale space filtering," in *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pp. 1019–1022, Kaufman, Karlsruhe, Germany, 1983.

[12] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, no. 5, pp. 363–370, 1984.

[13] P. J. Burt, "Attention mechanisms for vision in a dynamic world," in *Proceedings of the International Conference on Pattern Recognition*, vol. 1, pp. 977–987, The Hague, The Netherlands, 1988.

[14] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention," *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.

[15] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, article 13, 2006.

[16] M. G. Ramos and S. S. Hemami, "Eigenfeatures coding of video-conferencing sequences," in *Proceedings of the Visual Communications and Image Processing*, vol. 2727 of *Proceedings of SPIE*, pp. 100–110, Orlando, Fla, USA, March 1996.

[17] M. Isard and A. Blake, "CONDENSATION—conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[18] L. Sun and G. Liu, "Visual object tracking based on combination of local description and global representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 4, pp. 408–420, 2011.

[19] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 605–612, Madison, Wis, USA, June 2003.

[20] B. Chen and J. C. Principe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Processing Letters*, vol. 19, no. 8, pp. 491–494, 2012.

[21] Y. Wang and S. C. Zhu, "Modeling complex motion by tracking and editing hidden Markov graphs," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. I856–I863, July 2004.

[22] X. Xie, H. Liu, W. Y. Ma, and H. J. Zhang, "Browsing large pictures under limited display sizes," *IEEE Transactions on Multimedia*, vol. 8, no. 4, pp. 707–714, 2006.

[23] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[24] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[25] S. Qi and X. Huang, "Hand tracking and gesture gecogniton by anisotropic kernel mean shift," in *Proceedings of IEEE International Conference Neural Networks and Signal Processing (ICNNSP '08)*, pp. 581–585, June 2008.

[26] H. W. Lee, K. C. Hung, B. D. Liu, S. F. Lei, and H. W. Ting, "Realization of high octave decomposition for breast cancer feature extraction on ultrasound images," *IEEE Transactions on Circuits and Systems I*, vol. 58, no. 6, pp. 1287–1299, 2011.

[27] Z. H. Khan, I. Y. H. Gu, and A. G. Backhouse, "Robust visual object tracking using multi-mode anisotropic mean shift and particle filters," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 1, pp. 74–87, 2011.

[28] G. Tsagkatakis and A. Savakis, "Online distance metric learning for object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 12, pp. 1810–1821, 2011.