Traffic Video Image Segmentation Model Based on Bayesian and Spatio-Temporal Markov Random Field

Jun Zhou1,a, Xu Bao1,b, Dawei Li2,c and Yongwen Yin3
1Key Laboratory for Traffic and Transportation Security of Jiangsu Province, Huaiyin Institute of Technology, Huai’an, China
2Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic School of Transportation, Southeast University, Nanjing, China
3Transportation Bureau, Huai’an, China
Email: a joujou1980@163.com, b 112026245@qq.com, c 328650320@qq.com

Abstract. Traffic video image is a kind of dynamic image and its background and foreground is changed at any time, which results in the occlusion. In this case, using the general method is more difficult to get accurate image segmentation. A segmentation algorithm based on Bayesian and Spatio-Temporal Markov Random Field is put forward, which respectively build the energy function model of observation field and label field to motion sequence image with Markov property, then according to Bayesian’ rule, use the interaction of label field and observation field, that is the relationship of label field’s prior probability and observation field’s likelihood probability, get the maximum posterior probability of label field ‘s estimation parameter, use the ICM model to extract the motion object, consequently the process of segmentation is finished. Finally, the segmentation methods of ST-MRF and the Bayesian combined with ST-MRF were analyzed. Experimental results: the segmentation time in Bayesian combined with ST-MRF algorithm is shorter than in ST-MRF, and the computing workload is small, especially in the heavy traffic dynamic scenes the method also can achieve better segmentation effect.

1. Introduction
Segmentation of moving objects in complex scenes is one of the hotspots in computer vision research in recent years. Markov random field model (MRF) defined by Gibbs distribution is widely used in image segmentation. Such as: Pappas[1] used MRF model to segment the still image, Liu[2] used the method of MRF model combined with fuzzy clustering for color image segmentation. Andrey [3] used MRF for segmentation on video texture image, although these studies are very successful, and they all have adopted the model of 2D-MRF in the space images (such as still images). But for traffic videos, this kind of dynamic image sequences, segmentation effect of above-mentioned method is not as good as still images. This paper will be based on the 2D-MRF model, and extend the 2D-MRF to the ST-MRF model. The model not only can optimize the image distribution on the space but also optimize the image on the time axis. On the time axis, the image sequence of successive frames is associated on each pixel. ST-MRF takes this association into consideration, which will benefit the segmentation of dynamic images. For the lack of research on dynamic images of above-mentioned algorithm, ST-MRF combines still information and dynamic information, and this algorithm is less sensitive to light, so there is no need for images to denoise to detect moving targets.

Recently segmentation technology of the moving object in complex scene is one of the hot spots of computer vision.
2. ST-MRF Model

2.1 ST-MRF Theoretical Knowledge

Aimed at a series of points $S = \{s_1, s_2, \ldots, s_i\}$ on a set of 2D images, for these points, there is a corresponding prior observation field $D = \{d_1, d_2, \ldots, d_i\}$ (also called time field) and a corresponding label field $L = \{l_1, l_2, \ldots, l_i\}$, also known as the space field. Among them, the prior observation field $D$ of the image, which is the frame difference image, recording the relationship between pixels and its frame can be obtained according to the subtraction between two continuous frames of of continuous time:

$$D(t + 1) = |S(t + 1) - S(t)|$$

The label field records the relationship between each pixel and background as well as the neighborhood. In order to simplify the model, two mutually adjacent groups are only considered here. For the $S$ on the two-dimensional images, there are 8 adjacent points on the spatial domain and 18 adjacent points on the time domain. The spatio-temporal neighborhood system on the image $S$ can be described by the Gibbs random field with the ST-MRF distribution. In this paper, we assume that each frame of gray image on the video image is a ST-MRF. At this time, random variable $x_i$ has become the label value at $s_i$ of the pixel. (As the gray value), the definition of the energy function is: the sum of the potential of all groups in a neighborhood system of a pixel point $i$ in a certain frame image. The energy function consists of two parts, the formula is:

$$U(x) = U_{space}(x) + \alpha U_{time}(x)$$

Among them: $U_{space}(x)$ is the energy function of label field (spatial field). It records the relationship information between each pixel gray value and the background as well as the neighborhood on a frame image; $U_{time}(x)$ is the energy function of observation field (time field). It records the relationship information between the gray value of each pixel in the current frame and intercepted frame images among continuous frames. $\alpha$ is the weight coefficient for the two parts. In the image segmentation, single pixel cannot be used as a scene in ST-MRF (site), a larger group of pixel is needed. Therefore, we need to divide a 720 * 640 image into 90 * 80 blocks, as each block consists of 8 x 8 pixels; blocks between adjacent images are linked through their vector.

2.2 The Construction of the Energy function of ST-MRF

$$U(y) = \sum_k U(y_k)$$

$$U(y_k) = U_N(N_{yk}) + U_{pre}(D_{syk}, M_{syk})$$

$$= U_N(N_{yk}) + U_D(D_{syk}) + U_M(M_{syk})$$

$$= a(N_{yk} - \mu)^2 + b(M_{syk} - \mu)^2 + cD^2_{syk}$$

Among them, $G(t-1) = g$, $G(T) = h$: the value of image $G$ at the time of $t-1$ is $g$, at the time of $t$, the value is $h$. For each pixel, it can be expressed as: $G(t-1; i, j) = g(i, j)$, $G(t; i, j) = h(i, j)$;

$X(t-1) = x$, $X(t) = y$: the target map $X$ was detected that its label distribution is $X$ at the time of $t-1$, at the time of $t$, the label distribution was detected is $y$. For each block, it can be expressed as: $X_k(t-1) = x_k$, $X_k(t) = y_k$, $K$ for block number;

$U$: energy function;

$N_{yk}$: The number of the block with the same number in adjacent blocks of a block, if you consider
using the 8-neighborhood groups, then \( \mu N_{\mu} = 8 \) is the maximum value;

\[ D_{\mu} \]: Texture correlation between representative image \( G(t-1) \) and \( G(T) \);

\[ M_{\mu} \]: In the current target map \( X(T) = y \), the effective value of the previous map \( X(t-1) = x \), that is, the number of pixels in the occlusion part in the partial occlusion two blocks, obviously \( \mu M_{\mu} = 64 \) is the maximum value;

\[ U(y) \]: The energy functions of ST-MRF.

### 3. The Segmentation Method of the Bayesian Combined with ST-MRF

#### 3.1 The Bayesian Rules

Assuming that the random fields described by the image data is \( Y = y \), \( Y \) represents a random field known as the observation field and time field, \( y \) represents a reality of the random fields; Assuming that the random fields described by the image data is \( X = x \), \( X \) represents a random field known as label field or space field. \( X \) is a reality of the random fields that is a result of segmentation.

According to Bias theorem, the image segmentation problem is expressed as:

\[
P(X = x | Y = y) = \frac{P(Y = y | X = x)P(X = x)}{P(Y = y)}
\]  \( (5) \)

Among them, \( P(X = x | Y = y) \) is the posterior probability of the label field under the condition of given observation data \( Y = y \); \( P(Y = y | X = x) \) indicates the joint distribution of the observed field \( Y = y \) under the condition of a realistic condition of the given label field \( X = x \). \( P(Y = y) \) is the joint distribution of the observed field \( Y = y \).

According to the formula \( (5) \) for image segmentation directly, the calculation is very complex and it is not the correct result. In order to carry out the MAP (A Posterior Maximum) image segmentation under the Bayesian framework, two hypotheses are needed.

Hypothesis 1: Assume that each component of the observation field \( Y = y \) is independent under a realistic condition \( X = x \) in a given label field.

If the label field has a total of \( K \) markers, that is, each position of the label field has a total of \( K \) species, that is, all the observed field data will be divided into \( K \) species, which can be expressed as:

\[
P(X = x | Y = y) = \prod_{k=1}^{K} \frac{P(Y = y | X = x_k)P(X = x_k)}{P(Y = y)}
\]  \( (6) \)

Among them, \( P(Y = y | X = x_k) \) represents the probability distribution of the \( K \) species composition \( y^k \) of the observation field under the condition of \( X = x_k \).

Hypothesis 2: Assume that the same type of pixels in the image is subject to the same distribution, such as the Gauss distribution.

Under the assumption of 1 and 2, if the distribution function is expressed as \( P(y^k | X_s = k) \), then the image segmentation problem is expressed as:

\[
\hat{x}_s = \arg \max_{x_s} \{ P(y^k | X_s = k)P(X_s = k | X_N) \}
\]  \( (7) \)

Among them, \( P(X_s = k | X_N) \) represents the local probability of label field, \( N_s \) is the neighborhood location set for location \( s \). If the same type of distribution of pixels’ observation field is represented by the Gauss distribution, is expressed as:

\[
P(y^k | X_s = k) = \frac{1}{\sqrt{2\pi}\sigma^2_k} \exp\left[-\frac{(y^k - \mu_k)^2}{2\sigma^2_k}\right]
\]  \( (8) \)
3.2 Algorithm Estimation of Bias Combined with ST-MRF Model

For the target map \( X \), it is detected that label distribution is \( X \) at the time of \( t-1 \), it is detected that the label distribution is \( y \) at the time of \( t \), image \( G \) has the value of \( g \) at the time of \( t-1 \) and has the value of \( h \) at the time of \( t \). A posterior probability model can be established, and maximize it. A maximum possible realization of the label field can be obtained. The posterior probability can be expressed as Bias:

\[
P(X(t) = y) = \frac{P(G(t-1) = g, X(t-1) = x, G(t) = h | X(t) = y) P(X(t) = y)}{P(G(t-1) = g, X(t-1) = x, G(t) = h)}
\]

In the formula, \( P(G(t-1) = g, X(t-1) = x, G(t) = h) \) is constant, so the maximum value of the posterior probability is determined by the fractions of the formula (9).

\[
P(X(t) = y) = \prod_{k} \exp[-U_{P}(N_{\ast k})]/Z_{NN} = \prod_{k} \exp[-\frac{1}{2\sigma_{y}^{2}}(N_{\ast k} - \mu_{N})^{2}]/Z_{NN}
\]

In the formula, \( N_{\ast k} \) is the adjacent block of the block \( C_{k} \), the blocks all have the same vehicle label.

When \( N_{\ast} \) is 8, the value of the energy function \( U_{P}(N_{\ast k}) \) is the minimum, when \( N_{\ast} = 0 \), the value of the energy function \( U_{P}(N_{\ast k}) \) is the maximum. So the maximum value of the posterior probability is:

\[
P(G(t-1) = g, X(t-1) = x, G(t) = h | X(t) = y) = \prod_{k} \exp[-U_{P}(M_{y | k})]/Z_{DMK}
\]

\[
\prod_{k} \exp[-U_{M}(M_{y | k})]/Z_{MK} \cdot \prod_{k} \exp[-U_{P}(D_{y | k})]/Z_{OK}
\]

The image segmentation based on maximum a posteriori probability (MAP) criterion is to find the mark set \( x \), which makes the maximum of the posterior probability distribution about \( x \). Considering the problem of computational efficiency, we can use conditional iterative method (ICM). ICM algorithm is an iterative algorithm; maximizing the conditional probability point by point through the realization of the pixel value is updated.

3.3 The Realization of Maximum a Posteriori Probability

3.1.1 ICM iterative algorithm. ICM (condition model iterative) \(^{[6]}\) is a deterministic algorithm based on local conditional probability which was proposed by Besag in 1986, it complete the image segmentation through updating the image mask point by point.

Assuming that each pixel of the image data \( y = \{ y_{1}, y_{2}, \ldots, y_{n} \} \) is independent on the condition of giving segmentation result \( x \), and the conditional distribution of \( y_{i} \) on \( x \) depends only on its mark \( x_{i} \):
Therefore, the conditional distribution of $y$ on $x$ represents:

\[ f(y | x) = \prod_{i=1}^{n} f(y_i | x_i) \]  

(13)

Given the image data and the mark of the neighborhood of $i$, according to the Bayesian formula:

\[ P(x_i | y, x_{\sim i}) = f(y_i | x_i) P(x_i | x_{\sim i}) \]  

(14)

Maximize the formula (14); we can get the classification mark of the pixel:

\[ \hat{x}_i = \arg \max_{y_i} P(x_i | y, x_{\sim i}) (i = 1, 2, \ldots, n) \]  

(15)

3.1.2 The basic steps of ICM algorithm. If the Gauss mixed model is used to model the observation field data and Potts model is used to model the label field data, and the fixed potential function $\beta$ is used, so the basic steps of the ICM algorithm can be expressed as:

Step1 set categories $K$, the potential function $B(\beta)$ of the image;
Step2 uses the K-mean value algorithm to calculate the initial segmentation results;
Step3 estimation of field parameters $\mu$ and $\sigma^2$;
Step4 calculated the observed field energy;
Step5 calculate the label field energy;
Step6 according to the principle of minimum energy, estimate the new segmentation results;
Step7 according to current segmentation results of parameters and the iteration of the image, according to the formula (15), calculate the maximum possible category of each point;
Step8 determine the convergence; determine whether the $|\Delta U|$ is less than the set $t$. If the satisfaction is out; otherwise return to Step3 to the next iteration.

Figure 1 The Segmentation Interface of ICM
3.1.3 Segmentation Result. In Figure 1, the left image is a video image, and the right image is the segmentation result. In the segmentation process, set the image classification number \( K=5 \), classification number is the number of initialization classification, that is, the foreground category which is relatively easy to separate from background, the more the category, the segmentation is more detailed. Potential function or \( \beta = 0.9 \) and \( t \) which is threshold for the convergence of the energy function, that is the difference among two adjacent iterations of energy levels \( |\Delta U| \), its value is set to 0.05, the number of iterations of the algorithm is 612, the value of the minimization of the energy function is 978773, consuming time is 77882.4ms. Fig. 2 is the change curve of energy function.

4. Experimental results and analysis
The two Figure shown in the 3 picture ,their pixel size are 627 x 461 and 643 x 466 respectively. The segmentation results are shown in figure 4. We can conclude that the segmentation effect in ST-MRF methods and ST-MRF combined with Bias is almost on the static image. But for the segmentation data, we can conclude that the segmentation time in Bayesian combined with ST-MRF algorithm is shorter than in ST-MRF, and the computing workload is small, especially in the heavy traffic dynamic scenes the method also can achieve better segmentation effect.

(a) The 263rd frame  (b) the 287rd frame

Figure 3 The image to be segmented
The above-mentioned simulation results show that the segmentation method based on Bayesian combined with ST-MRF can better segment the moving objects from the complex background. But there are still many issues that need to continue to study. First of all, the most important problem is that the computing time is too long, because the parameter estimation of the model is complex, leading to the parameter estimation need to spend more time computing. The results will affect the entire image segmentation time. In order to improve the speed of operation, according to the experimental experience, join the artificial selection in the ICM iterative algorithm, that is, adjust it according to the characteristics of the image, it is conducive to the segmentation of dynamic video image.

5. Conclusion
Reliable image segmentation is an important prerequisite to achieve automatic detection of the video image, and it is significant for the development of automatic incident detection and traffic flow monitoring system. This paper is aimed at the problem of occlusion in image segmentation, establish the image segmentation algorithm of the Bayesian combined with ST-MRF model, The algorithm can not only optimize the spatial image distribution but also optimize the image distribution on the time axis. It can solve the problem of occlusion between targets effectively. The research results of this dissertation are: 1) establish energy function aimed at video image with Markov; 2) establish image segmentation model based on Bayesian and spatio temporal Markov random field; 3) Use the iterative conditional model (ICM) algorithm to achieve the maximum a posteriori (MAP) estimation problem to achieve the extraction of moving objects; 4) ST-MRF segmentation method and Bayesian combined with ST-MRF segmentation method are compared and analyzed.

This work was supported in part by National Natural Science Foundation of China( no. 51608115), Natural Science Foundation of the Jiangsu Higher Education Institutions of China( no.14KJB580002), Six Talent Peaks Project in Jiangsu Province ( no. 2015-XXRJ-017). Natural Science Foundation of Jiangsu Province(no.BK20150613), Jiangsu province natural science in colleges and universities research major projects, Projects of International Cooperation and Exchange of the National Natural Science Foundation of China (No. 51561135003), Philosophy and Social Science of Universities in Jiangsu(Grant no. 2016SJB790057), Department of housing and urban rural development(Grant no. 2015-R2-063), Social Science Foundation of Jiangsu
Province (Grant no. 14EYD010), Social Development of Huai’an(Grant no.HASZ201635).

6. References
[1]. XIANG R H, WANG R S. A Range Image Segmentation Algorithm Based on Gaussian Mixture Model[J]. Journal of Software, 2003,14(7):1250-1257.
[2]. Cressie N, Verzelen N. Conditional-Mean Least-Squares Fitting of Gaussian Markov Random Fields to Gaussian Fields[J]. Computational Statistics & Data Analysis, 2008, 52(5):2794-2807.
[3]. Andrey P. Tarroux P. Unsupervised Segmentation of Markov Random Field Modeled Textured Images Using Selectionist Relaxation[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1998, 20(3):252-262.
[4]. LIU G Y, MA G R, WANG L G. Image Modeling and Segmentation in the Wavelet Domain Based on Markov Random Field[M]. BEIJING: SCIENCE PRESS,2010.
[5]. LI X C, ZHU S A. A Survey of the Markov Random Field Method for Image Segmentation[J]. Journal of Image and Graphics, 2007, 12(5):789-798.
[6]. CAO J Z, SONG A G. Research on the texture image segmentation method based on markov random field[J]. Chinese Journal of Scientific Instrument,2015,36(4):776-786.
[7]. HUANG X W, ZHU L, ZHONG X R. A Novel Moving Object Segmentation Technology Based on Spatiotemporal Markov Random Field[J]. Journal of Electronics & Information Technology, 2006, 28(2):367-371.