Foreground Extraction from Dynamic Videos Based on Computational Verb Theory

Li Bin and Yang Tao

***Abstract*—In this paper, a new method of Gaussian Mixture Model algorithm is proposed due to the inspiration of knowledge-based on computer vision and model recognition which are based on computational verb theory. This algorithm takes the binary image profiles and contour shapes to fulfill the foreground extraction from dynamic videos. Experiments show that, with respect to the performance in the dynamic videos, our algorithm is better than the algorithms used widely in other experiments. It is more accurate and can easily follow the track of the moving objects in the videos.**

***Index Terms*—Foreground extraction, GMM, connectivity analysis, computational verb theory.**

I. Introduction

Computational verb theory (CVT for short) is an emerging theory of Artificial Intelligence. Computational verb systems are new systematic frameworks of artificial intelligence by embedding the knowledge and experiences which are expressed or implied by verbs into machine intelligence. It turns out that verbs are very important to describe the dynamics of complexities for the survival of human beings. Verbs are also very important building blocks for transferring and representing human experiences. The applications of verbs to modeling complexity are discussed. The dynamics of linguistic expressions are viewed as the usage of verbs. Nonlinear and linear dynamic systems are used to model and classify verbs. In-depth study of characteristics of human thinking, CVT reduces the complexity of computation comparing to conventional methods based one intuitions and experiences of dynamics. Since its invention in 1997, CVT has been widely applied to many industrial fields. One important application of CVT is digital image processing. In [1], the author did an overview of CV image processing and introduced some applications in industrial fields to guide a new direction. In [2], the authors applied CVT to image compression, and the author of [3] applied the clustering algorithm to improve the image compression algorithms based on CVT [4].

Nowadays, video monitoring systems are widely used in many public places, but to most systems, it is required to keep observing the surveillance videos. So how to realize the auto monitor and control of the monitoring system is a hot topic recently. One kind of the technology of auto monitoring systems is monitoring the new objects in one scene, firstly it is required to detect whether there is change in the video image sequences thus new objects, then segment and extract the objects in the video image sequences, and prepare for the target recognition and track to extract data. Therefore, the function of a video monitoring system directly influences the result of the foreground extraction.

Manuscript received April 4, 2014; revised June 10, 2014.

The authors are with the Electronic Engineering, school of Information Science and Technology, Xiamen University, Fujian, 361001, China (e-mail: Lbbsmile@126.com).

Nowadays, video monitoring systems are widely used in many public places, but to most systems, it is required to keep observing the surveillance videos. So how to realize the auto monitor and control of the monitoring system is a hot topic. One kind of the technology of auto monitoring systems is monitoring the new objects in one scene, firstly it is required to detect whether there is change in the video image sequences thus new objects, then segment and extract the objects in the video image sequences, and prepare for the target recognition and track to extract data. Therefore, the function of a video monitoring system directly influences the foreground extraction.

Because the background in the video monitoring system is fixed while moving objects like people and cars entering the scene as foreground, we can extract the foreground. GMM (Gaussian Mixture Model) come up with by Chris Stauffer and others is used as statistic model of background to construct different Gaussian model of different states in the same scene[5], [6].

This algorithm has adaption compared to other algorithms. This paper uses the simplified GMM to model the video image sequences, and realize the effect of foreground extraction in the videos.

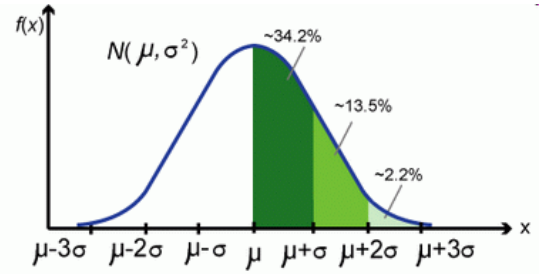

Fig. 1. The distribution curve of Gaussian distribution.

II. System Model

A. Gaussian Model Theory

Gaussian distribution is put forward by Gauss in 1809. For random variable x, the probability density is:

$$p(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad (\sigma>0) \qquad (1)$$

Parameter μ is the expectation of Gaussian distribution, and σ is the variance of Gaussian distribution. If a group of data suits Gaussian distribution, then most of the data will centralize in the section of [μ-2σ, μ+2σ]. The distribution

DOI: 10.7763/IJMO.2014.V4.402

curve of the function is showed in Fig. 1.

In our daily life, many events are accord with the Gaussian distribution. For the video image in a scene, if the background is still relatively, and there are no disturbing factors, then every pixel in the background in the arrangement of one time series can be described by one Gaussian distribution. But in practice, especially for the outdoor scene, because of the disturbs of all kinds of noisy, the variation of background is quite obvious, there will be the branch shaking, light illumination changing and so on. So the actual background could not be described only by one Gaussian distribution. In this case, we should take many Gaussian models to describe the dynamic background and different states. We define k as the number of Gaussian models, and the probability density of the pixel currently observed is:

$$p(x_t) = \sum_{i=1}^{k} w_{i,t} \times p(x_{i,t}, \mu_{i,t-1,k}, \sigma_{i,t-1,k}) \quad (2)$$

Parameter $w_{i,t}$ is the weight value of every single Gaussian model, and $\mu_{i,t-1,k}$ is the mean value o. the single Gaussian model of number i, $\sigma_{i,t-1,k}$ is the variance of the single Gaussian model of number i.

B. The Background Model and Update of GMM

The video images involve color components, the reference [7] adopts the method of covariance to calculate, but this algorithm takes too much calculated amount and has no obvious effect, it does not meet the demand of real-time. Therefore, we directly simplify the process above, and make the Gaussian model through converting color images to gray images. According the complexity of the scene, k can be evaluated within 3-7, thus the value of k greater, the scene is more comple x, and the calculation amount increases. In this algorithm, k equals 3.

There are three steps using GMM to make the Gaussian model: background training, template matching and background update.

1) Background training

Firstly, train a video, and then compute the mean value, variance and weight of the model in the training frame as the parameters in the background model. In the process of the training, it is not necessary to ensure every Gaussian model, to one pixel, if the variation is not obvious in the training time, thus within one or two models more than 90% of the pixels are included, then the greater and less variance can be valued to other models. In this algorithm the variance equals 30 and weight equals 0.005.

2) Template matching

After getting the background model, we can make the foreground extraction through template matching. Before template matching, we will rank three Gaussian models in the GMM, and judge which Gaussian model matches the background image. Because the variances of dots in the static region are less than those in the dynamic region, and the appearance of moving objects may lead in weight of the updated single Gaussian model reduces, hence the weight greater and the variance less, the degree of matching is higher. Match every single Gaussian model in the GMM after ranking the priority through the $e = w/\sigma$ formula.

The matching condition must satisfy:

$$|z - \mu| < m \times \sigma \quad (3)$$

Parameter z stands for the gray value of the pixel in the current frame, m values 2 to 2.5.

In this algorithm m is valued 2.5 for the consideration that $m \times \sigma$ could not be too great or too little, so we select a high-low threshold with the high is 30 and the low is 15. In the process of matching, it will quit when find out the matched template.

If one pixel in the current frame does not match to all Gaussian models, it can be taken as the foreground dot, while if it matches to one Gaussian model, it could not be taken as background dot yet because there may be some noisy and disturbing factors in the background. But noisy and disturbing factors do not stay in the background for a long time, the weight of them are little. In the process of ranking the Gaussian models, it is required to set a weight threshold T [8]. If the weight of first Gaussian model ranked is greater than the threshold, then B is 1, otherwise accumulate the weights of every Gaussian model until the weight is greater than the threshold, and value the type of the Gaussian model to B. B is:

$$B = Min(\sum_{i=1}^{k} w_t > T) \quad (4)$$

Parameter T is the threshold defined, in this algorithm it is 0.75, reflecting the percentage of the background element in the series observed.

In the successful matching templates, the model is background if its type is less than B, otherwise it is foreground.

3) Background update

Due to the continue variation in the scene environment, the background model set up through training may not adapt to the need of new background, hence it is required to update the background in time, ensuring the validity of the foreground extraction.

In the process of update of the background model, we bring in the learning rate named a, when learning rate is little, the ability to adapt to new environment is low, at this moment it is required sufficient time to update. On the contrary, when a is great, the ability to adapt to new environment is high and update the background model quickly, but for the object stays in the scene in a period of time, it is hard to learn. In this situation, we adjust the learning rate, thus set different learning rate in different places in the image to ensure the demand of variation in the scene. The variation field of learning rate is from 0 to 1.

To update Gaussian models with the amount of k, we adjust three parameters weight, mean value and variance in the GMM.

The update algorithm of weight is [8]:

$$w_{k,t} = (1-\alpha) \times w_{k,t-1} + \alpha \times M \quad (5)$$

Parameter M is the matched-degree, and it is 1 when match, otherwise it is 0. If pixel of current dot does not match all the

Gaussian distribution models, we will create a new Gaussian model to replace the last model ranked, and its distribution should has the same greater variance and less weight, while the mean value should be the value of the current pixel.

C. Moving Foreground Extraction

1) The CV connectivity analysis

The binary images extracted by GMM usually contain many noisy dots, so we often use expansion and erosion to eliminate the noisy dots [9]. This paper uses a connectivity analysis method based on computational verb theory (CV connectivity analysis) to eliminate the noisy dots, and make the foreground extraction.

The CV connectivity analysis the method of finding out every moving foreground piece in the binary images of GMM and mark them with different mark number, then return information of every region of the connected region. In every region there exist 4-interconnected or 8-interconnected ones.

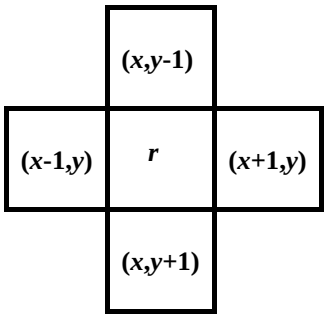

Fig. 2. 4-interconnected.

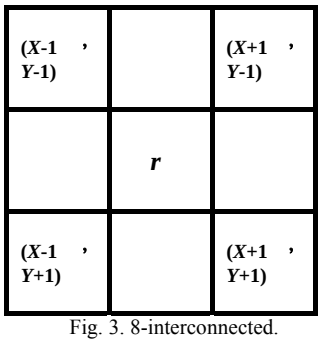

Fig. 3. 8-interconnected.

2) Noise elimination of images

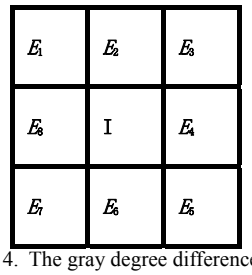

Fig. 4. The gray degree difference.

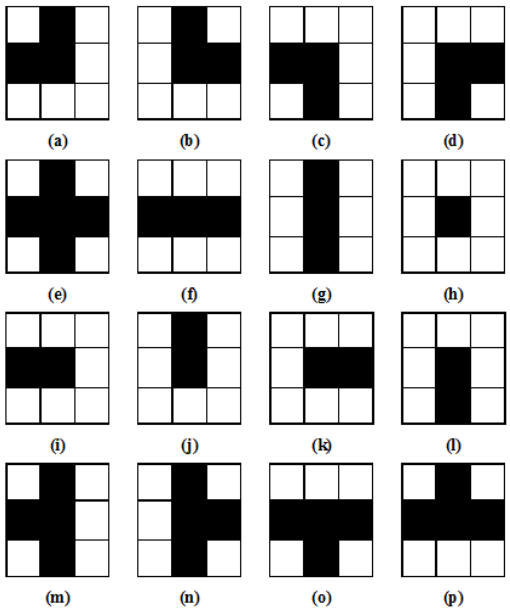

Fig. 5. 16 typical figurate verbs of the CV 4-connectivity analysis.

$p_0 = [1,0,1,1,1,1,1,0]$, $p_1 = [1,0,1,0,1,1,1,1]$,
$p_2 = [1,1,1,1,1,0,1,0]$, $p_3 = [1,1,1,0,1,0,1,1]$,
$p_4 = [1,0,1,0,1,0,1,0]$, $p_5 = [1,1,1,0,1,1,1,0]$,
$p_6 = [1,0,1,1,1,0,1,1]$, $p_7 = [1,1,1,1,1,1,1,1]$,
$p_8 = [1,1,1,1,1,1,1,0]$, $p_9 = [1,1,1,0,1,1,1,1]$,
$p_{10} = [1,1,1,1,0,1,1,1]$, $p_{11} = [1,1,1,1,1,0,1,1]$,
$p_{12} = [1,0,1,1,1,0,1,0]$, $p_{13} = [1,0,1,0,1,0,1,1]$,
$p_{14} = [1,1,1,0,1,0,1,0]$, $p_{15} = [1,0,1,0,1,1,1,0]$,

The bisect vector of the 16 typical figurate verbs are:

In the 16 typical figurate verbs, (d), (i), (j), (k) and (l) is noise, and according to the CV connectivity analysis, the noise will be eliminated.

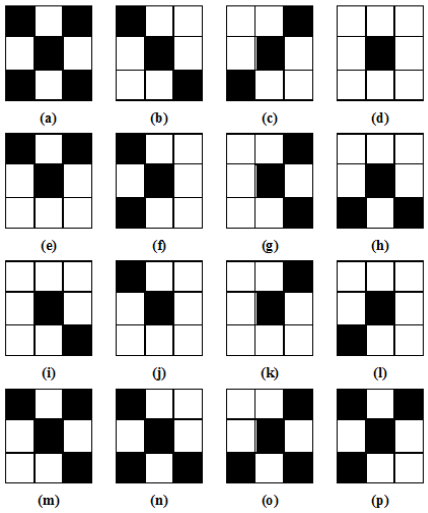

Fig. 6. 16 typical figurate verbs of the CV 8-connectivity analysis.

$p_0 = [0,1,0,1,0,1,0,1]$, $p_1 = [0,1,1,1,0,1,1,1]$,
$p_2 = [1,1,0,1,1,1,0,1]$, $p_3 = [1,1,1,1,1,1,1,1]$,
$p_4 = [0,1,0,1,1,1,1,1]$, $p_5 = [0,1,1,1,1,1,0,1]$,
$p_6 = [1,1,0,1,0,1,1,1]$, $p_7 = [1,1,1,1,0,1,0,1]$,
$p_8 = [1,1,1,1,0,1,1,1]$, $p_9 = [0,1,1,1,1,1,1,1]$,
$p_{10} = [1,1,0,1,1,1,1,1]$, $p_{11} = [1,1,1,1,1,1,0,1]$,
$p_{12} = [0,1,0,1,0,1,1,1]$, $p_{13} = [0,1,1,1,0,1,0,1]$,
$p_{14} = [1,1,0,1,0,1,0,1]$, $p_{15} = [0,1,0,1,1,1,0,1]$,

The bisect vector of the 16 typical figurate verbs are:

In the 16 typical figurate verbs, (*h*), (*i*), (*j*), (*k*) and (l) is noise, and according to the CV connectivity analysis, the noise will be eliminated.

The figurate verbs left are the ones of the connected analysis model, and should be used for the noise elimination of images.

III. Result and Analysis

This paper implements the experiment on the platform of vs 2010 and opencv 2.4.5, to analyze one video from the traffic camera. The picture resolution of the video in this paper is the format 352×288 and 5 frames/sec.

The system needs a period of time to train, then gets the GMM background model, and extracts the foreground shown in Fig. 7, 8, 9 and 10.

In the algorithm, the set of threshold is important, for different scene it is required to set different threshold, then will get the good foreground extraction effect, and it should be adjusted in practice. When setting the region threshold of area, it can help eliminate noise and disturbing factors if the

set is appropriate. Adjusting the learning rate can change the adaptation to different scenes.

In the real detecting systems, to ensure the requirement of real time and improve the efficiency, we can calculate the images by the format of 8×8 to present the foreground and output binary images of M/8×N/8 to analyze the connectivity.

Through experiments, in conclusion that foreground extraction based on computational verb theory is more effective in the aspect of real-time than the algorithms used before.

Fig. 7. Video image.

Fig. 8. Background extraction.

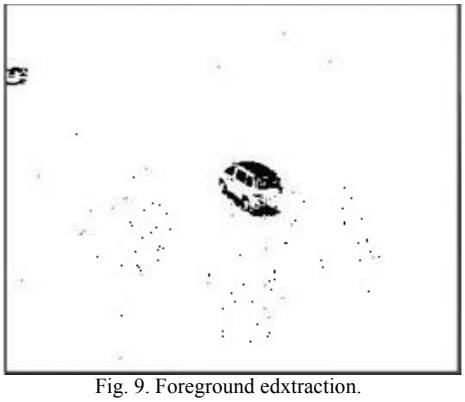

Fig. 9. Foreground edxtraction.

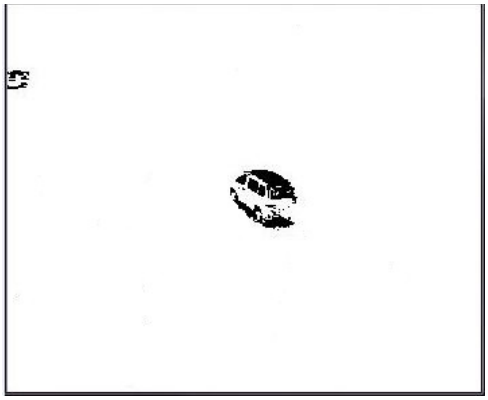

Fig. 10. foreground extraction based on CV connectivity analysis.

In the Fig. 7- Fig. 10, we get the background extraction, foreground extraction and foreground extraction based on CV connectivity analysis. We can clearly know that the significant effect of the CV connectivity analysis in the noise elimination of images. And there are the final figures below.

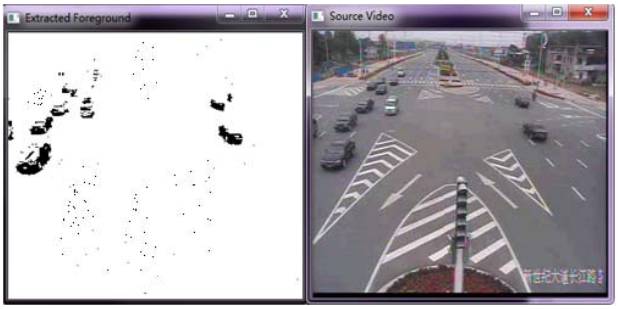

Fig. 11. Contrast of foreground extraction.

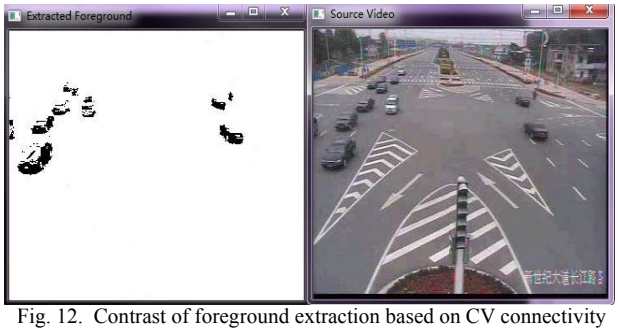

Fig. 12. Contrast of foreground extraction based on CV connectivity analysis.

IV. Conclusion

This paper tells the foreground extraction from dynamic videos, and implements it on the platform of vs2010 and opencv2.4.5. The experiment shows that the algorithm based on computational verb theory (CV) connectivity analysis has good results in most scenes, and the adaption and real-time performance are both good.

References

[1] T. Yang, *The Mathematical Principles of Natural Languages: The First Course in Physical Linguistics*, Yang's Scientific Press, Tucson, AZ, Dec. 2007, ISBN: 0-9721212-4-2.

[2] S. L. Wei and T. Yang, "Computational verb image compression," *International Journal of Computational Cognition*, vol. 7, no. 3, pp. 1-3, 2007.

[3] H. Q. Liu and T. Yang, "Computational verb clustering algorithm," *Compression and Its Open CV Implementation*, Xiamen University, pp.1-17, 2006.

[4] H. Q. Liu, Y. H. Liao, T. Yang, and C. Chen, "Image interpolation algorithm based on computational vberb theory," in *Proc. 2010 International Conference on Anti-Counterfeiting Security and Identification in Communication (ASID)*, pp. 269-272, 2010.

[5] C. Stauffer and G. Wel, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recgnition*, New York, 1999, pp. 246-225.

[6] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric Model background substraction," *Lecture Notes in Computer Science*, vol.1843, pp. 751-767 2002.

[7] S. L. Song, D. H. Liu, and L. Zeng, *A combined Image Segmentation and the Background Modeling of Moving Target Detection Algorithm*, 2002.

[8] S. Fang, F. Z. Xue, and X. H. Xu, *The Study of the Dynamic Target Detection Algorithm Based on Background Modeling and Simulation*, 2005.

[9] X. F. Wang, Z. Liu, and M. H. Cheng, *An Improved Model Based on Mixture Gaussian Distribution of Adaptive Background Elimination Algorithm*, 2008.

Li Bin received the bachelor degree in electronic and information engineering from Xidian University, Shanxi, China, in 2012. At present she is a graduate student in electronic engineering in in Xiamen University, Fujian, China. Her current interests are in machine learning, data mining, neural computing, pattern recognition, information retrieval, and image processing. In these areas she has taken part in several projects such as building video processing system and developing experimental boxes of Internet of Things and embedded system.

Yang Tao served as the chief scientist of the American Academy of Yang Sky Science in 2002, He is also the international computational cognitive magazine editor-in-chief, founder of the computational verb theory. He mainly engages in the computational verb theory, physics semantics, nonlinear electronic circuit, computational cognitive, visual chip structure and algorithm, and video understanding algorithm and architecture etc. He has a patent in the United States, has published 11 books, academic papers published over 100 articles.