Asian Journal of Information Technology, 4(11): 1098-1101, 2005

Cellular Neural Networks for Object Segmentation of Image Sequence

Yaming Wang, Weida Zhou and Xiongjie Wang
Research Center for Computer Vision and Pattern Recognition, Zhejiang Sci-Tech University,
Road no.2, Xia-Sha, Hangzhou 310018, P.R. China

Abstract: This study proposes a novel approach based on Cellular Neural Networks (CNN) is proposed for segmenting moving objects from monocular image sequence regardless of complex, changing background. First, a Gaussian distribution model for image pixel is proposed. The parameters contained in the model are adaptively updated based on the information from the current and historical frames. Based on this, every image frame is mapped from spatial domain to statistical domain. Then, a CNN framework is proposed for segmenting moving objects in statistical domain. The desirable feature of CNNs is that the processors arranged in the two dimensional grid only have local connections, which lend themselves easily to VLSI implementations. By modeling pixel interactions through using a spatial-temporal neighborhood of the CNN, sparse nosy pixel can be erased and robust segmenting results of moving objects can be achieved. Experimental results from two real monocular image sequences demonstrate the feasibility of the proposed approach.

Key words: Cellular neural networks, object segmentation, monocular image sequence, statistical domain

INTRODUCTION

The efficient segmenting of moving objects from image sequence is an important task in computer vision research area. It has many applications in surveillance, motion input for man-machine interaction, etc.[1,2]

Object segmentation from image sequence is a difficult task in case of complex and changing background. Ridder *et al.*[3] utilized a Kalman filter to model each image pixel, this approach made their system more robust to lighting changes in the scene. The limitation of this approach is that it recovers slowly and does not handle bimodal backgrounds well. Wang *et al.*[4] proposed a mutual-information based method to combine different segmentation model. This method repots good results. However, changes in scene lighting is not involved in this method. Friedman and Russell[5] proposed a pixel-wise EM framework for detection of vehicles. Their framework attempts to classify the pixel values into three different distributions corresponding to the road color, the shadow color and vehicle colors, respectively. The limitation of this framework is that it is not clear what behavior their system would exhibit for pixels which did not contain these three distributions.

This study addresses the problem of segmenting objects from image sequence with the complex and changing background. We present a Gaussian distribution model for each image pixel and map each image frame spatial domain to statistical domain. The parameters contained in the distribution model are adaptively updated based on the information from the current and historical frames. This is suitable for the problem of changing background. We then use a Cellular Neural Networks (CNN) framework to segment moving objects in statistical domain. A kind of CNN architecture has recently been designed by a group of researchers aiming at implementing the Bayesian image-processing task with real-time processing capability[6] The CNN only have local connections. By modeling pixel interactions through using a spatial-temporal neighborhood of the CNN, robust segmenting results of moving objects can be achieved.

MAPPING IMAGE FRAME FROM SPATIAL DOMAIN TO STATISTICAL DOMAIN

Consider the tth M×N frame

$$X^t = \left\{ x^t_{i,j} \middle| 1 \le i \le M; 1 \le j \le N \right\}$$

where $x^t_{i,j}$ is the value of pixel (i, j). P(V) is the mean size ration of the object in an frame. We assume that the object visits in position in a frame in a uniform manner in the image sequence. The object can be observed at every pixel with probability P(V) . $P^t_{i,j}(x)$ is the probability that pixel value x is observed at a pixel (i, j). We also define the distribution $P^t_{i,j}(x|V)$ for the object region with the value x at the pixel (i, j). For simplicity, we assume $P^t_{i,j}(x|V)$ is equal for every pixel position of the frame. Then $P^t_{i,j}(x|V)$ can be simply denoted as $P^t(x|V)$ Based on this, the probability $P^t_{i,j}(V|x)$ that a pixel (i, j) having the value x belongs to the object region can be calculated by

$$P^t_{i,j}(V|x) = \frac{P^t(x|V)P(V)}{P^t_{i,j}(x)} \qquad (1)$$

In real applications, it is necessary to know about $P^t_{i,j}(x)$ and $P^t(x|V)$ inadvance.Here,we consider calculating these two distributions from the input image sequence.

We assume the distribution of pixel value is Gaussian. The density function is

$$f^t_{i,j}(x)=\frac{1}{\sqrt{2\pi}\sigma^t_{i,j}}e^{\frac{(x-u^t_{i,j})}{(\sigma^t_{i,j})^2}} \tag{2}$$

where

$$\mu^t_{i,j}=\alpha x^t_{i,j}+(1-\alpha)\mu^{t-1}_{i,j} \tag{3}$$

$$(\sigma^t_{i,j})^2=\alpha\left(x^t_{i,j}-\mu^t_{i,j}\right)^2+(1-\alpha)(\sigma^{t-1}_{i,j})^2 \tag{4}$$

where is a constant and $0<\alpha<<1$. In first frame, the value of $\mu^1_{i,j}$ is selected as $x^1_{i,j}$, $(\sigma^1_{i,j})^2$ is set as a positive constant β. $P^t_{i,j}(x)$ is calculated by

$$P^t_{i,j}(x)=\int_{x-0.5}^{x+0.5}f^t_{i,j}(x)dx \tag{5}$$

In real applications $P^t_{i,j}(x)$, can be simply calculated by

$$P^t_{i,j}(x)=f^t_{i,j}(x)*1.0=f^t_{i,j}(x) \tag{6}$$

In order to calculate $P^t(x|V)$,we first calculated the difference frame between *t*th and (*t-s*)th frame by

$$D^t=X^t-X^{t-s}=\begin{cases}d^t_{i,j}=1, & \left|x^t_{i,j}-x^{t-s}_{i,j}\right|\geq L\\ d^t_{i,j}=0, & \text{otherwise}\end{cases} \tag{7}$$

where L is a threshold, *s* is the least number of frames, in which an identical mean size object in frame t and frame t-s are not overlapped. In case of $t\leq s$, D^t is calculated by

$$D^t=X^t-X^1=\begin{cases}d^t_{i,j}=1, & \left|x^t_{i,j}-x^1_{i,j}\right|\geq L\\ d^t_{i,j}=0, & \text{otherwise}\end{cases} \tag{8}$$

$P^t_{i,j}(x|V)$ can then be calculated by

$$P^t(x|V)=\alpha\frac{\sum_{\{i,j\,|x_{ij}=x\}}d^t_{i,j}}{\sum_{i,j}d^t_{i,j}}+(1-\alpha)P^{t-1}(x|V) \tag{9}$$

Substituting (6) and (9) into (1), we can map a frame from spatial domain to statistical domain.

MOVING OBJECTS SEGMENTATION IN STATISTICAL DOMAIN BASED ON THE CNN

A revolutionary role of CNNs[7] in neural networks is first of all in its local connectivity. Contrasting with the fully connected Hopfield Networks[8], it has local connectivity and dynamic circuit capability and is easy to integrate. This local connectivity and the adjusted weight of dynamic network are appropriate to implement in VLSI. At the same time, it provides a new way for high speed and parallel processing of large-scale images.

Any cell of CNNs is connected only to its neighbor cells interact directly with each other. Cells not in the immediate neighborhood have indirect effect because of the propagation effects of the dynamics of the network. The cell located in position (i, j) of a two-dimensional $M\times N$ is denoted by C(i, j) and its r-neighborhood $N_r(i,j)$ is defined by

$$N_r(i,j)=\{C(k,l)\,|\,\max\{|k-i|,|l-j|\}\leq r,1\leq k\leq M;1\leq l\leq N\} \tag{10}$$

Where the size of the neighborhood r is a positive integer number.

Each cell has a state x, a constant external input *u* and output y. The first-order non-linear differential equation defining the dynamics of a cellular neural network can be defined by;

$$C\frac{dx_{ij}(t)}{dt}=-\frac{1}{R_x}x_{ij}(t)+\sum_{C(k,l)\in N_r(i,j)}A(i,j;k,l)y_{kl}(t)+\sum_{C(k,l)\in N_r(i,j)}B(i,j;k,l)u_{kl}+I \tag{11}$$

$$y_{ij}(t)=\frac{1}{2}(|x_{ij}(t)+1|-|x_{ij}(t)-1|) \tag{12}$$

Where x_{ij} is the state of cell $C(i,j)$, $x_{ij}(0)$ is the initial condition of the cell, C and R_x conform the integration time constant of the system and I is an independent bias constant.

The constrain condition are

$$|x_{ij}(0)|\leq 1,\ |u_{ij}|\leq 1 \tag{13}$$

In addition, the assuming condition is

$$A(i,j;k,l)=A(k,l;i,j) \tag{14}$$

In our problem, the basic function of the CNN is mapping the input statistical date to the corresponding output image. Here we define the output value as 1 and −1, which represent object value and background value, respectively. In order to apply the CNN to the image processing, we use difference equations as follow:

$$\frac{C}{h}[x_{ij}((n+1)h)-x_{ij}(nh)=-\frac{1}{R_x}x_{ij}(nh)+\sum_{C(k,l)\in N_r(i,j)}A(i,j;k,l)y_{kl}(nh)+\sum_{C(k,l)\in N_r(i,j)}B(i,j;k,l)u_{kl}+I \tag{15}$$

$$y_{ij}(nh)=\frac{1}{2}(|x_{ij}(nh)+1|-|x_{ij}(nh)-1|)\equiv f[x_{ij}(nh)] \quad (16)$$

Where $1\le i\le M, 1\le j\le N$, h is the time length. Here we assume

$$I_{ij}=\sum_{C(k,l)\in N_r(i,j)} B(i,j;k,l)u_{kl}+I \quad (17)$$

Equation 15 can be reformulated as

$$x_{ij}(n+1)=x_{ij}(n)+\frac{h}{C}[-\frac{1}{R_x}x_{ij}(n)+ \sum_{C(k,l)\in N_r(i,j)} A(i,j;k,l)y_{kl}(n)+I_{ij}] \quad (18)$$

where $x_{ij}(n)\equiv x_{ij}(nh)$, $y_{ij}(n)\equiv y_{ij}(nh)$.

Equation 18 denotes a two-dimension filter that converts an statistical date into another image $x_{ij}(n+1)$. In Eq 16 when *h* is inclined to 0, the original system equation can be recovered. In order to see clearly the image switching system of CNN, the state equation of CNN can be expressed in integral as follows

$$x_{ij}(t)=x_{ij}(0)+\frac{1}{C}\int_0^t[-\frac{1}{R_x}x_{ij}(\tau)+f_{ij}(\tau)+g_{ij}(\tau)+I]d\tau \quad (19)$$

$$1\le i\le M, 1\le j\le N$$

Where $f_{ij}(t)=\sum_{C(k,l)\in N_r(i,j)} A(i,j;k,l)y_{kl}(t)$,

$g_{ij}(u)=\sum_{C(k,l)\in N_r(i,j)} B(i,j;k,l)u_{kl}$

Equation (19) denotes the graphic at the time t. It is decided by the initial image $x_{ij}(0)$ and the kinetic rules of CNN. So we can use CNN to get a dynamic switching of initial image at any time t. In our problem, Eq.18 can simply reformulated as:

$$x_{ij}(t+1)=Ay_{ij}(t)+Bu_{ij}(t)+I \quad (20)$$

RESULTS

Two real vehicle moving image sequences are used for moving object segmentation experiment. The 112th frame of the sequence1 and the 155th frame of the sequence 2 are shown in Fig. 1 2a, respectively. The parameters of the CNN is selected as

$$A=\begin{bmatrix}-1 & -1.5 & -1\\ -1.5 & 10.5 & -1.5\\ -1 & -1.5 & -1\end{bmatrix},\ B=\begin{bmatrix}-1 & -1 & -1\\ -1 & 6 & -1\\ -1 & -1 & -1\end{bmatrix},\ I=0.$$

Fig 1a: The 112th frame of sequence 1

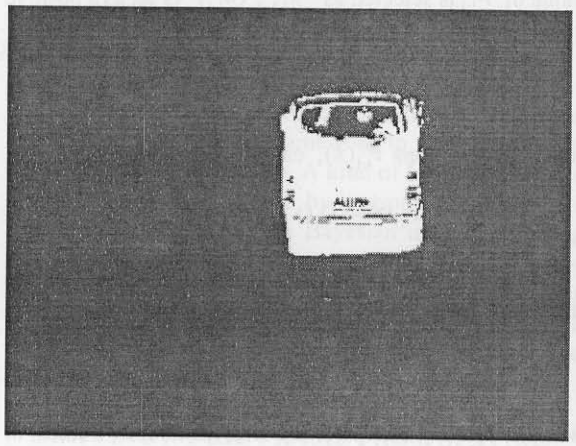

Fig. 1b: The segmentation result of Fig. 1a

Fig. 2a: The 155th frame of sequence 2

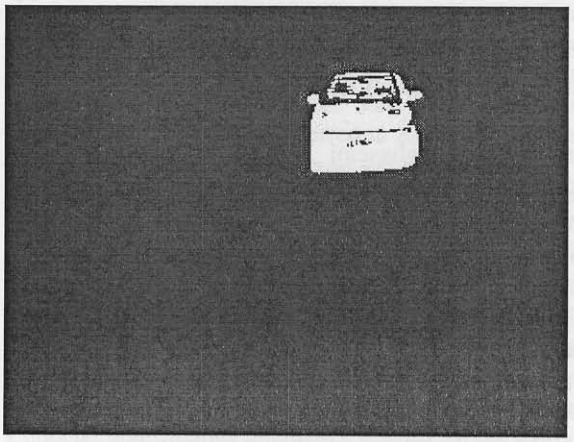

Fig. 2b: The segmentation result of Fig. 2a

The segmentation results are shown in Fig. 1b and 2b, respectively. It is obviously that the segmented object regions are compact and sparse nosy pixels are erased.

CONCLUSIONS

We have developed an approach to segmenting moving objects from image sequence based on the CNN. By using a statistical approach, the image frame is mapped from spatial domain to statistical domain. Thus the CNN is used to segment moving objects in statistical domain. By modeling pixel interactions through using a spatial-temporal neighborhood of the CNN, sparse nosy pixel can be erased and robust segmenting results of moving objects can be achieved.

ACKNOWLEDGEMENTS

This study is supported by the National Science Foundation of China under Grant 60473038 and Zhejiang Provincial Natural Science Foundation of China under Grant RC02064.

REFERENCES

1. Cai, Q. and J. K. Aggarwal, 1996. Tracking human motion using multiple cameras. In: Proceeding of 13th International Conference on Pattern Recognition, pp: 68-72.
2. Wren, C., A. Azarbayejani, T. Darrell and A. Pentland, 1996. Pfinder: Real-time tracking of the human body. In: SPIE Proceeding, 2615: 89-98.
3. Ridder, C., O. Munkelt and H. Kirchner, 1995. Adaptive background estimation and foreground detection using kalman-filtering. In: Proceeding of International Conference on Recent Advances in Mechatronics, pp: 193-199.
4. Wang, Y., Y. Zhang, L. Cao, W. Zhou and W. Huang, 2005. Segmentation of moving objects using mutual-information. Asian J. Inform. Technol., 4: 218-221.
5. Friedman, N. and S. Russell, 1997. Image segmentation in video sequence: A probabilistic approach. In: Proceeging of the 13th Conference on Uncertainty in Artificial Intelligence, pp: 175-181.
6. Lithon, F. and D. Dragomirescu, 1999. A cellular analog network for MRF-based motion detection, IEEE Trans. Circuits and System, 46: 281-293.
7. Chua, L.O. and L.Yang, 1988. Cellular neural networks theory and applications. IEEE Trans, Circuits and System, 35: 1257-1290.
8. Asai, H., K. Onodera, T. Kamio and H. Ninomiy, 1995. A study of Hopfield neural networks with external noises, In: Proceeding of IEEE International Conference on Neural Networks, pp: 1584-1589.