Multiple Objects Tracking Based on Snake Model and Selective Attention Mechanism

Haoting LIU, Guohua JIANG, Li WANG

Institute of Space Medico-Engineering
imkyran@hotmail.com

Abstract

In this paper we present a novelty multiple objects tracking (MOT) algorithm that relies on Snake and selective attention mechanism (SAM) for aerial image sequences. Snake model with five energy functions is used to segment contours of targets precisely. At the same time, in order to decrease the computational complexity, a SAM model is designed. In our method we view attention selection as a process of feature extraction and classification. So unlike most of other SAM models, which compute in pixels level absolutely, our approach focuses on the process of classifying high-level information about color, size, speed and relative location to introduce SAM. Binary BP classifier and linear criterion function are used to decide whether a target is noticed by camera or not. Simulation results show this approach can discriminate targets and segment their contours in different computational complexities under the assumption that no occlusions occur.

1 Introduction

Researches on tracking multiple non-rigid targets have been intensified in recent years because of their potential applications in military and business affairs. Accurate segmentation of contours and tracking them in sequences are helpful to analysis the partial characteristics of each target. As a result this technique has been emphasized broadly in many tracking applications. The aim of our paper is to design a MOT algorithm to track contours of multiple targets precisely with tolerable computational complexities for aerial image sequences. Some tracking applications have employed Snake model to segment contour [1,2,3]. In [2] Snake was used to segment contour of football players in game. In that method in order to decrease the processing time Snake model had to be set with small number of snaxels, which affected the calculation results apparently. At the same time when the number of snaxels or targets increased the segmentation results always became blurred, i.e. that system had to reduce accuracy and accept some incompetent results because of the large amount of computation. So the conflicts between computational complexity problems (CCP) and computational accuracy problems (CAP) still were prominent in that approach.

The design of SAM computational model is another research issue in vision or image understanding areas. SAM model always makes the machine vision system worked like an actual vision system and more effective. In general, all the researches in SAM [4,5,6] are mainly established in the process of discovering characteristics of biologic visual system and simulating them, which always focus on two issues: eyes movement and the resolution of retina. From engineering points of view, one of credible computation model for simulating human visual attention has been proposed by Koch and Ullman in 1985 [7,8]. This model has been great successful and become the prototype of many other SAM ones, but it still has a drawback: the large amounts of pixel-level calculation, which makes this framework break away from actual application in many situations.

In this paper a MOT approach, which works like a biologic vision system, is proposed. On the one hand, active contour model is employed to segment contours of moving objects. In that case, Snake model with five energy items (two external functions and three internal ones) is used to segment contours of targets precisely. Three internal energy functions that have always been used in medical applications are designed to improve the segmentation results. Meanwhile Split Template and Kalman filter are used to match and forecast the targets in sequences respectively. On the other hand, we hope to stress some important targets and distribute more computational resources on them. So a SAM computational model, which is implemented by binary BP classifier and Linear Criteria Function (LCF), is employed. Unlike the traditional SAM model, the eye movement and spot light phenomena [9] are computed according to three factors (dominant color, size and speed) and relative location information respectively. Here, dominant color, size and speed always can be seen as a kind of self-factor while location information represents the cluster-factor of targets, so a four-input with one-output SAM module is operated to identify targets and distributes different computational resources on them. So the final motivation of our model is to use high-level information to introduce SAM, which shares small amount of computation relatively.

The rest of this paper is organized in the following way. In section 2 the framework of our algorithm will be given. In section 3 main techniques for segmentation and track will be introduced. In section 4 SAM module will be discussed. In section 5 and 6 some experiment results and conclusions will be given.

2. Algorithm Framework

We aim to classify all the targets into two species and distribute different computational resources on them. The framework of our algorithm can be shown in fig1. Some basic digital image processing methods [10] such as motion analysis, Kalman filter, template matching techniques etc, are employed here. When the image data are fed into the model, after operations with features extraction, SAM module will calculate the features about color, size, speed, and location of each target to decide whether a target would be noticed or not by camera. If deemed as a noticed one, this target will be segmented contour by Snake model which has high computational complexity but a better segmentation result; when a target is judged as an unnoticed one, the algorithm will use a simple computational strategy to decrease the amount of computation. After that, this approach will forecast the location of targets by Kalman filter in next frame, then search and match them by split template.

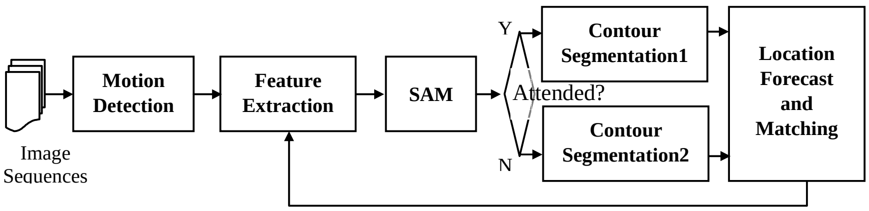


Fig. 1. Framework of algorithm.

3. Snake Model and Split Template

3.1 Snake Model

A. Energy Function Design

The energy function for Snake in this paper is composed of five items. Image gradient and color gradient are called external energy functions customarily, which guarantee Snake curve convergence to the gradient mutative regions of image. Besides, elastic, rigidity and constraint energy functions are regard as internal energy ones, which moderate Snake curve to maintain some kinds of geometrical shapes respectively.

$$E_{total} = (k_i \times E_{image} + k_c \times E_{color}) + (k_e \times E_{elastic} + k_r \times E_{rigidity} + k_{co} \times E_{constraint}) \qquad \textbf{(1)}$$

where $k_i, k_c, k_e, k_r, k_{co}$ are weights of energy function.

In this model, image and color energy function will be calculated in RGB color space [11]:

$$E_{image} = \sum_{i=0}^{n-1} \left\{ \nabla \left[I(v_i) \right]^2 \right\} \qquad E_{color} = \sum_{i=0}^{n-1} \sqrt{\lambda_+^{rgb} - \lambda_-^{rgb}} \qquad \textbf{(2)}$$

where $\lambda_\pm = \frac{g_{11}^{rgb} + g_{22}^{rgb} \pm \sqrt{(g_{11}^{rgb} - g_{22}^{rgb})^2 + 4(g_{12}^{rgb})^2}}{2}$, $g_{11}^{rgb} = \left|\frac{\partial r}{\partial x}\right|^2 + \left|\frac{\partial g}{\partial x}\right|^2 + \left|\frac{\partial b}{\partial x}\right|^2$, $g_{22}^{rgb} = \left|\frac{\partial r}{\partial y}\right|^2 + \left|\frac{\partial g}{\partial y}\right|^2 + \left|\frac{\partial b}{\partial y}\right|^2$,
$g_{12}^{rgb} = \frac{\partial r}{\partial x}\frac{\partial r}{\partial y} + \frac{\partial g}{\partial x}\frac{\partial g}{\partial y} + \frac{\partial b}{\partial x}\frac{\partial b}{\partial y}$, $r(R,G,B) = \frac{R}{R+G+B}$, $g(R,G,B) = \frac{G}{R+G+B}$, $b(R,G,B) = \frac{B}{R+G+B}$.

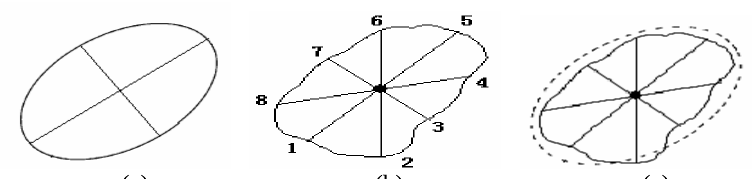
Internal energy can be defined as follows [12]:

$$E_{elastic} = \sum_{i=0}^{n-1} \left[l - \left| v_i - v_{i-1} \right|^2 \right] \quad E_{rigidity} = \sum_{i=0}^{n-1} \left| v_{i-1} - 2 \times v_i + v_{i+1} \right|^2 \quad E_{constraint} = \sum_{i=0}^{n-1} \left| v_i - v_c \right| \qquad \textbf{(3)}$$

where $l = \frac{1}{n}\sum_{i=0}^{n-1} \left| v_i - v_{i-1} \right|^2$, describes the distance between adjacent snaxels, and $v_c = \frac{1}{n}\sum_{i=0}^{n-1} v_i$, is the barycenter of Snake curve.

B. Initialization of Snaxels

With the hypotheses of small targets and without any occlusions, an ellipse initialization method is more suitable than a rectangle one [2]. When a new location of barycenter is given, an ellipse will be created in the same position. Length of major axis is 1.1 times as long as that of the line 1-5 in fig2 (b) and direction of created ellipse will parallel it in previous frame.



(a) (b) (c)

Fig. 2. Ellipse initialization method. (a) shows an ellipse prototype with its major and minor axes; (b) presents an arbitrary closed curve with its 8 average segmentation points; (c) is the curve of created ellipse in dash line

C. Optimization Method and Termination Conditions

In this paper, we employ n×n Fast Greedy Algorithm (FGA) [13] to calculate the optimal results of Snake model, i.e., each snaxel will be seen as a center while its (n×n-1) adjacent points will be taken into account at the same time.

A binary termination condition is considered: the maximum iteration number α ($\alpha>0$) of optimal computation and the fixed-point ratio β ($\beta\in[0,1]$) of snaxels. The first condition is reasonable enough, while the second one means: after iterations for some times, parts of snaxel will reach their minimum energy location while others are still in the courses of searching. So the second termination condition implies if the number of fixed-point is large enough, we can conclude the curve has reached an optimum state with some permitted error rates.

$$\beta = \text{fixed\_points}/\text{total\_points} \qquad \textbf{(4)}$$

3.2 Split Template

In fig3, a split template is displayed. An integrated template has 5 sub-templates; each sub-template shares the size of n×n.

	n×n Template3	
n×n Template4	**n×n Template1**	**n×n Template2**
	n×n Template5	

Fig. 3. Split Template

First of all, a sub-template (template 1 in fig.3) is created at the position of barycenter in target region, and then algorithm will calculate 4 distances between barycenter and boundaries of target region (right, top, left and bottom) in turns. If the distance is large enough, such as 1.5 times than n, a new sub-template will be created, i.e. there are at most 16 different kinds of way to create the split sub-template.

Template matching criteria can be defined as follows and the optimal match will be gotten when function D(A,B) reaches a maximum value.

$$D(A,B) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} R(A_{i,j}, B_{i,j}) \qquad \textbf{(5)}$$

$$R(A_{i,j}, B_{i,j}) = \begin{cases} 1 & |A_{i,j} - B_{i,j}| < T \\ 0 & \text{else} \end{cases} \qquad (6)$$

where $A_{i,j}, B_{i,j}$ is template and image blob respectively; T is threshold.

4. SAM Model with High-Level Information

4.1 Model of SAM

A BP classifier and a linear criterion function are designed in this section to introduce SAM. In fig4, among these four factors, color, size and speed always represent the self-attribution of each target, while the location factor describes the cluster-attribute of targets distinctly. So all these four factors are treated in different ways.

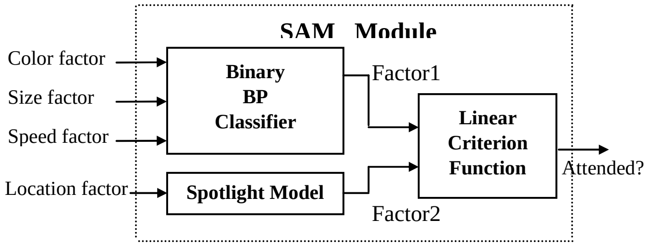


Fig. 4. SAM module. All these factors are divided into two classes and composed of input vector of the binary BP classifier and spotlight model respectively. Their outputs, Factor1 and Factor2, are the input values of the linear criterion function

4.2 Design of 3 Factors with BP Classifier

Three-layer modified BP net classifier is used in SAM module. The input and supervisory vector is $[Color, Size, Speed, Attention]^T$, while the output can be defined as

$Factor_i^1$ with its binary number 0.1 and 0.9. All the subscript (or superscript) i in this paper is sequence number of target i.

Training data of BP net are designed regularly while the supervising data are well organized at the same time. First, 9 dominant colors are selected from color space and mapped into data space [0.1,0.9]. Then the color of each target will be sampled and clustered to these centers. Size and speed factors are also be normalized into [0.1, 0.9] by certain means. All the supervising data (or teacher data) are created by subjects, which ensure the property of BP net more closed to the cognitive characteristics of human being. What's more, this model is more flexible than any other ones because it can be moderated (or be trained) for different purposes easily.

At last, when organizing the supervising data, our method supposes that among 3 factors, color has a higher priority than size, i.e., the color of target is more attractive than the size to our eyes. And so it is with the size factor and the speed factor.

A. Color Factor

Likes Content Based Image Retrieval (CBIR)[14] technique, dominant color of each target will be calculated and transformed into CIELab color space [15]. 9 dominant colors are selected out with their priority sequences as below: red, blue, amethyst, black, green, cyan, orange, yellow and white. The clustering criteria between color of target and cluster center in Lab color space can be described as:

$$| X - Z_i | = \sqrt{(L_x - L_{Z_i})^2 + (a_x - a_{Z_i})^2 + (b_x - b_{Z_i})^2} \tag{7}$$

where Z_i is clustering center, X is sampling color in target region.

Each color can get a corresponding data as displayed in table 1 and formula (8):

$$f_{color}^{i} = Data \tag{8}$$

Table 1. Dominate color and their corresponding values

Color	Red	Blue	…	White
Data	0.9	0.8	…	0.1

B. Size and Speed Factors

Size and Speed of each target can be calculated by Snake model and Kalman filter respectively. If the prior-information of size or speed can be obtained at the beginning, moderation of these data into [0.1, 0.9] is not a difficult task, but if not, a relative value will be available in formulas (9) and (10). In order to make training data proportional in data space, this approach also adopts 9-level data scale (0.1~0.9).

$$f_{size}^{i} = \frac{size_i}{size_{\max}} \times (0.9 - 0.1) + 0.1 \tag{9}$$

$$f_{speed}^{i} = \frac{speed_i}{speed_{\max}} \times (0.9 - 0.1) + 0.1 \tag{10}$$

where the subscript 'max' means the maximum of size or speed.

After calculating relative values, we can employ minimum distant principles to cluster those values into their 9 clustering centers (0.1, 0.2 …etc.).

4.3 Location Factor

When deal with location factor, a “spot light” computational model is designed. First a visual field center of all the targets is calculated, then each target will estimate the distance with it, if the distance is too large (small) this target will be regard as an unnoticed (noticed) target. Finally, we can calculate the barycenter of all the targets as Visual Field Center (VFC):

$$V_{VFC} = \frac{1}{M}\sum_{i=1}^{M} v_c^i \qquad \textbf{(11)}$$

where M is the number of targets, v_c^i is the barycenter of target i, $V_{VFC} = (V_{VFC}^x, V_{VFC}^y)$, which presents the coordinate of VFC.

So the location factor can be defined as follow:

$$f_{location}^i = \frac{D_i}{D_{\max}} \in (0,1) \qquad \textbf{(12)}$$

where $D_\bullet$ is the distance between V_{VFC} and barycenter of target i.

4.4 Linear Criterion Function

Before we discuss the linear criterion function, here we introduce two simulation results of Snake model at first: (1) increasing search region (i.e. the n) of FGA will improve the convergence results, (2) increasing maximum iteration number (α) or fixed-point ratio (β) of Snake model will also lead to some better results to some extent. So the finial formula of LCF can be defined in (13):

$$A_i = \gamma \times Factor_i^1 + \eta \times (1 - Factor_i^2) \qquad \textbf{(13)}$$

$$Factor_i^2 = f_{location}^i \qquad \textbf{(14)}$$

$$\mathrm{N}_i = \begin{cases} 1 & \text{if } A_i > \sigma \\ 0 & \text{else} \end{cases} \qquad \textbf{(15)}$$

where γ and η are parameters, σ is threshold.

In (15), if a target is attended by camera, i.e. N_i equal to 1, a larger neighbors searching strategy will be employed comparing with those unnoticed targets, at the same time, α and β of Snake model will be increased too.

5. Simulation Results

A new on-line training BP net, which has an ability to jump from the local minimum, is designed in this model. Table2 shows the primary parameters [16,17,18] and the average classification accuracy of BP nets. Table3 gives out the parameters of MOT algorithm.

Table 2. Primary parameters and performance of BP net. LA–Learning Rate, MP–Momentum Parameter, HLN-Hidden Layer Number, NE-Net Error, CA-Classification Accuracy

Weights	LA	MP	HLN	NE	CA
Values	0.01	0.95	10	0.03	0.93

Table 3. Parameters of MOT algorithm

Weights	N	α	β	γ	η	σ
Noticed	5	60	0.8	0.65	0.35	0.5
Unnoticed	3	50	0.6	＼	＼	＼

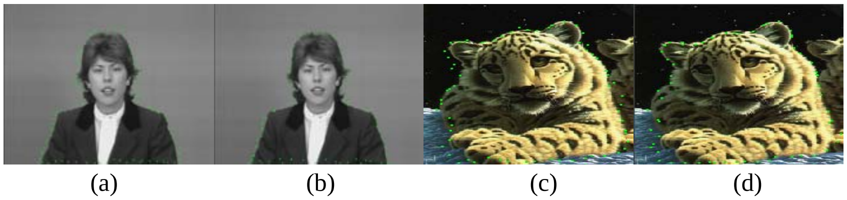


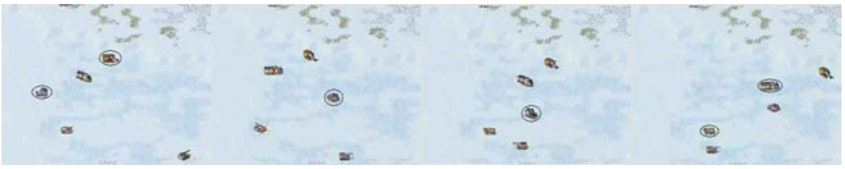
Fig. 5. Results with different computational strategies for Snake. (a) and (b) are gray images, while (c) and (d) are 24 bits color (RGB) ones

Table 4. Evaluation results of fig.5

Test		CCP		CAP
Image	N	α	β	Subjective Evaluation
(a) (c)	3×3	1000	0.8	Good
(b) (d)	5×5	10000	0.9	Better

Fig5 shows the segmentation results using different search and termination strategies. Table 4 represents the CCP and CAP of fig5. In general, it's a difficult task to compare the CCP in theory although large search strategy and termination condition seem stricter than weak ones. But most simulation results have shown the former occupies more computation resources than the latter. Meanwhile, between α and β, the maximum iteration number (α) is more efficient than the fixed-point ratio (β).

Fig.6 gives out the multiple objects tracking results of proposed algorithm. Parameters of Snake model for this experiment are shown in table 5. Test images in fig6 are cut from video of AVI format with their size of 352×240. The average speed is about 7~8 frames per second on our PC (Intel Pentium Ⅳ with 256MB RAM) relying on C implementation of the proposed approach.



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

Fig. 6. Tracking results in different frames. To make it clear, attended targets are tagged with a black ellipse

Table 5. Parameters of Snake model. Weight of constraint item is negative, which ensures snaxels contract to the barycenter of target

Weight	k_i	k_c	k_e	k_r	k_{co}
Value	-6	-6	1	8	-8

6. Conclusions

A multiple objects tracking algorithm based on Snake model and selective attention mechanism is proposed in this paper, which can be used to track aerial image sequences or remotely sensed image ones. With this method Snake model, Kalman filter and split template are designed to segment and track contour of targets precisely while four factors, color, size, speed and relative location are considered in SAM module to decrease the computational complexity. On the one hand, unlike traditional tracking approach, our algorithm has some advantages both in accuracy and complicity. On the other hand, comparing with the framework of SAM advised by Koch and Ullman, which is established in collecting and processing elementary image data, this approach focuses on the course of classifying high-level information, so it is still closed to the cognition process of human being but more efficient than other models.

Acknowledgements

Thank Dr. Xiaochen Bo from Academy of Military Medical Sciences for his advice in our research and efforts in amending this paper. Our thanks also go to my cousin Mai Li for her advice.

References

1. Zhimin FAN, Jie ZHOU, Dashan GAO and Zhiheng LI: Contour Extraction and Tracking of Moving Vehicles for Traffic Monitoring. The IEEE 5th International Conference on Intelligent Transportation Systems (2002) 84-87
2. Sébastien Lefèvre, Jean-Pierre Gérard, Aurélie Piron and Nicole Vincent: An Extended Snake Model For Real-Time Multiple Object Tracking. RFAI: International Workshop on Advanced Concepts for Intelligent Vision Systems (2002) 268-275
3. Sébastien Lefèvre, Cyril Fluck, Benjamin Maillard and Nicole Vincent: A Fast Snake-based Method to Track Football Players. RFAI: International Workshop on Machine Vision Applications (2000) 501-504
4. Michele Rucci and Paolo Dario: Selective Attention Mechanisms in A Vision System Based on Neural Networks. International Conference on Intelligent Robots and Systems (1993) 1742-1749
5. Albert Ali Salah, Ethem Alpaydin and Lale Akarun: A Selective Attention-Based Method for Visual Pattern Recognition with Application to Handwritten Digital Recognition and Face Recognition. Pattern Analysis and Machine Intelligence, Vol. 24, No. 3 (2002) 420-425
6. Tianfu Wu, Jun Gao and Qin Zhao: A Computational Model of Object-Based Selective Visual Attention Mechanism in Visual Information Acquisition. International Conference on Information Acquisition (2004) 405-409
7. C. Koch and S. Ullman: Shifts in selective visual attention toward the underlying neural circuitry. Human Neurobiology, Vol. 4, No. 4 (1985) 219-227
8. Laurent Itti and Christof Koch: A Saliency-based Search Mechanism for Overt and Covert Shifts of Visual Attention. Vision Research 40 (2000) 1489-1506
9. Marvin M. Chun and Jeremy M. Wolfe: Visual Attention. Blackwell Handbook of Perception, Oxford, UK: Blackwell Publishers Ltd (1999) 272-310
10. Ali Erkin Arslan and Mübeccel Demirekler: Visual Tracking with Group Motion Approach. Conference on Computer Vision and Pattern Recognition Workshops, Vol. 8 (2004)
11. Theo Gevers, Sennay Ghebreab and Arnold W. M. Smeulders: Color Invariant Snakes. British Machine Vision Conference (1998) 578-588
12. Senthil Kumar and Dmitry Goldgof: Automatic Tracking of SPAMM Grid and the Estimation of Deformation Parameters from Cardiac MR Images. Medical Imaging Vol. 13, No. 1 (1994) 122-132
13. K. M. Lam, H. Yan: Fast Greedy Algorithm for Active Contours. Electronics Letters, 6th, Vol. 30 No. 1 (1994) 21-23
14. Ahmet Ekin and A. Murat Tekalp: Shot Type Classification by Dominant Color for Sports Video Segmentation and Summarization. IEEE International Conference on Acoustics, Speech, and Signal Processing (2003) 173-176
15. Christine Connolly and Thomas Fliess: A Study of Efficiency and Accuracy in the Transformation from RGB to CIELAB Color Space. Image Processing, Vol. 6, No. 7 (1997) 1046-1048
16. Mehmet Ali Arslan: The BP Neural Networks with Data Clustering Enhancement – An Emerging Optimization Tool. The IEEE International Symposium on Intelligent Control (1996) 188-193
17. D. Randall Wilson and Tony R. Martinez: The Inefficiency of Batch Training for Large Training Sets. International Joint Conference on Neural Networks Vol. 2 (2000) 113-117
18. Pan Hao, Jing-Ling Yuan, Luo Zhong: Probing Modification of BP Neural Network Learning-Rate. The First International Conference on Machine Learning and Cybernetics (2002) 307~309

Haoting LIU, received his M.S. degree in man-machine-environment engineering from ISME (Insititute of Space Midico-Engineering), Beijing, China, 2005. He is currently a member of Vision Simulation and Virtual Reality Laboratory in CCA (Center of China Astronauts). His main area of interests are signal processing for tracking filter, stereo and motion analysis, and medical image segmentation.

Guohua JIANG, received his B.S. in biophysics from Nan Kai University, Tianjing, China, in 1989 and an MS in man-machine-environment engineering from ISME (Insititute of Space Midico-Engineering), Beijing, China, in 1992. He is currently a professor of CCA (Center of China Astronauts), majoring in Space Ergonomics and Training Simulation.

Li WANG, received her B.S. degree from China Medical University in 1993 and M.S. degree in man-machine and environment system engineering from Bei Hang University in 2003, respectively. Since then, she is an associate professor in CCA (Center of China Astronauts), specializing in space ergonomics.