

## 基于运动补偿的 Snake 视频对象跟踪算法\*

郭礼华, 李建华, 杨树堂

(上海交通大学 电子信息与电气工程学院, 上海 200030)

**摘要:**当对象移动幅度大时, Snake 视频对象跟踪算法中的曲线迭代过程易陷入局部最小, 为此, 提出一种运动补偿预处理的解决方法。此方法预先估计对象的运动信息, 然后对 Snake 曲线的初始轮廓位置进行运动补偿, 最后进行 Snake 跟踪。实验表明, 这种方法不仅跟踪效果好, 而且能有效地减少 Snake 曲线演化的迭代次数。

**关键词:** Snake 模型; 主动轮廓模型; 视频对象跟踪; 运动补偿; 运动估计

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1007-2276(2005)01-0093-05

### Video object tracking method based on motion compensation Snake model\*

GUO Li-hua, LI Jian-hua, YANG Shu-tang

(School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200030, China)

**Abstract:** If a somewhat fast moving object exists in successive images, Snake's nodes may fall into the local minima in their motion to the new positions of the target object in next image. In this paper, Snake tracking algorithm based on the motion compensation is presented to overcome this problem. Firstly, the motion information of object is analyzed. Secondly, the object's position is made compensated according to this motion information. Finally, the object is tracked by Snake model. The final experimental results prove that this method performances a good tracking result and can greatly decrease the number of the Snake's contour evolvement.

**Key words:** Snake model; Active contour model; Video object tracking; Motion compensation; Motion estimation

## 0 引言

在视频对象分割和跟踪算法中, Snake 模型是一种非常有效的工具和算法。这种模型也称为 Snake 模型, 它主要定义了 Snake 积分能量函数, 其能量函数

包括内部能量、图像能量和外部能量。对象跟踪的依据是使其所有的能量函数积分最小。Kass 等人<sup>[1]</sup>最早提出此模型并给出变分法来求解积分最小。由于变分法在实际的求解过程中, 数值解不稳定, 并且在复杂的图像拓扑结构中不能很好地进行曲线演化, 为

收稿日期: 2004-06-22; 修订日期: 2004-07-10

\* 基金项目: 国家 863 高科技资助项目 (2002AA145090)

作者简介: 郭礼华 (1978-), 男, 江西永新人, 博士生, 研究方向为视频内容分析与处理、图像分割、视频跟踪等。

此 Caselles<sup>[2]</sup>提出了 Geodesic 主动轮廓模型等。这种主动轮廓模型采用 Osher 和 Sethia<sup>[3]</sup>提出的 Level Set 方法作为曲线演化算法,把曲线演化过程模拟成一个曲面演化过程。该方法很好地解决了采用变分法进行数值求解所产生的问题,且对初始曲线的选择也不敏感。Level Set 方法由于这些特有的优势,现已广泛地应用于各种实际的 Snake 视频跟踪算法中<sup>[4-6]</sup>。

在视频对象跟踪技术中,Snake 模型虽然可以自动进行曲线演化变形,以适应视频对象的运动,可是如果所跟踪的对象位置飘忽不定,或者在相邻的视频帧内对象大幅度地运动,曲线叠代演变到新位置时将很容易陷入局部最小,此时不仅跟踪次数将大大增加,而且跟踪效果也不好。为此,提出了一种基于运动补偿的 Snake 视频对象跟踪算法,在运用 Snake 模型进行对象跟踪之前,对跟踪对象运动信息进行预估计,然后根据预估计的结果对视频对象的位置和大小进行运动补偿,最后再利用 Snake 模型进行变形跟踪。

## 1 运动估计和运动补偿

在视频编码中,运动估计方法一般都是采用块匹配的方法来实现,块匹配方法主要通过搜索范围内计算相邻帧块之间的绝对差(MAE),绝对差最小的位置被认为是块运动后的位置。其绝对差计算公式如下:

$$MAE(\Delta x, \Delta y) = \frac{1}{n \times n} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} |S_c(x+j, y+k) - S_c(x+\Delta x+j, y+\Delta y+k)| \quad (1)$$

式中  $S_c(x, y)$  表示当前帧内位置为  $(x, y)$  处的方块;  $S_c(x+\Delta x, y+\Delta y)$  表示参考帧内位置为  $(x+\Delta x, y+\Delta y)$  处的方块。  $(\Delta x, \Delta y)$  是其可能的运动矢量,它的范围是  $(-M \leq \Delta x \leq M, -M \leq \Delta y \leq M)$ 。块匹配的主要思想是在  $M \times M$  搜索范围内寻找最小的绝对差,从而找到当前  $S_c(x, y)$  块所对应的运动矢量。在实际的运动估计中,基于块匹配的方法受图像帧内的噪声干扰等各种因素的影响,运动估计很不准确。为了能够准确估计出实际跟踪对象的运动信息,利用特征点的匹配来估计对象的运动信息,目前有很多的特征点提取算法,采用参考文献[7]提出的 Harris 特征点提取方法,这种方法对比例变形、场景光照变化、视点改变和图像

噪声干扰都有很好的鲁棒性<sup>[8]</sup>。具体实现细节见参考文献[9]。

文中的特征点匹配方法主要思想源于参考文献[10],下面详细介绍其实现步骤。由于已经提取出来了视频帧  $I^n$  和帧  $I^{n+1}$  所对应的特征点,下面的工作就是在这两个相邻的视频帧内建立起特征点的匹配关系。为此,在搜索区域内计算其对应点之间的协方差,通过协方差最大来搜索其匹配的特征点。最后再利用互验证来最终实现特征点的匹配。具体实现如下:

设  $X^n = \{p_k^n \mid p_k^n = (x_k^n, y_k^n)\}_{k=1}^M$  和  $X^{n+1} = \{p_k^{n+1} \mid p_k^{n+1} = (x_k^{n+1}, y_k^{n+1})\}_{k=1}^M$  分别是视频帧  $I^n$  和帧  $I^{n+1}$  所对应的特征点集。搜索区域范围为  $S = \{(d_1, d_2) \mid d_1^{\min} \leq d_1 \leq d_1^{\max}, d_2^{\min} \leq d_2 \leq d_2^{\max}\}$ 。通过计算特征点  $p_k^n$  和特征点  $p_k^{n+1}$  之间的互协方差  $\psi(p_k^n, p_k^{n+1})$  来评估其相似程度。协方差计算公式为:

$$\psi(p_k^n, p_k^{n+1}) = \frac{\sum_{i=-N_x}^{N_x} \sum_{j=-N_y}^{N_y} [I^n(x_k^n, y_k^n) - u^n(x_k^n, y_k^n)][I^{n+1}(x_k^{n+1}, y_k^{n+1}) - u^{n+1}(x_k^{n+1}, y_k^{n+1})]}{(2N_x+1)(2N_y+1)\sqrt{\sigma^2(x_k^n, y_k^n)\sigma^2(x_k^{n+1}, y_k^{n+1})}} \quad (2)$$

式中  $u^n(x_k^n, y_k^n)$  和  $u^{n+1}(x_k^{n+1}, y_k^{n+1})$  分别是视频帧  $I^n$  和帧  $I^{n+1}$  所对应位置的均值;  $\sigma^2(x_k^n, y_k^n)$  和  $\sigma^2(x_k^{n+1}, y_k^{n+1})$  分别是视频帧  $I^n$  和帧  $I^{n+1}$  所对应位置的方差。相关匹配窗口为  $(2N_x+1) \times (2N_y+1)$ 。在这个匹配窗口中,只有当其对应的像素灰度值最相似,其互协方差才会最大。对于视频帧  $I^n$  内的每个特征点  $p_k^n \in X^n$ , 在这个特征点为中心的搜索区域  $S$  内寻找视频帧  $I^{n+1}$  对应的特征点  $p_k^{n+1} \in X^{n+1}$ , 并且使得其两点之间的互协方差最大。一般来说,可以用下面数学表达式来表示:

$$\forall p_k^n \in X^n: \max_{p_i^{n+1} \in X^{n+1}} \psi(p_k^n, p_i^{n+1}), p_i^{n+1} - p_k^n \in S \quad (3)$$

为了保证其匹配关系的可靠性,还需保证其计算的互协方差  $\psi(p_k^n, p_k^{n+1})$  大于一个域值,如:  $\psi(p_k^n, p_k^{n+1}) >$

$\rho$ 。  $\rho$  值取 0.8。

受图像的噪声和其他因素的影响,特征点之间可能会产生误匹配。为了进一步提高运动估计的准确性和消除误匹配的影响,采用互验证处理技术。由于已经建立视频帧  $I^n$ 、 $I^{n+1}$  中的特征点的匹配关系,利用前面同样的方法建立视频帧  $I^n$  和视频帧  $I^{n+1}$  的特征点的匹配关系。当这种匹配关系相互验证时,认为这个相互关系是可靠的。具体来说,如果特征点  $p_k^{n+1}$  是特征点  $p_k^n$  的最佳匹配点,而且特征点  $p_k^n$  又是特征点  $p_k^{n+1}$  的最佳匹配点,这种匹配关系得以保留,其他匹配关系将舍弃。通过相互验证,其匹配关系的可靠性得到很大的提高,这为后面正确估计对象的运动信息提供保障。

对准确的特征点匹配信息进行分析,从而计算出对象的运动类型。假设所有的运动都可以用运动模型来描述。通常,运动模型根据其复杂程度,分成 2 参数的平移模型和 6 参数的仿射模型。2 参数的平移模型主要针对对象在运动过程中不会发生形变的情况,其参数估计实现非常简单。而 6 参数的仿射模型可以模拟对象的形变,如扭曲、伸缩和变形等,这种模型的参数估计算法实现复杂。在本算法中,Snake 模型可以很好地处理对象的形变,在此采用最简单的 2 参数平移模型,减少运算复杂度。2 参数平移模型的变形方程如下:

$$\begin{aligned} (x_k^n, y_k^n)^T &= T_i (x_k^{n+1}, y_k^{n+1})^T = \\ & (x_k^{n+1}, y_k^{n+1})^T + (b_1, b_2)^T \end{aligned} \quad (4)$$

式中  $T_i$  是平移矩阵;  $(b_1, b_2)$  是平移偏移量。根据前面建立的特征点匹配关系,分别计算其相应的特征点平移偏移量参数。再利用参考文献[11]中的模糊 C 均值聚类方法实现平移偏移量参数的聚类工作。聚类后的每个类对应于一个视频对象,而其类中心就是对象的平移偏移量,可以根据这个平移偏移量,直接把前一帧所跟踪的对象平移到当前帧的相应位置,再对其进行下一步的 Snake 跟踪。

## 2 Snake 视频对象跟踪

选取了 Chan-Vese<sup>[12]</sup>模型为 Snake 跟踪模型,此

模型对视频跟踪建模成下面的两个能量积分之和,其公式如下:

$$\begin{aligned} F_1(C) + F_2(C) = & \int_{\text{inside}(C)} |u_0(x, y) - c_1|^2 dx dy + \\ & \int_{\text{outside}(C)} |u_0(x, y) - c_2|^2 dx dy \end{aligned} \quad (5)$$

式中  $C$  表示可变的曲线;常数  $c_1$  和  $c_2$  依赖于  $C$  的取值,  $c_1$  表示前一帧视频图像中可变曲线  $C$  包含的所有内部点的像素亮度均值,  $c_2$  则表示前一帧可变曲线  $C$  外区域的像素亮度均值;  $u_0(x, y)$  表示位置为  $(x, y)$  点处的像素亮度值。最简单的情况,当对象完全跟踪上后,其能量之和应该最小。

$$\inf_C \{F_1(C) + F_2(C)\} \approx 0 \approx F_1(C_0) + F_2(C_0) \quad (6)$$

如图 1 所示,可以看出,曲线在跟踪对象的外面时,  $F_1(C) > 0$ , 而  $F_2(C) \approx 0$ 。如果曲线跟踪到对象的内部时,  $F_1(C) \approx 0$ , 而  $F_2(C) > 0$ 。当曲线一部分在内部,一部分在外部时,  $F_1(C) > 0$ , 且  $F_2(C) > 0$ 。只有当完全跟踪上时,  $F_1(C) \approx 0$ , 且  $F_2(C) \approx 0$ 。从而两者之和才最小。

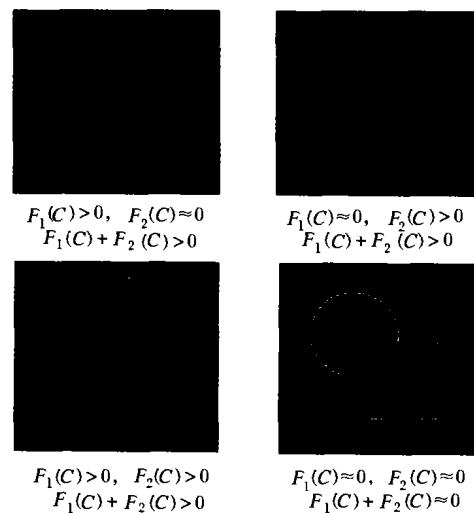


图 1 不同位置的曲线所对应的能量之和关系图

Fig.1 Total Snake's energy of the curve in different position

在实际的 Snake 模型实现中,还增加了一些约束量,如曲线的长度和曲线所围的内部面积。实际的能量函数定义如下:

$$F_1(C) + F_2(C) = \mu \text{Length}(C) + \gamma \text{Area}(\text{inside}(C)) +$$

$$\lambda_1 \int_{\text{inside}(C)} |u_0(x,y) - c_1|^2 dx dy + \lambda_2 \int_{\text{outside}(C)} |u_0(x,y) - c_2|^2 dx dy \quad (7)$$

式中  $\mu \geq 0; \gamma \geq 0; \lambda_1, \lambda_2 > 0$ 。在具体的数值求解过程中, 设定  $\gamma = \lambda_1 = \lambda_2 = 1, \mu = 0$ 。能量函数关系式(7)通过 Euler-Lagrange 方程求解<sup>[12]</sup>, 最终的曲线叠代演化方程如下:

$$\begin{cases} \frac{\partial \phi}{\partial t} = [-\gamma - \lambda_1 (u_0 - c_1)^2 + \lambda_2 (u_0 - c_2)^2] N \\ \phi(0, x, y) = \phi_0(x, y) \end{cases} \quad (8)$$

式中  $\phi_0(x, y)$  是初始叠代时的分布函数, 轮廓上的点, 其  $\phi$  值为 0; 轮廓内的点, 其  $\phi$  值大于 0; 而轮廓外的点, 其  $\phi$  值小于 0, 采用 Level Set 方法<sup>[3]</sup>实现曲线演化方程(8), 其数值解可见参考文献[12]。

### 3 实验结果和讨论

图 2 是运动估计性能评估图, 图(a)和(b)分别是

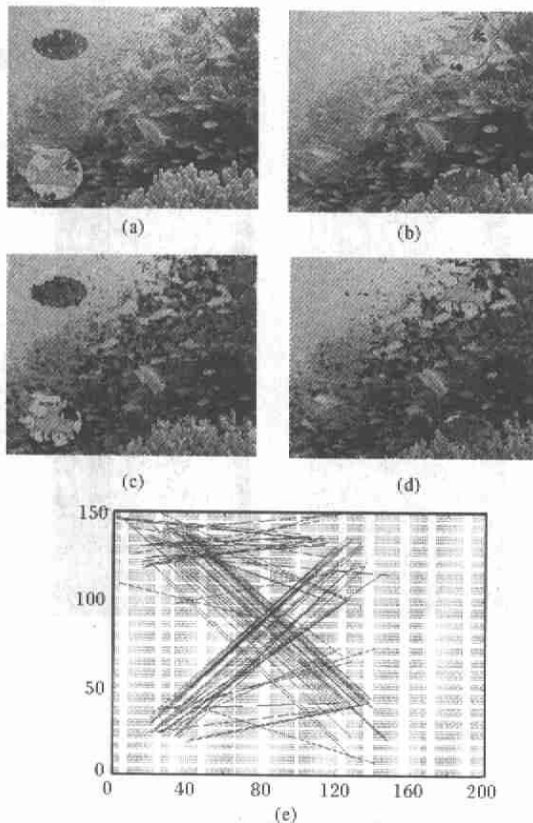


图 2 运动估计性能评估图

Fig.2 Motion estimation performance

测试视频源的视频帧  $I^n, I^{n+1}$ 。其中有两个特殊加入的视频对象, 一个是从左下角移到右上角, 另一个是从左上角移到右下角。如果只用 Snake 算法进行跟踪,

其曲线演化过程将陷入局部最小。图(c)和(d)中的黑点是其对应的特征点。由于图中纹理结构较复杂, 所以特征点相应也比较多。经过特征点匹配和互验证后, 其特征点的对应运动矢量如图(e)所示。在图(e)中, 两股明显的直线分别对应视频对象真实运动信息。由于场景比较复杂, 所以仍然还有一些错误的匹配关系, 最终通过模糊 C 均值聚类方法得到很好的改善。表 1 是模糊 C 均值聚类后的运动矢量和真实运动矢量对照表, 从表中可以看出, 运动的预估计还是比较准确的。

表 1 实际的运动矢量和预估计的运动矢量对照表

Tab.1 Real motion vector and estimated motion vector

|          | Real motion vector(pixels) | Estimated motion vector(pixels) |
|----------|----------------------------|---------------------------------|
| Object 1 | (200, -197)                | (190.7, -183.8)                 |
| Object 2 | (202, 200)                 | (198.4, 194.3)                  |

图 3 是室外的监控场景视频对象跟踪结果, 图(a)中最中间的那辆白车就是所监控跟踪的对象。图(a)是视频的初始帧, 在这个初始帧中, 其初始轮廓是人为设定的。图(b)~(f)分别对应视频序列中的第 5 帧、第 10 帧、第 15 帧、第 20 帧和第 25 帧的跟踪效果图。从这些图中可以看出, 此跟踪算法可以很好地适应车子的快速移动, 并且跟踪效果也比较好。

最后比较一下采用运动预估计 Snake 视频跟踪算法和普通 Snake 视频跟踪算法所对应的曲线演化

表 2 Level Set 曲线演化叠代次数对照表

Tab.2 Evolutional times of the curve using Level Set methods

|                           | Frame 5 | Frame 10 | Frame 15 | Frame 20 | Frame 25 |
|---------------------------|---------|----------|----------|----------|----------|
|                           | /time   | /time    | /time    | /time    | /time    |
| Without motion estimation | 49      | 56       | 47       | 72       | 55       |
| With motion estimation    | 8       | 11       | 8        | 13       | 9        |

过程中叠代次数的情况。这两种跟踪算法都是采用 Level Set 方法。表 2 是采用运动预估计和普通的视

频监控算法的叠代次数对照表。从表 2 可以清楚地看到,采用了运动估计后的 Level Set 方法中曲线演化过程的叠代次数明显少于不采用运动估计的叠代次数。

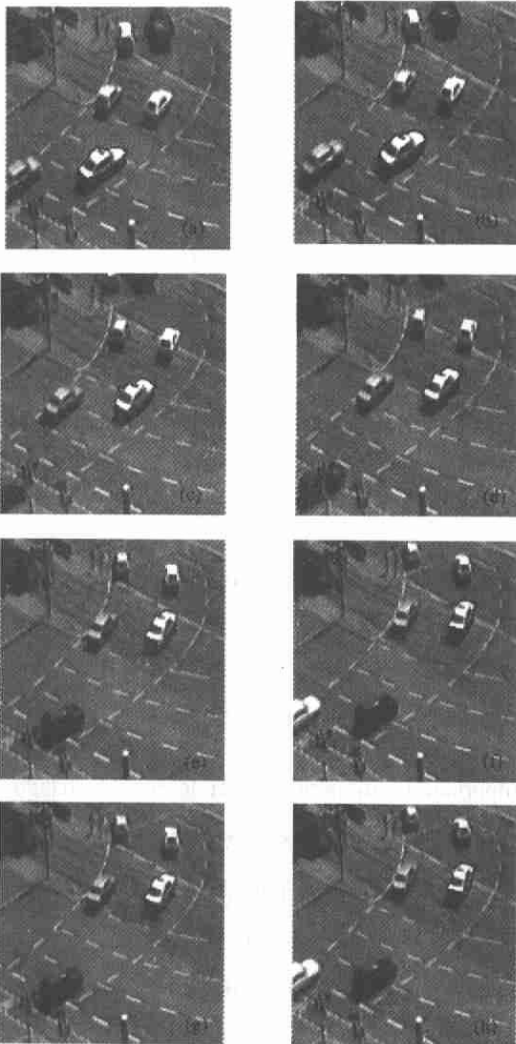


图3 实际室外监控场景中视频对象跟踪效果图

Fig.3 Video object tracking results in the real outdoor environment

#### 4 结束语

提出了一种基于运动补偿的 Snake 视频跟踪算法,此算法首先对所跟踪的视频对象进行运动估计,然后根据估计的结果对视频对象的位置和大小进行运动补偿,最后再利用 Snake 模型进行变形跟踪。实验证明,对跟踪对象进行运动估计和运动补偿处理后,跟踪效果得到了改善,并且 Snake 模型的曲线变形叠代次数也减少了很多。虽然提出的算法在对象跟踪方面有很好的效果,但是由于 Snake 模型主要的缺点是算法复杂、运算时间长,如果应用到实时

的视频对象跟踪中還有一定的距离,所以下一步的目标是如何提高 Snake 模型的运算速度和研究快速 Snake 跟踪算法。

#### 参考文献:

- [1] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models [A]. Proc First Intl Conf Computer Vision[C]. 1987. 259-268.
- [2] Caselles V, Kimel R, Sapiro G. Geodesic active contours[J]. IJCK, 1997, 22(1): 61-79.
- [3] Osher S, Sethian J A. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulation[J]. J Comput Phys, 1988, 79: 12-49.
- [4] Gastaud M, Barlaud M, Aubert G. Tracking video objects using active contours [A]. Proceedings of the Workshop on Motion and Video Computing[C]. 2002. 90-95.
- [5] Mansouri A R. Region tracking via level set PDEs without motion ion computation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 947-961.
- [6] Paragios N, Deriche R. Geodesic active contours and level sets for the detection and tracking of moving objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(3): 266-280.
- [7] Harris C, Stephens M. A combined corner and edge detector[A]. Proceeding of the 4th Alvey Vision Conference[C]. 1988. 147-151.
- [8] Schmid C, Mohr R, Bauckhage Ch. Comparing and evaluating interest points[A]. Proc IEEE Int Conf Computer Vision[C]. 1998. 230-235.
- [9] Konrad J, Lan Z D. Dense disparity estimation from feature correspondences[A]. Proc SPIE Stereoscopic Displays and Virtual Reality Systems VII[C]. 2000. 3957. 90-101.
- [10] Mansouri A R, Konrad J. Multiple motion segmentation with Level Sets[J]. IEEE Trans on Image Processing, 2003, 12(2): 201-220.
- [11] ARee S C, Lursinsap C, Sophasathit P, et al. Fuzzy C-Mean: a statistical feature classification of text and image segmentation method[J]. Intel Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 2001, 9(6): 661-671.
- [12] Chan T F, Vese L A. Active contours without edges[J]. IEEE Trans On Image Processing, 2001, 10(2): 266-276.