

## 基于粒子群的关键帧提取算法

张建明, 蒋兴杰, 李广翠, 姜 靛

(江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013)

(jxj40@163.com)

**摘 要:** 关键帧提取是基于内容的视频检索中的重要一步, 为了能够有效地提取出不同类型视频的关键帧, 提出一种基于粒子群的关键帧提取算法。该方法首先提取出视频中每帧的全局运动和局部运动特征, 然后通过粒子群算法自适应地提取视频关键帧。实验结果表明, 采用该算法对不同类型的视频提取出的关键帧具有较好的代表性。

**关键词:** 视频检索; 关键帧提取; 粒子群; 运动特征

**中图分类号:** TP391.41 **文献标志码:** A

## Key frame extraction based on particle swarm optimization

ZHANG Jian-ming, JIANG Xing-jie, LI Guang-cui, JIANG Liang

(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang Jiangsu 212013, China)

**Abstract:** Key frame extraction was an important step in video retrieval. In order to effectively extract key frames of different video types, a key frame extraction algorithm based on particle swarm was proposed in this paper. This method first extracted the global motion and local motion features in each frame, and video key frame was extracted by Particle Swarm Optimization (PSO) adaptively. The experimental results show that the key frame extraction algorithm for different types of video is more representative.

**Key words:** video retrieval; key frame extraction; Particle Swarm Optimization (PSO); motion characteristic

### 0 引言

近年来,随着多媒体技术的快速发展,产生了大量的视频数据。如何将这海量视频数据进行有效的归类和管理成为一个重要的研究课题。利用关键帧代表视频片段是一种常用的内容压缩和处理方法<sup>[1]</sup>,关键帧质量的好坏直接影响到视频检索的效果。对于每一个镜头而言,内部帧之间通常有比较大的信息冗余,这对于存储和管理视频很不利,所以选取的关键帧必须有代表意义,能够反映出镜头中的主要事件但是又不能太多。目前关键帧提取的主要方法包括以下4类。

1) 基于镜头边界的方法<sup>[2]</sup>。该类方法通常选择固定位置提取关键帧,比如每个镜头的第一帧、中间帧和末尾帧等。最大的优点是算法简单,操作容易;但缺点也比较明显,获得的关键帧代表性不强。

2) 基于运动分析提取关键帧<sup>[3]</sup>。典型算法是光流法,该方法能很好地表达视频内的全局性运动,如镜头平移、摇动等,但是忽略了更为重要的局部运动。

3) 基于图像信息提取关键帧<sup>[4]</sup>。一般是通过图像的颜色直方图等特征来描述视频变化,该方法相对于第二类方法计算复杂度有比较好的改善,但是运动描述能力有所下降,并且有可能对完全不同的图片产生相同的直方图,给关键帧提取带来一定的麻烦。

4) 基于聚类提取关键帧<sup>[5-6]</sup>。该类方法把视频中每一帧看做聚类算法中的一个点,提取特征如颜色直方图、光流、纹理等,然后用聚类算法计算聚类中心作为关键帧。该类方法最大的缺陷是聚类个数也就是关键帧个数不容易确定。

针对以上问题,提出了基于粒子群的关键帧提取算法,采用全局运动特征与局部运动特征作为视频的整体特征,同时提取典型视频的相应特征,通过粒子群算法自适应地提取视频关键帧。

### 1 粒子群算法

粒子群 (Particle Swarm Optimization, PSO) 算法是美国社会心理学家 James Kennedy 和电气工程师 Russell Eberhart 于 1995 年提出的一种演化计算技术,最初是为了模拟鸟类的捕食过程,后来发现 PSO 是一种非常好的优化算法。算法首先初始化一群随机解  $x[i]$ , 这些随机解称为粒子,每个粒子都有自己的速度  $v[i]$ , 由适应度函数  $f$  和目标向量  $A$  寻求粒子的个体最优位置  $b_g[i]$ , 同时寻求整个粒子群的最优位置  $p_g$ , 在找到了这两个最优位置以后粒子群通过下面的公式来更新自己的速度和位置。

$$v[i, j] = v[i, j - 1] + c_1 \text{rand}() (p_g - x[i, j - 1]) + c_2 \text{rand}() (b_g[i] - x[i, j - 1]) \quad (1)$$

$$x[i, j] = x[i, j - 1] + v[i, j] \quad (2)$$

其中:  $v[i, j]$  表示第  $i$  个粒子第  $j$  次迭代的速度,  $c_1, c_2$  是两个学习因子为常数,  $\text{rand}(\cdot)$  为一个随机数函数,取值在 0 到 1 之间,  $x[i, j]$  是第  $i$  个粒子在第  $j$  次迭代后的位置。一次执行完以后再通过适应度函数和目标函数更新各粒子位置和速度,重复上述迭代直至达到最大循环次数或者满足要求。

### 2 特征向量的提取

考虑到运动是导致关键帧提取的主要原因和视频运动

收稿日期: 2010-07-20; 修回日期: 2010-09-11。

基金项目: 国家自然科学基金资助项目 (60673190); 江苏省自然科学基金资助项目 (BK2009199)。

作者简介: 张建明 (1964-), 男, 江苏镇江人, 教授, 博士, 主要研究方向: 虚拟现实、图像处理、模式识别; 蒋兴杰 (1983-), 男, 江苏南京人, 硕士研究生, 主要研究方向: 图像处理、模式识别; 李广翠 (1986-), 女, 山东临沂人, 硕士研究生, 主要研究方向: 图像处理、模式识别; 姜靛 (1985-), 男, 江苏镇江人, 硕士研究生, 主要研究方向: 图像处理、模式识别。

的特殊性,将运动划分为两类:全局运动和局部运动。同时为了方便对不同类视频进行处理,首先要对视频图像进行归一化处理,归一化以后帧图像的大小为  $Row \times Col$ 。

### 2.1 全局运动特征的提取

全局运动主要是由于摄像机运动导致的,比如摄像机平移、旋转、拉伸等。摄像机运动给视频留下的痕迹就是光流<sup>[7]</sup>,因此用光流作为视频图像帧的全局运动特征是比较合适的。考虑到光流算法比较复杂,如果计算整幅图像的光流耗时严重,而且在大多数视频中左上角区域出现局部运动的情况比较少,故只提取该区域的光流用以代表整幅图像的全局运动,如图1所示。

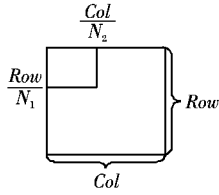


图1 全局运动特征提取区域

图1中,  $N_1$  和  $N_2$  是两个常数因子(本文都取10),用以确定特征区域的大小。由于光流法耗时严重,本文选择 Lucas & Kanade 算法,该方法具有计算速度快、应用简单的特点<sup>[8-9]</sup>。应用该光流法提取视频第  $k$  帧的全局运动特征  $Gd(k)$ 。

$$Gd(k) = \frac{1}{N_1 \times N_2} \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} gd(i, j) \quad (3)$$

其中:  $gd(i, j)$  表示像素点  $(i, j)$  处的光流大小。第  $k$  帧和第  $k+1$  帧之间的全局运动特征的距离用  $GD(k, k+1)$  表示,定义如下:

$$GD(k, k+1) = |Gd(k) - Gd(k+1)| \quad (4)$$

### 2.2 局部运动特征的提取

局部运动通常在图像的中间区域出现,一般是由于人物或者物体运动导致的,包含有大量的信息,所以关注度比较高;但局部运动形式复杂多样,用光流法来表示是非常困难的,而颜色直方图算法简单,却可以在一定程度上描述这样的变化。HSV 颜色空间符合人眼视觉特性,能较好反映人对色彩的感知和鉴别,能够分离颜色、亮度及饱和度信息,与人的视觉感知直接对应,非常适合于图像处理<sup>[10]</sup>,本文采用该空间模型计算局部运动特征。将色调空间  $H$ 、饱和度  $S$ 、亮度  $V$  空间按照人眼的颜色感知进行非等间隔划分<sup>[5]</sup>,分别划分成8份、3份、3份。这样可以将 HSV 颜色空间划分成72维子空间,通过式(5)把  $(H, S, V)$  合成一维向量。

$$I = 9H + 3S + V \quad (5)$$

在合成的一维向量中统计第  $k$  帧图像在各个子空间上的像素个数就得到了该帧图像的颜色直方图向量  $I[k, 0] \sim I[k, 71]$ , 即该帧图像的局部运动特征。其中  $I[k, 0]$  表示第  $k$  帧图像的颜色直方图向量中第0个分量的值,由于局部运动通常只发生在图像的中间区域,所以没有必要计算整幅图像的颜色直方图只需取出中间区域即可,如图2所示。

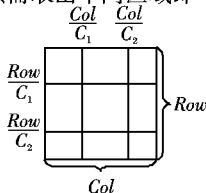


图2 局部运动特征提取区域

图2中  $C_1$  和  $C_2$  是两个常数因子,用以确定特征区域的大小,本文分别取4和4/3,即所取的中间区域占整个图像面积的1/4。第  $k$  帧和第  $k+1$  帧之间的局部运动特征距离用  $PD(k, k+1)$  表示,定义如下:

$$PD(k, k+1) = \frac{1}{\left(\frac{Col}{C_2} - \frac{Col}{C_1}\right) + \left(\frac{Row}{C_2} - \frac{Row}{C_1}\right)} \times \sum_{i=0}^{71} (I[k, i] - I[k+1, i])^2 \quad (6)$$

## 3 基于粒子群的关键帧提取

通常情况下,一段视频包括多个镜头。关键帧提取的目的是提取出能够代表视频主要内容的一帧或多帧图像,从表达意义角度考虑,视频关注局部运动更多,因为局部运动表达的都是相对重要的物体或者人物的运动,含有高层语义。本文算法中特征向量的形式为  $X[i] = (R_1, R_2, S_1, S_2, F)$ , 其中,  $F$  表示间隔帧数,为上一个关键帧其在视频序列中位置为  $m$  和下一个关键帧其在视频序列中位置为  $n$  之间的帧图像数。

$$R_1 = \frac{1}{n-m} \sum_{i=0}^{n-m-1} PD(m+i, m+i+1) \quad (7)$$

其中:  $R_1$  表示两个关键帧之间局部运动的平均值大小,该值越大表示视频局部运动越明显。

$$R_2 = \frac{1}{n-m} \sum_{i=0}^{n-m-1} |R_1 - PD(m+i, m+i+1)| \quad (8)$$

其中:  $R_2$  表示局部运动波动情况,该值越大说明视频内局部运动的波动性越大。

$$S_1 = \frac{1}{n-m} \sum_{i=0}^{n-m-1} GD(m+i, m+i+1) \quad (9)$$

其中:  $S_1$  表示两个关键帧之间全局运动的平均值大小,该值越大说明视频全局运动越明显。

$$S_2 = \frac{1}{n-m} \sum_{i=0}^{n-m-1} |S_1 - GD(m+i, m+i+1)| \quad (10)$$

其中:  $S_2$  表示全局运动的波动情况,该值越大说明视频内全局运动的波动性就越大。

在进行关键帧提取实验之前本文首先提取了5类典型视频的特征向量。第1类,全局运动和局部运动都不明显的视频特征向量,  $X[1] = (112, 108, 14, 3, 542)$ ; 第2类,全局运动较为明显,局部运动不明显的视频的特征向量,  $X[2] = (168, 123, 34, 17, 320)$ ; 第3类,全局运动较为明显,局部运动也较为明显的视频特征向量,  $X[3] = (390, 270, 59, 28, 169)$ ; 第4类,全局运动较为明显,局部运动非常明显的视频特征向量,  $X[4] = (895, 504, 107, 40, 107)$ ; 第5类,全局运动和局部运动都非常明显的视频特征向量,  $X[5] = (2108, 1309, 204, 75, 21)$ 。

### 3.1 视频预处理

获得了典型视频特征向量之后,对待检测视频进行归一化处理,归一化结束以后视频帧的大小为  $320 \times 240$ 。提取视频内所有图像的局部运动特征和全局运动特征,取适应度函数如式(11)所示。

$$f_i = \beta_1 \left( \alpha_1 \frac{R_{1Xi}}{R_{1A}} + \alpha_2 \frac{R_{2Xi}}{R_{2A}} \right) + \beta_2 \left( \alpha_1 \frac{S_{1Xi}}{S_{1A}} + \alpha_2 \frac{S_{2Xi}}{S_{2A}} \right) \quad (11)$$

其中:  $f_i$  表示特征向量  $A$  与  $X[i]$  的适应度,该值在0到1之间,越大表示相似度越高。  $R_{1Xi}$  表示  $X[i]$  向量中  $R_1$  分量的值,其余的类似。  $\alpha, \beta$  均为权值,且  $\alpha_1 + \alpha_2 = 1, \beta_1 + \beta_2 = 1$ 。由于局部运动和运动均值更能反映出视频的主要内容,所以

往往关注度比较高,本文取  $\alpha_1 = \beta_1 = 0.75$ ,  $\alpha_2 = \beta_2 = 0.25$ ,

若  $\frac{R_{1Xi}}{R_{1A}} \cdot \frac{R_{2Xi}}{R_{2A}} \cdot \frac{S_{1Xi}}{S_{1A}} \cdot \frac{S_{2Xi}}{S_{2A}}$  有大于 1 的则取倒数。

3.2 关键帧提取

单个关键帧的提取按以下步骤进行:

1) 以上一个关键帧的位置  $m$  为起始帧,计算  $m + 1, \dots, m + k$ , 共  $k$  帧图像的特征向量记为  $A$ ,本文中  $k$  取值为 10,如果不足 10 帧则算法结束。 $k$  的取值过小将导致图像特征向量不具有代表性,如果过大又会导致算法耗时严重。以典型视频特征向量  $X[1] \sim X[5]$  为粒子群初始粒子。各粒子初始速度为 0,初始最优位置为各粒子的当前位置。

2) 由适应度函数分别计算目标向量  $A$  和各粒子  $X[1] \sim X[5]$  的适应度,取适应度最大值  $f_i$  对应的粒子  $X[i]$  为粒子群最优解  $p_g$ 。取最优解向量对应的间隔帧数  $F_{X[i]}$  为目标向量  $A$  的间隔帧数  $F_A$ 。

4) 如果粒子群最优解  $X[i]$  和目标向量  $A$  的适应度  $f_i$  大于给定值本文取 0.98,则认为最优解与目标向量足够相似,下一个关键帧位置即为起始帧位置加上目标向量间隔帧数  $F_A$  对应的那一帧,如果该位置大于视频序列长度则算法结束,并且取消最后一个关键帧,否则转 5)。

5) 由当前全局最优解和各粒子的局部最优解计算各粒子的下一次运行速度,同时计算各粒子的下一个位置并且更新向量  $X[1] \sim X[5]$ ,转 2)。

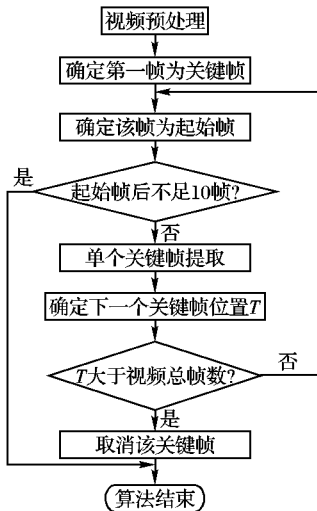


图 3 关键帧集合确定流程

4 实验结果及分析

本文实验运行于 P4 4 GHz 台式机上,内存为 1 GB,操作系统为 Windows XP,在 VC6.0 环境下实现。为了验证算法的有效性,选取了不同类型的视频,包括:足球、新闻、电影、演讲、MTV,视频帧序列从几十帧到几万帧不等,基本情况如表 1 所示。

表 1 实验数据

视频	帧数	本文算法提取出的关键帧个数
足球	4 536	112
新闻	7 847	76
电影	43 876	454
演讲	2 474	12
MTV	604	54

下面以 3 段具体视频镜头为例,说明本文算法有效性,分别为:全局运动较小的镜头、全局运动和局部运动都较大的镜头以及全局运动和局部运动都很小的镜头。

首先以电影《拆弹部队》中一段全局运动比较小的镜头为例,该镜头共有 68 帧,描述的是一颗弹壳落地的过程。图 4、5 分别为传统颜色直方图提取的关键帧和应用本文算法提取出的关键帧。

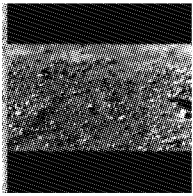


图 4 传统颜色直方图提取的弹壳落地关键帧

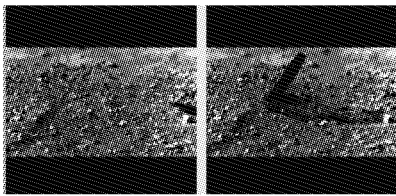


图 5 本文算法提取的弹壳落地关键帧

由图 4、5 可见,传统颜色直方图提取出的关键帧只反映出了背景的情况,并没有反映出弹壳落地的过程,即没有反映出镜头的主要内容。这是由于传统颜色直方图只考虑整个镜头颜色值的变化,而在该镜头中弹壳落地过程颜色值变化过于缓慢,导致传统方法提取特征时并不能反映出这样缓慢的颜色值变化,从而不能提取出表示弹壳落地的关键帧。而本文算法由于更加注重局部运动,虽然弹壳落地过程颜色值变化缓慢,但由于在权值分配上本文注重局部运动,所以提取的特征能够体现出该颜色值的变化,因此提取出了反映弹壳落地过程的关键帧。

再以电影《拆弹部队》中一段全局和局部运动都比较大的镜头为例,该镜头共有 530 帧。描述的是从爆炸开始到结束的整个过程中,一个人从站立—倒地—静止的状态变化。图 6、7 分别为传统颜色直方图提取的关键帧和应用本文算法提取出的关键帧。

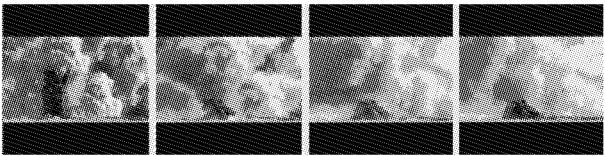


图 6 传统颜色直方图提取的爆炸关键帧

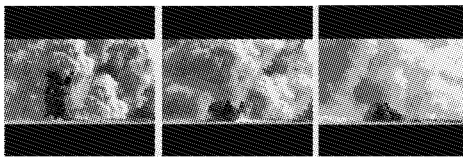


图 7 本文算法提取的爆炸关键帧

由图 6、7 可见,传统颜色直方图提取出的关键帧虽然能够反映出视频的主要内容,但是图 6 中第二和第三个关键帧反映的事件都是人物的倒地过程,存在冗余。这是由于在该镜头中全局运动相对局部运动更为激烈,而传统颜色直方图法并没有区分局部运动和全局运动的重要性,导致全局运动

过大时提取了过多的关键帧,造成了冗余。而本文算法在考虑全局运动和局部运动的同时更加注重局部运动,即使全局运动过大,由于在权值分配上全局运动所占比重较小,从而能够降低全局运动带来的干扰,使得本文算法提取出的关键帧不仅反映出视频的主要内容,而且没有冗余。

最后以一段全局和局部运动都很小的视频为例,该视频为2008CCTV杯英语演讲大赛视频中的一个镜头,长度为214帧,描述的是一个人物的演讲过程。图8、9分别为传统颜色直方图提取的关键帧和应用本文算法提取出的关键帧。

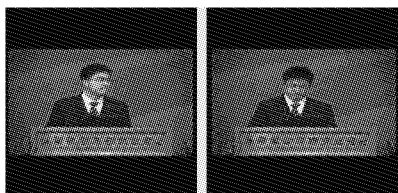


图8 传统颜色直方图提取的演讲关键帧



图9 本文算法提取的演讲关键帧

由图8、9可见,传统颜色直方图提取了两个关键帧来描述人物演讲这一事件,存在冗余,这是由于传统颜色直方图在计算颜色特征时,存在一个差异累加过程从而导致了对于这类运动小但是镜头长度长的视频提取了过多的关键帧。而本文算法采用的是运动信息提取关键帧,由于该镜头中全局和局部运动都很小,本文算法只提取了一帧关键帧,不仅反应出了该视频的主要内容且没有冗余,克服了传统颜色直方图算法中差异累加带来的缺陷。

通过以上结果可以看出,本文算法由于同时考虑了全局运动和局部运动并且更加突出局部运动,相对于传统颜色直方图过分强调全局运动特征,检测结果有了很好的改善,且不采用差异累加,对运动量小镜头长度长的镜头检测效果也令人满意。

## 5 结语

本文提出了一种可以有效地提取出不同类型视频关键帧的新方法,该方法利用视频本身的运动信息,比较准确地提取出了能够反映视频主要内容的关键帧,并且弥补了单一运动信息对视频内容描述不足的问题。同时不采用差异累加,因此能够很有效地解决关键帧冗余问题。

### 参考文献:

- [1] 蒋鹏,秦小麟. 基于视觉注意模型的自适应视频关键帧提取[J]. 中国图象图形学报, 2009, 14(8): 1650-1655.
- [2] ZHANG HONGJIANG, SMOLIAR S W, WU J H. Content based video browsing tools[C]// Proceedings of the SPIE Multimedia Computing and Networking. Washington, DC: IEEE Computer Society, 1995: 389-398.
- [3] WOLF W. Key frame selection by motion analysis[C]// Proceedings of the 1996 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC: IEEE Computer Society, 1996: 1228-1231.
- [4] SUN ZHONGHUA, JIA KEBIN, CHEN HEXIN. Video key frame extraction based on spatial-temporal color distribution[C]// International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Washington, DC: IEEE Computer Society, 2008: 196-199.
- [5] 王方石,须德,吴伟鑫. 基于自适应阈值的自动提取关键帧的聚类算法[J]. 计算机研究与发展, 2005, 42(10): 1752-1757.
- [6] 张婵,高新波,姬红兵. 视频关键帧提取的可能性 C-模式聚类算法[J]. 计算机辅助设计与图形学学报, 2005, 17(9): 2040-2045.
- [7] 贺玉文,赵黎,钟玉琢,等. 快速鲁棒的全局运动估计算法[J]. 软件学报, 2001, 12(8): 1220-1228.
- [8] BARRON J L, FLEET D J, BEACHEMIN S S. Performance of optical flow techniques[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 1992: 236-242.
- [9] LUCAS B D. Generalized image matching by the method of differences[D]. Pittsburgh: Carnegie Mellon University, 1985.
- [10] 陈允杰,张建伟,韦志辉,等. 基于 HSV 颜色空间的中国虚拟人脑图像自动分割方法[J]. 计算机研究与发展, 2007, 44(12): 2036-2043.

(上接第340页)

动距离,以及降低网络均匀度等方面,优于基本粒子群算法,能有效地实现使用有限的节点达到较高覆盖性能的目的。

### 参考文献:

- [1] 于宏毅,李欧,张效义. 无线传感器网络理论、技术与实现[M]. 北京: 国防工业出版社, 2008.
- [2] HUANG C-F, TSENG Y-C. The Coverage problem in a wireless sensor network [J]. Mobile Networks and Applications, 2005, 10(4): 519-528.
- [3] CARDEI M, WU J. Energy-efficient coverage problems in wireless Ad Hoc sensor networks [J]. Computer Communications, 2006, 29(4): 413-420.
- [4] WANG BANG, BENG L H, MA DI. A survey of movement strategies for improving network coverage in wireless sensor network [J]. Computer Communications, 2009, 32(13/14): 1427-1436.
- [5] 何璇,郝群,宋勇. 一种移动无线视频传感器节点的覆盖算法[J]. 传感技术学报, 2009, 22(8): 1163-1168.
- [6] WANG XUE, WANG SHENG. Dynamic deployment optimization in wireless sensor networks [C]// Intelligent Control and Automation, LNCS 344. Berlin: Springer, 2006: 182-187.
- [7] 林祝亮,冯远静. 基于粒子群算法的无线传感网络覆盖优化策略[J]. 计算机仿真, 2009, 26(4): 190-193.
- [8] LI SHIJIAN, XU CONGFU, PAN WEIKE. Sensor deployment optimization for detecting maneuvering targets [C]// Proceedings of 7th International Conference on Information Fusion. Washington, DC: IEEE Computer Society, 2005: 1629-1635.
- [9] 刘丽萍,王智,孙优贤. 无线传感器网络部署及其覆盖问题研究[J]. 电子与信息学报, 2006, 28(9): 1752-1757.
- [10] 李兵,蒋尉孙. 混沌优化方法及其应用[J]. 控制理论与应用, 1997, 14(4): 613-615.
- [11] 程志刚,张立庆. 基于 Tent 映射的混沌混合粒子群优化算法[J]. 系统工程与电子技术, 2007, 29(1): 103-106.