

文章编号:1001-9081(2010)06-1573-04

## 基于改进 Mean-Shift 与自适应 Kalman 滤波的视频目标跟踪

周尚波<sup>1,2</sup>, 胡 鹏<sup>1</sup>, 柳玉炯<sup>1</sup>

(1. 重庆大学 计算机学院, 重庆 400044;

2. 重庆市/信息产业部计算机网络与通信技术重点实验室, 重庆 400065)

(shbzhou@cqu.edu.cn)

**摘 要:**提出一种改进的 Mean-Shift 和自适应 Kalman 滤波器相结合的视频运动目标跟踪算法。对选定的跟踪目标,采用三帧差和区域增长法分割目标并得到主颜色信息。在跟踪过程中,利用自适应的 Kalman 滤波器估计每一帧的起始迭代位置,再利用改进的 Mean-Shift 算法得到跟踪位置并作为测量值反馈给自适应 Kalman 滤波器,并引入遮挡率因子以自适应地调节 Kalman 估计参数。实验结果表明,该算法能对视频中的运动目标实现检测和连续跟踪,对遮挡也有较好的鲁棒性。

**关键词:**Mean-Shift 算法;Kalman 滤波;目标跟踪;遮挡

**中图分类号:**TP391.9 **文献标志码:**A

## Target tracking based on improved Mean-Shift and adaptive Kalman filter

ZHOU Shang-bo<sup>1,2</sup>, HU Peng<sup>1</sup>, LIU Yu-jiong<sup>1</sup>

(1. College of Computer Science, Chongqing University, Chongqing 400044, China;

2. Chongqing City / Ministry of Information Industry Computer Networks and Communications Technology Laboratory, Chongqing 400065, China)

**Abstract:** In this paper, a target tracking algorithm was proposed by combining the improved Mean-Shift algorithm with the adaptive Kalman filter. For a selected moving object, frame difference and region growing methods were used to segment target, and the dominant color was extracted. In the tracking process, the initial iterative position was obtained by adaptive Kalman filter in every frame, and the tracking result obtained by the improved Mean-Shift was fed back to the adaptive Kalman filter as the measurement for correction. The estimate parameters of adaptive Kalman filter were adjusted by occlusion ratio adaptively. The experimental results demonstrate that the proposed algorithm can detect and track the moving object consecutively in video and has better robustness to occlusion.

**Key words:** Mean-Shift algorithm; Kalman filter; target tracking; occlusion

### 0 引言

在计算机视觉的应用中,运动目标的识别和跟踪是一项重要的研究课题,在诸如人机交互、视频监控、智能建筑、司机助手和医学图像处理等领域得到了广泛的应用<sup>[1]</sup>。国内外有许多学者对这个领域进行了研究,提出了许多有效的目标跟踪方法,主要有基于运动模型的<sup>[2]</sup>、基于光流的<sup>[3]</sup>、基于特征的<sup>[4-6]</sup>跟踪等几类方法,另外也有一些方法是基于图像区域信息的<sup>[5]</sup>。其中,基于特征的跟踪方法利用了目标特征的位置变化来跟踪目标,常用的目标特征有颜色、纹理、角点、边缘和轮廓、团块和区域等<sup>[4]</sup>。与其他方法相比,基于特征的跟踪方法的实时性和鲁棒性较好,应用最为广泛。

基于特征的跟踪方法的典型代表是基于颜色特征的 Mean-Shift 算法。文献[6]把 Mean-Shift 算法成功地运用到目标跟踪领域。该算法作为一种高效的模式匹配算法,基于目标区域的灰度或色彩分布特征,利用梯度优化方法实现目标快速定位,能够对目标实时跟踪,对目标的变形、旋转等运动有较好的适用性,但是 Mean-Shift 算法在目标跟踪过程中没有利用目标在空间中的运动方向和运动速度信息,当周围环

境存在干扰或者目标持续地受到其他物体遮挡时,跟踪窗口将不能正确对目标进行跟踪,甚至会产生误跟踪<sup>[4]</sup>。本文在研究传统 Mean-Shift 目标跟踪算法的基础上,提出了一种改进的 Mean-Shift 算法进行运动目标跟踪。使用自适应的 Kalman 滤波器处理遮挡问题,并引入了一个遮挡率因子来动态调整 Kalman 滤波器的参数,使改进后的 Mean-Shift 算法具有对目标物体后继状态进行预测估计的能力,从而在目标发生短时间的遮挡后依然可以对目标进行准确跟踪。

### 1 Mean-Shift 结合 Kalman 滤波仿真

本文算法主要包括3个步骤:1)初始化、分割运动目标;2)提取运动目标的特征;3)使用改进的 Mean-Shift 算法结合自适应的 Kalman 滤波器进行目标跟踪。系统结构见图1。

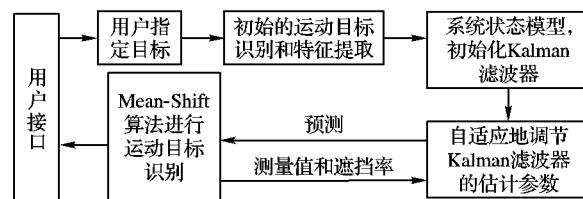


图1 系统结构

收稿日期:2009-12-22;修回日期:2010-03-09。

基金项目:重庆市/信息产业部计算机网络与通信技术重点实验室资助项目(CY-CNCL-2008-02)。

作者简介:周尚波(1963-),男,广西宁明人,教授,博士生导师,主要研究方向:混沌神经网络、图像处理;胡鹏(1984-),男,湖南华容人,硕士研究生,主要研究方向:图像处理;柳玉炯(1983-),男,河南南阳人,硕士研究生,主要研究方向:人工神经网络、图像处理。

## 1.1 初始的运动目标分割

### 1.1.1 运动目标的分割目的及方法

运动目标分割的目的是分离出用户感兴趣的目標并提取其特征。帧差法在视频运动目标分割中是一种简单的方法。针对本文的应用情况,使用传统的帧差和区域生长算法能有效地分割出运动目标。

本文采用文献[7]的方法,由人工指定一个运动物体作为待跟踪目标并分割出来。因为需要人工来指定目标,所以这是一种半自动跟踪系统。对于连续的视频帧,由于两帧之间的时间间隔很短(大约为30 ms),在这么短的时间内,背景的变化可以忽略。本文中假设初始状态下的背景是静止的,利用连续帧的差分信息区分出运动目标<sup>[8-9]</sup>。

### 1.1.2 运动目标的分割过程

用户指定一个运动物体作为目标后,这个目标被 $t-1, t, t+1$ 三帧的差异分割出来。使用连续帧的差分信息检测不同帧的改变区域。差分操作定义为:

$$F(x, y, t) = \begin{cases} 0, & |f(x, y, t+1) - f(x, y, t)| \leq \alpha \\ 1, & \text{其他} \end{cases} \quad (1)$$

其中: $f(x, y, t)$ 和 $f(x, y, t+1)$ 分别代表第 $t$ 帧和第 $t+1$ 帧 $(x, y)$ 位置处的像素值。取适当的阈值 $\alpha$ ,如果两帧差的绝对值小于 $\alpha$ ,则认为该像素为背景,反之为目标。本文中,经过实验仿真分析,取 $\alpha = 80$ 。为了增加分割的精确性,采用三帧差算法来进行目标分割,分别计算出 $t-1, t, t+1$ 帧的帧差,然后取其交集,定义为:

$$M(x, y, t) = F(x, y, t-1) \cap F(x, y, t)$$

最终分割出的运动目标定义为:

$$TAR = \{(x, y) | M(x, y, t) = 1\} \quad (2)$$

然后,对该二值图像做一次膨胀操作,以用户的选取点为中心,进行区域增长,可得到比较准确的运动目标模板。初始目标的分割结果如图2所示。



(a) 用户指定目标 (b) 目标分割结果

图2 初始目标的分割

## 1.2 运动目标的特征提取

把运动目标的模板图像由RGB颜色空间转换到HSV图像空间,取其 $H$ 分量计算模板图像的灰度直方图,并找出其最大的 $H$ 值,作为该模板的一维特征。即计算反向投影,把实际的图像转换为概率分布图像,并根据反向投影结果确定出运动目标。运动目标的特征提取结果如图3所示。



(a) 目标模板的反向投影 (b) 最终框定的目标

图3 运动目标的特征提取结果

## 1.3 改进 Mean-Shift 算法

改进 Mean-Shift 算法的基本思想是每次计算搜索框的重

心,然后判断该区域的概率分布图像是否符合模板的反向投影特征,如果符合,停止计算;否则修正 $x, y$ 和搜索框大小,直到两次移动的距离小于某一阈值。

该算法与Dorin Comaniciu的算法有所不同,Dorin Comaniciu的Mean-Shift算法采用Bhattacharyya系数来度量模板图像和目标图像的相关性,而本文提出的Mean-Shift是在二维概率分布图像中,不断计算图形某区域的重心,并调整位置直到重心达到汇聚。假设区域 $I$ 的宽度为 $m$ ,高度为 $n$ ,计算该区域内的0阶矩和两个1阶矩:

$$\begin{cases} M_{00} = \sum_{x=0}^m \sum_{y=0}^n I(x, y) \\ M_{10} = \sum_{x=0}^m \sum_{y=0}^n x \cdot I(x, y) \\ M_{01} = \sum_{x=0}^m \sum_{y=0}^n y \cdot I(x, y) \end{cases} \quad (4)$$

重心计算公式:

$$\begin{cases} X_c = \frac{M_{10}}{M_{00}} \\ Y_c = \frac{M_{01}}{M_{00}} \end{cases} \quad (5)$$

改进的Mean-Shift算法流程如图4所示。

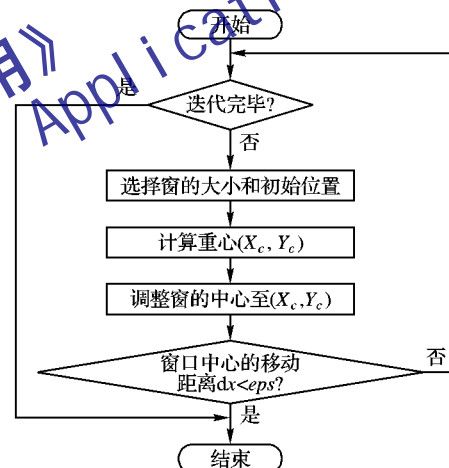


图4 改进的Mean-Shift算法流程

## 1.4 自适应Kalman滤波

如果目标的运动速度过快,Mean-Shift算法的跟踪效果不好,如果存在遮挡或者光线变化较大的情况,还可能丢失目标。本文引入Kalman滤波器来解决这个问题。Kalman滤波器是一种对动态系统的状态序列进行线性最小方差估计的算法,通过状态方程和观测方程来描述一个动态系统,它可以从任意一点作为起点开始检测,采用递归滤波的方法计算,具有计算量小、可实时计算的特点。

### 1.4.1 递归滤波

从第2帧到第 $t-1$ 帧,目标中心的位置信息形成它的运动轨迹,将这些中心位置信息作为Kalman滤波器的观测值。首先用Kalman滤波器预测 $t$ 帧目标的位置,作为Mean-Shift算法中目标的起始位置,Mean-Shift会在这个位置的邻域内找到目标最优的位置,再以这个目标位置作为Kalman滤波器的观测值,进行下一帧的运算。

### 1.4.2 目标的状态模型

假定运动目标中心在 $x, y$ 轴上的运动均为由随机加速度扰动的匀速直线运动,即目标的运动状态参数为某一时刻目标的位置和速度。由于相邻两帧的时间较短,目标的运动状态

变化较小,所以可假设目标在单位时间间隔内是匀速直线运动,建立如式(6)~(7)的运动模型和测量模型。

$$\mathbf{X}_t = \Phi \mathbf{X}_{t-1} + \mathbf{W}_t \quad (6)$$

$$\mathbf{Z}_t = \mathbf{H} \mathbf{X}_t + \mathbf{V}_t \quad (7)$$

其中: $\Phi$ 和 $\mathbf{H}$ 分别是状态转移矩阵和观测矩阵, $\mathbf{W}_t$ 和 $\mathbf{V}_t$ 均为零均值的高斯白噪声。假定其满足式(8)的统计特性:

$$\begin{cases} E[\mathbf{W}_t] = 0 \\ E[\mathbf{W}_t \mathbf{W}_t^T] = \mathbf{Q}_t \delta_{tt} \\ E[\mathbf{V}_t] = 0 \\ E[\mathbf{V}_t \mathbf{V}_t^T] = \mathbf{R}_t \delta_{tt} \\ E[\mathbf{W}_t \mathbf{V}_t^T] = 0 \end{cases} \quad (8)$$

其中: $\mathbf{Q}_t$ 是系统过程噪声 $\mathbf{W}_t$ 的 $p \times p$ 维对称非负定方差矩阵, $\mathbf{R}_t$ 是系统观测噪声 $\mathbf{V}_t$ 的 $m \times m$ 维对称正定方差矩阵,而 $\delta_{tt}$ 是Kronecker- $\delta$ 函数<sup>[10]</sup>。

#### 1.4.3 Kalman 滤波跟踪过程

本文中,只对运动目标的位置,即坐标 $(x, y)$ 进行估计。定义Kalman滤波器系统状态 $\mathbf{X}$ 为一个四维向量 $(x_t, y_t, vx_t, vy_t)$ ,其中, $x_t, y_t, vx_t, vy_t$ 分别为目标在 $x$ 轴和 $y$ 轴方向上的位置和速度。由于假设目标是在单位时间间隔内做匀速运动,所以状态转移矩阵 $\Phi$ 定义为( $\Delta t$ 为两帧的时间差):

$$\Phi = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

观测矩阵 $\mathbf{H}$ 为:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

通过多次实验,确定 $\mathbf{W}_t$ 和 $\mathbf{V}_t$ 的初始协方差矩阵为:

$$\mathbf{Q}_t = \text{diag}(2\mathbf{E} - 6, 2\mathbf{E} - 6, 2\mathbf{E} - 6, 2\mathbf{E} - 6)$$

$$\mathbf{R}_t = \text{diag}(1\mathbf{E} - 2, 1\mathbf{E} - 2)$$

预测步骤:

$$\begin{cases} \hat{\mathbf{X}}_{t, t-1} = \Phi \hat{\mathbf{X}}_{t-1} \\ \mathbf{P}_{t, t-1} = \Phi \mathbf{P}_{t-1} \Phi^T + \mathbf{Q}_{t-1} \end{cases} \quad (9)$$

修正步骤:

$$\begin{cases} \mathbf{K}_t = \mathbf{P}_{t, t-1} \mathbf{H}^T [\mathbf{H} \mathbf{P}_{t, t-1} \mathbf{H}^T + \mathbf{R}_t]^{-1} \\ \hat{\mathbf{X}}_t = \hat{\mathbf{X}}_{t, t-1} + \mathbf{K}_t [\mathbf{Z}_t - \mathbf{H} \hat{\mathbf{X}}_{t, t-1}] \\ \mathbf{P}_t = [\mathbf{I} - \mathbf{K}_t \mathbf{H}] \mathbf{P}_{t, t-1} [\mathbf{I} - \mathbf{K}_t \mathbf{H}]^T + \mathbf{K}_t \mathbf{R}_t \mathbf{K}_t^T \end{cases} \quad (10)$$

跟踪过程中,定义在 $t$ 帧的运动目标区域的像素数和 $t-1$ 帧的运动目标区域的像素数的比例为目标遮挡率因子。引入遮挡率因子以自适应地调节Kalman滤波器的估计参数。

$$\alpha_t = \begin{cases} \left| \frac{N_t}{N_{t-1}} - 1 \right|, & \left| \frac{N_t}{N_{t-1}} - 1 \right| \leq 1 \\ 1, & \text{其他} \end{cases} \quad (11)$$

其中 $N_t$ 和 $N_{t-1}$ 分别为第 $t$ 帧和第 $t-1$ 帧中运动目标的像素数目。当 $\alpha_t$ 取1时,表示出现完全遮挡。本文中为 $\alpha_t$ 定义一个阈值 $\beta$ ,以自适应地调整过程噪声和观测噪声的协方差矩阵,当 $\alpha_t < \beta$ 时, $\mathbf{Q}_{t-1} = 1 - \alpha_t$ , $\mathbf{R}_t = \alpha_t$ ;否则 $\mathbf{Q}_{t-1} = 0$ , $\mathbf{R}_t = \infty$ 。

## 2 实验结果与分析

本文提出的算法在Windows XP系统下,使用VC++ 6.0以及OpenCV计算机视觉库编程实现。分别采用Mean-Shift

算法和本文提出的算法对一段羽毛球比赛视频中的运动员进行跟踪。序列总长200帧,每帧像素为 $720 \times 576$ ,其中有短时间的遮挡出现。

图5显示了Mean-Shift算法的跟踪结果(第111、128、130、132、134、142帧是遮挡处理过程),可以看到第130帧的目标被完全遮挡,然后出现跟踪失败,直到序列最后都无法正确跟踪目标。

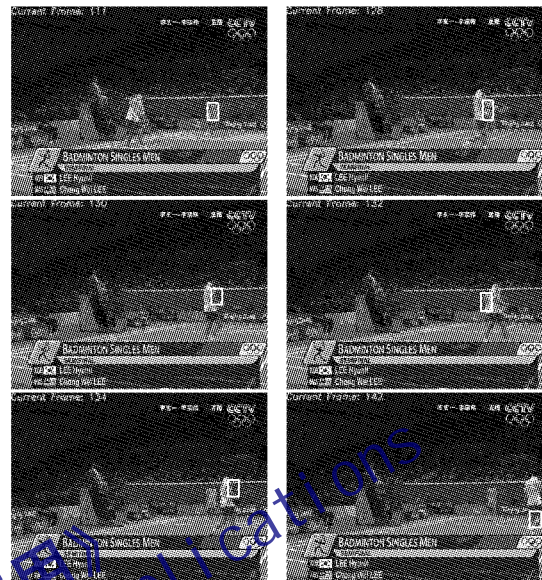


图5 Mean-Shift 算法的跟踪结果

图6显示了本文算法的跟踪结果(使用了同样的6帧)。当第130帧出现遮挡时,遮挡率因子 $\alpha_t = 1$ ,调节 $\mathbf{Q}_{t-1}$ 和 $\mathbf{R}_t$ 分别为0和系统极大值。此时的跟踪结果中,预测值起主导作用,而Mean-Shift反馈的观测值对结果影响较小。直到目标重新出现后, $\alpha_t$ 减小,重新调节 $\mathbf{Q}_{t-1}$ 和 $\mathbf{R}_t$ ,系统可以有效的对目标继续跟踪。



图6 本文算法的跟踪结果

使用Matlab画出跟踪结果的 $x$ 坐标(见图7)。从图7中可以看出,在本文改进算法中,Kalman滤波(KF)预测值和Mean-Shift(MS)算法确定的观测值共同作用,最终结果是两者的综合。其中,第130帧发生完全遮挡时,Mean-Shift算法确定的观测值误差很大,最后的结果与预测值一致。

图8显示了Mean-Shift和本文算法跟踪结果中 $x$ 坐标的残



差,从图8中很容易看出,在第130帧发生完全遮挡时候,Mean-Shift算法已经完全不能跟踪目标,之后再也不能正确跟踪到目标。而改进的Mean-Shift算法误差较小,而且由于Kalman滤波的作用,在遮挡过程完成之后又正确跟踪到目标。

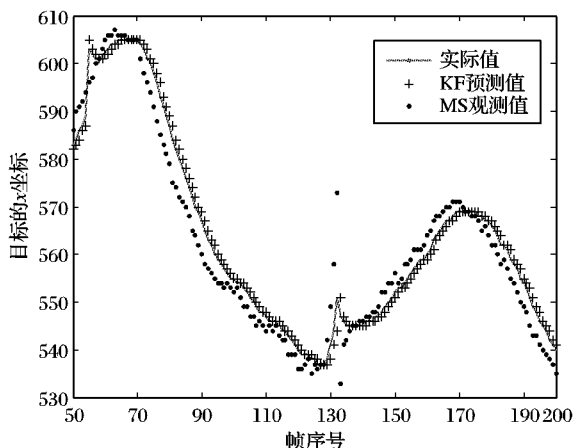


图7  $x$  坐标示意图

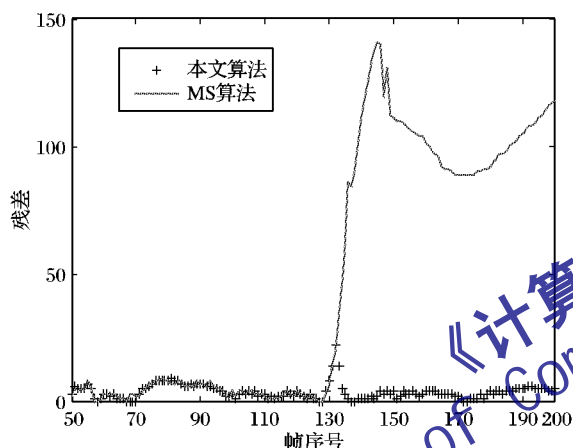


图8 MS算法和本文算法 $x$ 坐标残差示意图

对比两种算法的实验结果,可以看出本文提出的算法在处理遮挡方面比传统Mean-Shift算法更加有效,对运动目标跟踪的稳定性更好。

### 3 结语

本文提出一种自适应Kalman滤波和改进的Mean-Shift

跟踪算法。采用人工点选目标,再使用三帧差检测方法得出运动目标的信息,并提出了一种改进的Mean-Shift算法结合Kalman滤波进行有效的目标跟踪,即基于目标以前的位置信息,采用Kalman滤波预测目标在当前帧的可能位置,然后利用Mean-Shift算法在预测位置附近找到最佳的匹配位置,并在跟踪过程中根据遮挡率因子动态调整Kalman滤波器参数。实验结果表明,提出的算法在处理遮挡方面比传统Mean-Shift算法更加有效,能对运动目标实现检测和连续跟踪,具有较好的鲁棒性。

### 参考文献:

- [1] 王江涛,杨静宇. 遮挡情况下基于Kalman均值偏移的目标跟踪[J]. 系统仿真学报, 2007, 19(18): 4216-4220.
- [2] 左军毅,梁彦,赵春晖,等. Mean-Shift跟踪算法中尺度自适应策略的研究[J]. 中国图象图形学报, 2008, 13(9): 1750-1757.
- [3] LIU P R, MENG M Q H, LIU P X, *et al.* Optical flow and active contour for moving object segmentation and detection in monocular robot [C]// Proceedings 2006 IEEE International Conference on Robotics and Automation. Washington, DC: IEEE, 2006: 4075-4080.
- [4] 曾伟,朱桂斌,李瑶. 基于Kalman点匹配估计的运动目标跟踪[J]. 计算机应用, 2009, 29(6): 1677-1682.
- [5] 罗嘉,韦志辉. 基于几何活动轮廓模型的目标跟踪与快速运动估计[J]. 中国图象图形学报, 2009, 14(7): 1361-1368.
- [6] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using Mean-Shift [C]// IEEE Computer Vision and Pattern Recognition. Washington, DC: IEEE, 2000: 142-149.
- [7] PARK D K, YOON H S, WON C S. Fast object tracking in digital video[J]. IEEE Transactions on Consumer Electronics, 2000, 46(3): 785-790.
- [8] WENG S K, KUO C M, TU S K. Video object tracking using adaptive Kalman filter[J]. Journal of Visual Communication and Image Representation, 2006, 17(6): 1190-1208.
- [9] KIM C, HWANG J-N. Fast and automatic video object segmentation and tracking for content-based applications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(2): 122-129.
- [10] 付梦印,邓志红,张继伟. Kalman滤波理论及其在导航系统中的应用[M]. 北京: 科学出版社, 2003: 20-25.

(上接第1546页)

- [2] KOHONEN T. Self-organizing neural projections [J]. Neural Networks, 2006, 19(6/7): 723-755.
- [3] ROJAS I, GONZALEZ J, POMARES H, *et al.* Multidimensional and multideme genetic algorithms for the construction of fuzzy systems [J]. International Journal of Approximate Reasoning, 2001, 26(3): 179-210.
- [4] SHAPIRO A F. The merging of neural networks, fuzzy logic, and genetic algorithms [J]. Insurance: Mathematics and Economics, 2002, 31(1): 115-131.
- [5] LIN C T, LEE C S G. Neural-network-based fuzzy logic control and decision system [J]. IEEE Transactions on Computers, 1991, 40(12): 1320-1336.
- [6] YU W D, SKIBNIEWSK M J. A neuro-fuzzy computational approach to constructability knowledge acquisition for construction technology evaluation [J]. Automation in Construction, 1999, 8(5): 539-552.
- [7] KAWAKAMI J, HOSHI K, ISHIYAMA A, *et al.* Application of a self-organizing map to quantitative structure-activity relationship analysis of carboquinone and benzodiazepine [J]. Chemical and Pharmaceutical Bulletin, 2004, 52(6): 751-756.
- [8] FABRY-ASZTALOS L, ANDONIE R, COLLAR C J, *et al.* A genetic algorithm optimized fuzzy neural network analysis of the affinity of inhibitors for HIV-1 protease [J]. Bioorganic & Medicinal Chemistry, 2008, 16(6): 2903-2911.
- [9] KOHONEN T, SOMERVUO P. How to make large self-organizing maps for nonvectorial data [J]. Neural Networks, 2002, 15(8/9): 945-996.
- [10] SAMECKA-CYMERMAN A, STANKIEWICZ A, KOLON K, *et al.* Self-organizing feature map (neural networks) as a tool in classification of the relations between chemical composition of aquatic bryophytes and types of streambeds in the Tatra national park in Poland [J]. Chemosphere, 2007, 67(5): 954-1013.