

8-20-2020

## Gesture Segmentation and Recognition Based on Kinect Depth Data

Yanming Mao

*Department of Computer, Ningde Normal University, Ningde 352100, China;*

Liliang Zhang

*Department of Computer, Ningde Normal University, Ningde 352100, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Gesture Segmentation and Recognition Based on Kinect Depth Data

### Abstract

**Abstract:** Aiming at the problems that gesture recognition required high environmental background, segmented gesture usually contained wrist data and closing fingers easily caused false recognition, a gesture recognition method based on depth data was proposed. It captured gesture depth map, and it used Hands Generator to obtain the information of palm for gesture segmentation, in order to remove redundant wrist data, the constraint of the palm which looks like a square was added. The number of all the other four fingers except thumb could be acquired with the use of the scanline method, therefore, the width ratio of the adjacent fingers and the peculiarity of the thumb position were integrated to achieve the gesture recognition. Experimental results show that the average recognition rate of the method is 98.4%, and the method has good real-time performance and robustness in the different illumination conditions and complex backgrounds.

### Keywords

Kinect sensor, depth data, gesture segmentation, gesture recognition

### Recommended Citation

Mao Yanming, Zhang Liliang. Gesture Segmentation and Recognition Based on Kinect Depth Data[J]. Journal of System Simulation, 2015, 27(4): 830-835.

# 基于 Kinect 深度信息的手势分割与识别

毛雁明, 章立亮

(宁德师范学院计算机系, 宁德 352100)

**摘要:** 针对手势识别对环境背景要求高、分割的手势往往包含手腕及手指靠拢易造成误识别等问题, 提出了一种基于深度信息的手势识别方法。采集手势深度图; 接着利用手生成器获取手心信息分割手势, 为了去除多余的手腕信息, 增加了手掌近似正方形的约束条件; 利用扫描线法获取拇指外其它四指数, 并结合相邻手指宽度比例及拇指位置的特殊性实现手势的识别。实验结果表明, 该方法的平均识别率达到 98.4%, 实时性好, 且对不同光照和复杂背景具有很好的鲁棒性。

**关键词:** Kinect 传感器; 深度信息; 手势分割; 手势识别

中图分类号: TP391.41      文献标识码: A      文章编号: 1004-731X (2015) 04-0830-06

## Gesture Segmentation and Recognition Based on Kinect Depth Data

Mao Yanming, Zhang Liliang

(Department of Computer, Ningde Normal University, Ningde 352100, China)

**Abstract:** Aiming at the problems that gesture recognition required high environmental background, segmented gesture usually contained wrist data and closing fingers easily caused false recognition, a gesture recognition method based on depth data was proposed. It captured gesture depth map, and it used Hands Generator to obtain the information of palm for gesture segmentation, in order to remove redundant wrist data, the constraint of the palm which looks like a square was added. The number of all the other four fingers except thumb could be acquired with the use of the scanline method, therefore, the width ratio of the adjacent fingers and the peculiarity of the thumb position were integrated to achieve the gesture recognition. Experimental results show that the average recognition rate of the method is 98.4%, and the method has good real-time performance and robustness in the different illumination conditions and complex backgrounds.

**Keywords:** Kinect sensor; depth data; gesture segmentation; gesture recognition

## 引言

手势识别是人机交互领域中热门的研究方向之一, 以人的手部动作直接控制计算机的输入, 相比传统的键盘、鼠标等具有操作便捷、自然、直观

等优点。手势识别技术涉及人工智能、模式识别、机器学习、计算机图形学等众多学科, 其应用也深入到虚拟现实、手语识别、电脑游戏以及机器人控制等方面<sup>[1]</sup>。

传统的基于数据手套的手势识别方法是利用数据手套反馈各关节的数据, 并通过位置跟踪器获取人手的三维坐标, 从而来测量手势在空间运动中的轨迹和时序信息<sup>[2]</sup>。该方法的优点是获取的数据精度高、不受光照和复杂背景的影响且手势识别率高, 但缺点是需要用户穿戴复杂的数据手套和位置



收稿日期: 2013-12-14      修回日期: 2014-05-12;  
基金项目: 福建省自然科学基金(2011J01358);福建省教育厅 A 类科技项目(JA13337);宁德师范学院服务海西基金项目(2012H311);  
作者简介: 毛雁明(1982-), 男, 福建福安人, 硕士, 讲师, 研究方向为计算机视觉、人工智能、模式识别; 章立亮(1963-), 男, 福建宁德人, 本科, 教授, 研究方向为计算机图形图像。

<http://www.china-simulation.com>

• 830 •

跟踪器, 从而阻碍了手势的自然特性, 并且数据手套价格昂贵也不利于普及推广。

近年来随着技术的发展, 使用网络摄像头捕获手势的计算机视觉技术成为手势识别的主流方式<sup>[3-5]</sup>, 基于视觉的手势识别方法利用单个或多个摄像头采集手势信息, 先经计算机系统处理分割出手势, 然后采用特定的方法实现手势的识别。目前常用的手势分割方法主要有:(1)基于徒手表现特征的手势分割方法(徒手的表现特征指手的肤色、纹理、指尖、手的轮廓、手型和手的大小等), 通常都是利用采集的肤色信息, 在 YUV, HSV 或 YCbCr 等颜色空间下进行建模来对手势进行分割<sup>[6-8]</sup>, 该方法受外界环境的干扰较大, 如光照影响, 类肤色的干扰等。(2)人为增加限制的方法, 如使用黑色和白色的墙壁、深色的服装等简化背景或要求人手佩戴特殊的手套等强调前景, 以简化手区域与背景区域的划分<sup>[9]</sup>, 该方法虽然可以解决类肤色的干扰, 但对环境要求较为苛刻, 使系统的实时性和手势识别的自然性大幅降低。(3)建立手势形状数据库的方法, 如密西根州立大学计算机系 Cui Yuntao 建立了一种数据库, 包括各种手势类在不同时刻、不同位置、不同比例时的手型图像, 将它作为基于模板匹配识别的模板<sup>[10]</sup>, 建立手势形状数据库的方法考虑了各种手势的形状, 以便于不同手势的模板匹配, 但其计算量随之增加, 不利于系统的快速识别。以上基于视觉的手势分割均无法克服外界因素的影响, 虽然这些方法人机交互较自然、直接, 摄像头价格相对较低, 但是计算过程复杂, 对光照条件和复杂背景非常敏感, 识别率和实时性均较差。

本文是利用微软新近推出的 Kinect 传感器获取深度信息, 由于这种深度信息往往包含很大的噪声, 因此在手势识别预处理阶段首先对深度信息进行平滑去噪以提高手势分割的质量, 然后利用 OpenNI 提供的 HandsGenerator 生成器获取手心位置, 再根据手心的深度进行手势的分割, 但此时分割得到的手势往往会包含手腕, 这将影响后续的手势识别效果, 因此在手势分割阶段增加了手掌近似

正方形的约束条件以去除手腕的干扰; 最后利用扫描线方法并加入手指宽度比例及拇指位置的特殊性实现了对 5 种手势 1~5 的识别。实验结果表明, 该手势识别方法实时性好, 对复杂背景和光照条件也具有很好的鲁棒性, 识别率达到 98.4%。

## 1 手势分割

手势分割是把人手区域从图像中分割开来, 是手势识别中最重要也是最基础的一个步骤, 手势分割的好坏直接影响后续手势识别的效果。本文重点研究对 1~5 这 5 种手势的识别, 为了满足算法需要, 假定手的位置距离 Kinect 摄像头最近, 此外还要求手势的手心面向摄像头, 展开手掌, 指尖向上, 手势平面与摄像头平面尽量保持平行。

### 1.1 深度图像预处理

利用 Kinect 传感器获取深度信息, 并把深度信息转换为灰度深度图像, 使得后续对深度的处理转换为对深度图像像素值的处理。由于 Kinect 传感器是采用结构光编码技术获取深度信息, 获取的深度图像实际上是不稳定的, 存在很多空洞及噪声点, 特别是边缘处的深度值比较不稳定, 直接进行处理会对后续的分割和识别造成一定的影响, 因此在手势分割之前需对深度图像进行平滑去噪以提高手势分割的效果。

Kinect 传感器获取的深度图像边缘性较强, 且当深度差大于 3~4 mm 时被认为具有不同的深度, 为了保持边缘信息以和算法的实时性, 本文采用形态滤波算法对深度图像进行预处理, 先利用数学形态学的闭运算填充场景中的细小空洞, 再利用开运算去除场景中的噪声, 同时保留了目标边缘信息。

### 1.2 手心位置获取

通过 Kinect 传感器对手心进行跟踪, 应用程序采用 OpenNI 中间层定义的手部分析中间件提供的 HandsGenerator 生成器, 该生成器可实现快速准确地完成对人体手部的跟踪, 从而获取手心在整幅

图像中的位置信息和深度信息。图 1 为通过 OpenNI 获取手势的深度和彩色图像,其中彩色图像中的红点为通过手部分分析与视觉开发包(OpenCV)相结合的方法检测并标示出的手心位置,记为  $P(x_{Palm}, y_{Palm}, z_{Palm})$ ,其中  $(x_{Palm}, y_{Palm})$  代表手心在二维图像中的位置信息,  $z_{Palm}$  代表手心距离深度传感器的深度信息,这里通过回调函数 `HandUpdate()` 返回手心位置的三维数据。



图1 深度图像与彩色图像(右图中红点表示手心)

### 1.3 手势提取

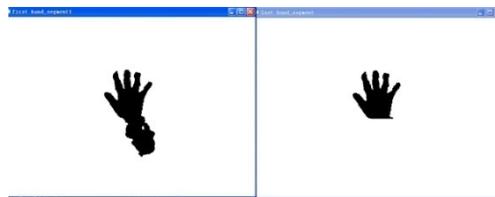
我们限定人手的位置总是在摄像头的最前面,采用基于深度阈值的分割方法来提取手势。从图 1 中已经得到手心的位置,而深度图中每个点的像素值恰好表示了场景中每个点到摄像头的相对距离,因此使用下式(1)可以分割出手势,其分割的手势如图 2(a)所示。

$$seg(i, j) = \begin{cases} 0 & \text{if } z_{Palm} - 5 \leq Dep(i, j) \leq z_{Palm} + 5 \\ 1 & \text{otherwise} \end{cases} \quad (1)$$

其中,  $Seg(i, j)$  表示分割后的图像在像素点  $(i, j)$  的灰度值,  $z_{Palm}$  表示手心的深度值,  $Dep(i, j)$  表示深度图像中像素点  $(i, j)$  的灰度值。

基于深度阈值的手势分割虽然能把手势分割出来,但得到的手势往往包含手腕(如图 2(a)),这将大大影响后续手势识别的效果。为了去除手腕的干扰,通过对手掌形状的分析发现,当五指伸开时,手掌近似正方形,因此在深度阈值分割中增加了手掌近似正方形的约束条件。首先通过扫描图像获取手势图像最左边、最上面和最右边的三个像素坐标点分别记为  $leftM$ ,  $topM$ ,  $right$ , 接着计算正方形边长  $d = right.x - left.x$ , 再把手势像素点的纵坐标大于  $topM.y + d$  的像素点的灰度值赋值为 255, 得到图

2(b)所示的手势分割图。从实验结果可以看出,本文的手势分割算法能够很好的把手势分割出来并去除冗余的手腕信息。



(a) 深度分割得到的手势 (b) 本文算法得到的手势

图2 手势分割对比图

## 2 手势识别

手势识别是整体设计中最关键的一个环节,手势识别的方法有很多种,不同的识别方法所建立的模型不尽相同。相对于常见且较为耗时的神经网络法<sup>[11-12]</sup>、隐马尔科夫法<sup>[13-15]</sup>和传统的模板匹配法<sup>[16]</sup>,本文提出了一种新的手势识别方法:“扫描线法”。“扫描线法”是利用扫描线在手势区域最上方开始往下扫描,扫描线将交割于除拇指以外的其他四指,由此识别出四指伸出的个数。当存在手指靠拢和拇指伸出时,会造成误识别,因此增加相邻手指宽度比例和拇指位置的特殊性以实现对手势的最终识别。该方法无需训练样本,且能够识别出手指靠拢的情况,具有适应性强、运算速度快、实时性好等优点。

### 2.1 手势分析

本文只针对 1~5 这 5 种手势进行识别,图 3 中(a)~(e)为应用以上介绍的手势分割算法得到的 5 种手势的分割图。可以发现总存在一条直线段能穿过除拇指以外的其它伸出的手指,若能找到这条直线段,就可以通过直线段与手指的第一个交点数来判断手指伸出的个数。需要考虑以下问题:(1)如图 3(f)所示,如何找到这样的直线段满足上述要求;(2)如图 3(g)所示,当两根手指靠拢时,会把手势 4 和手势 5 误识别为手势 3;(3)如图 3(f)所示,上述方法不能判断有无拇指伸出,此时会把手势 5 误识别为手势 4。

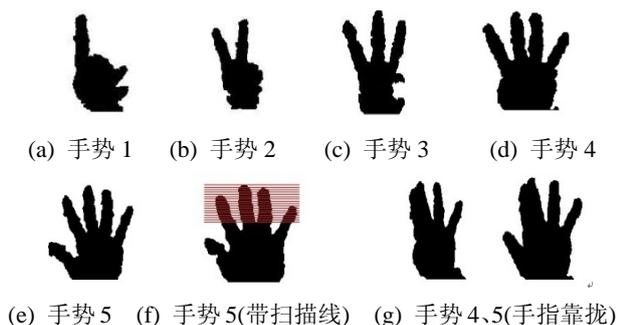


图3 手势1~5的X-Y平面示意图

应用手势分割算法可以得到手势的最左边、最右边、最上面和最下面的四个像素坐标点, 分别记为  $leftM$ ,  $rightM$ ,  $topM$ ,  $bottomM$ , 由此计算出手势的宽度  $width = right.x - left.x$ 。对上述 3 个问题, 可用以下方法解决: 对问题(1), 采用扫描线方法, 每一次扫描线从点  $(leftM.x, topM.y + 1/8 * width + i)$  开始往右扫描到点  $(right.x, topM.y + 1/8 * width + i)$  结束, 第一次扫描  $i=0$ , 往后  $i+=3$ , 扫描过程中需记录手指数和扫描线穿过每根手指的宽度, 一直到扫描线的纵坐标的值大于  $topM.y + 0.4 * d$  时结束, 把获取最大手指数  $fingerNum$  作为识别的结果, 对应识别得到的每根手指宽度保存在数组  $count[j]$  中,  $j$  表示第  $j$  根手指。对问题(2), 由于要求做手势时手指尽量张开, 因此在做手势 1 和 2 时基本不存在手指靠拢的现象, 做手势 3~5 时, 有靠拢的基本上就是食指与中指。经试验发现, 当手指没有靠拢时, 各手指的宽度都比较接近, 当两根手指靠拢而被误判为一根手指时, 其宽度往往是单个手指宽度的 1.5 倍以上, 通过上述分析, 当  $count[1]/count[2] \geq 1.5$  时, 表示有手指靠拢, 此时手指数  $fingerNum$  加 1。对问题 3, 只有当做手势 5 时, 拇指才伸出, 这时手势的最左坐标点  $leftM$  在拇指上, 因此当  $leftM.y > topM.y + 0.4 * d$  时, 表示有拇指伸出, 此时手指数  $fingerNum$  加 1。

## 2.2 手势识别算法

结合上面的介绍, 具体的基于 Kinect 深度信息的手势分割与识别算法描述如下:

步骤 1 利用 Kinect 传感器采集手势的彩色图

和深度图。

步骤 2 采用形态滤波对深度图预处理。由于 Kinect 传感器获取的深度图是不稳定的, 存在很多空洞和噪声, 而形态学滤波可以去除目标内部的空洞和噪声。

步骤 3 利用 OpenNI 提供的 HandsGenerator 生成器对手势进行检测和跟踪并获取手心位置。

步骤 4 根据获取的手心位置采用双阈值法对手势进行分割, 此时分割得到的手势图像往往包含手腕信息。

步骤 5 根据手掌近似正方形的约束条件去除手腕干扰。

步骤 6 利用扫描线法从上往下扫描分割的手势, 直到扫描的距离为  $0.4 * d$ , 其中  $d$  为手势宽度, 把扫描线与每根手指的第一个交点数的最大值  $fingerNum$  作为识别结果, 及相应的各手指的宽度保存在数组  $count[i]$  中。如果  $fingerNum=1$ , 转到步骤 9, 否则往下执行步骤 7。

步骤 7 如果  $fingerNum=2$  或 3 并且  $count[1]/count[2] \geq 1.5$ , 则  $fingerNum++$ , 否则转到步骤 9。

步骤 8 如果  $fingerNum=4$  且手势最左纵坐标的值大于  $topM.y + 0.4 * d$ , 表示有拇指伸出, 此时  $fingerNum++$ 。

步骤 9 识别结束,  $fingerNum$  为识别结果。

## 3 实验结果与分析

为了验证本文算法的有效性, 以 VS2008 为开发工具, 基于 OpenNI 和 OpenCV 编写了一个基于 Kinect 深度信息的手势分割与识别程序, 识别速度为 25f/s, 达到了手势识别实时性的要求。

图 4 为在强光、黑暗及复杂背景下的手势分割图, 其中图 4(a)为强光条件下的手势分割效果图, 图 4(b)为黑暗条件下的手势分割效果图, 图 4(c)为复杂背景条件下的手势分割效果图。从效果图可以看出本文的手势分割方法对光照条件和复杂背景具有很好的鲁棒性并能够有效地去除手腕干扰。



(a) 强光下手势分割效果图



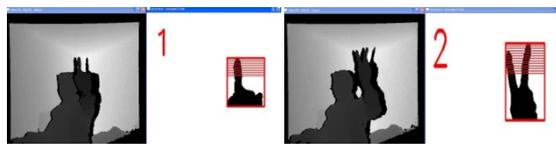
(b) 黑暗下手势分割效果图



(c) 复杂背景下手势分割效果图

图4 手势分割效果图

图 5 为手势识别实验结果图, 从图 5(a)~(c)、(e)、(g)可以看出, 当手指不靠拢时, 能够实现对手势 1~5 的正确识别; 从图 5(d)、(f)、(h)可以看出, 即使存在两指靠拢, 也能够正确识别出手势。



(a) 手势1

(b) 手势2



(c) 手势3

(d) 手势3(两指靠拢)



(e) 手势4

(f) 手势4(两指靠拢)



(g) 手势5

(h) 手势5(两指靠拢)

图5 典型手势识别测试结果图

表 1~2 分别为采用文献[17]和本文算法采集 10 个人分别对手势 1~5 各做 10 次的识别结果。从表 1 可以看出, 文献[17]的平均识别率为 95%, 且不能很好的处理手指靠拢的情况; 从表 2 可以看出, 本文算法的平均识别率为 98.4%, 对手势 1、3 的识别率达到了 100%; 手势 2 被识别为手势 1 是因为两手指紧紧靠拢, 一般情况下, 我们做手势 2 时不会把两手指靠拢, 即使两手指靠拢, 只要两指之间存在间隙, 利用我们的方法也能够识别; 手势 4 会被识别为手势 3 是因为做手势时, 手平面弯曲, 使得扫描线不能穿过小指造成的, 手势 4 会被识别为手势 5 是因为在做手势 4 时拇指没有往里收或食指朝外, 使得最左坐标不在食指上, 造成了误识别; 手势 5 会被识别为手势 3 是因为扫描线没能穿过小指且没有去判断有无拇指伸出造成的。从表 1、表 2 的实验结果可以看出, 本文的手势识别算法能够很好的实现对手势 1~5 的识别且能够处理手指靠拢的情况。

表 1 应用文献[17]算法得到的手势识别结果

输入	输出1	输出2	输出3	输出4	输出5	识别率%
手势1	95	2	3	0	0	95
手势2	3	93	4	0	0	93
手势3	5	3	92	0	0	92
手势4	0	0	0	98	2	98
手势5	0	0	0	3	97	97

表 2 应用本文算法得到的手势识别结果

输入	输出1	输出2	输出3	输出4	输出5	识别率%
手势1	100	0	0	0	0	100
手势2	1	99	0	0	0	99
手势3	0	0	100	0	0	100
手势4	0	0	2	95	3	95
手势5	0	0	2	0	98	98

## 4 结论

本文主要研究了基于 Kinect 深度信息的手势分割与识别方法。在手势分割阶段, 利用 OpenNI 提供的 HandsGenerator 生成器获取手心位置进行手势分割, 此时分割的手势往往包含手腕, 通过加入手掌近似正方形的约束条件有效地去除了手腕;

在手势识别阶段, 利用扫描线扫描分割的手势获取手指指数, 接着加入手指宽度比例以及拇指位置的特殊性实现对手势的最终识别。该方法使用设备简单、计算量少、效率和识别率高, 且对复杂背景和光照具有很好的鲁棒性。本方法的不足之处是识别类型局限在 1~5 这 5 种手势, 当手势方向与 Kinect 摄像头垂直方向夹角达到某个值时会造成误识别, 后续工作将增加对手势 6~9 的识别, 并考虑即使手势倾斜时也能够识别。

### 参考文献:

- [1] J P Wachs, M Kolsch, H Stern, Y Edan. Vision-based hand-gesture applications [J]. Communications of the ACM (S0001-0782), 2011, 54(2): 60-71.
- [2] 冯志全, 蒋彦. 手势识别研究综述 [J]. 济南大学学报(自然科学版), 2013, 27(4): 336-341.
- [3] A Erol, G Bebis, M Nicolescu, *et al.* Vision-based hand pose estimation: A review [J]. Computer Vision and Image Understanding (S1077-3142), 2007, 108(1/2): 52-73.
- [4] S Mitra, T Acharya. Gesture recognition: A survey [J]. IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews (S1094-6977), 2007, 37(3): 311-324.
- [5] G R S Murthy, R S Jadon. A review of vision based hand gesture recognition [J]. International Journal of Information Technology and Knowledge Management (S1461-4111), 2009, 2(2): 405-410.
- [6] 杨波, 宋晓娜, 冯志全, 等. 复杂背景下基于空间分布特征的手势识别算法 [J]. 计算机辅助设计与图形学学报, 2010, 22(10): 1841-1848.
- [7] 谭同德, 郭志敏. 基于双目视觉的人手定位与手势识别系统研究 [J]. 计算机工程与设计, 2012, 33(1): 259-264.
- [8] 张秋余, 王道东, 张墨逸, 等. 基于特征包支持向量机的手势识别 [J]. 计算机应用, 2012, 32(12): 3392-3396.
- [9] 汤勇, 顾宏斌, 周来. 交互系统中实时手势分割及指尖检测方法 [J]. 光电工程, 2010, 37(7): 145-150.
- [10] 吕颖. 基于摄像头的手势交互系统的研究 [D]. 青岛: 青岛大学, 2009.
- [11] 谭昶, 肖南峰. 基于改进RCE和RBF神经网络的静态手势识别 [J]. 计算机工程与应用, 2011, 47(7): 172-176.
- [12] 李文生, 解梅, 邓春健, 等. 基于 Hermite 神经网络的动态手势学习和识别 [J]. 计算机工程与科学, 2012, 34(2): 116-122.
- [13] 于美娟, 马希荣. 基于 HMM 方法的动态手势识别技术的改进 [J]. 计算机科学, 2011, 38(1): 251-252.
- [14] 江超, 艾娇燕. 基于 OpenCV 的摄像头动态手势轨迹识别及其应用 [J]. 计算机应用, 2012, 32(S1): 128-133.
- [15] 陶丽君, 李翠华, 张希倩, 等. 基于 Kinect 传感器深度信息的动态手势识别 [J]. 厦门大学学报(自然科学版), 2013, 52(4): 493-497.
- [16] 李文杰. 基于骨架化和模板匹配的交通指挥手势识别 [D]. 杭州: 浙江工业大学, 2011.
- [17] 邓瑞, 周玲玲, 应忍冬. 基于 Kinect 深度信息的手势提取与识别研究 [J]. 计算机应用研究, 2013, 30(4): 1263-1265.

## 《系统仿真学报》荣获“2014 中国最具国际影响力学术期刊”证书

由中国学术期刊(光盘版)电子杂志社与清华大学图书馆联合成立的中国学术文献国际评价研究中心, 以美国汤森路透 Web of Science 收录的 1.2 万余种期刊为引文统计源, 首次研制发布了 2012《中国学术期刊国际引证年报》(CAJ-IJCR 年报)。第一次给出了我国 5600 余种中外文学学术期刊总被引频次、影响因子、半衰期等各项国际引证指标, 并采用了新的国际影响力综合评价指标 CI 对期刊排序, 发布了“中国最具国际影响力学术期刊”(排序 TOP5%)和“中国国际影响力优秀学术期刊”(排序 TOP5-10%), 在国内外学术界产生了较大反响。之后, 2013 年版年报, 将引文统计源期刊扩展到 1.44 万多种。目前, 2014 版国际、国内年报与 TOP5%和 TOP5-10%期刊的遴选业已完成, 《系统仿真学报》被列入“2014 中国最具国际影响力学术期刊”行列。

我学报连续 2 年被列入 TOP5%国内一流的中国最具国际影响力学术期刊, 走向世界, 进入国际一流, 指日可待!