Journal of Computer Applications
计算机应用,2013,33(2):503 -506
ISSN 1001-9081
CODEN JYIIDU
2013-02-01
http://www.joca.cn

文章编号:1001 -9081(2013)02 -0503 -04
doi:10.3724/SP.J.1087.2013.00503

基于细胞神经网络的快速手语视频分割方法

张爱华[1,2*],雷小亚[2],陈晓雷[1],陈莉莉[2]
(1.兰州理工大学 电气工程与信息工程学院,兰州 730050; 2.兰州理工大学 计算机与通信学院,兰州 730050)
(*通信作者电子邮箱 lutzhangah@163.com)

摘 要:为实现感兴趣区手语视频编码,提高通话效率,提出一种基于细胞神经网络(CNN)的快速手语视频分割方法。该方法首先利用肤色信息特征进行基于CNN的肤色检测,检测出手语视频中的肤色区域;然后对肤色检测结果,利用帧差法进行基于CNN的运动检测,获得初始的手势区域;最后采用形态学处理方法进行空洞填充和边界平滑,实现了手语视频图像序列中的面部和手部区域的分割。研究结果表明,该方法能够快速准确地进行手语视频分割。

关键词:细胞神经网络;手语;视频分割;肤色检测;运动检测
中图分类号:TP391.41 **文献标志码**:A

Fast segmentation of sign language video based on cellular neural network

ZHANG Aihua[1,2*], LEI Xiaoya[2], CHEN Xiaolei[1], CHEN Lili[2]
(1. *School of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou Gansu* 730050, *China*;
2. *School of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu* 730050, *China*)

Abstract: To achieve sign language video coding of region of interest, and improve call efficiency, a fast segmentation methodology of sign language video based on Cellular Neural Network (CNN) was proposed. Firstly, the skin regions of sign language video were detected through corresponding CNN templates by using the skin color information characteristics. Secondly, CNN based motion detection was carried out on the skin detection results by using inter-frame difference algorithm, and then the initial gesture region could be obtained. Finally, morphological processing methods were employed to fill small holes and smooth the boundaries of regions, and eventually the segmentation of the face and hands regions of sign language video image sequence was realized. The results show that the method can rapidly and accurately segment sign language video.

Key words: Cellular Neural Network (CNN); sign language; video segmentation; skin color detection; motion detection

0 引言

手语是由手形、手臂运动并辅之以表情、唇动以及其他体势表达思想的视觉语言,是聋哑人进行信息交流的最自然方式[1]。聋哑人理解手语信息的速度是理解文字信息速度的7到10倍[2],所以对于聋哑人而言文字短信方式并不是有效的移动通信方式。

随着移动通信快速发展和具备视频摄取及播放功能手机的日益普及,聋哑人有望和正常人一样利用手机进行实时双向视频通话,提高通话效率。但是,由于移动通信网络带宽有限且手语视频数据量大,要实现手语视频通话,就必须在保证手语视频可理解性的同时最大化压缩手语视频以满足带宽要求,为此研究人员提出了感兴趣区手语视频编码方法[3-5]。这种方法的主要思想是:提高手语视频中聋哑人关注区域(面部和手部)的视觉质量,同时降低聋哑人不关注区域(背景区)的视觉质量,从而在保证面部和手部编码质量的同时,提高手语视频压缩率。感兴趣区手语视频编码方法的前提条件是快速分割出手语视频中的面部、双手和背景区域。

细胞神经网络(Cellular Neural Network, CNN)[6-7]是一种结构形式为局部连接的神经网络。CNN由于其固有的特点,被广泛应用于图像和视频处理领域[8-14]。文献[15]提出了一种基于CNN的视频分割算法并在基于人眼视觉的系统中得到成功应用,该算法能够快速分割出感兴趣区,而且在精确度上也取得了较好的效果,但对于变化比较快的视频对象,其分割精度并不是很好;文献[16]针对头肩序列的分割提出了一些改进算法,得到了较好的分割结果,但是,对于复杂背景的头肩序列的分割效果并不是很好;文献[17]提出了实时性相对较好的一种针对头肩序列的基于CNN的视频序列分割方法,但是从仿真的结果来看,效果并不是很理想,在抗噪声、阈值喧杂、遮挡处理、阴影处理、多目标分割以及算法速度等方面都有待改进;文献[18]提出了一种基于CNN的视频分割算法,虽然具有较好的分割性能和应用优势,但是由于算法复杂度较高,在实时性方面依然没有取得较好的效果。

为此,本文提出了一种基于CNN的信息融合手语视频分割方法,该方法利用CNN模板,首先在YCbCr空间利用肤色信息初步分割出手语视频中的肤色区域,然后结合运动信息分割出手语视频中的面部和手部。实验结果表明,本文方法能够快速准确地进行手语视频分割。

1 基于细胞神经网络的手语视频分割

首先,在进行肤色检测前,要使肤色能够适应不同光照的变化,对输入的每一帧手语视频采用gray world方法进行颜色

收稿日期:2012-08-22;**修回日期**:2012-10-06。
基金项目:国家自然科学基金资助项目(11072099);甘肃省自然科学基金资助项目(112RJZA033)。
作者简介:张爱华(1964 -),女,河北永年人,教授,博士,主要研究方向:机器视觉信息获取与处理、检测技术、智能信息处理; 雷小亚(1987 -),女,甘肃天水人,硕士研究生,主要研究方向:数字图像处理、图像通信; 陈晓雷(1979 -),男,甘肃兰州人,博士研究生,主要研究方向:数字图像处理、智能信息处理; 陈莉莉(1986 -),女,甘肃武威人,硕士研究生,主要研究方向:数字图像处理、智能信息处理。

均衡;然后,考虑到 YCbCr 是亮度与色度分离的颜色空间,而利用色度分量将会使肤色具有很好的聚类性,因此在 YCbCr 颜色空间进行基于 CNN 的肤色检测;由于背景中类肤色点的存在,接下来需要对肤色检测完的结果进行基于 CNN 的运动检测,以有效滤除背景中类肤色点的干扰;但提取出的视频对象轮廓容易出现“空洞”和“重影”现象,很难保证其完整性,因此需要再对其进行中值滤波;最后则采用形态学处理方法进行空洞填充和边界平滑,最终得到手语视频中人的面部和手部区域。基于 CNN 的信息融合手语视频分割流程如图 1 所示。

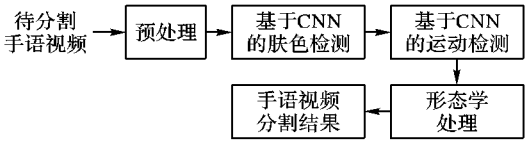


图 1 基于 CNN 的信息融合手语视频分割流程

1.1 基于 CNN 的肤色检测

1.1.1 算法原理

利用 Y、C_b 和 C_r 分量提取模板将输入的 YCbCr 手语视频分为 Y、C_b 和 C_r 三个分量,然后对 C_b 和 C_r 分量进行阈值判断($80 \leqslant C_b \leqslant 130$ 并且 $140 \leqslant C_r \leqslant 180$)将手语视频分为肤色区域和非肤色区域;接着用膨胀、腐蚀和区域填充模板对得到的肤色区域进行优化,最终得到肤色分割结果。该算法的流程如图 2 所示。

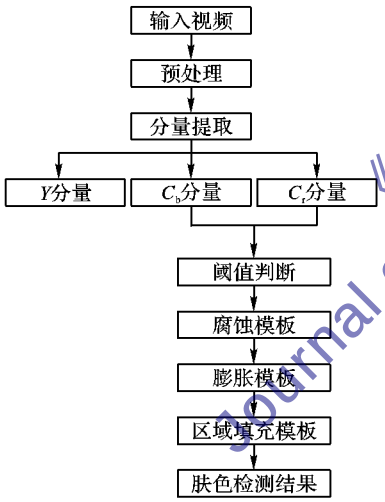


图 2 基于 CNN 的肤色检测流程

设输入的当前帧为 $f(x,y)$,通过阈值判断之后得到的二值图像为 $g(x,y)$,最终肤色检测完的结果为 $d(x,y)$,基于 CNN 的肤色检测模板运算过程为:

$$d(x,y) = \mathrm{FILL}[\mathrm{DILT}[\mathrm{EROISON}[g(x,y)]]] \tag{1}$$

1.1.2 模板设计

肤色检测算法用到的模板有:Y、C_b 和 C_r 分量提取模板、腐蚀模板、膨胀模板以及区域填充模板。Y、C_b 和 C_r 分量提取模板是 CNN 模板库中的自带模板,可以直接使用。本文主要设计了腐蚀模板、膨胀模板和区域填充模板。

1)腐蚀模板。CNN 腐蚀模板属于稳态输出型模板,是针对图像局部特征运算的模板,其重要作用是去除简单的背景。

腐蚀模板的输入为

$u_{ij} = f(x,y), f(x,y) \in \{-1,1\}; 0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

初始状态为

$x_{ij}(0) = 0$

输出为腐蚀后的二值图像。

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

CNN 腐蚀模板的结构为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad I = -4$$

2)膨胀模板。同腐蚀模板,CNN 膨胀模板也属于稳态输出型模板,是针对图像局部特征运算的模板,它的主要作用是连通运动区域间断的边缘,为下一步的操作做好准备。

膨胀模板的输入为

$u_{ij} = f(x,y), f(x,y) \in \{-1,1\}; 0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

初始状态为

$x_{ij}(0) = 0$

输出为膨胀后的二值图像,使 $t \to T = 1$。

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

CNN 膨胀模板的结构为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 \end{bmatrix}, \quad I = 1$$

3)区域填充模板。CNN 区域填充模板是二值图像处理模板,它的功能是使由纯黑色像素包围的区域中的所有纯白色像素变为纯黑色,从而达到闭合区域内部填充的目的。

CNN 区域填充模板的输入为

$u_{ij} = f(x,y), f(x,y) \in \{-1,1\}; 0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

初始状态为

$x_{ij}(0) = 1$

输出为填充后所得的二值图像,使 $t \to \infty$。

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

CNN 区域填充模板的结构为

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I = 3$$

1.2 基于 CNN 的运动检测

1.2.1 算法原理

首先利用负片模板对肤色检测结果进行处理,然后利用标量加合模板实现差分运算,得到差分灰度图像;接着用二值化模板进行阈值判断,得到二值图像;最后利用几种形态学处理模板消除孤立噪声点,最终得到手语视频分割结果。该算法流程如图 3 所示。

设 $f(x,y,k)$ 表示当前帧,$f(x,y,k+1)$ 表示后一帧,$D_{(k+1,k)}$ 为最后运动检测完的结果,则基于 CNN 的运动检测模板运算过程如下所示:

$$\begin{aligned} D_{(k+1,k)} = \mathrm{FILL}[\mathrm{EDGE}[\mathrm{DILT}[\mathrm{AND}[[\mathrm{THH}[\mathrm{ADD}[f(x,y,k+1), \\ \mathrm{REV}[f(x,y,k)]]]], \mathrm{THL}[\mathrm{ADD}[f(x,y,k+1), \\ \mathrm{REV}[f(x,y,k)]]]]]]]] \end{aligned} \tag{2}$$

1.2.2 模板设计

运动检测算法用到的模板有负片模板、标量加合模板、二值化模板、逻辑与模板、膨胀模板、边缘检测模板和区域填充

模板,本文主要设计了负片模板、标量加合模板、二值化模板和逻辑与模板。

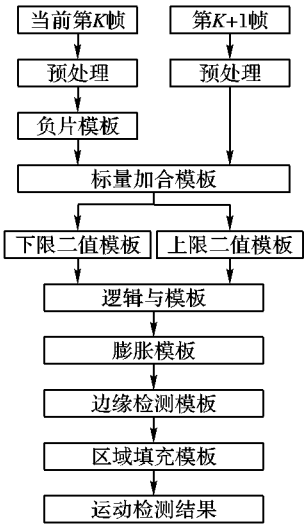


图 3　基于 CNN 的运动检测流程

1)负片模板。负片模板运算记为 REV[·],即 REV[$f(x,y)$] = $-f(x,y)$,其中负片模板的输入为灰度图像:

$u_{ij} = f(x,y)$; $f(x,y) \in [-1,1]$

初始状态为

$x_{ij}(0) = 0$

输出为

$y_{ij}(t) = -f(x,y)$

模板在图像边界时的条件为固定条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

CNN 负片模板可以表示为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I = 0$$

2)标量加合模板。标量加合模板记为 ADD[·],其作用是与负片模板相配合,实现差分运算,对于灰度图像可以表示为:ADD[$f_1(x,y), f_2(x,y)$] = $f_1(x,y) + f_2(x,y)$。

CNN 标量加合模板的输入为

$u_{ij} = f_2(x,y), f_2(x,y) \in [-1,1]$;
$0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

初始状态为

$x_{ij}(0) = f_1(x,y), f_1(x,y) \in [-1,1]$;
$0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

输出为

$y_{ij}(t) = f_1(x,y) + f_2(x,y), y_{ij}(t) \in [-1,1]$

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

标量加合模板的结构为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I = 0$$

3)二值化模板。二值化模板的功能是将输入的灰度图像以一定的阈值转化成二值图像。

CNN 二值化模板的输入为

u_{ij} = 任意值

初始状态为

$x_{ij}(0) = f(x,y), f(x,y) \in [-1,1]$;
$0 \leqslant x \leqslant M, 0 \leqslant y \leqslant N$

输出为

$$y_{ij} = \begin{cases} -1, & f(x,y) \leqslant t \\ 1, & f(x,y) > t \end{cases}$$

其中 t 为阈值。

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

二值化模板的结构为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 200 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$I = -200I^*$; $-1 < I^* < 1$

在这里,上下限二值化模板的阈值分别为 $I = -20$ 和 $I = 4$。

4)逻辑与模板。逻辑与模板的功能是针对两幅二值图像,将其对应像素点值均为 1 的点标记出来。

CNN 逻辑与模板的输入为

$u_{ij} = f_1(x,y), f_1(x,y) \in \{-1,1\}$

初始状态为

$x_{ij}(0) = f_2(x,y), f_2(x,y) \in \{-1,1\}$

输出为

$$y_{ij}(t) = \begin{cases} -1, & f_1(x,y) = -1 \text{ 且 } f_2(x,y) = -1 \\ 1, & \text{其他} \end{cases}$$

模板在图像边界的条件为固定边界条件,即

$x_{i^*j^*}(t) = 0, u_{i^*j^*} = 0$; i^*j^* 表示边界胞元

CNN 逻辑与模板为:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad I = 1$$

2　实验结果

本文利用细胞神经网络工具箱 MatCNN 在 Matlab 7.8 上选取 Irene 和 Silent 两个 QCIF 分辨率为 176 × 144 的标准手语视频进行实验,图 4 为 Irene 检测结果,图 5 为 Silent 检测结果。从实验结果可见:仅采用肤色检测,背景中将存在很多的孤立噪声点;结合运动信息之后,这些噪声点被明显滤除。在相同的实验环境下,对采用肤色和运动信息相融合的文献[19]的算法也进行了测试,结果如图 4(d)和 5(d)所示。可以看出,本文算法效果明显优于文献[19]算法的结果。

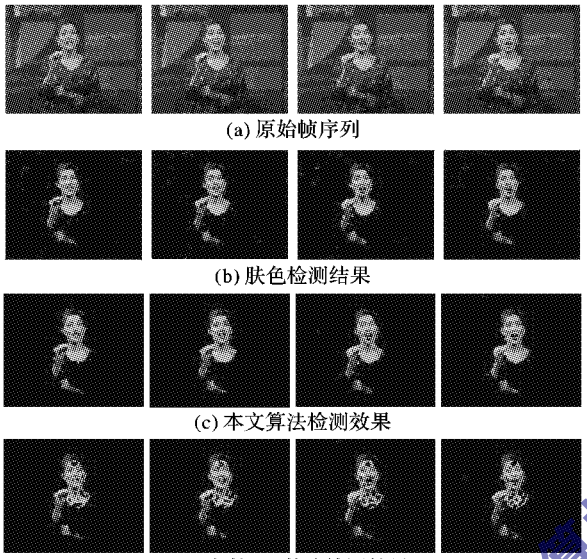
另外表 1 和表 2 分别给出了肤色检测和运动检测所使用模板在 ACE4K 芯片上的运算时间估算值。从实验结果可知,采用本文方法,分割一帧手语视频的时间总计约为 132.41 μs,这是其他方法无法达到的,所以,本文算法完全可以满足实时性要求。

表 1　肤色检测部分各个模板物理运行时间

CNN 模板	运行一次的时间/μs	迭代次数	总运行时间/μs
分量提取模板	0.20	3	0.60
腐蚀模板	0.83	5	4.15
膨胀模板	0.83	10	8.30
区域填充模板	0.83	50	41.50
总计	—	—	54.55

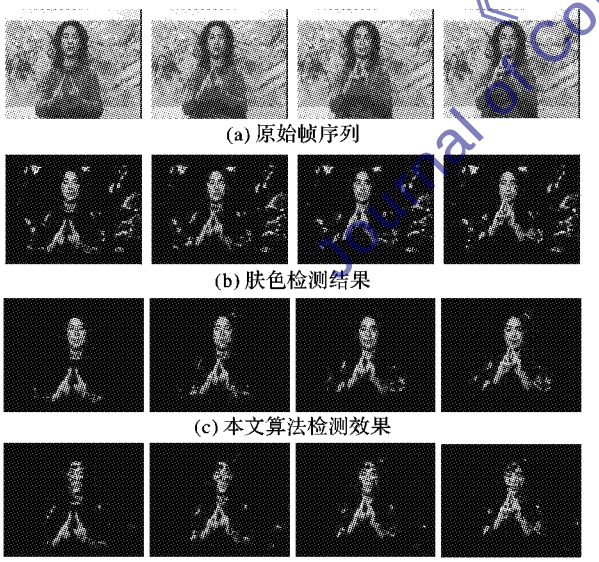
表 2 运动检测部分各个模板物理运行时间

CNN 模板名称	运行一次的时间/μs	迭代次数	总运行时间/μs
负片模板	6.60	1	6.60
标量加合模板	6.60	1	6.60
二值化模板	6.60	2	13.20
逻辑与模板	0.83	1	0.83
膨胀模板	0.83	10	8.30
边缘检测模板	0.83	1	0.83
区域填充模板	0.83	50	41.50
总计	—	—	77.86



(a) 原始帧序列 (b) 肤色检测结果 (c) 本文算法检测效果 (d) 文献[19]算法检测效果

图 4 手语视频 Irene 的检测结果



(a) 原始帧序列 (b) 肤色检测结果 (c) 本文算法检测效果 (d) 文献[19]算法检测效果

图 5 手语视频 Silent 的检测结果

3 结语

本文提出了一种基于细胞神经网络的手语视频分割方法,该方法将肤色检测和运动检测相结合,有效地去除了背景区域中的干扰;同时,由于 CNN 的引人,使得计算速度更快,具有很好的实时性。实验结果表明,本文提出的方法能够快速准确地分割出手语视频中的面部和手部。

参考文献:

[1] von AGRIS U, ZIEREN J, CANZLER U, *et al*. Recent developments in visual sign language recognition[J]. Universal Access in the Information Society, 2008, 6(4): 323 – 362.

[2] LADNER R E. Communication technologies for people with sensory disabilities[J]. Proceedings of the IEEE, 2009, 100(4): 957 – 973.

[3] SAXE D M, FOULDS R A. Robust region of interest coding for improved sign language telecommunication[J]. IEEE Transactions on Information Technology in Biomedicine, 2002, 6(4): 310 – 316.

[4] HABILI N, LIM C C, MOINI A. Segmentation of the face and hands in sign language video sequences using color and motion cues [J]. IEEE Transactions on Circuits Systems Video Technology, 2004, 14(8): 1086 – 1097.

[5] 曹昕燕, 赵继印, 李敏. 基于肤色和运动检测技术的单目视觉手势分割[J]. 湖南大学学报: 自然科学版, 2011, 38(1): 78 – 83.

[6] CHUA L O, YANG L. Cellular neural networks: theory[J]. IEEE Transactions on Circuits and Systems, 1988, 35(10): 1257 – 1272.

[7] CHUA L O, YANG L. Cellular neural networks: applications[J]. IEEE Transactions on Circuits and Systems, 1988, 35(10): 1273 – 1290.

[8] KIM H, ROSKA T, CHOU L O, *et al*. Automatic detection and tracking of moving image target with CNN-UM via target probability fusion of multiple features[J]. International Journal of Circuit Theory and Applications, 2003, 31(4): 329 – 346.

[9] COSTANTINI G, CASALI D, PERFETTI R. Analogic CNN algorithm for estimating position and size of moving objects[J]. International Journal of Circuit Theory and Applications, 2004, 32(6): 509 – 522.

[10] ANZALONE A, BIZZARRI F, STORACE M, *et al*. A cellular non-linear network for image fusion based on data regularization [J]. International Journal of Circuit Theory and Applications, 2006, 34(5): 533 – 546.

[11] KOSKINENA L, PAASIOB A, HALONEN K. CNN-type algorithms for H.264 variable block-size partitioning[J]. Signal Processing: Image Communication, 2007, 22(9): 797 – 808.

[12] YU S-N, LIN C-N. An efficient paradigm for wavelet-based image processing using cellular neural networks[J]. International Journal of Circuit Theory and Applications, 2010, 38(5): 527 – 542.

[13] 孙继平, 吴冰, 刘晓阳. 基于膨胀/腐蚀运算的神经网络图像预处理方法及其应用研究[J]. 计算机学报, 2005, 28(6): 985 – 990.

[14] 姜庆玲, 刘万军, 张闯. 基于 CNN 的分块自适应彩色图像边缘检测的研究[J]. 计算机应用研究, 2009, 26(3): 1131 – 1134.

[15] KARABIBER F, GRASSI G, VECCHIO P, *et al*. Implementation of a cellular neural network-based segmentation algorithm on the bio-inspired vision system [J]. Journal of Electronic Imaging, 2011, 20(1): 122 – 128.

[16] 王慧. 基于细胞神经网络的视频对象分割算法的研究[D]. 上海: 上海大学, 2003.

[17] 张庆利. 视频对象自动分割技术及其细胞神经网络实现方法的研究[D]. 上海: 上海大学, 2005.

[18] 蒋文艳. 一种基于 CNN 的视频运动对象分割的研究[D]. 南宁: 广西大学, 2007.

[19] AKYOL S, ALVARADO P. Finding relevant image content for mobile sign language recognition[C]// Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Application. [S.l.]: IASTED, 2001: 48 – 52.