DOI:10.15961/j.jsuese.2014.02.027

第46卷 第2期 四川大学学报(工程科学版) Vol. 46 No. 2
2014年3月 JOURNAL OF SICHUAN UNIVERSITY (ENGINEERING SCIENCE EDITION) Mar. 2014

文章编号:1009-3087(2014)02-0066-06

适应手脸遮挡手语视频的手势检测算法

陈晓雷[1],张爱华[1*],林冬梅[1],杨欣翥[2]

(1. 兰州理工大学 电气工程与信息工程学院,甘肃 兰州 730050; 2. 兰州交通大学 机电工程学院,甘肃 兰州 730070)

摘 要: 针对手脸遮挡条件下的手语视频手势检测问题,提出一种基于力场(force field)转换的手势检测算法。首先分别计算手脸遮挡帧和纯脸部帧的力场图像,然后将力场图像分块并统计各分块直方图特征,再将相同空间位置的分块直方图对应相减,得到各分块直方图灰度分量差,最后将各分块直方图灰度分量差与灰度阈值进行比较获得手部位置。实验证明该算法能够实时进行手脸遮挡条件下的手势检测。

关键词: 手势检测; 手语视频; 力场转换; 直方图

中图分类号: TP391 **文献标志码:** A

Hand Detection Algorithm During Hand over Face Occlusion in Sign Language Video

CHEN Xiaolei[1], *ZHANG Aihua*[1*], *LIN Dongmei*[1], *YANG Xinzhu*[2]

(1. School of Electrical and Info. Eng., Lanzhou Univ. of Technol., Lanzhou 730050, China;
2. School of Mechatronic Eng., Lanzhou Jiaotong Univ., Lanzhou 730070, China)

Abstract: Based on image force field transformation, a novel algorithm was proposed to track the hand during hand over face occlusion in sign language video. First, the frames with a hand occluding the face and those with only a face were transformed to force field images. Then the force field images were partitioned into sub-images and the histograms of each sub-image were calculated. For each sub-image, the histogram of frame with only a face was subtracted from the frame with a hand occluding the face to get the difference histogram. Finally, for each sub-image the difference histogram was compared to threshold to get the position of the hand. Experimental results showed that the proposed algorithm is capable of real-time tracking of hand.

Key words: hand detection; sign language video; force field transformation; histogram

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。针对手语视频的手势检测是手语识别、移动手语视频通信[1]中极为关键的问题,手势检测的好坏直接影响这些系统的工作性能。近年来,国内外研究人员提出了一些针对手语视频的手势检测方法。Habili等[2]提出利用肤色和运动信息进行手势分割,曹昕燕[3]在分析肤色信息特征和手势运动特性基础上,构建了一种基于单目视觉的手势检测方法,张爱华等[4]提出了一种基于细胞神经网络(cellular neural network,CNN)的快速手语视频检测方法。但是,这些方法仅适用于手脸未遮挡条件下的手语视频手势检测,存在一定的局限性。手语视频存在手脸遮挡现象,当手脸遮挡时,由于手部和脸部具有相同的肤色和纹理,并且手部具有多种形态,所以要从手脸遮挡手语视频中检测出手部难度较大。Holden等[5]通过结合运动信息和Snake算法来检测发生手脸遮挡时的手部,但是该方法只有在相邻视频帧之间的手部形状变化非常小的情况下才能获得好的性能,并不适用于手部形状变化非常大的手语视频。Gonzalez等[6]同时利用肤色和边缘信息来解决手脸遮挡条件下的手势检测问题,但是该方法分割后的图像存在空洞和方块效应。Smith等[7]利用图像力度场(force field)转换和高斯混合(mixture of Gaussians,MoG)模型来解决手脸遮挡问题,但是该方法计算量大,不能实时分割手脸

收稿日期: 2013-07-04

基金项目: 国家自然科学基金资助项目(61302116; 61365003); 教育部博士点基金资助项目(20116201110002); 甘肃省自然科学基金资助项目(1212RJYA026; 1212RJZA050)

作者简介: 陈晓雷(1979—),男,讲师,博士. 研究方向: 图像处理; 机器视觉. E-mail: chenxl703@gmail.com

* 通信联系人

遮挡视频,并且手部的形态仅限于完全张开的双手。Hussain 等[8]提出图像力度场结合局部二进制模式(local binary patterns,LBP)的手脸遮挡条件下的手势检测算法,与文献[7]相比提高了算法的实时性。作者在前人研究的基础上,提出一种新颖的手脸遮挡条件下的手势检测算法。实验结果表明,本文算法的检测准确率高,并且检测速度更快。

1　图像力场转换

力场转换由 Hurley 等[9]提出并主要用于人耳识别[10-13],利用它进行手势检测。力场转换的思想是把图像视为一个能量场,像素对其他像素之间有力的作用并且符合引力定律,即大小正比于该像素的灰度值,反比于像素之间距离的平方。每一个像素所受的力是其他所有像素对其作用力的矢量和。

由于手部和脸部具有不同的区域结构,当手部遮挡脸部时区域结构的变化形成图像灰度分布的变化,将分辨率为 $M \times N$ 的视频帧看作由 $M \times N$ 个像素点组成的矩阵,假设每个像素点的灰度值均对图像中其他像素点的灰度值产生影响,即对图像其他点处的能量有所贡献,为图像能量的源。则位于图像矢量位置 $\boldsymbol{r}_i$ 的源像素点灰度值 $I(\boldsymbol{r}_i)$ 对位于图像矢量位置 $\boldsymbol{r}_j$ 的像素所产生的能量,可用式(1)描述:

$$E_i(\boldsymbol{r}_j) = \frac{I(\boldsymbol{r}_i)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1}$$

因为每个像素点均为能量源,那么位于图像矢量位置 $\boldsymbol{r}_j$ 的像素点将受到其他 $M \times N - 1$ 个像素点灰度值的共同作用,即 $\boldsymbol{r}_j$ 之外的任何像素点均对 $\boldsymbol{r}_j$ 处的能量有贡献,则 $\boldsymbol{r}_j$ 处的总能量为:

$$E(\boldsymbol{r}_j) = \sum_{i=0}^{M\times N-1} E_i(\boldsymbol{r}_j) = \sum_{i=0}^{M\times N-1} \frac{I(\boldsymbol{r}_i)}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{2}$$

根据上述图像能量方程,借鉴分子力学原理,将 $\boldsymbol{r}_j$ 之外的任何像素点对 $\boldsymbol{r}_j$ 点所产生的作用力表示为:

$$F(\boldsymbol{r}_j) = \frac{\mathrm{d}E(\boldsymbol{r}_j)}{\mathrm{d}\boldsymbol{r}_j} = \sum_{i=0}^{M\times N-1} I(\boldsymbol{r}_i)\ \frac{\boldsymbol{r}_i - \boldsymbol{r}_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3} \tag{3}$$

对图像中所有像素点应用式(3)进行转换,即得到整幅手语视频帧的矢量力场,矢量大小经量化归一到灰度后可以得到力场转换手语视频。图 1 是原始手语图像和经过力场转换后的手语图像。可以看出,经过力场转换之后的手语图像有明显的滤波平滑效果,这是因为利用式(3)进行引力计算时,像素的灰度值是其周围所有像素灰度值的加权和,权值的大小与像素之间的距离成反比,因而距离较近的像素起主要作用。当图像的某一区域变化平缓时,像素周围的灰度值接近,受力方向对称、互相抵消,最终所受合力较小,从而产生滤波平滑效果。

(a) 原始图像　　(b) 力场转换后图像

图 1　手语图像进行力场转换的结果

Fig. 1　Result of sign language image force field transformation

2　基于手语力场图像直方图特征的手语视频手势检测算法

2.1　基于分块的手语力场图像直方图特征

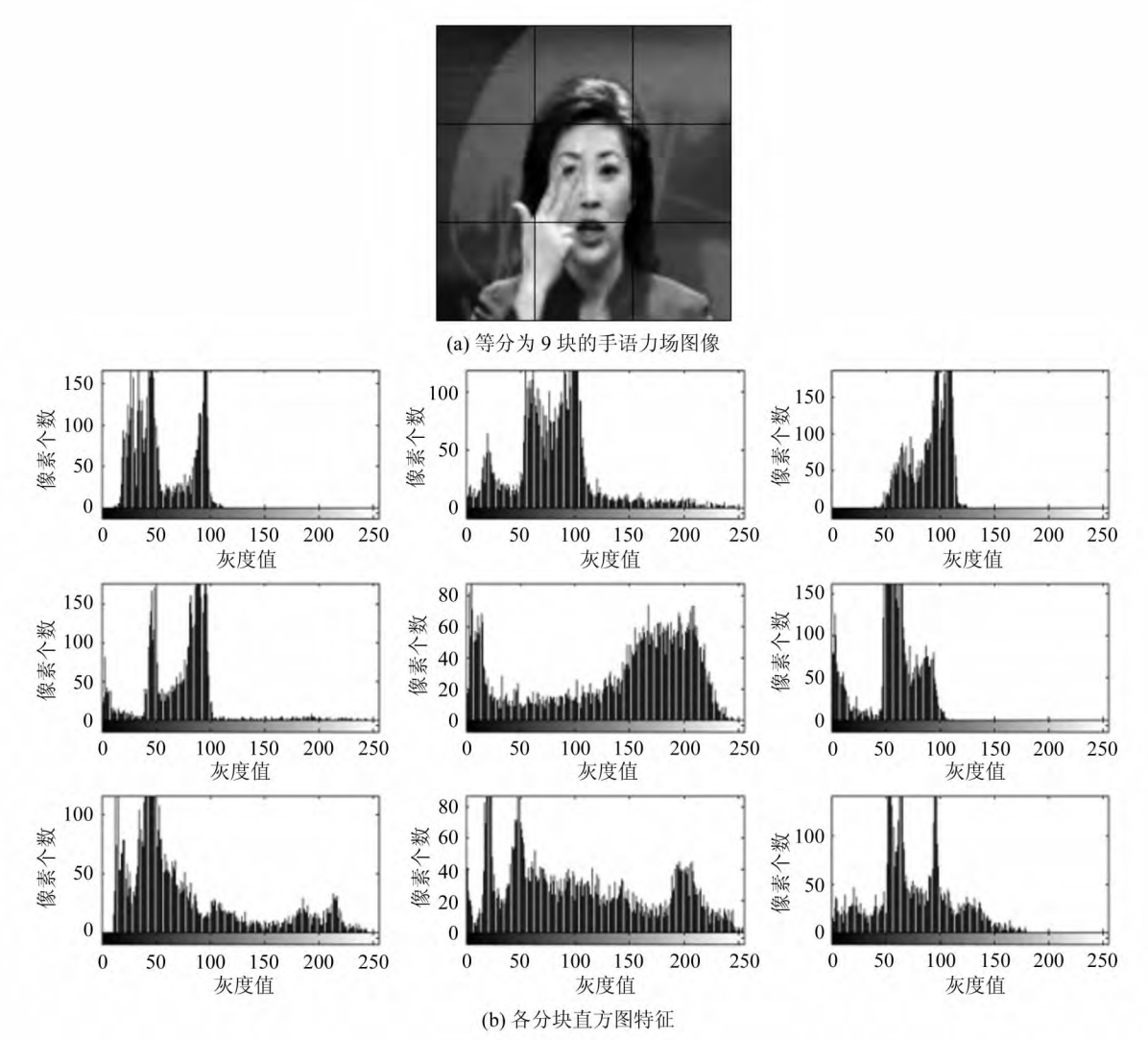
得到手语力场图像后,就可以用直方图作为特征向量来描述该图像,即将力场图像中的所有像素,按照灰度值的大小,统计其所出现的个数,从而得到对其灰度信息分布的统计。然而直方图特征的缺点是它只反映图像中不同灰度值的出现次数,而不反映某一灰度值像素所在位置。也就是说,它只包含了图像中某一灰度值的像素出现次数,而丢失了其所在位置信息。为了保持直方图特征向量的空间结构关系,采用了分块策略,即对力场图像,首先将它等分成小块的图像区域,然后在每一个局部的小块中统计直方图信息。图 2 给出了提出的基于分块的力场图像直方图特征。可见,由于采取分块的策略,手语力场图像全局的空间结构位置关系都被有效的保留下来。

2.2　手脸遮挡手语视频的手势检测算法

通过对手语视频进行分析可以发现其具有如下特征: 聋哑人进行手语会话过程中,一开始是先出现纯脸部图像,然后才使用单手或双手打出手语,在打手语过程中,手部形状和位置变化非常大,脸部表情和位置变化小,发生手脸遮挡时的脸部表情和位置与初始时刻纯人脸图像中的脸部表情和位置几乎完全相同。充分利用手语视频的上述特征,提出了如图 3 所示的手脸遮挡手语视频的手势检测算法。

算法步骤如下:

Step 1: 对手脸遮挡图像和纯脸部图像分别进行力场转换得到力场图像。



(a) 等分为9块的手语力场图像

(b) 各分块直方图特征

图2 力场图像分块及各分块直方图特征

Fig. 2 Partition of a force filed image and histogram features of each sub-image

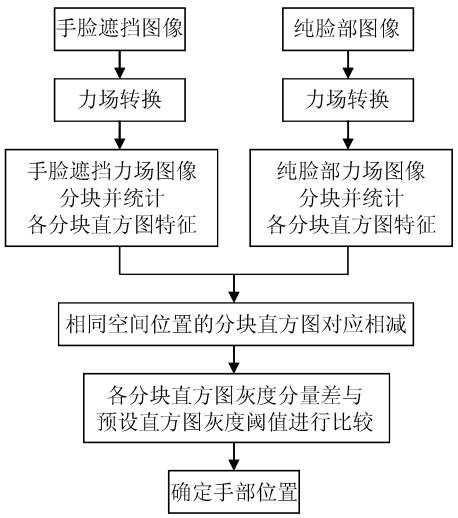


图3 手脸遮挡手语视频的手势检测算法

Fig. 3 Hand detection algorithm during hand over face occlusion in sign language

Step 2: **将手脸遮挡力场图像和纯脸部力场图像等分为9个分块,统计各分块直方图特征。**

Step 3: **将手脸遮挡力场图像和纯人脸力场图像相同空间位置的分块直方图对应相减,得到分块直方图灰度分量差。设初始时刻 t_0 未发生手脸遮挡,将该时刻纯脸部力场图像第 n 个区域的直方图灰度分量记为 $f_{H(n,t_0)}$,将 k 时刻手脸遮挡力场图像第 n 个区域直方图灰度分量记为 $f_{H(n,t_k)}$,定义分块直方图灰度分量差如下:**

$$\Delta f_n = |f_{H(n,t_k)} - f_{H(n,t_0)}| \tag{4}$$

其中,$n \in [0,9]$。

Step 4: **各分块直方图灰度分量差 Δf_n 和预设的直方图灰度阈值 f_T 进行比较,如果某分块直方图灰度分量差 Δf_n 大于预设的直方图灰度阈值 f_T,则表明该区域出现手部。相反,如果某分块直方图灰度分量差 Δf_n 小于预设的直方图灰度阈值 f_T,则表明该区**

域未出现手部。即：

$$P(n) = \begin{cases} 1, & \Delta f > f_T; \\ 0, & \text{else} \end{cases} \tag{5}$$

其中，$P(n) = 1$ 表示该区域出现手部，$P(n) = 0$ 表示该区域未出现手部。

3　实验结果与分析

3.1　手脸遮挡手语视频库及算法测试硬件环境

由于目前国内外尚缺少可使用的手脸遮挡手语视频库，所以以《大家学手语》视频录像[14]为基础自建了手脸遮挡手语视频库。该手脸遮挡手语视频库以中国手语为准，分 22 个方面的内容介绍 100 句日常生活用语的手语打法。视频帧率为 20 帧/s，分辨率为 256 × 256，总时长为 1 h 27 min，总帧数为 104 400帧，其中，75% 以上视频帧存在手脸遮挡现象。图 4 为从自建手语视频库中选取的部分手语图像，系第 4 课“赞美”视频流的第 832、858、944、957、960、985、1 036、1 194 帧。

本文算法测试的主要硬件环境是 Intel Core3 Duo CPU，主频 1.7 GHz，内存 4 G DDR3，AMD Radeon HD 7570M 显卡。

3.2　本文算法参数分析实验

选取 3 组测试视频比较了不同的力场图像分块数对手部检测准确率和检测时间的影响。实验结果如表 1 所示。

表 1　不同分块方式下的算法性能

Tab. 1　Performance of proposed algorithm at different partitions

测试视频	力场图像分块数	手部检测准确率/%	检测一帧平均时间/s
第 3 课 欢迎（共 780 帧）	2 × 2	75	0.272 4
	3 × 3	92	0.581 4
	4 × 4	95	0.930 6
第 4 课 赞美（共 1 640 帧）	2 × 2	88	0.279 0
	3 × 3	95	0.584 3
	4 × 4	98	0.931 2
第 9 课 道别（共 2 460 帧）	2 × 2	78	0.776 3
	3 × 3	94	1.709 8
	4 × 4	96	2.764 0

从实验结果可以看出，力场图像分块数对手部检测准确率及检测时间有一定影响，当分块数增加时，手部检测准确率随之提高，这说明分块数增多能够更好的保持直方图特征向量的空间结构关系，有利于提高检测准确率。但是随着分块数的增加，检测时间也随之增加。综合考虑检测准确率和检测时间，最终选择分块数为 3 × 3，即将图像等分为 3 × 3 共 9 个区域。

图 4　手脸遮挡手部视频库部分图像

Fig. 4　Part of images of the constructed hand-over-face gestures database

图 5 给出了当力场图像分块数为 9 时，本文算法的检测结果。由图 5(d) 各分块直方图灰度分量差可以看出，区域 4、5、7 和 8 的灰度分量差大于阈值 150，区域 1、2、3 和 6 的灰度分量差小于阈值 150，依据本文算法可以判断出手部出现在区域 4、5、7 和 8，这种判断结果与图 5(a) 中手部区域出现的位置相吻合，这说明本文算法是有效的。对图 5(d) 各分块直方图灰度分量差进一步分析可以得到：区域 8 直方图灰度分量差的灰度级值最大，接近250；区域7直方图灰度分量差的灰度级

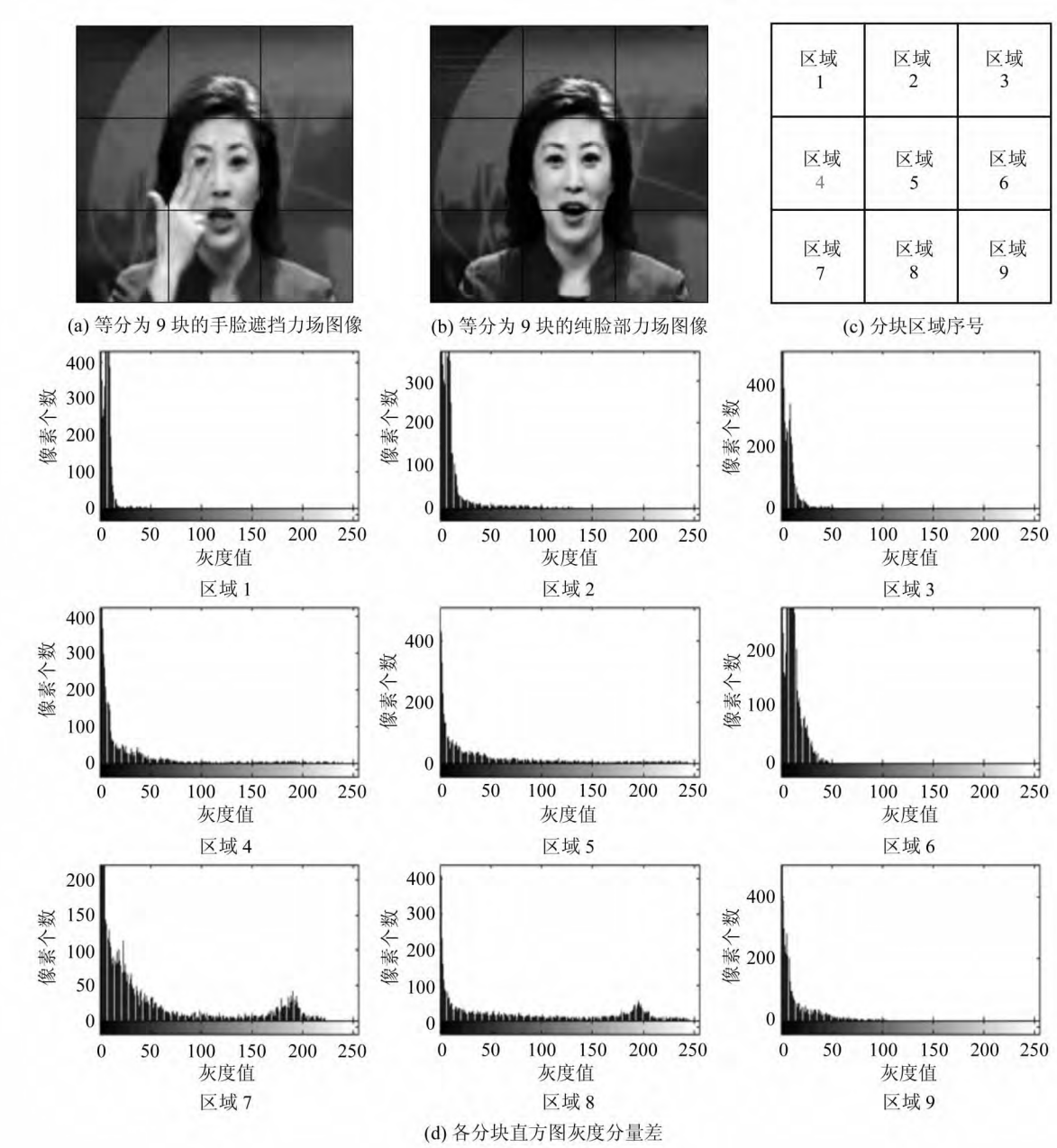


图5 本文算法的实验结果

Fig. 5 Experiment results of the proposed algorithm

值小于区域8,接近225;区域5直方图灰度分量差的灰度级值虽然与区域8相似,也接近250,但其直方图灰度分量差中具有高灰度值(灰度值为200)的像素个数小于区域8和区域7;区域4直方图灰度分量差的灰度级值小于区域5,这就说明手部区域在各小块中遮挡面积的大小顺序为区域8、7、5、4,与图5(a)中所反映信息也吻合。

3.3 本文算法和其他算法的比较

比较本文算法与文献[7]和[8]所提算法的性能。为公平起见,所有算法都在相同的硬件环境下进行测试,并且都没有做算法优化。实验结果如表2所示。

表2 本文算法与其他算法的比较

Tab. 2 Performance comparison of different methods

检测算法	手部检测准确率/%	检测一帧平均时间/s
文献[7]	80	9.473 5
文献[8]	97	0.658 6
本文算法	95	0.584 3

从表2中实验结果可见,与文献[7]相比,文献[8]算法与本文算法都能以更快的检测速度获得更高的手部检测准确率。与文献[8]相比,本文算法的手部检测准确率降低2%,检测速度提高12.7%。

4 结　论

手部和脸部具有相同的肤色和纹理，并且手部具有多种形态，要从手脸遮挡手语视频中检测出手部难度较大。为了解决这一难题，利用手部和脸部具有不同的区域结构，而力场可以凸显图像中区域结构的特性，提出了一种新的手脸遮挡条件下的手势检测算法。实验表明，本算法的检测准确率和检测时间可以满足实时的手势检测要求，说明本算法具有较好的实用性。如何自适应地确定预设灰度阈值将是下一步的研究方向。

参考文献:

[1] Ciarmello F M，Hemami S S. A computational intelligibility model for assessment and compression of American sign language video [J]. IEEE Transactions on Image Processing，2011，20(11)：3014 – 3028.

[2] 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 and Systems for Video Technology，2004，14(8)：1086 – 1097.

[3] Cao Xinyan，Zhao Jiyin，Li Min. Monocular vision gesture segmentation based on skin color and motion detection [J]. Journal of Hunan University: Natural Sciences，2011，38(1)：78 – 83. [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.]

[4] Zhang Aihua，Lei Xiaoya，Chen Xiaolei，et al. Fast segmentation of sign language video based on cellular neural network [J]. Journal of Computer Applications，2013，33(2)：503 – 506. [张爱华，雷小亚，陈晓雷，等. 基于细胞神经网络的快速手语视频分割方法 [J]. 计算机应用，2013，33(2)：503 – 506.]

[5] Holden E，Lee G，Owens R. Australian sign language recognition [J]. Machine Vision and Applications，2005，16(5)：312 – 320.

[6] Gonzalez M，ollet C，Dubot R. Head tracking and hand segmentation during hand over face occlusion in sign language [J]. Lecture Notes in Computer Science，2012(6553)：234 – 243.

[7] Smith P，Lobo N，Shah M. Resolving hand over face occlusion [J]. Image and Vision Computing，2007(25)：1432 – 1448.

[8] Hussain A，Abbasi A R，Afzulpurkar N. Detecting and interpreting self-manipulating hand movements for student's affect prediction [J]. Human-centric Computing and Information Sciences，2012，2(1)：14.

[9] Hurley D J，Nixon M S，Carter J N. Force field feature extraction for ear biometrics [J]. Computer Vision and Image Understanding，2005，98(3)：491 – 512.

[10] Zhu Haihua，Li Yajuan，Song Zhijian. Ear recognition based on image force field transformation [J]. Acta Automatica Sinica，2006，32(4)：512 – 518. [朱海华，李雅娟，宋志坚. 基于图像力场转换的耳廓图像识别 [J]，自动化学报，2006，32(4)：512 – 518.]

[11] Mo Xingjun. Research of universal gravitation application on ear image recognition [D]. Chongqing: Chongqing University，2007. [莫兴俊. 万有引力在人耳图像识别中的应用研究 [D]. 重庆: 重庆大学，2007.]

[12] Dong Jiyuan，Mu Zhichun，Wang Yu. Multi-pose ear recognition based on force field convergence feature [J]. Application Research of Computers，2009，26(6)：2370 – 2375. [董冀媛，穆志纯，王瑜. 基于力场收敛特征的多姿态人耳识别 [J]. 计算机应用研究，2009，26(6)：2370 – 2375.]

[13] Nixon M S，Liu X U，Direkoglu C，et al. On using physical analogies for feature and shape extraction in computer vision [J]. The Computer Journal，2011，54(1)：11 – 25.

[14] 葛玉红. 大家学手语 [EB/OL]. (2008 – 10 – 01) [2013 – 06 – 28]. http://www.youku.com/playlist_show/id_1370463.html.

(编辑　杨　蓓)