2018 32nd International Conference on Advanced Information Networking and Applications Workshops

Application of Deep Learning for Infant Vomiting and Crying Detection

Chuan-Yu Chang, Fu-Ren Chen
Dept. of Computer Science and Information Engineering
National Yunlin University of Science and Technology
Yunlin, Taiwan
chuanyu@yuntech.edu.tw

***Abstract*—Obstruction of the nose and mouth by vomit or quilts is the main cause of sudden infant death. Different baby sleeping postures will increase the difficulty of face detection. Therefore, a deep learning neural network is adopted to detect baby faces and a novel algorithm is proposed to detect vomiting of infants in this paper. The SSD + mobilenet network in Teansorflow is adopted. Experimental results show reasonable detection results of the proposed method.**

Keywords—deep learning, face detection, infant vomiting detection

I. INTRODUCTION

As newborn babies cannot take care of themselves, they are easily exposed to dangerous conditions and accidents because of caregivers' negligence. The most common causes of infant accidents are asphyxiation of babies being covered with quilts during sleep, and choking of the trachea when drinking milk. These will cause the baby's brain damage, or even death. Therefore, the prevention of neonatal accidents is very important.

In order to correctly detect the baby's mouth, most of the methods will first retrieve the face, and then extract the eyes, nose and mouth feature points for subsequent identification. Face detection methods can be divided into: skin color detection [1, 2], feature extraction [3, 4], and machine learning [5]. Skin and geometric feature locations are often used to locate facial features [6]. However, the methods of skin color detection and feature extraction are often limited by the influence of the environment, good results are difficult to obtain if lighting conditions cannot be fixed. In recent years, deep learning technologies have developed rapidly. A number of studies have shown that the features captured by deep learning are better than the traditional ones, and the trained models are also more accurate than the previous methods.

This paper uses a deep learning neural network to train a baby face detection model to identify the baby's face, and then extract the baby's mouth with the skin color features and geometric positions to determine whether quilts obscure mouth. The findings of this study can help newborns' parents, nurses and paramedics reduce the burden on newborns.

II. METHODOLOGY

In order to detect whether the baby's nose and mouth is covered with vomit or quilts, we need to detect the position of the baby's face first. As babies retch, the vomit often blocked the nose or mouth and cause suffocation [7]. Therefore, this paper proposed an infant vomiting detection system. Figure 1 shows the system flow chart of the proposed system.

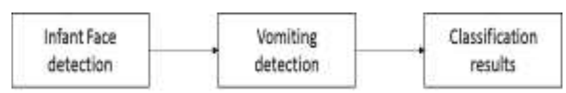

Figure 1. System flow chart of the proposed method

A. Infant Face detection

Neonatal sleeping most of the time, in order to prevent the baby vomiting, parents usually let the baby supine sleep. Different baby sleeping postures will increase the difficulty of face detection. The face detection technique proposed by Viola and Jones [5] has a good performance on the front face of adult, but less effective for the baby's face detection. Therefore, the deep learning is adopted to detect baby faces.

The Object Detection in Teansorflow is used in this paper, which contains the pre-training weight of five network structures including (1) SSD+mobilenet, (2) SSD+inception_v2, (3) R-FCN+resnet101, (4) faster RCNN+resnet101, and (5)faster RCNN+inception+ resnet101[8]. Due to the limitation of the detection speed, the SSD+ mobilenet network architecture [10] is finally adopted for training. The SSD+mobilenet is divided into two parts, of which mobilenetis is for object prediction, and Single Shot MultiBox Detector(SSD) is to determine the classification results. The network architecture is shown in Figure2.

The public face database-WIDERFACE dataset [11] is used for training. Then uses the pre-trained model to detect the infant faces. Figure 3 shows the infant face detection result.

978-1-5386-5395-1/18/$31.00 ©2018 IEEE
DOI 10.1109/WAINA.2018.00158

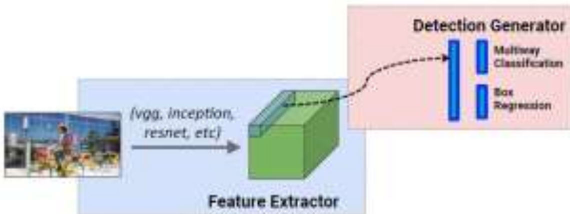

Figure 2. SSD architecture diagram [8]

Figure 3. infant face detection result.

B. Vomiting detection

The Gaussian filtering is firstly applying to the imported images, then color images are converted from RGB to HSV color space. After analyzing the real data, we found that the color of mouth and face are obviously different. The red component is relatively high in mouth. We can use this feature to locate the approximate position of the mouth. A closing operation is applied to refine the detected mouth region. Based on the difference of the consecutive frames to determine whether there is vomit around the mouth.

1) Gaussian filter

Since the influence of external factors, the image will have a lot of noise, resulting in the system cannot accurately capture the location of the mouth. Therefore, Gaussian filter is used to de-noise to improve the accuracy of the captured mouth.

2) Color space

According to the previous research, the vector space formed by RGB can't deal with the image brightness. In order to obtain the distribution of the mouth under various lighting conditions, we convert the color space RGB of the image into the HSV space. The conversion formula is represented as follows:

$$H = \begin{cases} \dfrac{60(G-B)}{V-\min(R,G,B)} & if \quad V = R \\ 120 + \dfrac{60(B-R)}{V-\min(R,G,B)} & if \quad V = G \\ 240 + \dfrac{60(R-G)}{V-\min(R,G,B)} & if \quad V = B \end{cases} \tag{1}$$

$$S = \begin{cases} \dfrac{V-\min(R,G,B)}{v} & if \quad v \neq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

$$V = \max(R,G,B) \tag{3}$$

where R, G, B represents the red, green, and blue components of a color image.

3) Color detection

As the babies' mouth are red, we defined a red range in the HSV space to detect the red range in the infant's image. The detection result is shown in Figure. 4 (a), the detected red range is displayed as white.

4) Closing operation

Infants often open their mouths, and the shadow in the mouth will affect the accuracy of the mouth detection. In order to avoid this problem, the closing operation is used to fill the shadows. Figure 4(a) shows the detected mouth before closing operation. Figure 4 (b) shows the filled mouth area.

(a) (b)

Figure 4. (a) Detected mouth before closing operation, (b)Detected mouth after closing operation

5) Calculate difference value

Figure 5 shows the system flow of vomiting detection. When the baby spits milk, the mouth has a clear white liquid covering the mouth, resulting in a smaller area of the mouth. The detected size of the mouth can be used to judge vomiting or not.

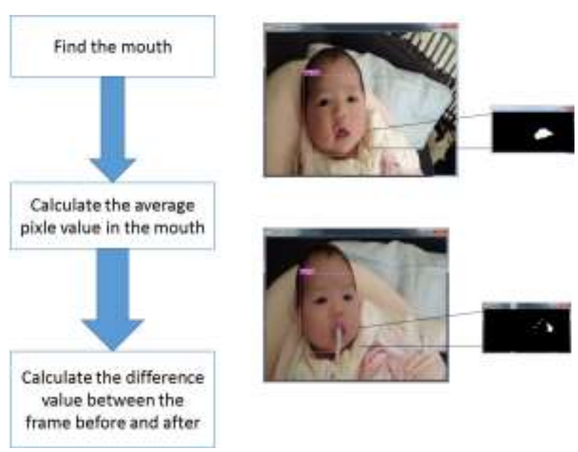

Figure 5. The flow chart of vomiting detection

Calculating the mouth area of each frame, the variation of the mouth area r is obtained by

$$r = \frac{t_2}{t_1} \tag{4}$$

there t_2 and t_1 is the detected mouth area of the previous and current frame. When the r value is less than 0.5, indicates that the vomiting or mouth covering by milk had occurred and the system will issue a warning. If the r value is greater than 1.5, indicates mouth may be open or shaking and the system does not response.

Since babies may fall asleep and their bodies may move occasionally, the proposed detection method is to find a reference images every two seconds, and detect vomit for every 10 images.

III. Experimental Results

In order to verify the performance of the proposed system, two experiments were conducted: (1) baby face detection, (2) infant vomiting detection. Due to the front face provides more information, the proposed method can accurately detect it even in complex environments. Figure 6 (a) shows a baby lying on a patterned pillow. Figure 6 (b) show a baby lying on a white pillow. In both cases the baby's face can be correctly detected. Figure 7 show the detection results of two different shooting angles. The system can correctly detect baby faces under different shooting angles.

(a) (b)

Figure 6. Baby front face detection results, (a) in a complex environment, (b) in a simple environment

(a) (b)

Figure 7. Baby face detection under different angles, (a) Horizontal shot of baby face, (b) Baby face from other angles

We used a six-second video to test the effectiveness of vomiting detection. In the 79th frame, the system correctly detected the baby's face. At the 129th frame, the baby's hand touched the face and the system was judged as normal. At the 172 frames, spit milk occurred and the system detected an abnormality and the status changed to vomiting.

Figure 8 The test result for vomiting detection

IV. Conclusion

The baby's face can be detected only under the simple background and stable illumination condition in traditional methods. This paper uses Tensorflow for infant vomiting detection. Experimental shows the proposed method can detect the baby's face effectively under the complex background and various illumination conditions. The system can help parents monitor the baby whether his/her mouth and nose is covered with vomit or quilt. The system can real-time notification to parents, to improve infant safety and reduce the burden on parents.

References

[1] R. L. Hsu, A. M. Mohamed, and A. K. Jain, "Face Detection in Color Images", *IEEETrans.on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 696-706, 2002.

[2] D. Chai and K. N. Ngan, "Face Segmentation Using Skin-Color Map in videophone Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, pp. 551-564, 1999..

[3] J. M. Lee, J. H. Kim, and Y. S. Moon, "Face Extraction Method Using Edge Orientation and Face Geometric Features",. *Proc. of the International Conference on Convergence Information Technology,* pp.1292 - 1297, 2007.

[4] J. G. Wang and T. N. Tan, "A new face detection method based on shape information," *Pattern Recognition Letters*, Vol. 21, pp. 463-471, 2000.

[5] Paul Viola, M. J. Jones. "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol 57, pp. 137-154, 2002.

[6] M. Hassaballah and Shun Ido, "Eye Detection Using Intensity and Appearance Information," *Proc. of the IAPR Conference on Machine Vision Applications*, pp. 20-22, May, 2009.

[7] http://www.setn.com/News.aspx?NewsID=243952

[8] Huang, Jonathan and Rathod, Vivek and Sun, Chen and Zhu, Menglong and Korattikara, Anoop and Fathi, Alireza and Fischer, Ian and Wojna, Zbigniew and Song, Yang and Guadarrama, Sergio and others, "Speed/accuracy trade-offs for modern convolutional object detectors, " arXiv 2016

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.- Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. *In European Conference on Computer Vision*, pages 21-37. Springer, 2016. 1, 2, 3, 4, 5

[10] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Effi- cient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017

[11] WIDER FACE:A Face Detection benchmark http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace/.