

---

# An Implementation of Cloud Based Camera Monitoring System for Healthcare

Feng Gao\*, Honggang Wang\*, Hua Fang+

\*Department of Electronical and Computer Engineering  
+ Department of Computer and Information Science  
University of Massachusetts Dartmouth

**Abstract.** New babies bring numerous happiness to parents as well as responsibilities. Parents would better fix their eyes on babies to guarantee everything run smoothly during their growing process. But it's hard to handle all other issues meantime looking after baby, especially for working couples. Baby Care Cloud Monitoring System help to solve such problem, then parents could get baby status anywhere anytime. It detects baby's sleeping status through OpenCV computer vision method and baby's crying through SVM Machine Learning method combined with MFCC acoustic character, assisting pediatrician to make a precise diagnose. The whole system consists of wireless terminal running on Openwrt Operating System and the Media Cloud Server built on AWS platform. The terminal captures and sends the live steaming complied by ffmpeg multimedia lib with the RTMP protocol to server. Media Cloud Server analyzes the received video and audio streaming separately, displays result and store the useful information.

**Keywords:** MFCC, SVM, OpenCV, Machine Learning, ffmpeg, RTMP Live Streaming, Cloud, Openwrt, AWS.

## 1 Introduction

Annually, over 12,000 children die caused by unintentional injuries and 9.2 million children with nonfatal injuries are treated for unintentional injuries in United States [1]. The children younger than 4 who occupy over half of those victim are the most vulnerable group [2]. Though caregivers make such efforts to take care of babies, they could not follow every aspect at every moment and carelessness still exist during babies' growth process. Sort of baby care surveillance system is a way to assist the babysitting work.

Baby Care Cloud Monitoring System (BCSS) takes visual and acoustic two aspects to capture and analyze above-mentioned key factors, store videos and analysis result for further use by pediatricians and caregivers. Live streaming captured by wireless web camera is transmitted to cloud server, then video and audio streaming are detected separately, finally processed live streaming is stored or contributed to requested devices by cloud server.

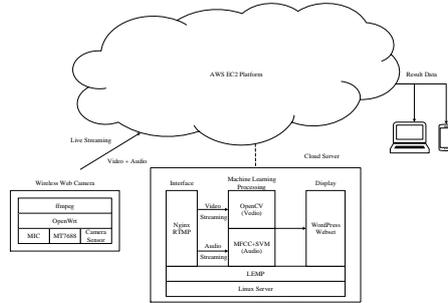
## 2 Methodology

### 2.1 BCCS System Structure

The whole BCCS basically could be divided into two parts: one is the Wireless Web Camera, the other is Cloud Server shown in Fig. 1.

Wireless terminal is built up with three-layer hierarchical structure. Microphone, Camera sensor and MPIS processor MT7688 make up the hardware layer. Based on the hardware layer, the Linux embedded operating system Openwrt is built up. Then the application FFmpeg [3] as a multimedia handling lib converts, packages and sends the live streaming which includes video and audio sub streaming to indicated port of cloud server.

Cloud server takes LEMP (Linux + Nginx + MySQL + PHP) frame set up on EC2 (Elastic Compute Cloud) instances of AWS (Amazon Web Services) Cloud Platform. RTMP port receives live streaming, then divide it into video and audio streaming for further processing. OpenCV (Open Source Computer Vision) lib analyzes the Video streaming. MFCC of audio streaming is extracted and then judged by the pre-trained SVM module to distinguish baby crying from other noises. Live streaming including analysis result is displayed on website built on the server. Any remote devices could access to the analysis data with caregiver account though internet.



**Fig. 1.** BCCS system block diagram.

## 2.4 Baby Crying Detection

To detect a specific sound, first extract the unique property of sound. MFCC (Mel-frequency cepstral coefficients) which could represent the uniqueness of the shape of vocal tract [4]. Extraction process of MFCC complies with following four steps:

- a) Pre-emphasize, Frame blocking and Hamming window

Pre-emphasizing assists to supplement the attenuation of high frequency part of input signal. Hamming window is added to attenuate the discontinuousness at window edges [5].

- b) DFT

Discrete Fourier Transform changes input signal to the format of multiplication of high frequency signal and low frequency signal in frequency domain.

- c) Mel filterbank

Human ears function like filters, it deploys more filters on the low frequency parts and less on the high frequency regions. It is spaced non-uniformly and follows the equation [6]:

$$Mel(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) \quad (1)$$

where common input frequency  $f$  is transformed into Mel frequency in which human has linear perception of pitch.

---

d) DCT

The 12 cepstral coefficients are achieved through series of transforms like logarithmics and discrete cosine transform, which describes the spectral envelope the low frequency part of spectral [7].

The second phrase of baby crying detection is classification. SVM (Supported Vector Machine) module is selected as the classifier. Over 300 baby crying samples and other noises with pre-set labels are used to train this module in this project.

## 2.5 Baby Sleeping Detection

As the initial step of further analysis of baby movements, related joint points of human body are specified as the target points, then located and tracked. Haar-like feature and Adaboost method algorithm [8] is utilized to catch these points. Combined with pre-trained computer vision classifiers of face, hands and other parts of human body from OpenCV lib, specific target parts of human body could be detected. Then Lucas-Kanade optical flow [9] is utilized to track target points. This method bases on the assumption that the intensity of the target pixels does not change between consecutive frames. If the target in the image moves by distance (dx, dy) after dt time, the assumption could be indicated as below.

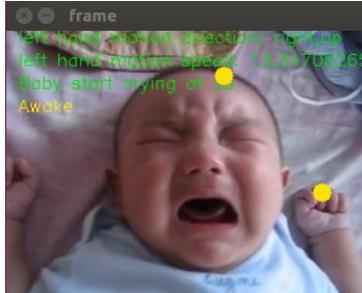
$$f(x,y,t) = f(x+dx,y+dy,t+dt) \quad (2)$$

Every frame of input video streaming is analyzed to detect if baby's sleeping. In detail, sleeping detection algorithm is implemented in following steps:

- a) Default initial status is that baby is sleeping;
- b) If face and open eye are detected, baby is awake;
- c) If movement of any other target points is detected he or she is awake;
- d) Any time, if crying is detected, baby is awake;
- e) Otherwise, baby stays in sleeping status.

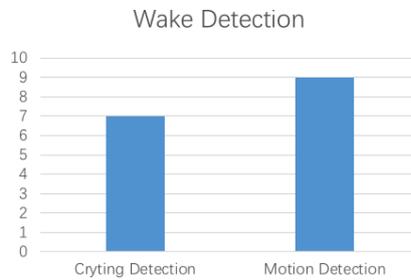
## 3 Result analysis and verification

One of video test results of baby sleeping status is demonstrated in Figure 2. Yellow points indicate the tracking points. The direction of baby left hand movement is right and up, and speed is approximately 13 pixels per second. Audio signal is analyzed per second and start point of baby crying is recorded. Crying status is figured out through audio processing with motion judgement simultaneously. Combined with all these status, program provides the result baby is awake.



**Fig. 2.** Final judgement result (source video from Youtube)

Shown in Figure 3, eleven random videos are utilized to test this system. The crying status could be detected correctly in 7 videos and motion detection effects in 9 videos.

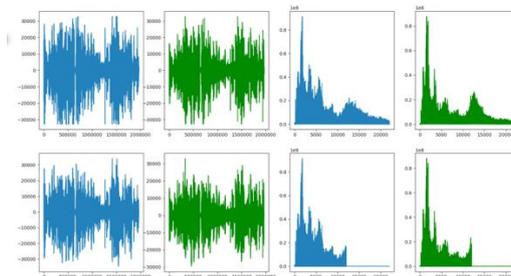


**Fig. 3.** Detection accuracy of sleeping status

Parts of possible reason are the crying sound unsuccessfully filtered out from the mixed other loud noises and low location accuracy under complex background of image. To verify this assumption, audio and video noise filtering is analyzed and processed as below.

### 3.1 Audio noise analysis and filter

It's known that common baby crying signal frequency range spreads from 4KHz to 17KHz [11], to filter out the high frequency noise, audio signal Fast Fourier Transmission (FFT) [10] are processed separately and 12KHz is selected as the threshold frequency value in this project. After filtering, the time-domain signal and frequency-domain signal of left and right channels are shown in Figure 4. Then reshaped audio signal after Inverse Fast Fourier Transmission (IFFT) [12] is tested and crying signal in noisy environment is recognized



**Fig. 4.** FFT filtering of left channel (blue) and right channel (green)

### 3.2 Image noise analysis and filter

Majority of image noises is caused by background of tracking target, which may result in the fault judgement of Haar-cascade filter. Thus, one reasonable method is to extract the region of interest (ROI) accurately. Figure 5 demonstrates the basic processes of ROI extraction. Sub figure ③ represents the subtraction result between sub figure ② and background sub figure ①. Sub figure ④ is achieved through gray scale processing of sub figure ③. Then contour of the tracking object is found, illustrated by the red boarder in sub figure ⑤. Bounding rectangle shown in green boarder is determined by the maximum points of the contour with largest area in sub figure ⑥.



Fig. 5. ROI basic extraction processes

Though background interference is filtered out, the final location shown in Figure 6.

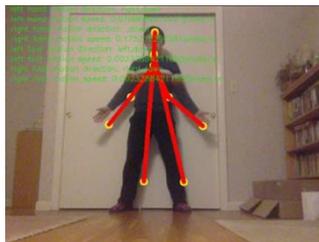
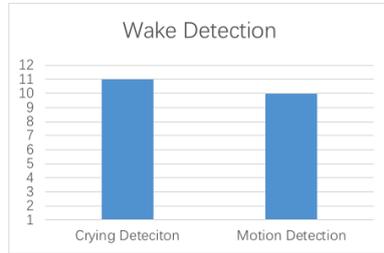


Fig. 6. Relocation after outside background exclusion from ROI

Adding preprocessing of filtering on both video and audio streaming, original eleven random videos are tested again. Result is shown in Figure 7. All the crying status could be detected in all videos and motion detection functions in 10 videos. Compared with result in Figure 4, there's obvious rising of detection rates in crying detection with filtering preprocessing step and meaningful improvement in motion detection.



**Fig. 7.** Detection factor accuracy of sleeping status after filtering

## 4 Conclusion

Infant care means huge time and monetary spending, and great responsibilities for caregivers while facing various pressure from different aspects of daily life. The BCCS system in this paper provides a feasible plan to share and release this kind of pressure through capturing and sending video and audio live streaming to caregivers while meaningful data being analyzed and stored in the cloud server. Crying is detected with high accuracy and sleeping status could be surveilled efficiently. Cloud server of the BCCS system ensures a safety, reliable platform for users' data. Regarding the complexity and facing potential challenges of this system, further optimization for performance and real-time response ability will be done in future work.

## 5 Conclusion

The research is partly supported by NSF grants (award number #1401711, 1429120 and 1451629).

## References

- [1] National Center for Injury Prevention and Control, "CDC Childhood Injury Report:Patterns of Unintentional Injuries among 0-19 Year Olds in the United States," Centers for Disease Control and Prevention, Atlanta, GA, 2008.
- [2] Child Welfare Information Gateway, "Child Abuse and Neglect Fatalities 2015: Statistics and Interventions," U.S. Department of Health and Human Services, Children's Bureau., Washington, DC, 2017.
- [3] F. Sonnati, "FFmpeg – the swiss army knife of Internet Streaming," 11 July 2011. [Online]. Available: <https://sonnati.wordpress.com/2011/07/11/ffmpeg-the-swiss-army-knife-of-internet-streaming-part-i/>.
- [4] P. Cryptograph, "Mel Frequency Cepstral Coefficient (MFCC) tutorial," 2013. [Online]. Available: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/#steps-at-a-glance>.

- 
- [5] Badlaoui and A. Hammouch, "Phonocardiogram classification based on MFCC extraction," 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Annecy, France, 2017, pp. 217-221.
- [6] S. Molau, M. Pitz, R. Schluter and H. Ney, "Computing Mel-frequency cepstral coefficients on the power spectrum," 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221), Salt Lake City, UT, 2001, pp. 73-76 vol.1.
- [7] K. Prahallad, "Speech Technology: A Practical Introduction Topic: Spectrogram, Cepstrum and Mel-Frequency Analysis".
- [8] D. Peleshko and K. Soroka, "Research of usage of Haar-like features and AdaBoost algorithm in Viola-Jones method of object detection," 2013 12th International Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), Polyana Svalyava, 2013, pp. 284-286.
- [9] Y. Douini, J. Riffi, M. A. Mahraz and H. Tairi, "Solving sub-pixel image registration problems using phase correlation and Lucas-Kanade optical flow method," 2017 Intelligent Systems and Computer Vision (ISCV), Fez, 2017, pp. 1-5.
- [10] J. Půčík, P. Kubinec and O. Ondřáček, "FFT with modified frequency scale for audio signal analysis," 2014 24th International Conference Radioelektronika, Bratislava, 2014, pp. 1-4.
- [11] Digital Human Modeling Applications in Health, Safety, Ergonomics and Risk Management, Springer, 2016.
- [12] D. Marelli, M. Aramaki, R. Kronland-Martinet and C. Verron, "An Efficient Time-Frequency Method for Synthesizing Noisy Sounds With Short Transients and Narrow Spectral Components," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 4, pp. 1400-1408, May 2012.