An Intelligent Baby Care System Based on IoT and Deep Learning Techniques

Chinlun Lai, Lunjyh Jiang

Abstract**—**Due to the heavy burden and pressure of caring for infants, an integrated automatic baby watching system based on IoT smart sensing and deep learning machine vision techniques is proposed in this paper. By monitoring infant body conditions such as heartbeat, breathing, body temperature, sleeping posture, as well as the surrounding conditions such as dangerous/sharp objects, light, noise, humidity and temperature, the proposed system can analyze and predict the obvious/potential dangerous conditions according to observed data and then adopt suitable actions in real time to protect the infant from harm. Thus, reducing the burden of the caregiver and improving safety efficiency of the caring work. The experimental results show that the proposed system works successfully for the infant care work and thus can be implemented in various life fields practically.

Keywords**—**Baby care system, internet of things, deep learning, machine vision.

I. Introduction

OBSERVING the frequent accidents in caring for new born babies, it is necessary to develop an intelligent baby caring/watching system [1]-[3] to prevent the infant from various harms including accidents that can occur in their surrounding such as falling, stab and wound, and suffocation by covering mouth and nose; physical contingencies such as sleeping on the stomach, milk vomiting, and so on. This paper proposed an intelligent infant monitoring system to reduce the caring burden of their parents. By integrating the IOT smart sensor and the machine learning techniques [4]-[8], infant information such as face, heartbeat, fever, surrounding abnormal objects, temperature and humidity, and so on can be acquired and analyzed in real time to predict the potential danger condition and then push a notification to a portable device, or take emergency actions automatically to prevent the infant from injury.

II. Methodology

The design principle and the functional blocks of the proposed system are depicted in Fig. 1 and explained as follows. First, image analysis function can be divided into two sub-groups which are face image analysis and environment image analysis, respectively, and are described in Fig. 2. In the face image analysis block, the face image is captured and undergoes face detection using a Haar-like classifier [5]. [6]. Once the infant face is detected, the output ROI image is further fed forward to a convolutional neural network (CNN) [7], which is trained previously by a series of abnormal events images such as suffocation by covering mouth/nose, sleeping on the stomach gesture or milk vomiting, to determine whether there are abnormal events. The structure of the adopted *CNN* is the same as in [7], which is reproduced here in Fig. 3 and is described as follows. The CNN structure is composed of five layers filters, of which, the first one layer filters the 224*224*3 input image with 96 kernels, while each kernel is composed of 11*11*3 sub-blocks. The second layer filters the output of the first one layer with 256 kernels of size 5*5*48. In contradict to the first two layers; the convolutional layers 3 to 5 are connected to each other without any intervening pooling or normalization layers. The third layer has 384 kernels of size 3*3*256 to filter the outputs of the second layer, while the fourth and the fifth layer has 384 kernels of size 3*3*192, and 256 kernels of size 3*3*192, respectively.

At the same time, the surrounding environment of the baby is monitored by a detector with Gaussian Mixture Model (*GMM*) [8] to determine whether there is something adverse near the baby. To reduce the false alarm of the detector due to the influence of light, the difference image of the time sequence should be analyzed further, thus to remove the independent influence from the background lighting. The results of the above two image analysis sub blocks are combined as an integrated recognition system and push the warning signal, including notification and real-time image, to the mobile device. On the other hand, the physiological information of the baby is also gathered via various sensors connected each other. The IoT system is composed of a central process unit (Arduino based) and smart sensors including humidity, temperature, heartbeat, breathing, as well as detectors for the surrounding light, and/or noise. All the sensor signals are forwarded to the process unit via wireless module and analyzed mutually to predict the potential danger conditions, and then push a warning notification to the mobile device, via the structure shown in Fig. 4, to help the babysitter understand what is happening to the baby in real time.

In order to reduce the false alarm results of the image analysis functions, a time series analysis method like [9] can be adopted to improve the system stability. By analyzing a series of captured image sequences, the false alarm rate can be reduced significantly thus the false alarm of the proposed system approaches near zero. For example, a total of f frames (or f / f_s seconds) were chosen for continuous sequence analysis. The captured image is considered object detected if over x frames, where $x \geq f / 2$, are identified as positive. The

Chinlun Lai is with the Communication Engineering Department, Oriental Institute of Technology, Taipei Taiwan (e-mail: fo001@ mail.oit.edu.tw).

Lunjyh Jiang is with the Nursing Department, Oriental Institute of Technology, Taipei Taiwan (e-mail: fk004@ mail.oit.edu.tw).

theoretical probability, P_T, which is defined as a warning condition being identified, can be described as:

$$P_T = \sum_{k=x}^{f} C_k^f (P_c)^k (P_e)^{f-k} \quad (1)$$

where $P_e = 1 - P_c$ is the error probability corresponding to which a real alarm case is not identified correctly.

According to (1), if only one frame is referred and $P_c = 0.85$, the final correct identification rate P_T become 0.95 if 10 frames are used, which is much better than that of determine by one frame. Accordingly, some practical adopted examples (under $x \geq f/2$, and $P_c = 0.854$ conditions) are listed in Table I. On the other hand, to further reduce the burden of the babysitter, the warning notifications are classified into four degrees according to the event characteristic, which represents the emergency level from low to highest. For a warning at the highest level, the necessary actions should be adopted immediately to protect the baby from the threat of death; such warnings include a possible suffocation event. On the other hand, for events of low level warning, such as abnormal environmental temperature, humidity, or lighting conditions, the babysitter only need to keep watching the captured event images until the event is confirmed to be safe. The detail warning classifications of the notification level are described in Table II, where 0 represents that nothing has happened and 1 represents that an abnormal event is detected.

TABLE I
SOME PRACTICAL VARIABLES AND THE CORRESPONDING RESULTS

f	x	P_T
10	8	83.05%
10	7	95.42%
10	6	99.12%
15	12	83.40%
15	11	94.41%
15	10	98.52%
15	9	99.69%
15	8	99.95%

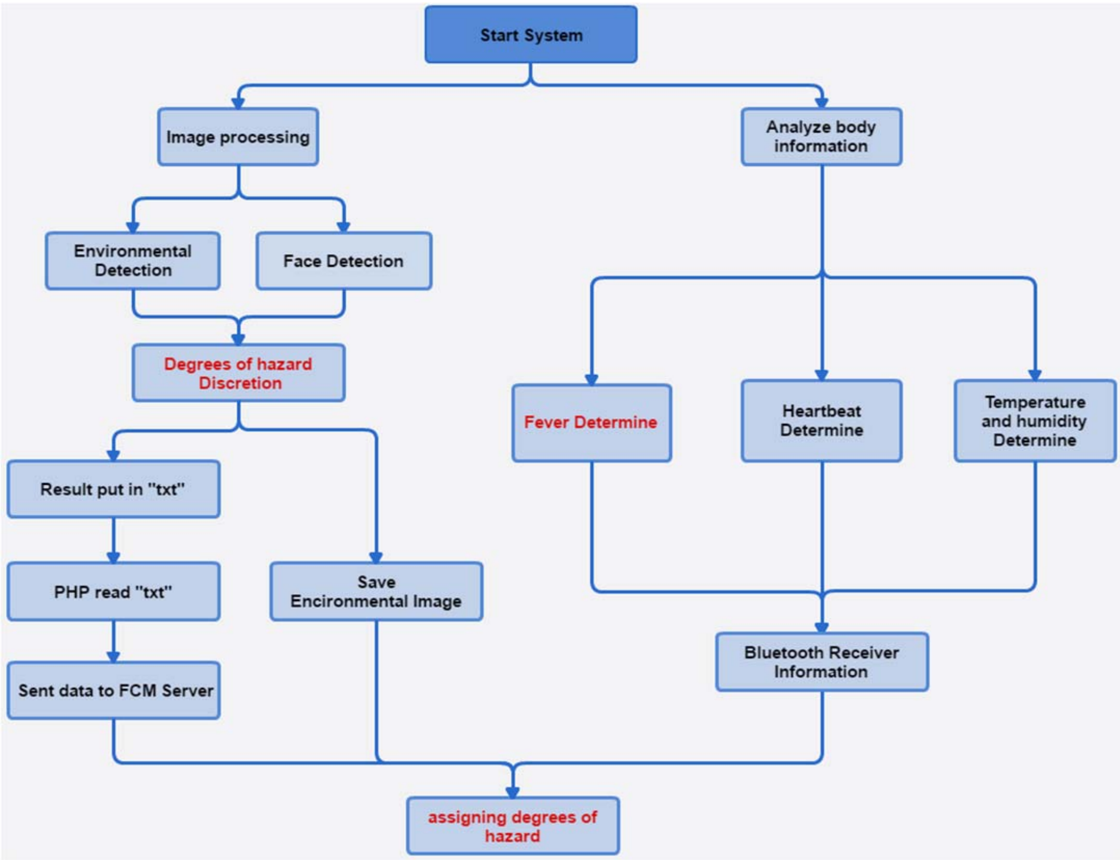

Fig. 1 The functional blocks of the proposed system

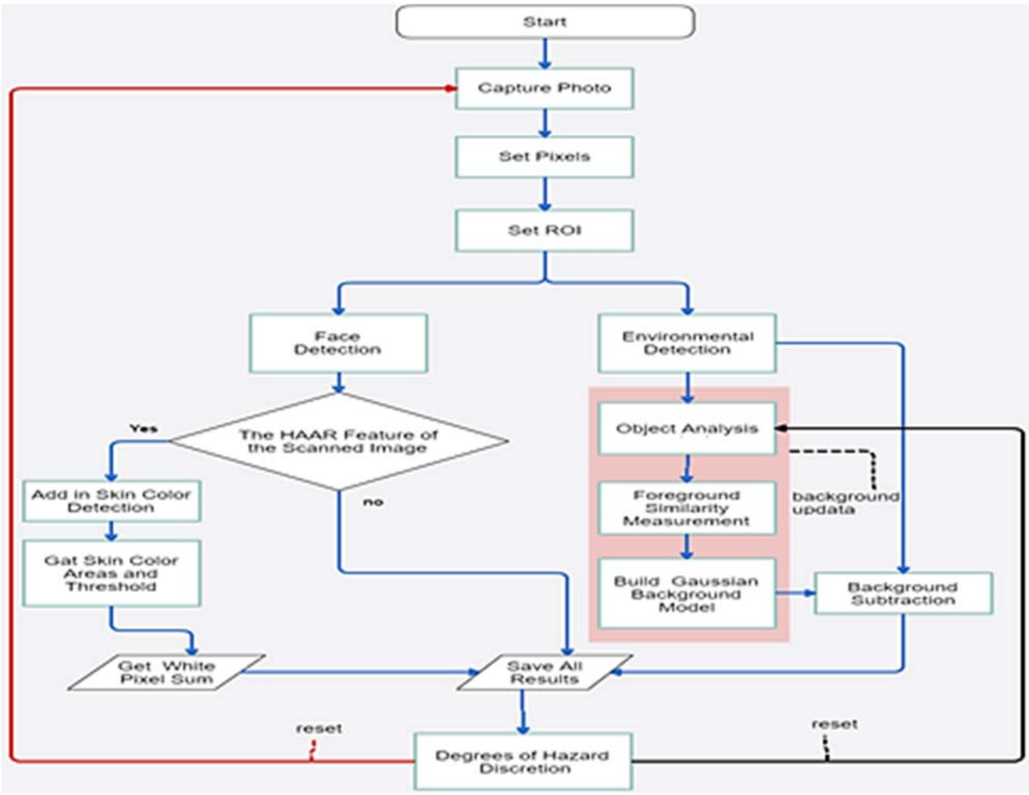

Fig. 2 Flowchart diagram of the image analysis functions

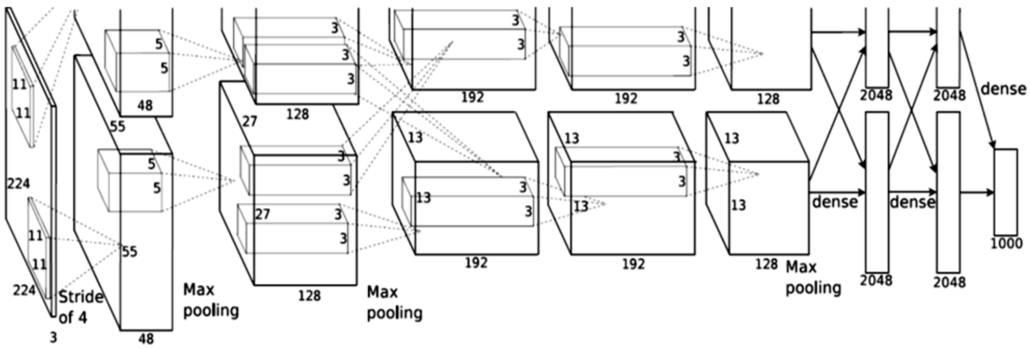

Fig. 3 Illustration of the proposed *CNN* architecture in environmental image analysis, this network structure is similar to A. Krizhevsky's work [7]

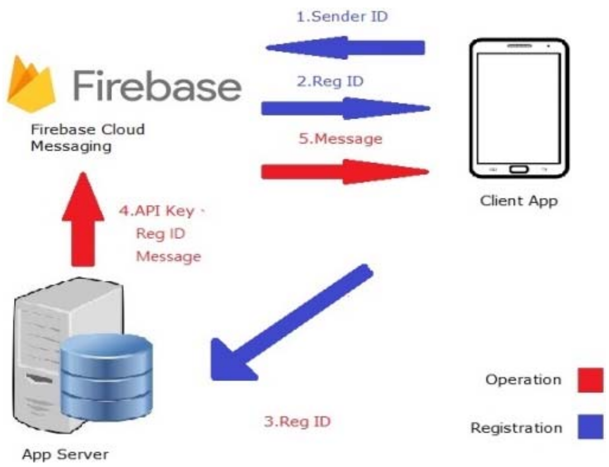

Fig. 4 Structure of the push notification app

TABLE II
THE WARNING CLASSIFICATIONS OF NOTIFICATION LEVELS

Case#	Skin color	Face	Upper body	Lower body	Warning level
1	0	0	0	0	-
2	0	0	0	1	low
3	0	0	1	0	middle
4	0	0	1	1	high
5	0	1	0	0	-
6	1	1	0	0	highest
7	0	1	0	1	mid
8	1	1	0	1	highest
9	0	1	1	0	mid
10	1	1	1	0	highest
11	0	1	1	1	high
12	1	1	1	1	highest

Fig. 5 Example of emergency alarm condition

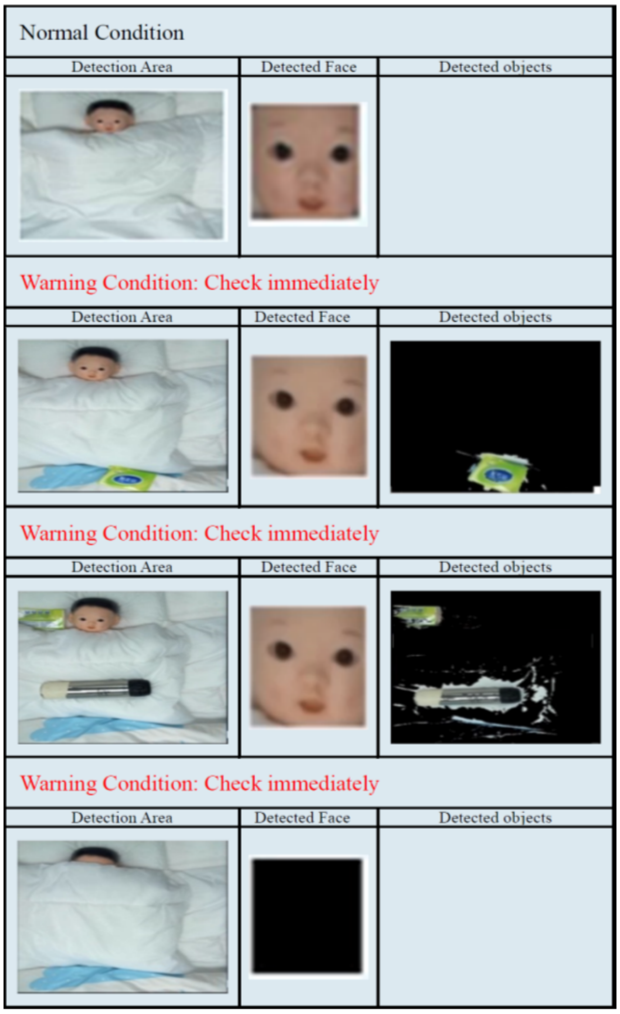

Fig. 6 Examples of abnormal events detected by the image analysis functions

III. Experimental Results and Discussions

The proposed system is completely implemented by a laptop PC with webcam, an Arduino platform with various sensors, and a mobile phone with Android OS. The notebook is used to implement the image recognition functions, while the physical sensors are activated and communicated via the Arduino platform. All the environmental signals are integrated into the notebook, where the analyzing system determines a final notification and then pushes it to the mobile phone as an alarm. Some of the simulation results, using the simulated infant with various added events as test samples, are shown in Figs. 5 and 6. In these examples, some abnormal events, which are determined as emergency cases including lost/mask the face, are detected and the alarm signals are pushed to the babysitter's smart phone. The total response time of the proposed system from when the events occurs to receiving the alarm signal is within about 3-4 seconds, which is good enough to be determined as processed in real-time for the baby caring application. On the other hand, the false alarm rate is about only 0.5%, while the detection failure rate is zero (among all the test samples). It is observed from the experimental results that the proposed baby care system works effectively and efficiently, and thus can be implemented and applied successfully to real life.

IV. Conclusion

Caring an infant brings extreme pressure and burden for the babysitter, and it may result in serious injury in the case of negligence. However, by the assistance of the proposed baby care system, it can detect (or even predict) the potential dangerous events immediately and 24/7, and thus prevent the baby from possible harm or death. By integrating the deep learning technique in image analysis as well as the IoT technique in environment understanding, baby care work can be much easier and more effective. Furthermore, via the time series analysis method, the false alarm rate can be reduced significantly near to zero, thus it makes the proposed system practical. The experimental tests show that the system cannot only detect abnormal events precisely but with extremely high efficiency. That is, the proposed system can be a good infant care assistant for families with new born babies, and thus, improves their life quality and protects infants from harm caused by accidental events.

ACKNOWLEDGMENT

We would like to extend our sincerest gratitude to the students in the Department of Communication Engineering and the Department of Nursing, Oriental Institute of Technology, for their assistance in this paper, including Wei-Ching Lai, Tsune-Han Lin, and Yan-Hong Guo.

REFERENCES

[1] R. Palaskar, S. Pandey, A. Telang, A. Wagh, and R. M. Kagalkar, "An Automatic Monitoring and Swing the Baby Cradle for Infant Care", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 12, pp. 187-189, Dec. 2015.

[2] C. Yadav, A. Pandey, M. Saraogi, and S Tribhuwan, "Active RFID and ESP8266 based Child Monitoring System", *International Journal of Computer Applications*, vol. 139, no. 12, pp. 22-25, April 2016.

[3] The Infant Care System ICS - A Baby Wearable <https://www.indiegogo.com/projects/the-infant-care-system-ics-a-baby -wearable#/>

[4] S. Hussain. "Machine learning methods for visual object detection", *Ph.D thesis of Laboratoire Jean Kuntzmann*, 2011.

[5] Jui-Hung Fang(2011), "An Approach to Age Estimation Based on Facial Images and AdaBoost Algorithm" master thesis, National Taipei University.

[6] Paul Viola, and Michael Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[7] Krizhevsky, A., Sutskever, I., and Hinton, G. E. "ImageNet classification with deep convolutional neural networks", In *NIPS*, pp. 1106–1114, 2012.

[8] Shr-Chi Jeng(2005), "A GMM-based Method For Dynamic Background Image Model Construction with Shadow Removal", master thesis, National Chiao Tung University.

[9] ChinLun Lai, and ChiuYuan Tai, "A Smart Spoofing Face Detector by Display Features Analysis," *Sensors*, no. 1136, vol. 16, issue 7, pp.1-15, 2016(SCI).