A Monitoring System Based on Behavior Analysis

C. Y. Fang*, B. R. Shao, and S. W. Chen

Department of Computer Science and Information Engineering

National Taiwan Normal University

Corresponding Author: violet@csie.ntnu.edu.tw

ABSTRACT

This paper presents a vision-based infant-monitoring system that adopts an infant behavior analysis approach to reduce infant injuries. In our study, the video camera is set above the crib to capture infant sequences. The system first preprocesses the input sequence to filter out the noise and reduce the effects of lights and shadows. Then, the infant's head and limbs are detected from the input frames and compared with pre-defined posture maps to select the most similar map. A posture map describes the current infant posture; the selected posture map can be regarded as a node to be linked by the occurrence order to construct a dynamic behavior graph that describes infant behaviour captured over time. If an input posture map does not exist in the dynamic behavior graph, this means that an unexpected situation has occurred and the system would then alert the baby sitter to attend to the infant immediately. A weighted dynamic behavior graph adjustment algorithm is used to accomplish the behavior analysis. Since infants grow very quickly and their growth processes may differ, the dynamic behavior graph should be continuously updated to fit the current behavior of infants. The experimental results show that the proposed method is able to perform robustly in real-time.

Keyword: Vision-based infant monitoring system, posture maps, dynamic behavior graphs, behavior analysis

1. Introduction

It is easy for infants to fall ill or be injured due to the negligence of baby sitters. Baby sitters, especially the tenderfoots, need to work under pressure to keep infants under close watch and out of danger. However, the average sleep time of an infant is very long (about 16 to 20 hours per day) and so it is unreasonable to request the baby sitters to monitor the infant all day long. Thus, this study develops a vision-based infant monitoring system to monitor the infants when they are left alone for a short time period and to ensure that they are safe.

Many research works s (Bobick 2001; Juang 2009; Kirishima 2005; Kosta 2008; Li 2007) have been proposed to detect or analyze human behavior. Kosta and Benoît (Kosta 2008) used the multilevel probabilistic finite state machine to analyze the movements of the human body limbs in order to detect and analyze simple actions such as walking and running. Bobick and Davis (Bobick 2001) utilized temporal templates to recognize human movement. Temporal templates are constructed using motion-energy images (MEIs) and motion-history images (MHIs) of the movement sequences, and matched with the stored instances of views of known actions for movement recognition. Juang *et al*. (Juang 2009) proposed a vision-based human body posture estimation method using body silhouette and skin color information. Body shape characteristics and skin color regions are used to locate significant body points and these points can be used to construct the structure of the human body to estimate the posture.

Most of the above research focuses on detecting the postures or motions of the adults but not of infants. Some researchers have also applied their methods to home care applications. For example, Juang and Chang (Juang 2007) proposed a home care system that uses a neural fuzzy network method to classify human body postures. The system can detect the motion of a person falling down and then sitting on the floor. However, infants can only lie on the bed or the crib, they cannot walk, run, jump, or stand like adults and thus, their behaviors are very different from adults. This paper proposes a method to detect and analyze infant behavior for home care applications. In this paper, an infant refers to a baby whose age is between one to six months.

In this study, the video camera is set above the crib to capture the video sequence for an infant. The video camera's field of vision covers the whole body of the infant (from head to foot), as shown in Figure 1. Although the video camera is set indoors and the light sources are fixed, some problems should still be considered and resolved in the vision-based monitoring system. For example, the motion of the baby sitter will cast moving shadows on the crib or even on the baby's face, and would affect the intensity values of the input images. In addition, since infants grow up quickly, the dynamic behavior graphs used to describe their behavior should be updated progressively. Infants also grow up at different rates and thus, each infant should have his/her own dynamic behavior graph.

Figure 1: The input frames of the infants.

2. System Flowchart

The flowchart of the proposed infant monitoring system is shown in Figure 2. Initially, once a video sequence has been input into the system, a detection technique is applied to the input sequence to detect motions after the preprocessing stage. If no motion has been detected, then the system detects the skin color regions from the frames of the input sequence. If the system does not detect any skin color

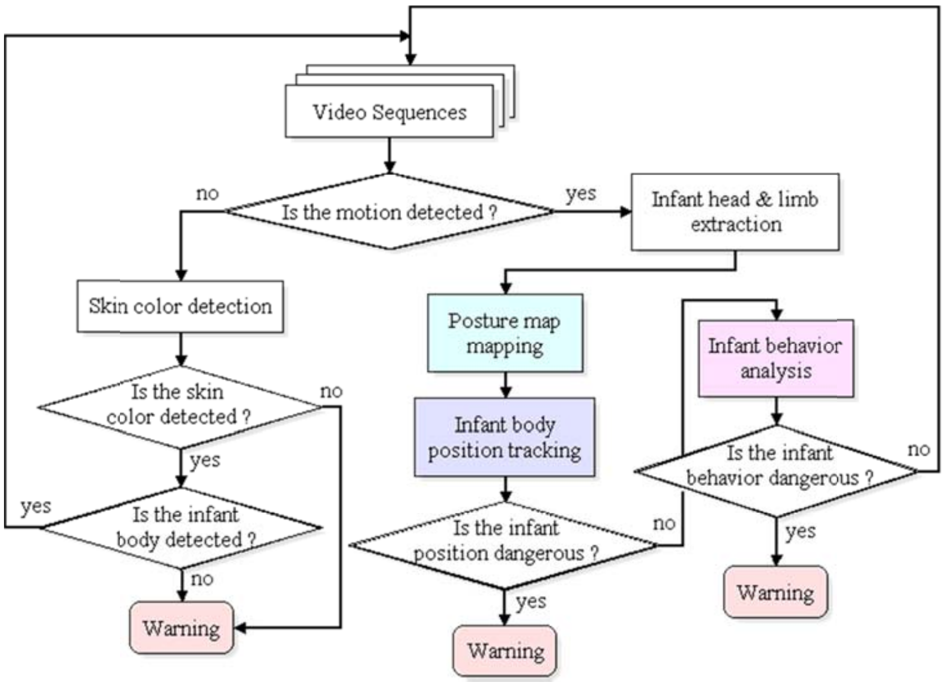


Figure 2: The flowchart of the infant monitoring system.

region, it means that the crib may be empty or the input frames are abnormal and the system would output the warning message. Otherwise, if skin color regions have been detected, then it means that the infant may be sleeping peacefully, i.e. the system identifies the infant by detecting the infant body and returns to the initial step.

If the system detects any motion from the input sequence initially, we assume that the infant is lying in the crib and the system extracts the regions showing the infant’s head and limbs. The extracted regions will all be traced in the following frames, regardless of whether they are occluded or not. However, we assume that the system can detect the head and each limb of the infant successively at the beginning of the sequence. This means that the infant’s clothes cannot cover all the five skin color regions of the infant head and limbs and these five regions cannot occlude each other initially. This initial constraint may be resolved in future work.

Following this, the system can locate the current posture and position of the infant by mapping the extracted skin color regions to the predefined posture maps. If the position of the infant is dangerous, then the system outputs the corresponding warning message. Otherwise, the system analyzes the infant’s current motion using a dynamic behavior graph, and decides whether the infant’s current motion is dangerous or not. If the infant’s current motion is considered dangerous, then the system again outputs the warning message.

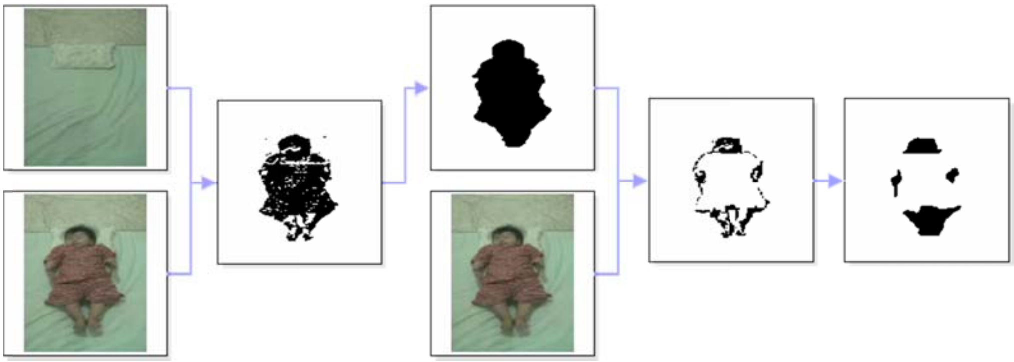
3. Preprocessing Stage

The preprocessing stage includes two main steps, namely foreground extraction, and skin color detection. Figure 3 shows an example preprocessing stage in detail. Firstly, the system translates the input frames from the RGB model into the HSI color model. Let B and I be the background image and the input image respectively. Then, the foreground image F can be obtained by

$$F(x,y)=\begin{cases}0 & \text{if } |I(x,y)-B(x,y)|\geq T_f \\ 1 & \text{otherwise}\end{cases},$$

where T_f is a threshold.

In the example shown in Figure 3, the system can obtain the foreground image F shown in Figure 3(b) if the background image B and the input frame I are as shown in Figure 3(a). The system then eliminates noise by morphological erosion, connects the foreground components by morphological dilation, and extracts the largest foreground component as the infant. The result is shown as the upper image in Figure 3(c). Using the predefined skin color range, the system extracts the skin color regions inside the largest foreground component to avoid the noise introduced by the background. The skin color regions inside the largest foreground component are shown in Figure 3(d). Moreover, the morphological erosion and dilation is applied to Figure 3(d) to obtain more complete skin color regions as shown in Figure 3(e).



(a) (b) (c) (d) (e)

Figure 3: An example of the preprocessing stage. (a) The upper one is the background image, and the lower one is the input frame (b) The foreground image (c) The upper one shows the largest foreground component, and the lower one is the input frame (d) The skin color regions of the largest foreground component (e)The skin color regions after the application of the morphological technique.

4. Infant Posture Mapping

4.1 The Classes of Infant Posture Maps

In this study, we comprehensively define 203 infant posture maps to describe various infant postures, including three main classes. The first class contains 31 normal and complete posture maps, where all five skin color regions of the infant head and limbs have been identified. These skin color regions may be occluded by others. Some examples of these posture maps and their corresponding infant postures are shown in Figure 4. The yellow node label *H* in the map represents the relative position of the infant head region. The orange nodes having a dotted line circle with labels *R* and *L* indicate the relative positions of the left and right hands, respectively. Similarly, the green nodes with labels *R* and *L* indicate the relative positions of the left and right legs, respectively. The first map in Figure 4 is called the initial posture map since the initialization of the system will be successful when this map is selected.

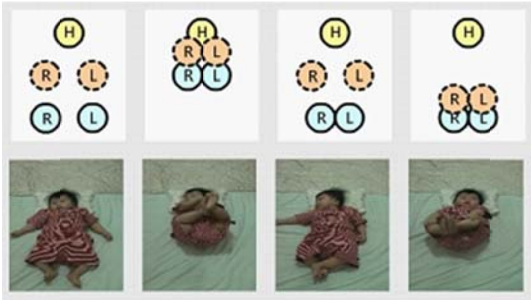


Figure 4: The upper row shows a few examples of the first class of the infant posture maps and the lower row shows their corresponding input frames.

(a) (b)

Figure 5: The second class of infant posture maps. (a) A few examples of the second class of the infant posture maps and (b) one corresponding input frame.

The second class contains 137 normal but incomplete posture maps where one or more skin color regions of the infant head and limbs have not been identified. These skin color regions may be occluded by the quilt or other objects. For example, Figure 5(b) shows an example where the system can only detect the head of the infant and the first map shown in Figure 5 (a) becomes its corresponding posture map.

Figure 6: A few examples of the third class of infant posture maps.

The third class contains 35 abnormal posture maps, which rarely occur. A few examples of these posture maps are shown in Figure 6. For example, the first map shown in Figure 6 shows that the positions of the infant's legs are closer to that of the head than the hands. The third map indicates that the infant's left leg occludes his/her head, and the other three parts occlude each other. In short, the postures in this class are seldom exhibited by infants.

4.2 Posture Map Mapping

As mentioned above, once a video sequence is input into the system, the skin color regions are extracted first and these skin color regions can be identified as the infant's head and limbs initially. The regions initially identified will be continuously processed until the five skin color regions have been detected separately and correctly. Its corresponding posture map is the initial posture map shown in Figure 4. Subsequently, the system tracks these regions and selects the next corresponding posture map from each frame. The flowchart of region tracking is shown in Figure 7.

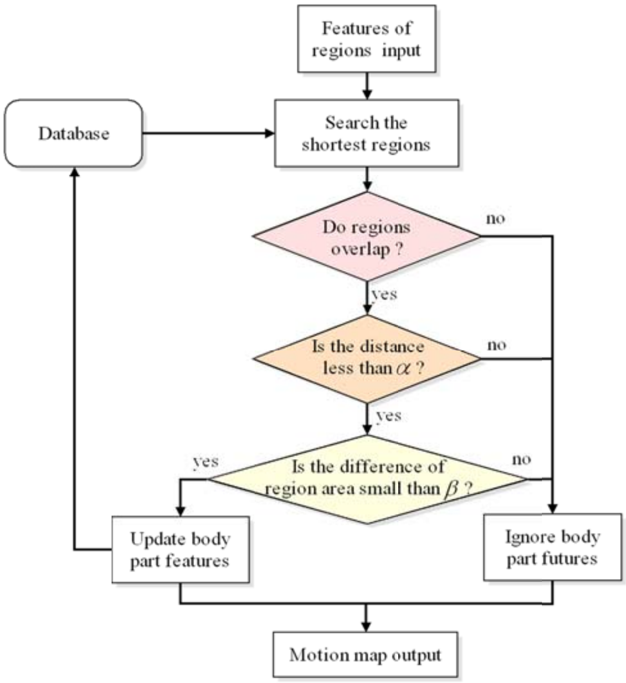


Figure 7: The flowchart depicting region tracking.

The center position and area of the region are defined as the features of the regions here. In Figure 7, once the features of a region R_t at time t are input into the system, its nearest region R_{t-1} at time t-1 can be found. If these two regions overlap each other and their center positions and region sizes are close enough, then the feature values of region R_{t-1} will be replaced by those of region R_t. If these two regions do not overlap or their center positions are not close enough, then the system retains the feature values of region R_{t-1} in the database. Conversely, the feature values of region R_t are used to decide whether or not the regions occlude with each other.

Figure 8 shows an example of region tracking. In Figure 8, the skin color regions extracted from the frames at time t-1 and time t, respectively, are shown in the first row, and they are overlapped and compared to trace the skin color regions. Here, we can observe that the regions obtained from the frame at time t-1 are shown in orange, and the regions obtained from the frame at time t are shown in blue, respectively. In this example, the right hand region of the infant is tracked successfully and the feature values of the regions will be updated in the database, depicted in the left image at the bottom in Figure 8. However, since the left and right leg regions of the infant are merged at time t, the feature values of these regions at time t-1 would not be able to find the suitable corresponding regions at time t. In such a case, the feature values of these regions at time t-1 will be preserved in the database, while the feature values of these regions at time t will be used to select a suitable posture map.

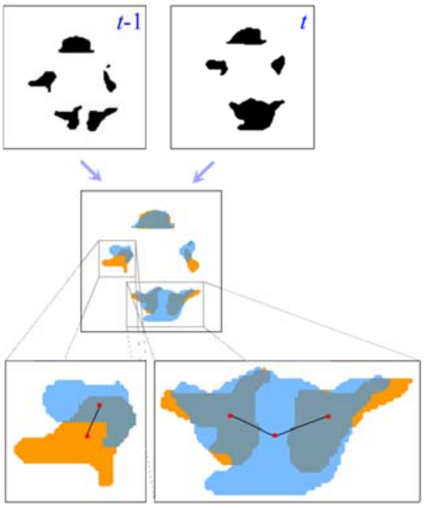


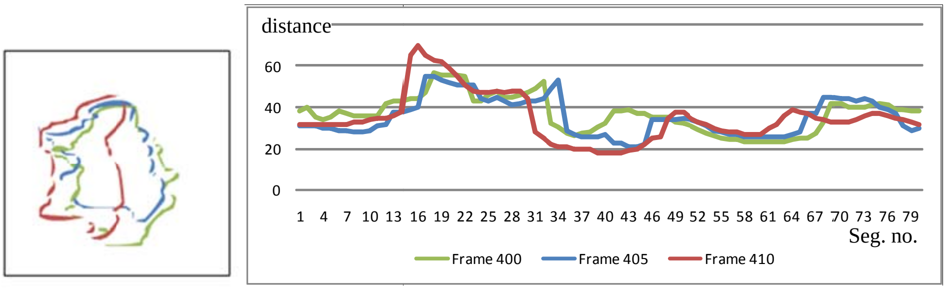
Figure 8: An example of region tracking.

4.3 Infant Head Position Location

The posture map is the first source of information used to analyze infant behavior, and in addition, information regarding the infant head position is required. As mentioned in Section 3, the system can extract the largest foreground component, shown as the upper image in Figure 3(c). The contour of the infant can be extracted by applying the Sobel filter on the image of the largest foreground component. Figure 10(a) shows three contours of different frames in one sequence, including frames 400(green), 405(blue) and 410(red).

Let the center of the largest foreground component be the origin. The system divides the contour into 80 segments counterclockwise by equal angles, where each segment is 4.5 degrees. Following this, the system then calculates the distances between the center of the largest foreground component and the contour pixels in the segments.

The minimum distance in a segment is defined as the segment distance. These distances are shown in Figure 10(b). The contour segment that has the maximum segment distance is defined as the position of the head.

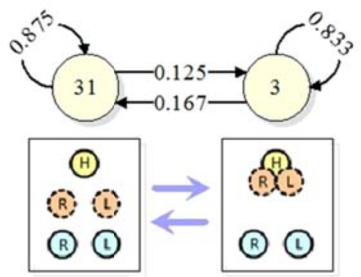


(a) (b)

Figure 10: An example of infant head position decision/identification.

5. The Dynamic Behavior Graph

The dynamic behavior graphs in the database are used to represent dangerous or safe behavior exhibited by the infant. We believe that each infant should have his/her own dynamic behavior graph because infants grow up at different rates and they also have different orders of motion behavior. The dynamic behavior graph contains nodes and weighted arcs. Each node corresponds to a posture map, and the arcs between the nodes indicate a path that the infant can change his/her posture. The weight associated with a link indicates transition probabilities, whereby the infant changes his/her posture from one posture to another.



(a) (b)

Figure 11: An example of the dynamic behavior graph.

Figure 11 shows a simple example of the dynamic behavior graph. This graph contains two nodes, whose corresponding maps are map 31 and map 3 respectively, and four links. The corresponding maps of the nodes are shown in Figure 11(a) and (b) respectively. In this example, we can observe the transition probability that the infant changes his/her posture from map 31 to map 3 is 0.125, and the transition probability that the infant stays at map 31 is 0.875. Similarly, the transition probability that the infant changes his/her posture from map 3 to map 31 is 0.167, and the transition probability that the infant would stay at map 31 is 0.833. This dynamic behavior graph indicates that the infant usually stays at the same posture.

Given an input sequence which includes t frames, $f_1, f_2, f_3, \ldots, f_t$, let $w_t(n, n')$ be the transition probabilities from nodes n to n', and $N_{nn'}^{t}$ indicate the occurring times until frame t during which the infant changes his/her posture from maps n to n'. The transition probabilities can be initialized using the following equation:

$$w_t(n,n') = \frac{N_{nn'}^{t}}{\sum_{m=1}^{t} N_{nm}^{t}}$$

Suppose the corresponding map of the frame at time t is map m_t whose corresponding node is n, and the corresponding map of the frame at time t+1 is map m_{t+1} whose corresponding node is n', the transition probabilities from node n to node n' will be increased and updated by the following function:

$$w_{t+1}(n,n') = \frac{N_{nn'}^{t}+1}{\sum_{m=1}^{t} N_{nm}^{t}+1}$$

On the other hand, the transition probabilities from node n to node n" (where $n' \neq n''$) should be decreased and updated using the following function:

$$w_{t+1}(n,n'') = \frac{N_{nn''}^{t}}{\sum_{m=1}^{t} N_{nm}^{t}+1}$$

We believe the transition probabilities will become stable when the length of the input sequences is sufficiently long.

6. Infant Behavior Analysis

The flowchart of infant behavior analysis is shown in Figure 12. Firstly, a current dynamic behavior graph of the infant is stored in the behavior database. Once a current posture map is input into the system, the system will search for the current map in the dynamic behavior graph within the database.

If the search is unsuccessful, it means that the current posture has never occurred before and the warning message will be output. If the search is successful, then the system will update the weight of the link between the previous posture map and the current map, and calculate the danger level of the current behavior. If the current infant behavior is dangerous, then the system outputs the warning message, else, the updated weights are stored in the behavior database. It should be noted that the infant head position location is used to correct the posture map mapping.

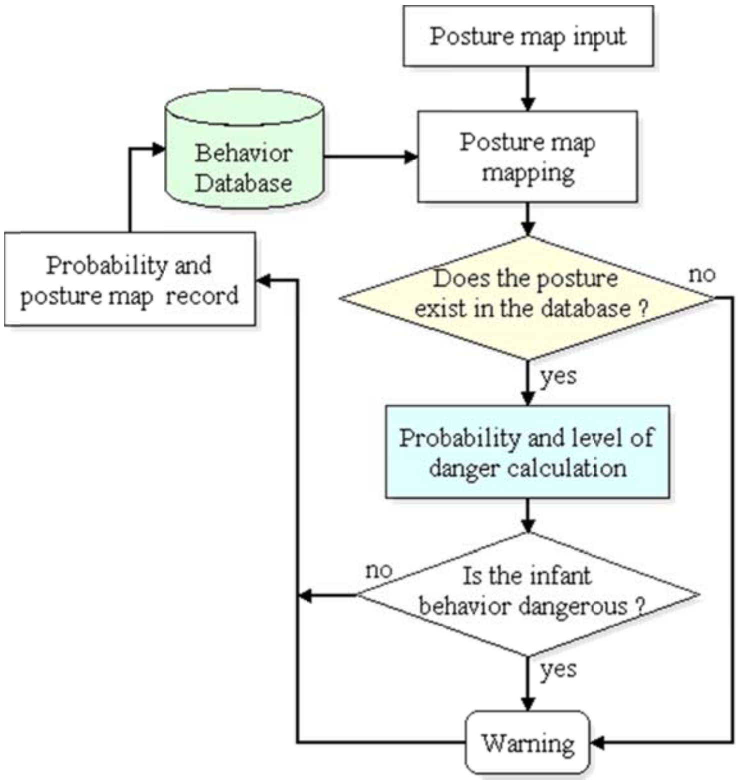


Figure 12: The flowchart of infant behavior analysis.

7. Experimental Results

The input data for our system was acquired using a video camera mounted above the crib and processed on a PC with an IntelRCore™ 21.86GHz CPU. The input video sequences were recorded at a rate of 30 frames/second.

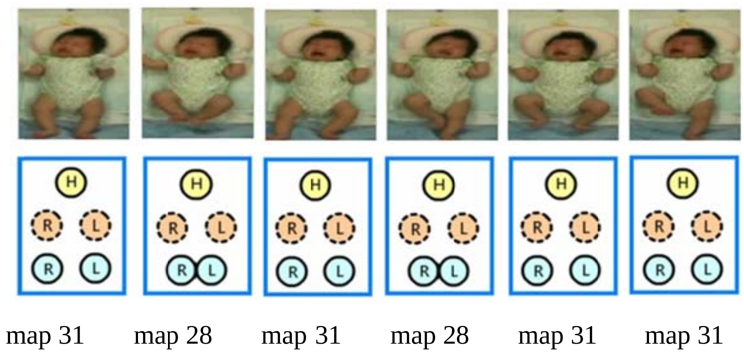


Figure 13: A two-month infant test.

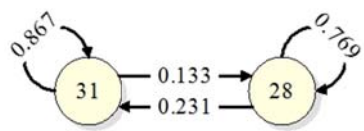


Figure 14: An example dynamic behavior graph.

Figure 13 shows the first test conducted on a two month old infant. The length of the input sequence is five minutes, and only six frames of the sequence are shown in the upper row of Figure 13. Their corresponding maps are shown in the lower row. In this test, only two maps (maps 31 and 28) are selected to construct the dynamic behavior graph. The transition probabilities are shown in Figure 14. We can observe that the dynamic behavior graph of a two-month infant is simple.

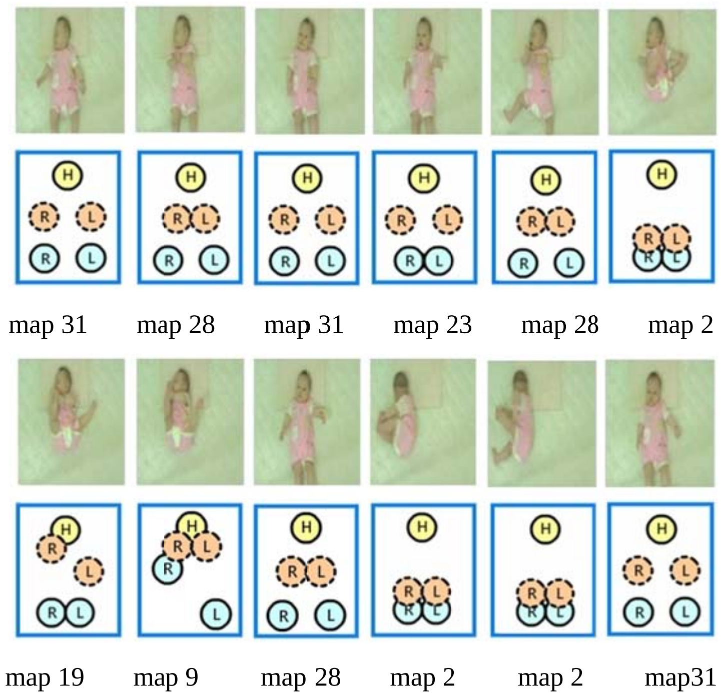


Figure 15: A four-month infant test.

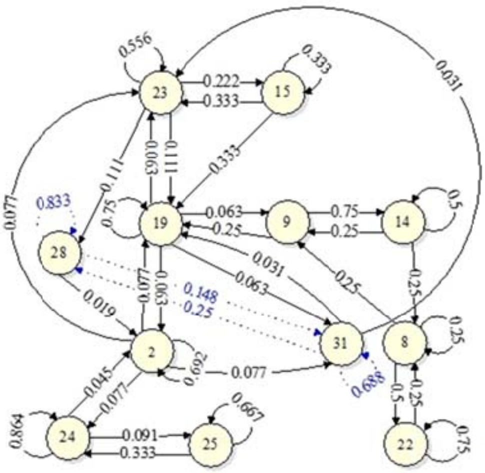


Figure 16: The dynamic behavior graph of test two.

The second test is shown in Figure 15. The infant is four months old and the length of the input sequence is twenty minutes. Only 12 frames of the sequence and their corresponding maps are shown in Figure 15. Figure 16 shows the dynamic behavior graph of this four-month old infant. In this example, 12 maps including maps 31 and 28 are selected to construct the dynamic behavior graph. The transition probabilities are also shown in Figure 16. It can be observed that the dynamic behavior graph of a four-month old infant is more complex than that of a two-month old infant. If more time is taken to construct the dynamic behavior graph, we believe that the nodes and the transition probabilities may become more stable. However, once the infant grows up, the nodes and the transition probabilities of the dynamic behavior graph will be updated progressively.

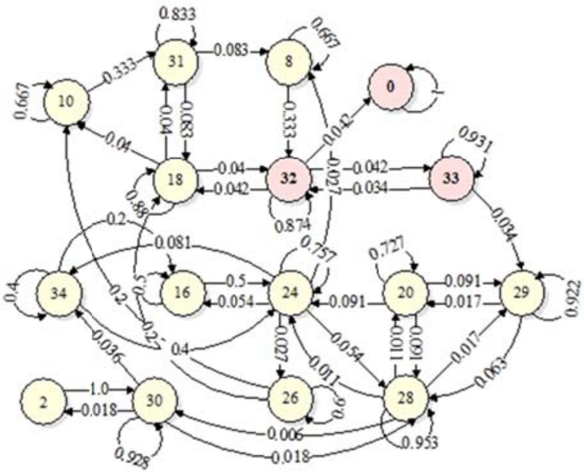
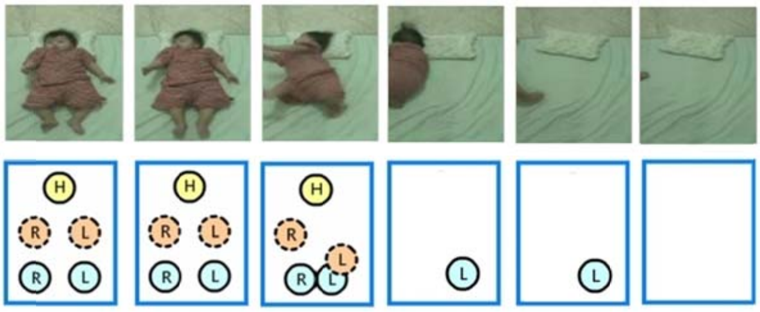


Figure 17: The dynamic behavior graph of a six-month old infant.

Figure 17 shows the dynamic behavior graph of a six-month old infant. In this example, 16 maps are selected to construct the dynamic behavior graph. The transition probabilities are also shown in Figure 17. Notice that three nodes (maps 0, 32, and 33), which belong to the second class of the infant posture maps, are marked using the pink color in this graph. If these nodes identify the infant's motion, it indicates that the infant may be outside the field of view of the camera. These situations are dangerous, thus the system will output the warning message to the baby sitter.

One example depicting the infant moving out of the field of view of the camera is shown in Figure 18. The length of the input sequence is two minutes, and only 6 frames of the sequence and their corresponding maps are shown in Figure 18. In this example, the system will output the warning message once map 32 has been selected.



map 31 map 31 map 18 map 32 map 32 map 0

Figure 18: An example which depicts the infant moving out of the field of view of the camera.

When an input posture map does not exist in the dynamic behavior graph, it means that an unusual situation may have occurred and the system would alert the baby sitter to attend to the infant immediately. However, the growth of the infant will also result in the occurrence of an unusual map. In this case, the system will add the map into the dynamic behavior graph as a new node, construct relative links, and assign suitable weights.

8. Conclusions and Future Work

This paper presents a system based on an infant behavior analysis approach to monitor various infant motion situations. A video camera is set above the crib to capture the infant sequences. Then, the infant's head and limbs are detected from input frames and used to select pre-defined posture maps. These selected posture maps are used to construct the individually weighted dynamic behavior graph of the infant. Once a current posture map is input into the system, the system can then decide whether or not the current posture is dangerous.

For future work, it can be noted that the initialization of the system is successful only when all the five skin color regions of the infant head and limbs have been detected. This is a constraint that may be addressed in the future. It can also be noted that the hands or legs of the infant may interfere with the location of the infant head position. The method to locate the infant head position should be improved in the future to avoid this kind of interference. More infant sequences will also be collected for future experiments.

Acknowledgment

The authors would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC-98-2221-E-003-014-MY2 and NSC 98-2631-S-003-002.

References

Bobick, A. and Davis, J., The Recognition of Human Movement Using Temporal Templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, pp.257 – 267, 2001.

Juang, C. F., Chang, C. M., Wu, J. R., and Lee, D., Computer Vision-Based Human Body Segmentation and Posture Estimation, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 39, pp.119-133, 2009.

Juang, C. F. and Chang, C. M., Human Body Posture Classification by a Neural Fuzzy Network and Home Care System Application, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 37, pp.984-994, 2007.

Kirishima, T., Sato, K., and Chihara, K., Real-time Posture Recognition by Learning and Selective Control of Visual Interest Points, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, pp.351 – 364, 2005.

Kosta, G. and Benoit, M., Group Behavior Recognition for Posture Analysis, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 18, pp.211-222, 2008.

Li, X., Yan, S., Tao, D., and Xu, D., Gait Components and Their Application to Gender Recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 38, pp.145-155, 2007.