**Infant Facial Expression Analysis: Towards a Real-Time Video Monitoring System Using R-CNN and HMM** (Cheng Li et al, 2021)

The proposed system is based on combining expression detection using Fast R-CNN with a compensated detection using analysing information from the previous frame and utilizing a Hidden Markov Model.

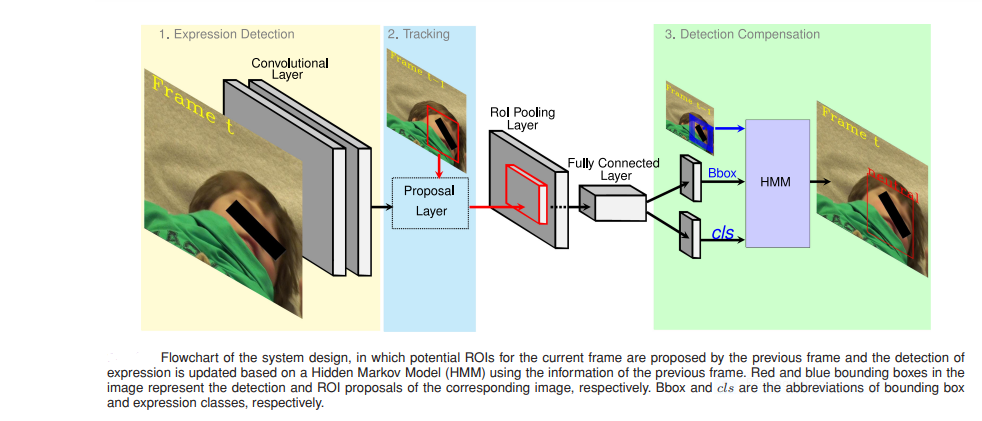
* **Dataset used** – Infant COPE/iCOPE dataset and YouTube dataset.

1. The Infant COPE Database, contains 204 facial images of 26 neonates experiencing the pain of a heel lance and three non-pain stressors: transport from one crib to another (a stressor that triggers crying that is not in response to pain), an air stimulus on the nose (a stressor that provokes eye squeeze), and friction on the surface of the heel (a stressor that produces facial expressions of distress that are similar to the expressions of pain). In addition to these four facial displays, the database includes images of the neonates in the neutral state of rest.
2. YouTube dataset- Data collected for each included video comprised date of upload, number of views, age of infant (approximately 0, 2, 4, 6 or 12 months, as assessed by researcher), sex of infant (as per the data collector’s judgment), number of injections, sex of main caregiver in the video, as well as the number of other people in the room, their sex and approximate age (child or adult), and the number of comments on each video. A total of 142 videos were included in the systematic review. A total of 120 videos were assessed for pain management strategies after the completion of the injection. The remaining 22 videos ended immediately after the completion of the injection. Crying duration is measured in seconds, and rated pain using the FLACC (Face, Legs, Activity, Cry, Consolability) tool.

* **Methods**

The framework of the system is divided into three stages, direct expression detection, object tracking and detection compensation. Expression detections are realized by Fast R-CNN and VGG-net. Hidden Markov Model (HMM) on the detection results of the current frame using the information from the previous frame. VGG-Net is also trained with ImageNet dataset. Inter frame technique (video-based) is used to track infant facial expression. Faster R-CNN is used is used for ROI tracking.

Detection Compensation - Hidden Markov Model (HMM) is used for modeling the dynamics of expression changes in a video sequence, to reduce false positives and enhance the stability of the monitoring.



* **Results** - The experimental results show a mean average precision of 81.9% and 84.8% for 4 infant expressions and 3 states evaluated with both clinical and daily datasets. Meanwhile, the average precision for discomfort detection achieves up to 90%. Expressions such as Discomfort, joy, neutral, sleep, open mouth, unhappy, pacifier expressions are detected.

**References**

[1] M. Sullivan and M. Lewis, “Emotional expressions of young infants and children,” Infants Young Child., vol. 16, no. 2, pp. 120–142, Apr.–Jun. 2003.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Washington, DC, USA: IEEE Comput. Soc., 2014, pp. 580–587. [Online]. Available: [https://doi.org/10. 1109/CVPR.2014.81](https://doi.org/10.%201109/CVPR.2014.81)

[3] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, “Single-shot refinement neural network for object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4203–4212.

[4] J. Li et al., “Facial expression recognition with faster R-CNN,” Procedia Comput. Sci., vol. 107, pp. 135–140, Dec. 2017.

[5] R. B. Girshick, “Fast R-CNN,” 2015 IEEE Int. Conf. Comput. Vis., pp. 1440–1448, 2015.

**A Novel Approach on Infant Facial Pain Classification using Multi Stage Classifier and Geometrical-Textural Features Combination** (Yosi Kristian et al,2017)

In this paper, for face detection and Active Shape Model (ASM) to detect facial landmark point on infant face. Features extraction is conducted and tested a few feature combinations between geometrical and textural to find the most suitable set for infant facial pain classification. Single stage classifier and two stage classifier (preceded by a cry detection) using SVM is tested.

**Dataset-** Hanindito dataset consist of 46 videos from 23 infants. From 44 video ,3 frames from each video and selected various crying expression if the videos contain any crying session.

**Methods-** Viola-Jones algorithm is used face detection. For Facial landmark detection Active Shape Modelling(ASM) is used.

Feature Extraction - Two sets of geometrical feature are extracted which is a normalized distance between landmark points found by ASM. NPPD: Normalized Point Pair Distance ENPPD: Extended Normalized Point Pair Distance is used to extract geometrical features. Local Binary Pattern (LBP) is used to extract texture features. SVM classifier is used for image classification Radial Basis Function (RBF) and polynomial kernel is used.

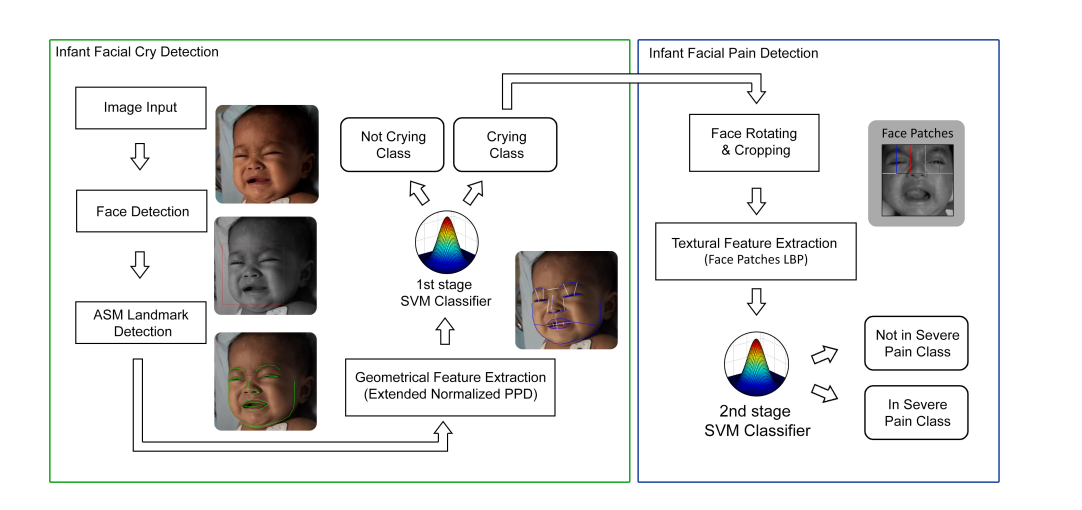


Figure Block diagram of proposed system

**Results –** Thecombination of geometrical feature and textural feature give the best result 88.7% accuracy with 0.871 average F1 Score. From results it is seen that geometrical feature is more suitable for cry detection, and for pain classification, the textural feature performs better than other features and even better than combination between textural and geometrical.

**References**

[1] K. M. Prkachin, N. A. Currie, and K. D. Craig, “Judging nonverbal expressions of pain.,” Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, vol. 15, no. 4, p. 409, 1983.

[2] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, “Automatically detecting pain in video through facial action units,” Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 41, no. 3, pp. 664–674, 2011.

[3] C. M. Lilley, K. D. Craig, and R. E. Grunau, “The expression of pain in infants and toddlers: developmental changes in facial action,” Pain, vol. 72, no. 1, pp. 161–170, 1997.

[4] S. W. Krechel and J. BILDNER, “Cries: a new neonatal postoperative pain measurement score. initial testing of validity and reliability,” Pediatric Anesthesia, vol. 5, no. 1, pp. 53–61, 1995.

[5] T. Voepel-Lewis, J. R. Shayevitz, and S. Malviya, “The flacc: a behavioral scale for scoring postoperative pain in young children,” Pediatr Nurs, vol. 23, pp. 293–7, 1997.

**Automated Infant Monitoring based on R-CNN and HMM (Cheng Li et al, 2021)**

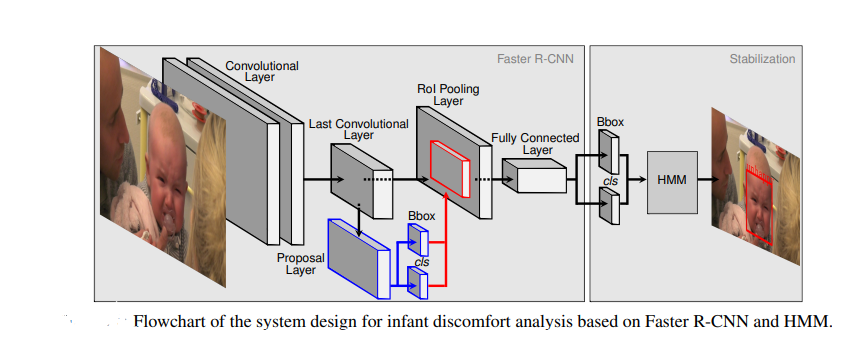
This work proposes a near real-time video-based infant monitoring system for the analysis of infant expressions. The system consists of two components: expression classification and expression state stabilization.

* **Dataset used** – Infant COPE/iCOPE dataset and YouTube dataset.

1. The Infant COPE Database, contains 204 facial images of 26 neonates experiencing the pain of a heel lance and three non-pain stressors: transport from one crib to another (a stressor that triggers crying that is not in response to pain), an air stimulus on the nose (a stressor that provokes eye squeeze), and friction on the surface of the heel (a stressor that produces facial expressions of distress that are similar to the expressions of pain). In addition to these four facial displays, the database includes images of the neonates in the neutral state of rest.
2. YouTube dataset- Data collected for each included video comprised date of upload, number of views, age of infant (approximately 0, 2, 4, 6 or 12 months, as assessed by researcher), sex of infant (as per the data collector’s judgment), number of injections, sex of main caregiver in the video, as well as the number of other people in the room, their sex and approximate age (child or adult), and the number of comments on each video. A total of 142 videos were included in the systematic review. A total of 120 videos were assessed for pain management strategies after the completion of the injection. The remaining 22 videos ended immediately after the completion of the injection. Crying duration is measured in seconds, and rated pain using the FLACC (Face, Legs, Activity, Cry, Consolability) tool.

**Methods-** The framework of the system is divided into three stages, direct expression detection, object tracking and detection compensation. Expression detections are realized by Fast R-CNN and VGG-net. Hidden Markov Model (HMM) on the detection results of the current frame using the information from the previous frame. VGG-Net is also trained with ImageNet dataset. Inter frame technique (video-based) is used to track infant facial expression. Faster R-CNN is used is used for ROI tracking.

Detection Compensation - Hidden Markov Model (HMM) is used for modeling the dynamics of expression changes in a video sequence, to reduce false positives and enhance the stability of the monitoring. VOC2007 metrics is used to evaluate the accuracy of infant expression detection. VOC2007 is specifically designed for detection and classification tasks.



**Results-** The experimental results show a mean average precision of 82.3% and 83.4% for 7 different expression classifications, and up to 90% for discomfort detection, evaluated with both clinical and daily datasets. Moreover, when adopting temporal analysis, the false expression changes between frames can be reduced up to 65%, which significantly enhances the consistency of the system output.

**Reference**

1. Brahnam, S., Chuang, C.-F., Shih, F. Y., and Slack, M. R. (2006). Svm classification of neonatal facial images of pain. In Bloch, I., Petrosino, A., and Tettamanzi, A. G. B., editors, Fuzzy Logic and Applications, pages 121–128
2. Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., and Darrell, T. (2017). Long-term recurrent convolutional networks for visual recognition and description. IEEE Trans. on Pattern Anal. Mach. Intell., 39(4):677–691.
3. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International Journal of Computer Vision, 88(2):303–338.
4. Fotiadou, E., Zinger, S., Tjon a Ten, W. E., Oetomo, S., and de with, P. H. N. (2014). Video-based facial discomfort analysis for infants. In Proc. SPIE 9029, Visual Information Processing and Communication V, 90290F.
5. Harrison, D., Sampson, M., Reszel, J., Abdulla, K., Barrowman, N., Cumber, J., Li, C., Nocholls, S., and Pound, C. M. (2014). Too many crying babies: a systematic review of pain management of practices during immunizations on youtube. BMC Pediatrics, 14:134.