2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Ideal Modified Adachi Chaotic Neural Networks and Active Shape Model for Infant Facial Cry Detection on Still Image

Yosi Kristian*‡
*Informatics Engineering Department
*Sekolah Tinggi Teknik Surabaya (STTS)
Surabaya, Indonesia
Email: yosi@stts.edu, yosik81@gmail.com

Mochamad Hariadi‡, Mauridhi Hery Purnomo‡
‡Electrical Engineering Department
‡Institut Teknologi Sepuluh Nopember (ITS)
Surabaya, Indonesia
Email: {mochar,hery}@ee.its.ac.id

***Abstract*—In this paper, we develop a pattern recognition system to detect weather an infant is crying or not just by using his facial feature. The system must first detect the baby face by using the Haar-like feature, then find the facial component using trained active shape model (ASM). The extracted feature then fed to Chaotic Neural Network Classifier. We designed the system so that when the testing pattern is not a crying baby the system will be chaotic, but when the testing pattern is a crying baby face the system must switch to being periodic. Predicting whether a baby is crying based only on facial feature is still a challenging problem for existing computer vision system. Although crying baby can be detected easier using sound, most CCTV don't have microphone to record the sound. This is the reason why we only use facial feature. Chaotic Neural Network (CNN) has been introduced for pattern recognition since 1989. But only recently that CNN receive a great attention from computer vision people. The CNN that we use in this paper is the Ideal Modified Adachi Neural Network (Ideal-M-AdNN). Experiments show that Ideal-M-AdNN with ASM feature able to detect crying baby face with accuracy up to 93%. But nevertheless this experiment is still novel and only limited to still image.**

***Index Terms*—Chaotic neural networks, active shape model, chaotic pattern recognition, ideal modified adachi neural network, infant facial cry detection.**

I. Introduction

Computer vision researches on infant face are starting to gain attention, like on [1] and [2] their research are related to infant face recognition, while [3] tries to detect if infants are in pain, and in [4] they develop a system to detect foreign object on baby face such as vomit.

In this research, we develop a system to detect a crying infant face. The first step is for the face detection, we use haar-like feature for the face detection [5]. After the face region discovered we match a trained ASM [6] to find face component location. From the face marking found by trained ASM we generate the NN feature to feed the Ideal-M-AdNN [7]. The flowchart of this infant facial cry detection system can be seen on Fig. 1

II. Chaotic Neural Networks

In common usage, "chaos" means "a state of disorder" when the present determines the future, but the approximate present

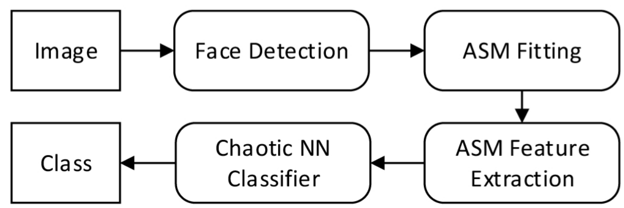


Fig. 1. Flowchart of Infant Cry Detection

does not approximately determine the future [8]. Based on [9] for a dynamical system to be classified as chaotic, it must have the following properties: 1) It must be sensitive to initial conditions; 2) It must be topologically mixing; and 3) Its periodic orbits must be dense. A Chaotic NN must also satisfy these properties.

The first model of Chaotic Neural Network was introduced by Aihara in 1989 [10]. Then in 1997 Adachi and Aihara propose a modified model an associative network that is constructed with chaotic neuron models and interconnected through auto-associative matrix known as Adachi Neural Network (AdNN) [11].

In 2007, based on the AdNN, Calitoiu introduce a modifications to the basic AdNN connections and called it Modified AdNN (M-AdNN) [12]. More recently in 2012 Qin and Oommen propose a tuning up of the M-AdNN to improve its pattern recognition ability, we will refer to this improvement as Ideal-M-AdNN [7]. The last model is the model we use to predict whether a facial feature is belong to a crying infant or not.

A. Adachi's Model of Chaotic Neural Networks: AdNN

The AdNN, actually is a Hopfield-like model, which composed of N neurons, and topologically arranged as a completely connected graph [11]. Each neuron communicates with every other neuron, including itself.

The AdNN is modelled as a dynamical associative memory, by means of the following equations relating the two internal states $\eta_i(t)$ and $\xi_i(t), i = 1..N$, and the output $x_i(t)$ as:

$$x_i(t+1) = f(\eta_i(t+1) + \xi_i(t+1)) \quad (1)$$

$$\eta_i(t+1) = k_f \eta_i(t) + \sum_{j=1}^{N} w_{ij} x_j(t) \quad (2)$$

$$\xi_i(t+1) = k_r \xi_i(t) - \alpha x_i(t) + a_i \quad (3)$$

- $x_i(t)$ is the output of the neuron i which has an analog value in $[0, 1]$ at the discrete time t.
- f is the logistic function with the steepness parameter satisfying:

$$f(y) = \frac{1}{1 + exp(-y/\varepsilon)} \quad (4)$$

- k_f is the decay parameters for the feedback inputs and and k_r is the refractoriness.
- w_{ij} are the synaptic weights to the i^{th} constituent neuron from the j^{th} constituent neuron.
- a_i denotes the temporally constant external inputs to the i^{th} neuron (=bias).

The feedback interconnections are determined according to the following symmetric auto associative matrix of the p stored patterns as in:

$$w_{ij} = \frac{1}{p} \sum_{s=1}^{p} (2x_i^s - 1)\left(2x_j^s - 1\right) \quad (5)$$

where x_i^s is the i^{th} component of the s^{th} stored pattern, and the bias is:

$$a_i = 2 + 6x_i^s \quad (6)$$

B. A modified Adachi Neural Networks: M-AdNN

Calitoiu et al on [12] and [13] propose a modification of Adachi's chaotic neural networks to be able to perform pattern recognition task. The M-AdNN introduce two global internal states used for all neurons, which are $\eta(t)$ and $\xi(t)$ obeying:

$$\eta_i(t+1) = k_f \eta(t) + \sum_{j=1}^{N} w_{ij} x_j(t) \quad (7)$$

$$\xi_i(t+1) = k_r \xi(t) - \alpha x_i(t) + a_i \quad (8)$$

After each step $t+1$, the global states are updated with the values of $\eta_N(t+1)$and $\xi_N(t+1)$.

On M-AdNN, every time instant, when computing a new internal state, we only use the contents of the memory from the internal state for neuron N. This is in contrast to the AdNN in which the updating at time $t+1$ uses the internal state values of all the neurons at time t.

The second change is for the weight assignment rule. In M-AdNN, instead of using equation 5, they use the classical variant:

$$w_{ij} = \frac{1}{p} \sum_{s=1}^{p} (x_i^s)\left(x_j^s\right) \quad (9)$$

In AdNN, if for any $i, j(1..N)$ the value $x_i^s = x_j^s = 0$ for all $s(1..p)$, then w_{ij} will be unity. However, for the M-AdNN, the value of w_{ij} will be zero in the identical setting. Clearly, the M-AdNN has a smaller self-feedback effect than the AdNN and the duration of the transitory process will be short if the coupling between the neurons is small.

The third modification is applying external input by increasing the biases, a_i, from 0 to unity whenever x_i^s = 1, keeping the other biases to be 0 whenever x_i^s = 0. The M-AdNN is more sensitive to the external input than the AdNN because in M-AdNN $a_i = x_i^s$ as opposed to the AdNN in which $a_i = 2 + 6x_i^s$. The range of input values is between 0 and 1 in the M-AdNN, in contrast with the range of input values being between 2 and 8 in the AdNN. Thus, the M-AdNN will be more receptive to external inputs, leading to, hopefully, a superior Pattern Recognition (PR) system.

Designing PR systems based on the brain model is not an easy task. On typical PR system, when presented by a testing sample, a decision to identity the class of the sample is made using the corresponding discriminant function, and this class is proclaimed by the system as the identity of the pattern. Contrary, Chaotic PR systems do not report the identity of the testing pattern with such a proclamation, chaotic PR system continuously demonstrate chaos as long as there is no pattern to be recognized, or whenever presented by an unrecognized pattern. But, when presented by a recognized pattern, the proclamation of the identity to be made by requiring the system simultaneously resonates sympathetically.

C. Fine tuning Modified Adachi Neural Networks: Ideal-M-AdNN

Calitoiu et. al. on [12], were the first researchers who recorded the potential of chaotic NNs to achieve PR. But unfortunately, their model, named the M-AdNN, was not capable of demonstrating all the PR properties mentioned before. A chaotic PR system is not intended to report the identity of a testing pattern with a class "proclamation" as in a traditional PR system. The system must yield a strong periodic signal when a trained pattern, which is to be recognized, is presented. Further, between two consecutive recognized patterns, none of the trained patterns must be recalled. On the other hand, and most importantly, if an untrained pattern is presented, the system must give a chaotic signal.

The primary aim of Ideal-M-AdNN paper is to show that the M-AdNN, when tuned appropriately, is capable of demonstrating ideal PR capabilities. There are three modification made by Qin and Oommen in [7]. The first is to return the weight initialization to the original ACNN algorithm on equation 5. Second are finding the Steepness Parameters ε. Signicance of ε for the ACNN: The next issue to consider concerns the value of the steepness parameter, ε, of the output function. As explained earlier, the output function is defined by the Logistic function which is a typical sigmoid function. One can see that ε controls the steepness of the output. If ε = 0.01, then f(x) is a normal sigmoid function. If ε is too small, for example, 0.0001, the logistic function almost degrades to become a unit step function. The value of ε as set in Calitoiu reserach [13] to be ε = 0.00015, is not appropriate. Rather, to develop the

Ideal-M-AdNN, they have opted to use a value of ε which is two orders of magnitude larger, i.e., ε = 0.015.

III. Feature Extraction

A. Haar Like Features

Haar-like features [5] and [14] is one of the well known methods for face detection. This method works by changing a part of image (Region of Interest) into a value. This method is one of the common methods for face detection with high speed and accuracy. They owe their name to their intuitive similarity with Haar wavelets and commonly used in the real-time face detector.

The term haar-like features origins from the calculation of Viola and Jones which similar with haar wavelet transform method . Basically, haar-like features working with changing a region near an image into pixel area based on its classifier [15]. the next step will be to calculate the intensity difference of each pixel area. The resulting difference is used to categorize each area for an object detection.

A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. Basically, haar-like features working with changing a region near an image into pixel area based on its classifier. The next step will be to calculate the intensity difference of each pixel area. The resulting difference is used to categorize each area for an object detection. Example of Haar-like feature implementation for face detection can be seen on Fig. 2.

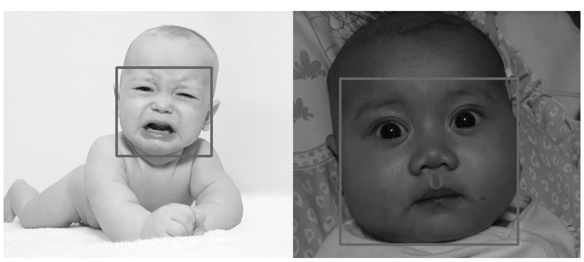


Fig. 2. Infant Face Detection

B. Active Shape Model

Active Shape Model (ASM) [6] [16] is a method where the model is iteratively changed to match that model into a model in the image. This method is using a flexible model which acquired from a number of training data sample. Given a guessed position in a picture, ASM iteratively will be matched with the image. By choosing a set of shape parameter b for Point Distribution Model, the shape of the model can be defined in a coordinate frame which centered in the object. Instance X in the images model can be constructed by defining position, orientation, and scale.

The ASM starts the search for landmarks from the mean shape aligned to the position and size of the face determined by previous face detector. It then repeats the following two steps until convergence:

1) Suggest a tentative shape by adjusting the locations of shape points by template matching of the image texture around each point.
2) Conform the tentative shape to a global shape model. The individual template matches are unreliable and the shape model pools the results of the weak template matchers to form a stronger overall classifier.

We train our ASM using 40 manually marked infant face, 20 of them are crying face, the rest are on normal expression. Fig. 3 shows ASM search on crying face, while Fig 4 shows ASM search on normal expression face.

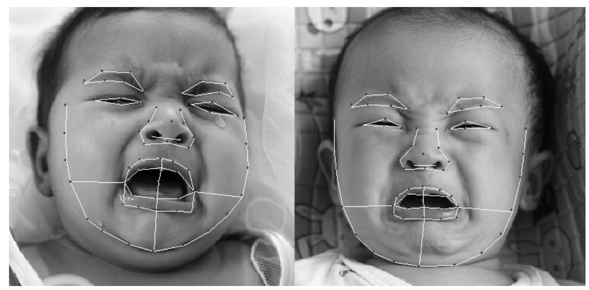


Fig. 3. Crying Infant Face ASM

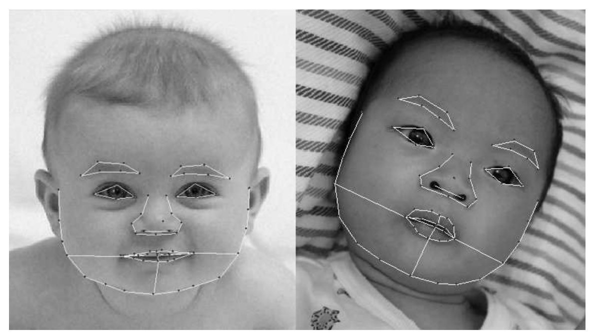


Fig. 4. Normal Expression Infant Face ASM

In this paper we only use geometrical feature extracted from ASM. Form [17] the goal of ASM is to generalize and detect features in new object in the wild, ”new” meaning not in the training set. We choose ASM instead of AAM because detecting infant face feature from a new picture is very close to the ASM goal. ASM also used in many expression research like in [18] and [19]. In this paper we use 68 landmark point for the face ASM same as in FG-NET aging dataset [20]. We use 15 of the normal expression face samples from FG-NET ageing dataset. We only selected images which have label 0 for the age.

C. Extracting Features From Active Shape Model

Extracting feature from facial landmark point were quite straightforward. We discover that crying infant have different shape on mouth, chin shape, eyes, and eye brows distance. Facial landmark points discovered by ASM were converted to distances to the center of mouth, eyes and nose. On Fig 5 we can see the complete feature drew in lines. We extracted 45 feature from the ASM.

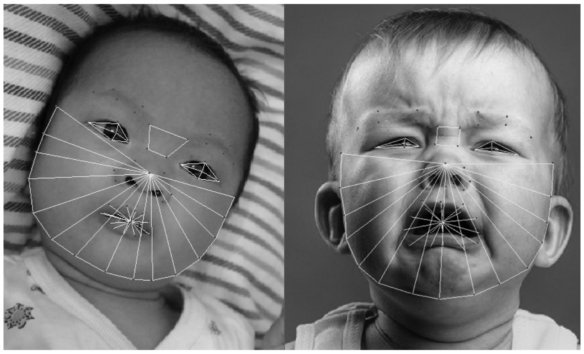


Fig. 5. Distance of Facial Landmark

IV. Chaotic NN Classifier

We design the Chaotic NN to be trained only by using crying faces. When the testing pattern is a crying face the system will be periodic, but when the testing pattern is not a crying baby the system must switch to being chaotic. We implement training and testing algorithm based on [12] then modified it based on [7].

Because there are 45 feature from ASM, we create a CNN classifier with 45 recurrent neurons. We trained our CNN only with crying face feature because the characteristic of CNN that output a periodic signal only when the input pattern are close to the trained pattern. So when we fed the CNN with a non-crying baby face feature the CNN will output a chaotic signal, contrary when we fed the CNN with a crying baby face feature the CNN will output a periodic signal. Fig 6 show CNN output for crying and non crying image.

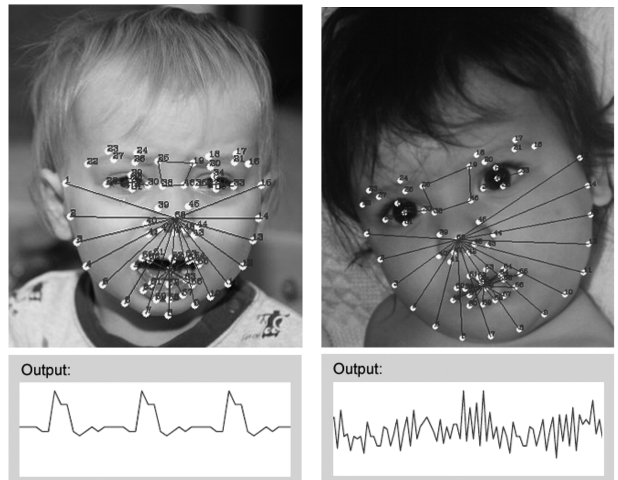
Since the goal of this research is to detect weather the input was a crying baby image or not, we did not design the CNN the other way around (trained them with just normal expression). If we trained the CNN with just the normal baby face feature then when the baby make other expression such as happy the CNN will also output a chaotic signal. The other reason behind our design, is because we are preparing to use video as input on our next research.

We achieve the detection property by running the CNN contiguously. Whenever the system output a periodic signal, means in the input image there is a crying baby face detected.

V. Experimental Result

The CNN classifier are tested using 30 images (10 crying and 20 normal or smiling). The result are quite fascinating, from 10 crying image all of them are detected as crying, and from 20 normal or smiling image 18 of them are detected as non crying, only two are miss classified as crying. The confusion matrix is shown on Table I and the miss classified images are shown on fig 7.

The performance of our CNN classifier compared to well known classifier such as SVM and MLP are shown on Table II. For SVM we use Radial Basis Function kernel and SMO optimization [21]. While for MLP we use backpropagation learning and logistic activation function.



(a) Crying Face (b) Non Crying Face

Fig. 6. (a) shows that CNN classifier will produce a periodic signal when the input image contains crying baby and (b) shows that CNN classifier will produce a chaotic signal when the baby face on the image is not crying.

TABLE I
Confusion Matrix for Ideal-M-AdNN Classifier

		Prediction	
		Crying	Non-Crying
Actual	Crying	10	0
	Non-Crying	2	18

TABLE II
Accuracy Comparison

Classifier	Training Data		Accuracy
	Crying	Non-Crying	Crying
SVM (SMO)	20	20	100%
MLP (Backpropagation)	20	20	97%
CNN (Ideal-M-AdNN)	20	0	93%

Fig. 7. Missclassified Images

VI. Conclusion and Future Work

In this paper, we have discussed the possibility to detect a crying baby based only on facial feature. With classification accuracy up to 93%, we can conclude that the geometrical facial feature discovered by ASM combined with Chaotic NN is made a good baby cry detection. The lower accuracy on our classification system is the result of our system did not trained using the non-crying image. For future works we planned to develop a system that can detect crying baby from video or real time camera stream.

Acknowledgement

This work was partially funded by Directorate General of Higher Education, Ministry of Education and Culture of Indonesia.

References

[1] D. Wen, C. Fang, X. Ding, and T. Zhang, "Development of recognition engine for baby faces," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 3408–3411, IEEE, 2010.

[2] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and S. K. Singh, "Face recognition for newborns: A preliminary study," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pp. 1–6, IEEE, 2010.

[3] M. Mansor, M. Rejab, S.-F. Jamil, A.-F. Jamil, A. Junoh, and J. Ahmad, "Fast infant pain detection method," in *Computer and Communication Engineering (ICCCE), 2012 International Conference on*, pp. 918–921, 2012.

[4] W.-L. Ou, S.-M. Yu, J.-W. Chang, and C.-P. Fan, "Video based vomit and facial foreign object detections for baby watch and safety," in *Orange Technologies (ICOT), 2013 International Conference on*, pp. 219–222, 2013.

[5] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511, IEEE, 2001.

[6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[7] K. Qin and B. Oommen, "Ideal chaotic pattern recognition using the modified adachi neural network," in *Chaotic Modeling and Simulation (CMSIM) 2012*, pp. 701–710, 2012.

[8] S. H. Kellert, *In the wake of chaos: Unpredictable order in dynamical systems*. University of Chicago Press, 1994.

[9] B. Hasselblatt and A. Katok, *A first course in dynamics: with a panorama of recent developments*. Cambridge University Press, 2003.

[10] K. Aihara, T. Takabe, and M. Toyoda, "Chaotic neural networks," *Physics letters A*, vol. 144, no. 6, pp. 333–340, 1990.

[11] M. Adachi and K. Aihara, "Associative dynamics in a chaotic neural network," *Neural Networks*, vol. 10, no. 1, pp. 83–98, 1997.

[12] D. Calitoiu, B. J. Oommen, and D. Nussbaum, "Periodicity and stability issues of a chaotic pattern recognition neural network," *Pattern analysis and applications*, vol. 10, no. 3, pp. 175–188, 2007.

[13] D. Calitoiu, *Chaotic Systems for Pattern Recognition and Brain Modeling*. PhD thesis, School of Computer Science Carleton University, Ottawa, Ontario, December 2006.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[15] E. Setyati, Y. Kristian, Y. M. Pranoto, and D. Alexandre, "Augmented reality 3d eyeglasses frame simulator using active shape model and real time face tracking," in *Proceedings of 2013 Eleventh International Conference on ICT and KE*, IEEE, November 2013.

[16] T. F. Cootes and C. J. Taylor, "Active shape models-smart snakes," in *BMVC92*, pp. 266–275, Springer, 1992.

[17] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Comparing active shape models with active appearance models.," in *BMVC*, vol. 99, pp. 173–182, 1999.

[18] E. Setyati, Y. Suprapto, and M. Purnomo, "Facial emotional expressions recognition based on active shape model and radial basis function network," in *Computational Intelligence for Measurement Systems and Applications (CIMSA), 2012 IEEE International Conference on*, pp. 41–46, 2012.

[19] C.-C. Hsieh and M.-K. Jiang, "A facial expression classification system based on active shape model and support vector machine," in *Computer Science and Society (ISCCS), 2011 International Symposium on*, pp. 311–314, 2011.

[20] "The FG-NET Aging Database." http://www-prima.inrialpes.fr/FGnet/. Accessed: 2012-10-12.

[21] J. Platt *et al.*, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.