Facial Expression Recognition by Active Basis Model

Chih-Yu Hsu[#1], Kuang-Yo Jeng[*2]

[*]*Department of Information and Communication Engineering, ChaoYang University of Technology*
168 Jifeng E. Rd., Wufeng, Taichung County, 41349,Taiwan

[1]Corresponding Author: tccnchsu@cyut.edu.tw
[2]jasperlovetina@gmail.com

***Abstract*—This paper proposed an image recognition method for distinguishing eight facial expressions. When the eyes and the mouth of a person are open or closed, there are eight possible facial expressions to be distinguished. The Active Basis Model is used to construct basis templates and deformation templates for the eyes and mouth. The characteristics of the basis template are used as features for training the Support Vector Machines. The eye and mouth areas of the test images are extracted and used to generate the basis and deformation template. After three Support Vector Machines were trained, eight facial expressions can be classified due to the statuses of two eyes and mouth. The accuracy of facial expression recognition is 93.076% for testing 310 facial images. The results show the proposed method for facial expression recognition is very effective.**

***Keywords*—Active Basis Model, Support Vector Machines, Facial Expression Recognition, Basis and Deformation Template**

1. INTRODUCTION

It is more and more convenient to get more images by common photographic equipments like video and camera. However, Image Classification technology has been widely attention because of these great numbers of images. Image recognition is a technology which is looking for the feature of objects from the images, and making analysis and classification to identify and describe images in the image processing process.

Image recognition is a common research agenda in the computer vision field. Humanity can easily find the difference between images, but, how computer can identify differences in the great numbers of images? It is also very time-consuming by using eyes to classify a large number of images, so facial expression feature extraction and recognition has become an important issue.

Jun-Su Jang [1][2]proposed Rectangle feature, Integral image, Proxy connection configuration and AdaBoost learning, these methods have fast computation and high detection rate. It was improved data and removed the adverse classification in 2008.

However, YN Wu [3] proposed a novel detection method in 2007, which using Active Basis Model to construct active template. After that, YN Wu proposed Learning active basis model for object detection and recognition[4] , it can detect and recognize objects quickly.

By facial feature recognition, it can further development of the face recognition system or monitoring systems and other applications. [5][6] construct models by using Active Basis Model, and make use of different classification to recognize mouth opening and closing state. Feature Extraction not only applied to the lips open and close state of recognize, but also classification of the eyes. Making use of changes in the eye and the lips can show more expression of diverse.

We can know people's emotions from changes in facial expression, therefore, Rizon et al [7] using changes of lips to guess the emotion. In his method, he asked the Japanese as his object of study emotion to make lips affective response recognition; he observed 6 different emotional characteristics of the lips, and to classify with the observed characteristics from these lips. Previous studies have been to identify and track the lips, [8][9] recognized the lip with your mouth to identify the words that you speak, then used TTS system to corresponding to the sound issue.

Opening and closing of the lips can also be used in the recognize emotional, to use the mouth to identify the person's emotional expression must have to detect the mouth at first. The easiest way is by the mouth opening and closing to identify whether people laugh. It is very rich of human expression, by eyebrows, eyes, mouth and facial muscles changes can produce a wide range of expression.

It is very important that in the study of human face emotional changes from lips opening and

closing state, however, we can not effectively know the true expression of emotion by recognizing changes in the lips, so this study use of the lips with eyes open and close state to recognize facial expression.

In this study, in order to make faster and efficient classification from the large number of lips and eyes pictures, in this paper taking LIBSVM[10] by C. C. Chang proposed in 2001, taking LIBSVM used in Active Basis Model to classify opening and closing of eyes and lips, the trained classifier can make the subsequent recognize quickly.

2. TRAINING PROCESS OF CLASSIFIER

The first step is to read the image. The second step is to use Movellan et al. [11] that he proposed the eye detection method to detect eyes. The third step is to calculate the proportion of eyes and lips to select cutting range, and the fourth step is to cut a part of the left eye, right eye and mouth's images, then construct the basis template and the deformation template by using Active Basis Model[3][4].And making use of features values of each model input to the SVM training, trained three SVM classifiers on the left eye, right eye and mouth opening and closing state. Classifier training process will be used to test the accuracy in the next section.

3. EXPERIMENTAL METHODS

This section will be detailed describe the step of classifier training process.

3.1. Capture the left eye, right eye and lips of the block images

If you want to extract images of eyes and lips, first must find the location of the eye, Fig. 1(a) is the original image, Fig. 1(b) is to use the Movellan et al. [11] proposed the eye detection method, by using Wavelet theory to find the location of the eye. Extraction of the left eye and right eye in Fig. 2(a), 4D is the distance between the eyes, 2D is the cutting width of the eye, and height is D. By this means the cut left eye and right eye, the size shown in Fig. 2(b).

Mouth part of the extraction method shown in Fig. 2(c), extending down from the eyes part, the distance is 5D, the cutting width of the mouth is 4D, the height is 2D, and the actual crop size as shown in Fig.2(d). As the proportion of different sizes cut will cause increased active basis model building defects, resulting decline in classification accuracy. Therefore, in order to reduce unnecessary errors, we will cutt the proportion of initial setting in this study.

Fig. 1 Use eye detection method to find the eye position. (a) original image (b) after detecting will put markers on the eye.

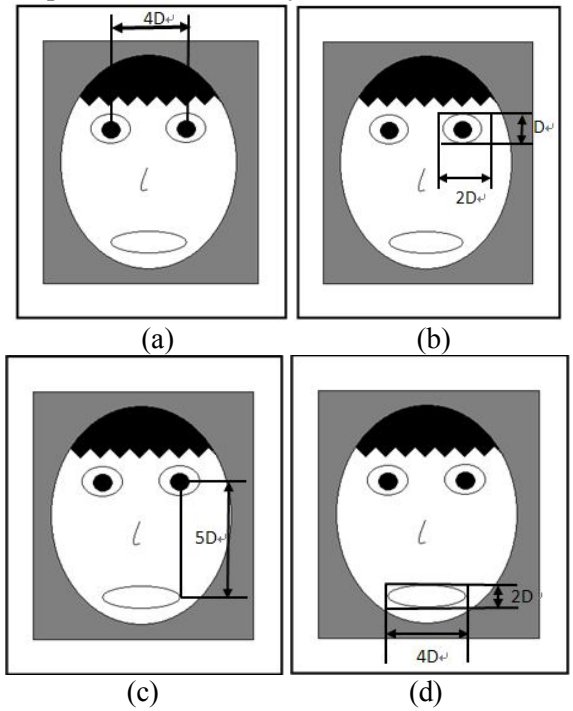


Fig. 2 Cutting left eye, right eye and mouth size ratio diagram. (a) The distance between the eyes after detecting (b) Cutting ratio of left and right eye (c) Set the cutting position of mouth (d) Cutting ratio of mouth.

3.2. Active Basis model (From Gabor element to the deformable template)

Active Basis Model was proposed by Y.N Wu [3][4]. Active Basis Model, which can according to version of image to construct a moving patterns, it can be applied to object detection and recognition. This model is the use of Gabor wavelet elements [3], which is constructed by using in the different direction and location for small perturbations of the linear combination, as shown in Fig. 3.

The Gabor wavelet elements can be rotated, pan and zoom. Gabor wavelet elements like

strokes, a text composed by the number of strokes;

An active basis model is formed by the small amount of Gabor wavelet element, each oval represents an Active basis in the picture. Each Active basis (Gabor wavelet element) through translation or rotation or translation and rotation and other disturbances to fit the edge of the object.

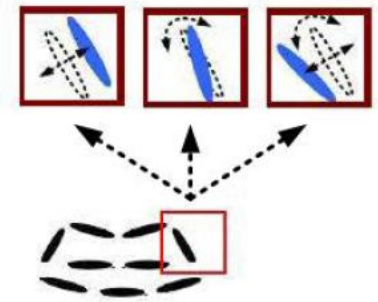


Fig. 3 Active basis model and Active basis disturbance diagram.

3.3 Sared Setch Agorithm

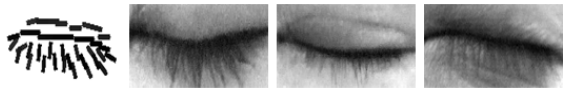
The first step is to construct Active basis model of the deformation template. Shared sketch algorithm is to train deformation template, deformation template of the Active basis is through all the training images chosen by the shared sketch algorithm. All training images using Kullback-Leibler divergence [12] compared the edge of the object (for example, the edge of the left eye) and the degree of difference between the background pixels, and select the top few(Active basis) have the largest difference between background pixels. Taking these basis elements to compose an active basis model, it is the deformation template.

The following example is left eye in Fig. 3, Fig. 4 (a) is a deformation template, Fig. 4 (b) (c) (d)are training images. Deformation template is the use of all training images to learn it, using of shared sketch algorithm to find out the Active basis for all training images. Then, taking these basis elements for the portfolio to obtain deformation template.

It is a following formula about Active basis model combined with Gabor wavelets,

$$I_m = \sum_{i=1}^{n} c_{m,i} B_{m,i} + U_m \quad (1)$$

$B_{m,j}$ represent the m image choose the basis element, $C_{m,j}$ is the weight of each basis element. The weight depends on the direction of the base element, the size of angle and zoom, the same position, angle and zoom factor, with the same weight. U_m represent the background residue of the basis element.



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

Fig. 4 is learning for deformation template by using shared sketch algorithm (a) is deformation template,(b)(c)and(d) all are closed left eye images.

3.4. Basis Template

Active basis model composed the deformation template by chosen Active basis at first. The second step is to use the deformation template of Active basis and SUM-MAX architecture [xx] to test object image of the edge perturbed and construct a basis template.

As an example, the image with the left eye, Fig.5 (a) is the active basis model constructs a deformation template at first step, part (b) is SUM-MAX by using the Active basis deformation template corresponding to the test image edge, then through the SUM1 and MAX1 to construct a basis template.

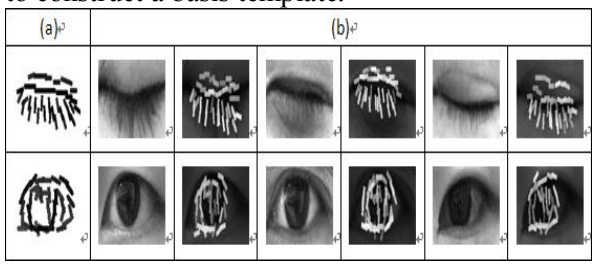


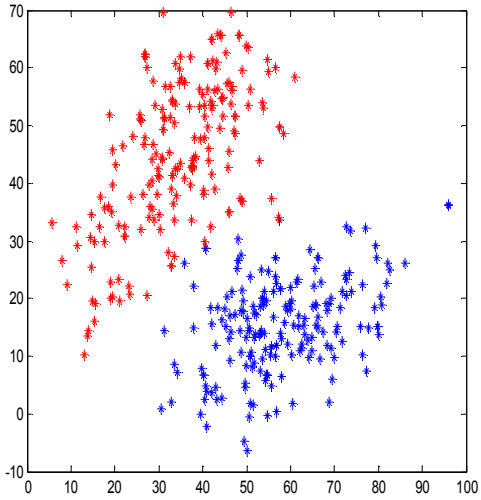
Fig. 5 By Using training images to construct a deformation template, and through the SUM-MAX to construct a basis template.

3.5. Feature of vectors

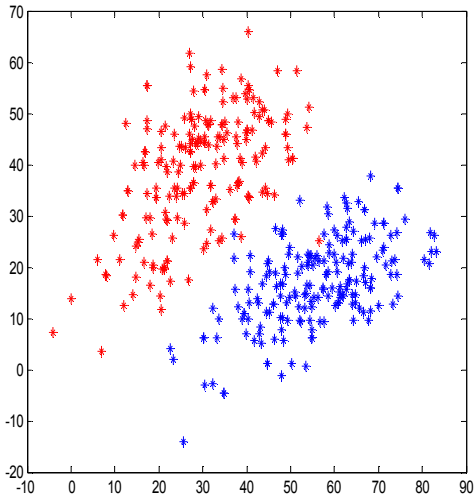
After finishing deformation template and basis template, we calculate the log-likelihood ratio of deformation template and basis template, such as formula (2). In each image, basis template p is corresponding to deformation model q. $r_{m,I}$ is the parallelism basis elements of basis template and deformation template. Baseis template for each image respectively compare with open and closed eyes of deformation template. Therefore, each image will get two vectors. These two vectors can be obtained for one point that represents the image in two-dimensional space. In this article, we use support vector machine for the classification of these points.

$$\log[p(\mathbf{I}_m \mid \mathbf{B}_m)/q(\mathbf{I}_m)] = \sum_{i=1}^{n} \log \frac{p(r_{m,i})}{q(r_{m,i})} \quad (2)$$

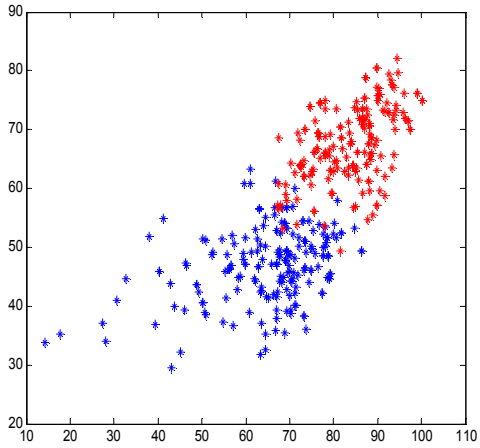
For Fig. 6 (a)(b)(c) as examples. We construct active basis mode (deformation template and basis template)on the left eye, right eye and mouth individually. After using equation (2) calculate the eigenvalues can be expressed vector of points in two-dimensional space. For 390 images as examples, such as Fig. 6 (a) left eye, Fig. 6 (b) right eye and Fig. 6 (c) mouth , the distribution of eigenvalue vectors, we can clearly distinguish the two groups from the distribution of red and blue clusters.



(a)



(b)



(c)

Fig. 6 There are feature vectors after training by active basis model, expressing the distribution of each sample,(a) left eye, (b) right eye (c) mouth.

3.6. Support Vector Machine

Both Support Vector Machines and Neural Networks can be used as classifiers. SVM is to find a optimal hyperplane, and to classify two different sets. In this thesis, we use LIBSVM [8] that develop by Professor Lin Zhiren. Therefore, we use (Active basis model) obtain feature vectors, then input feature vectors to LIBSVM for training, it can get effective classifier and provide quick sort for test data.

3.6.1. Not fault-tolerant linear

We input training information, assuming information is clear that simple linear separable. The main function of SVM is to find the best plane (can be called optimal hyperplane), such as Fig. 7. Using optimal hyperplane will be divided into two sets of training data, and as far as possible let optimal hyperplane away from the width of the the two edges the longer the better. In this way, we can clearly Identify which set of data are, and to train effective classifier.

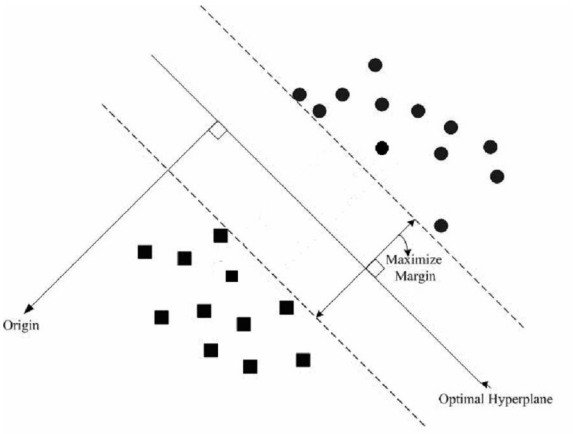


Fig. 7 Schematic diagram of two-dimensional classification SVM - Linear separable.

3.6.2. Linear fault-tolerant

This section introduces the support vector machine under conditions of linear fault-tolerant. Because of training information will be error, it is hard to find optimal hyperplane. It's more common to linearly inseparable. For example, Fig. 8, such information is usually overlapped in the border areas. These orderless informations can not find the support of a particular hyperplane and optimal hyperplane, and separate data specifically. Therefore, in order to deal with these data that overlapped the boundary, Vapnik et al [XX] propose to add the slack variables (ξ) and penalty weight (*C*). ξis represents a non-negative variables, classification error of the sample slack variable is greater than zero; The sample is no classification error equal to 0. Of course, we hope the range as small as possible to slack variables (ξ). Therefore, according to slack variables is equal to zero or not to zero, it will decided to give the corresponding penalty weight (C). And get a new objective function (3) and constraints (4).

$$\text{Min}\frac{\|w\|^2}{2}+C\sum_i \xi_i \tag{3}$$

limited

$$y_i(x_i \cdot w-b)-1+\xi_i \geq 0 \quad and \ \xi_i > 0 \quad \forall i \tag{4}$$

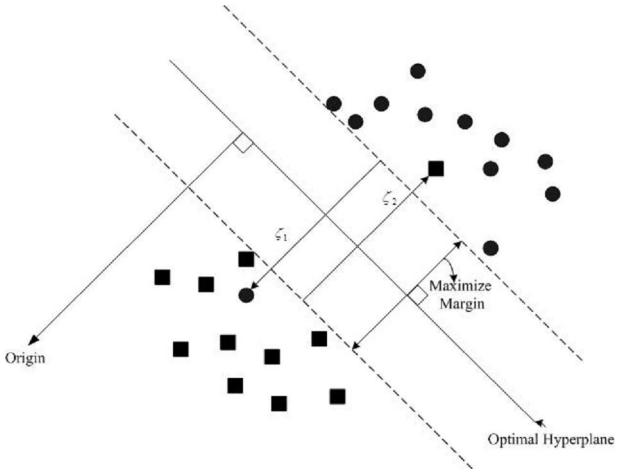


Fig. 8 Classification of two-dimensional diagram of SVM - Linear is not separable.

3.9. Experimental of results

It has to note number of basis element and scaling factor when using active basis model to construct the image templates.

Different number of basis element and scaling factor will impact on the model is good or bad, and affect the final results of the classification.

Different Image content and image size to set the parameters of training is not the same. It ‘s use 5 different number of active basis with 4 scaling factors to describe characteristics of the left eye closed in Fig. 9. We can know that if the number of active basis is too few, it will makes the important features can not fully describe, and scaling factor is too high will make the primitives overlap.

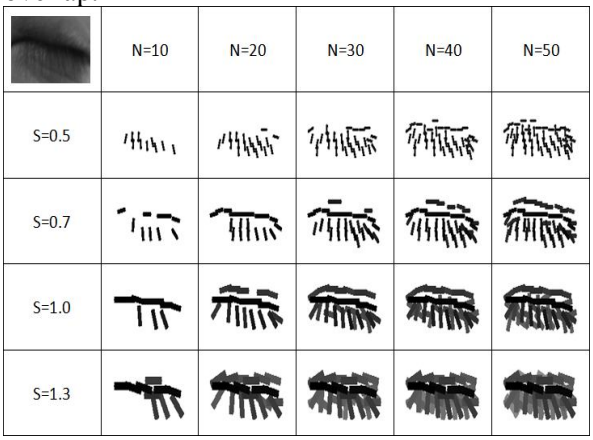


Fig. 9 Number of different Active basis reference different Scaling factor. (N is the number of basis element; S is scale size)

With left eye as an example, in order to training the best basis template makes basis template closest to features of left eye in image. We use the number of active basis and scaling factor to execute Cross validation to get the best training parameters. We let the number of lip

image sample with number of basis element 10, 20, 30, 40, 50, and scale size are 0.5, 0.7, 1.0, 1.3 to execute scaling factor. Such as Fig. 10, after many experiments, we can understand that left eye and right eye, the basis element number is 40 and scale size is 0.7 can construct the best template for the best accuracy.

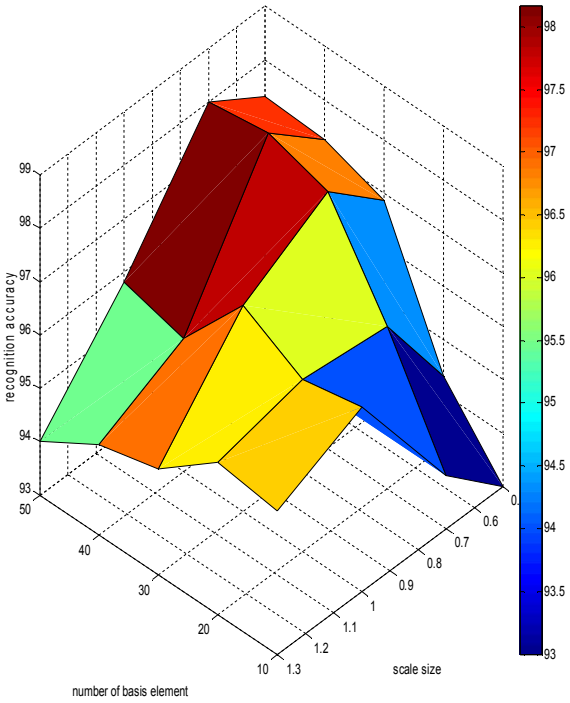


Fig. 10 The result about different number of basis element and scaling factor to cross validation. Test images are 200 open left eye images and 190 closed left eye images.

Experiment consists of 390 face images to be cut, there were 3 images ,left eye, right eye and lips, and each type of image with different types of open and closed. First of all, using 80 images as training images to construct deformation template, then using deformation template to construct 390 basis templates and get vector of feature value. After finishing that, we use SVM to classify. We trained three SVM classifiers.
The classifier SVM_L is used for the classification of the left eye. The classifier SVM_R is used for the classification of the right eye. The classifier SVM_M is used for the classification of the mouth.

The images of left eye, right eye and lips are training the SVM classifiers. It use 80 images, and use 390 images to test. The results show the correct use of the research methods as high as 93.076% rate, such as Table 1, Table 2.

Table 1 It is calculate the correct rate of recognition that after Active basis model construct the template and through SVM compare with original image.

No.	True			Class			Results
	L.E	Mouth	R.E	L.E	Mouth	R.E	
Test_000_001	0	0	0	0	0	0	Correct
Test_000_002	0	0	0	0	0	0	Correct
Test_000_003	0	0	0	0	0	0	Correct
⋮							
Test_100_207	1	0	0	1	1	0	Error
Test_100_208	1	0	0	1	0	0	Correct
Test_100_209	1	0	0	1	1	0	Error
⋮							
Test_111_390	1	1	1	1	1	1	Correct

Annotation: 0 represent closed eyes and lips; 1 represent open eyes and lips

Table 2 Left eye, right eye, mouth and the sum of the recognition accuracy.

Left eye recognition accuracy	mouth recognition accuracy	right eye recognition accuracy	sum of the recognition accuracy
98.717	95.384	98.974	93.076

4. CONCLUSIONS

Eight kinds of expressions are recognized according to the statuses of the eyes and lip. By using active basis model, the deformation and basis template are constructed for two types of image because the left eye, right eye and the mouth can be open or closed. The characteristics of basis template are used as features for the inputs of Support Vector Machines. Experimental results show facial expression recognition accuracy is 93.076%. In the future, the proposed method can offer a basis for the recognition of emotional.

ACKNOWLEDGEMENTS

The authors would like to thank the National Science Council, ROC, for partially supporting this work under grant number NSC 99-2115-M-324-001.

REFERENCES

[1] P. Viola and M. Jones, (2002), "Robust real-time object detection," *International Journal of Computer Vision*.

[2] Evolutionary Pruning for Fast and Robust Face Detection Jun-Su Jang and Jong-Hwan Kim, *Member, IEEE*.

[3] Wu, Y. N., Shi, Z., Fleming, C., & Zhu, S. C. (2007). Deformable template as active basis. *Proceedings of international conference on computer vision*.

[4] Wu, Y.N., Si, Z.Z., Gong. H., Zhu, S.C., "Learning active basis model for object detectionand recognition," International Journal of Computer Vision, in press, 2010.

[5] Hao Tang, Yun Fu, Jilin Tu, Thomas S. Huang, and Mark Hasegawa-Johnson, "EAVA: A 3D Emotive Audio-Visual Avatar," 2008 IEEE Workshop on Applications of Computer Vision (WACV'08), Copper Mountain, CO, January, 2008

[6] Hao Tang, Yun Fu, Jilin Tu, Mark Hasegawa-Johnson, and Thomas S. Huang, "Humanoid Audio-Visual Avatar with Emotive Text-To-Speech Synthesis," IEEE Transactions on Multimedia (T-MM), Volume: 10, Issue: 6, pp. 969-981, October, 2008

[7] Mohamed Rizon, M. Karthigayan, Sazali Yaacob and R. Nagarajan, "Japanese face emotions classification using lip features," *Geometric Modeling and Imaging*, 2007, pp.140-144.

[8] *Hao, T., F. Yun, J. Tu, S. Thomas .S. Huang, and Mark Hasegawa-Johnson, (2008). EAVA: A 3D Emotive Audio-Visual Avatar, 2008 IEEE Workshop on Applications of Computer Vision (WACV'08), 1-6.*

[9] Hao, T, F. Yun, J. Tu, Mark Hasegawa-Johnson, and Thomas S. Huang, (2008). Humanoid Audio-Visual Avatar with Emotive Text-To-Speech Synthesis, *IEEE Transactions on Multimedia (T-MM)*, vol. 10, no. 6, 969-981.

[10] C. C. Chang and C. J. Lin, (2001),"LIBSVM:a library for support vector machines,", *Software available at* http://www.csie.ntu.edu.tw/~cjlin/libsvm

[11] J. R. Movellan, B. Fortenberry and I.Fasel, "A Generative Framework for Real-Time Object Detection, Ucsd Mplab Technical Report, 2003.

[12] Kedar A. Patwardhan, Guillermo Sapiro, Vassilios Morellas"Robust Foreground Detection in Video Using Pixel Layers", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 30, NO. 4, APRIL 2008