A Robust Learning Algorithm Based on SURF and PSM for Facial Expression Recognition

Jinhui Chen*, Xiaoyan Lin †, Tetsuya Takiguchi*, Yasuo Ariki*
ianchen@me.cs.scitec.kobe-u.ac.jp, {takigu,ariki}@kobe-u.ac.jp
†Graduate School of Engineering, Kobe University, Kobe, 657-8501, Japan
*Graduate School of System Informatics, Kobe University, Kobe, 657-8501, Japan

***Abstract*—This paper proposes a novel machine-learning framework for facial-expression recognition, which is capable of processing images fast and accurately even without having to rely on a large-scale dataset. The framework is derived from Support Vector Machines (SVMs) but distinguishes itself in three key ways. First, the measure of the samples normalization is based on the Perturbed Subspace Method (PSM), which is an effective way to improve the robustness of a training system. Second, the framework adopts SURF (Speeded Up Robust Features) as features, which is more suitable for dealing with real-time situations. Third, we use region attributes to revise incorrectly detected visual features (described by invisible image attributes at segmented regions of the image). Combining these approaches, the proposed method has the following beneficial properties. First, the efficiency of machine learning can be improved. Experiments show that the proposed approach is capable of reducing the number of samples effectively, resulting in an obvious reduction in training time. Second, the recognition accuracy is comparable to state-of-the-art algorithms.**

Keywords: facial expression recognition; SVMs; SURF; region attributes.

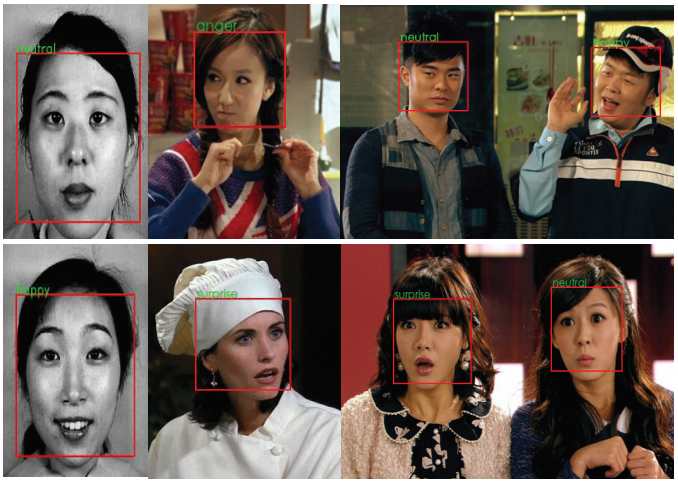


Fig. 1: Examples of Recognition Results

I. Introduction

Facial-expression recognition is a typical multi-class classification problem in computer vision. Furthermore, since it is one of the most significant technologies for auto-analyzing human behaviors, it can be widely applied in many domains. Therefore, the need for this kind of technology in various different areas keeps propelling research forward every year.

As the main detectors, AdaBoost and SVMs, etc. are widely used in this field of research. In 1995, Freund and Schapire [1] supplied the AdaBoost algorithm for realizing the learning framework of Boosted Trees, which could be derived from Probably Approximately Correct (PAC) learning proposed by Valiant [2]. Since then great advances have been made based on AdaBoost, especially milestone work by Viola and Jones [3]. But some ideal strong classifiers are usually required a large number of training samples and very time-consuming training experiments. Even recently, many researchers are trying to solve these problems. Li *et al.* [4] proposed a new learning SURF cascade for ameliorating boosting cascade frameworks. It improved the training efficiency, but the need for large-scale data gathering and extensive preparations create a critical bottleneck. On the other hand, similar problems also exist in methods based on SVMs, because of the limitation of length (they will be not enumerated here). Therefore, collecting many training samples and the long training time leads to considerable work and difficulty for researchers in the field of pattern recognition. Since training is a critical infrastructure for recognition engines, the research on training is significant for learning machines. Hence, there is a great need to solve the problem mentioned above.

This paper brings together new normalization measures, visual features and image attributes to construct a framework for facial-expression recognition. As almost all of the approaches relate to vector processing, the classifier of our proposed method is based on SVMs. There are three main approaches with emphasis on reducing training samples and improving the efficiency of learning machines. First, PSM is used to extend the training data, which allows for the generation of ideal strong classifiers without having to collect a large number of training samples. Second, the features are described by local multi-dimensional SURF descriptors [5], which are spatial regions with windows and are good at processing real-time scenes. Moreover, SURF is much faster and more efficient than most of the existing local features algorithms. Third, the region attributes of images are adopted to revise incorrect detection of classifiers relying on visual features, which are represented by feature vectors in a segmented region. Therefore, the distinctive discriminative capability can guarantee the proposed framework will be more robust.

In experiments, we implemented all training and detection models in C++ on RHEL (Red Hat Enterprise Linux) 6.5 OS

978-1-4799-2186-7/14/$31.00 ©2014 IEEE

platform. The proposed method is designed for dealing with 4-type facial expressions (Neutral, Happy, Angry and Surprised), and some examples of recognition result are shown in Fig. 1. The experimental results show that although using a mini-sized database of training samples, our approach can also construct a robust recognition system, which is comparable to the state-of-the-art method.

In the rest of the paper, we first revisit related works in section 2, then we describe the normalization of samples in section 3 and the classifying framework in section 4. Section 5 elaborates on region attributes estimation. Section 6 shows the experiments, and conclusions are drawn in section 7.

II. Related Work

Facial-expression recognition is a hot research topic in computer vision due to its many applications, and many researchers attach great importance to this field. For instance, Lyons *et al.* [6] adopted PCA and LDA to analyze facial expressions through closed experiments, and they achieved 92% accuracy on JAFFE [7]; Bartlett *et al.* [8] proposed a Gabor feature based AdaSVM method for expression recognition, obtaining a good performance based on the use of the Cohen-Kanade expression database [9]. However, it has been shown that the processing speed of these approaches is too slow to deal with real-time scenes. More recently, Anderson *et al.* [10] and Chen *et al.* [11] proposed their approaches for real-time expression recognition severally, but their methods required a large amount of data for training.

Our approach enables competition that complements a recent line of papers that use third-party software tools to obtain mirror images of samples for training in their facial-expression recognition systems, which we briefly review here. To the best of our knowledge, our approach is the first to employ the PSM directly for detectors training, but not use any tool. Experiments show it has the greatest impact on the performance of training efficiency because time can be saved, which would be spent on collecting vast amounts of data from the Internet or using third-party software to deal with the samples for getting mirror images of these samples. In addition, in our framework, the detector based on SVMs and the classification function are ameliorated, which can guarantee the results will be more reliable.

III. PSM for Samples Normalization

In this paper, PSM is applied to the normalization of samples. The images with visible condition changes, such as direction and illumination changes, can be predicted by this approach. In other words, through normalizing these samples using our approach, we can obtain the same results as when training a mass of their mirror images. Moreover, it is not necessary to waste a lot of time on using third-party software to deal with the samples, and the training time of the samples mirror images can also be saved. Therefore, it is an effective way to improve the robustness of the training system.

A. Training-sample Normalization

In order to reduce the noise, the size of the images is unified by $m \times n$ pixels, and the original samples are normalized by the mean value and variance of pixels transformation. Therefore, the images after the normalization can be obtained according to the following equation:

$$I'(x,y) = a\frac{I(x,y)-\mu}{2\sqrt{2}\sigma} + b \tag{1}$$

Here σ is the standard deviation, and

$$\sigma = \sqrt{\frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n}(I(x,y)-\mu)^2} \tag{2}$$

(a,b) is used to adjust the value of pixels (In this paper, we used regular samples in the experiments, therefore, a was set as 1, b was set as 0). μ is the mean value of pixels, and it can be computed through image traversal using the following equation:

$$\mu = \frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n} I(x,y) \tag{3}$$

B. Changing Orientational Factors

Algorithm 1 Reconstruct Three-dimensional Face

Require:

Input: two-dimensional shape vector: $S_{2D} \in R^2$

Output: three-dimensional shape vector: $S_{3D} \in R^3$

Initialization: set $\beta_0 = 0$, $i = 0$

while $i < K$ or $E_r \leq \varepsilon$ **do**

1. Let
$$S_{3D} \Leftarrow s_0 + \sum_{i=1}^{m}\beta_i s_i$$
2. Alignment: S_{2D} is aligned with the two-dimensional shape, which is obtained by projecting the frontal three-dimensional shape (s_i) onto the $x-y$ plane.
3. Minimize
$$\|P(R_\theta S_{3D} + T) - S_{2D}\|^2$$
4. Reconstruct $(S_{3D})_i$ using the shape parameter β_i.
5. Update R_θ and T with the fixed shape parameter and
$$E_r \Leftarrow \|P(R_\theta S_{3D} + T) - S_{2D}\|^2$$
6. Let
$$i \Leftarrow i + 1$$

end while

7. Reconstruct three-dimensional shape using the final shape parameters.
8. Output S_{3D}.

After the preparation in last subsection, we can thus extend the subspace of samples through changing the facial directions of the images. In this paper, we use the method proposed by Chen *et al.* [11] to reconstruct three-dimensional model and obtain three-dimensional data, and we indicate it in Algorithm 1. When E_r is below a threshold ε (e.g. in Ref. [12], Gower suggested setting ε =10^{-4}), or K landmarks are processed over, the while loop would be stopped, and the three-dimensional data will be output. Here β =

$(\beta_1, \beta_2, \cdots, \beta_m)^T$ is the shape parameter and m is the dimensionality of the shape parameter, which is used to adjust three-dimensional shape data. S_{3D} is a $3 \times n$ matrix, P is a 2×3 orthographic projection matrix, T is a $3 \times n$ translation matrix consisting of n translation vectors $t = [t_x, t_y, t_z]^T$, and R_θ is a 3×3 rotation matrix where the yaw angle is θ. In this paper, θ is set as $\pm 15^\circ$, $\pm 30^\circ$, $\pm 60^\circ$. Thus, through Algorithm 1, we can obtain the three-dimensional data $X = (x, y, z)^T$ from the original images. Hence, according to the transformation matrix formula:

$$X' = T_z \cdot T_y \cdot T_x \cdot S \cdot R_z \cdot R_y \cdot R_x \cdot X \tag{4}$$

we can convert the facial directions to extend the subspace of the training samples. Here T and R are the shear mapping transformation matrix and the rotation matrix respectively, and S is represented by the scaling matrix.

C. Changing Illumination Attributes

The illuminative change is conducted according to the following equation:

$$V_2^{(n)} = V_1^{(n)} + \sum_{m=1}^{K} w_m \cdot e_m^{(n)} \tag{5}$$

where V_1 is the changing feature, V_2 is the result after the changes, n is the dimensionality of the feature vector, w is the weight coefficient, and e is the basis of illumination-change-factor vectors.

In this paper, e is obtained through processing the luminance normalized rendering images by principal component analysis (PCA), wherein, m is the principal component ($m = 1, \cdots, 8$). The rendering images are gained by the treatment of three-dimensional images obtained in subsection 2.2.

IV. Classifying Framework

This section will provide the framework used for SVM machine learning through adopting SURF features. Moreover we will also employ the region attributes of image to revise incorrect detection of classifiers relying on visual features. We will describe them separately in this section.

A. Feature Description

SURF is a scale- and rotation-invariant interest point detector and descriptor. It is faster than SIFT [13] and more robust against different image transformations. In this paper, we adopt an 8-bin T2 descriptor to describe the local feature because it was successfully used in [4], and its more robust representation capacity was also demonstrated.

The descriptor can be computed quickly based on sums of two-dimensional Haar wavelet responses and we can make an efficient use of Integral Images [3]. Suppose d_x as the horizontal gradient image, which can be obtained using the filter kernel $[-1, 0, 1]$, and d_y is the vertical gradient image, which can be obtained using the filter kernel $[-1, 0, 1]^T$; Define d_D as the diagonal image and d_{antiD} as the anti-diagonal image, both of which can be computed using two-dimensional filter kernels $diag\ (-1, 0, 1)$ and $antidiag\ (-1, 0, 1)$. Therefore, 8-bin T2 is able to be defined as $v = (\sum(|d_x| + d_x), \sum(|d_x| - d_x), \sum(|d_y| + d_y), \sum(|d_y| - d_y), \sum(|d_D| + d_D), \sum(|d_D| - d_D), \sum(|d_{antiD}| + d_{antiD}), \sum(|d_{antiD}| - d_{antiD}))$.

The local candidate region of the features is divided into 4 cells. The descriptor is extracted in each cell. Hence, concatenating features in 4 cells together yields a 32-dimensional feature. About feature normalization, we use the same measure with [14].

B. Classifier Construction

The classifier of our framework is built based on One-Versus-Rest SVMs (OVR-SVMs). OVR strategy consists of constructing one SVM per class, which is trained to distinguish the samples of one class from the samples of all remaining classes. Normally, classification of an unknown object is carried out by adopting the maximum output among all SVMs. The proposed method is based on OVR-SVMs classifier, and implemented by re-developing liblinear SDK [15].

Usually, most of researchers estimate posterior probability by mapping the outputs of each SVM into probability separately. The method was proposed by Platt [16] . It applies an additional sigmoid function:

$$H(\omega_j | f_j(x)) = \frac{1}{1 + exp\ (c_j f_j(x) + d_j)} \tag{6}$$

$f_j(x)$ denotes the output of the SVM trained to separate the class ω_j from the other classes (total samples are M). Then, for each sigmoid the parameters c_j and d_j are optimized by minimizing the local negative log-likelihood:

$$-\sum_{k=1}^{N} \{p_k log(h_k) + (1 - p_k) log(1 - h_k)\} \tag{7}$$

here are N outputs of the sigmoid function, where h_k is the output of the sigmoid function with the probability p_k event. In order to solve this optimization problem, [16] applied a model-trust minimization algorithm based on the Levenberg-Marquardt algorithm. But in [17], Lin $et\ al.$ pointed out that there are some problems in this method, meanwhile they proposed another minimization algorithm based on Newton's method with backtracking line search.

But unfortunately, there is nothing to guarantee that:

$$\sum_{j=1}^{M} H(\omega_j | f_j(x)) = 1 \tag{8}$$

For this reason, maybe it is necessary to normalize the probabilities as following:

$$H(\omega_j | x) = \frac{H(\omega_j | f_j(x))}{\sum_{j'=1}^{M} H(\omega_{j'} | f_{j'}(x))} \tag{9}$$

Thus, we try to use another approach to estimate posterior probability, using OVR-SVMs to exploit the outputs of all SVMs to estimate overall probabilities. In order to achieve this purpose, we apply the softmax function to be regarded as

generalization of sigmoid function for the multi-SVMs case. Thus, in the spirit of improved Platt's algorithm [18], this paper applies a parametric form of the softmax function to normalize the probabilities by:

$$H(\omega_j|x) = \frac{exp\ (c_j f_j(x) + d_j)}{\sum_{j'=1}^{M} exp\ (c_{j'} f_{j'}(x) + d_{j'})} \tag{10}$$

and here the parameters c_j and d_j are optimized by minimizing the global negative log-likelihood

$$-\sum_{k=1}^{N} log(H(\omega_k|x_k)) \tag{11}$$

Optimizing the parameters c_j and d_j is intention of obtaining the lowest error rate on testing dataset. The reason of why we use the negative log-likelihood is not only it can optimize the parameters c_j and d_j , but also it can be used for comparing the various probability estimates, in other words, it can evaluate the error rate on machine learning and reject some unsatisfactory candidate expression regions described by SURF features.

V. Region Attributes Estimation

After the OVR-SVMs model classifying, we can obtain the recognition result classified by classifiers based on visual features. But it is necessary to make further efforts on reducing miss-recognition, thus, we use invisible image attributes to guarantee this purpose.

The detected face region is divided into 9×10 blocks, and the feature vector of each block is computed. We named it region attributes. It can be obtained after the normalization by equalizing value and variance of the luminance, meanwhile, the norm is set as 1. The region attribute is estimated by the following score equation.

$$d = \left\|X - \bar{X}\right\|^2 - \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_i + \delta^2} (\varphi_i (X - \bar{X}))^2 \tag{12}$$

here φ is eigenvector and λ is eigenvalue, δ^2 is the image noise correct divisor. When $\delta^2 = 0$, it means that the distances of all feature vectors of the current image projecting into subspace are unified, in the other words, the noise is negligible. X is estimated image region attributes, and $\bar{X}$ is the average feature vector of samples. The value of distance is smaller, the score is higher, namely, the probability of miss-detection is lower.

VI. Experiments

In this section we will show the details of implementation, dataset, and evaluation results. The proposed method is designed for Neutral-, Happy-, Angry- and Surprised-expression recognition, and the recognition results are shown in Fig. 1. We implemented all training and detection programs in C++ on RHEL (Red Hat Enterprise Linux) 6.5 OS. In expression recognition, the facial recognition part used the source code of Open CV, which was based on the Viola and Jones framework [3]. And the expressional recognition part was implemented based on the proposed framework. The experiments were done on the PC with Core i7-2600 3.40 GHz CPU and 8 GB RAM, and the training procedure was fully automatic. For SURF extraction, we adopted Integral Image to speedup the computation as described in section 3.1. For machine learning, we built the OVR-SVMs through re-developing liblinear software [15].

A. Experimental Dataset

In the training stage, it is necessary to construct the mini-size training set for machine learning, which will be applied to fix the parameters of sigmoid and softmax function. In the testing stage, we also need to build the testing set for evaluation. The easiest way to do this is to apply the same dataset to both of the training and testing stages. But, as pointed out by Platt [16], using the same data twice can sometimes lead to a disastrously biased estimate. Moreover, the wide practicability cannot be proved. Therefore, in experiments, we used different datasets to training and testing stage separately. The details of training set and testing set are shown as follows:

Training Database Set We used the Cohn-Kanade expression database (CK+) [9], which is a set of frontal face images posed by 123 people, as the training database, but not all of the people posed each type of expression we need. Therefore, we also collected some samples online using an image search engine. Finally, we obtained 240 initial facial samples for each type of emotion. All of facial samples were normalized to 90×100-pixel patches and processed by histogram equalization, no color information was used.

Testing Database Set The testing sets included two parts. One was obtained from soap operas with a total of 10 persons whose facial expressions were similar to the training samples. These images of these actors and actresses are on 8 video clips having a length of 120 seconds. We marked this set as Test Set A. The other one was the JAFFE database [7], whose facial samples are totally different from the CK+ database. We mixed 213 JAFFE images randomly and one image could be reused multiple times, which were made into 8 120-second-long videos, and we marked this set as Test Set B.

B. Experimental Evaluation

Training Experiments The training database of all methods was mentioned above, but only the proposed method did not adopt any process to obtain masses of mirror samples. Hence, it reduced a large number of samples and took only 49.8 min to complete the whole process. Besides, the training procedure was fully automatic. The relative data are shown in Table 1.

TABLE I: Training Efficiency Evaluation Results

Method	Proposed	K-means [19]	LUT_Ada [11]
Time cost	49.8 min	1,589 min	172.5 min

However, in order to enhance the generalization performance of comparison method [19] and comparison method [11], we had to deal with the images by some transformations: 1) mirror reflection; 2) rotate the images by horizontal and vertical angle $\pm 15°$, $\pm 30°$, $\pm 60°$, finally, we obtained each calss 30,720, total 122,880 facial samples for training classifiers. Therefore, they are very time-consuming tasks.

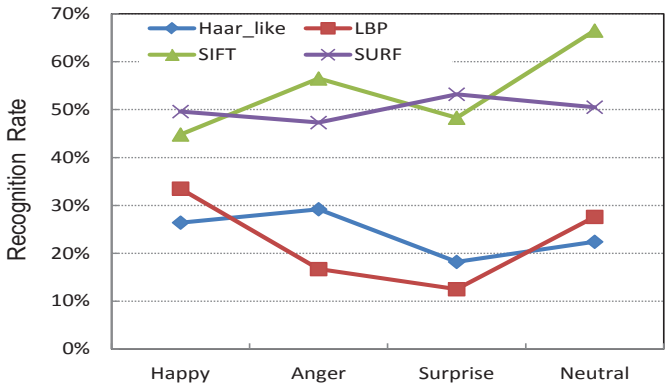


Fig. 2: Green: Recognition rate for OVR-SVMs with SURF. Purple: Recognition rate for OVR-SVMs with SIFT. Blue: Recognition rate for OVR-SVMs with LBP. Red: Recognition rate for OVR-SVMs with Haar-like. Feature detectors using SURF and SIFT obtained the more accurate recognition rate, but the feature extraction speed of SIFT was low.

Testing Experiments Fig. 2 indicates the expression-recognition rate for different feature detectors based on our ameliorated SVMs detector. The aim of this experiment was evaluating the performance of the proposed detector using different methods of feature extraction. Hence, this experiment was done without a PSM model. Feature detectors using SURF and SIFT obtained the more accurate recognition rates, but the average speed of the SIFT detectors version was only 16.8 FPS. In comparison, the speed of the SURF's version reached 39.4 FPS. Theoretically, 16.8 FPS is too slow to deal with complex scenes, such as real-time scenes; thus, SURF was selected as the feature detector.

In Fig. 3, the method choices are compared, which is the reason why the OVR-SVMs+PSM+SURF model was eventually decided on as the proposed method. We also used the other candidate approaches to do many experiments, but this model is the most accurate version of our detector.

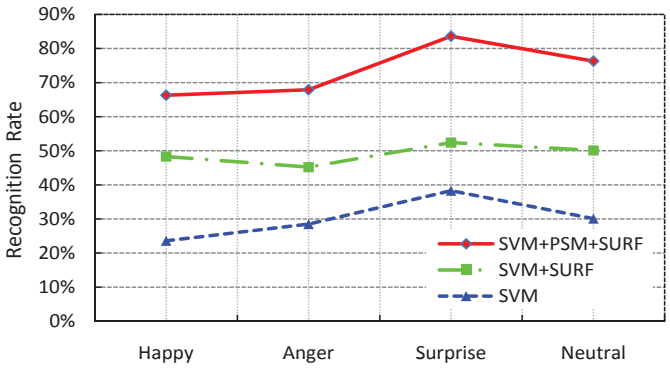


Fig. 3: Top: Recognition rate for proposed method. Middle: Proposed method without PSM; i.e., the OVR-SVMs+SURF model. Bottom: Only OVR-SVMs. OVRSVMs+PSM+SURF (proposed method) is the most accurate version of our detector.

Fig. 4 shows the results of the evaluation experiments for expressional region attributes. Fig. 4. was the result based on testing the Test Set A videos, and Fig. 4. (b) shows the results for Test Set B. In the experiments, we found that after introducing the region attributes model, the recognition accuracy of Test Set A improved approximately 7%. On the other hand, the results of Test Set B were almost unchanged, since the videos in Test Set B consisted of JAFFE images, and these images had been normalized by the supplier [7]. But the videos of Test Set A were used without any normalization. Therefore, this approach is capable of dealing with original images better; i.e., it is good at processing real-life videos.

TABLE II and TABLE III indicate the recognition accuracies, and they show the performance of the proposed method compared to other classifiers ([11] is one of the latest methods for facial expressions recognition, and it was based on AdaBoost; [19] is a typical expressions recognition method using K-means). TABLE II shows the recognition rate of evaluation experiments for Test Set A. Since the races and facial expressions of Test Set As people were similar to those of the training samples, the region attributes model was effective for Test Set A in which there are videos from real life. Consequently, its accuracy was quite better than the result shown in Table 3. The maximum recognition precision of the proposed method was 86.3%, and the worst result was 69.3%.

TABLE II: Experimental Results for Test Set A

	Proposed	K-means [19]	LUT_Ada [11]
Happy	69.3%	52.0%	61.2%
Anger	70.9%	64.5%	50.9%
Suprise	86.3%	42.8%	68.6%
Neutral	78.3%	37.1%	65.6%

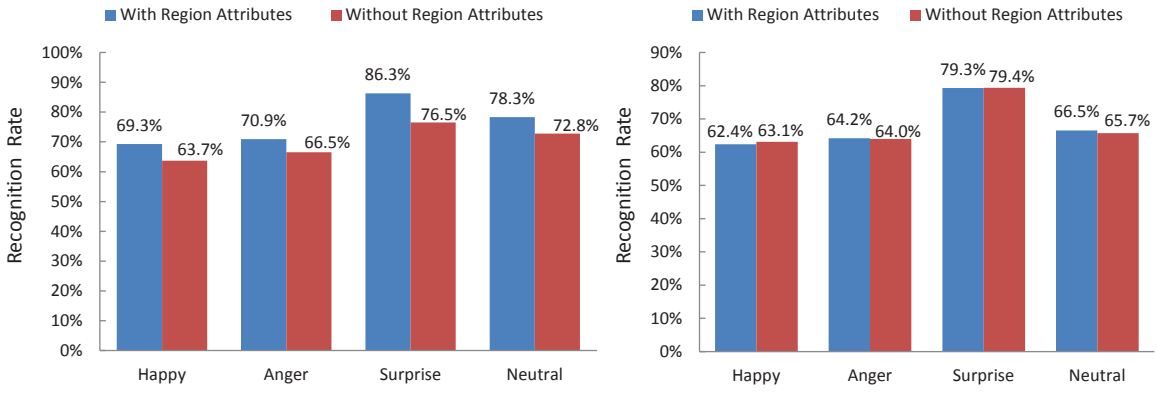
On other hand, TABLE III indicates the recognition accuracies for Test Set B. Due to the variation and complexity of the facial expression across different cultures and races, the region attributes model was not effective. The results of this test set were not better than Test Set B's. But on the whole, the results of both test sets showed that the proposed method was the more accurate version among these methods. Note that the proposed method used training samples without any image-mirror processing. Therefore, based on the mini-sized training set, the proposed method can also obtain a better result, thus this model allows for generating ideal, strong classifiers without the need for a large amount of training samples. Hence, under these experimental conditions, the validity of the proposed approach was proved.

TABLE III: Experimental Results for Test Set B

	Proposed	K-means [19]	LUT_Ada [11]
Happy	62.4%	55.3%	57.7%
Anger	64.2%	59.5%	48.2%
Suprise	79.3%	44.8%	68.4%
Neutral	66.5%	32.6%	71.6%

VII. Conclusion

This paper brings together new normalization measures, visual features and image attributes to construct a novel framework that minimizes the amount of training data needed but improves the training efficiency. It may well have broader application in machine learning.



(a) Region Attributes Estimation for Test Set A

(b) Region Attributes Estimation for Test Set B

Fig. 4: Evaluation Results for Expressional Region Attributes

PSM is an effective approach for alleviating the trouble of collecting large amounts of training samples. By carrying out a large number of experiments, we found that SURF is the most suitable feature descriptor for our detector, and the region attributes of images can revise some incorrectly detected classifiers caused by visual features. Combining these approaches together, a robust expression recognition framework can be constructed, but due to the variation and complexity of facial expressions across different cultures and races, there are many difficult challenges involved with using mini-sized training sets to obtain high recognition precision. Therefore, we have to do more work.

In future research, considering a possible implementation in a real-life scenario, we are inclined to consider these points: 1) We will try to use region attributes as binary latent variables, which are incorporated into the SVMs model for inference, and 2) we will ameliorate approaches on the construction of SVMs to improve accuracy and to make our method capable of handling more complex tasks.

REFERENCES

[1] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*, 1995, pp. 23–37.

[2] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[3] 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. IEEE, 2001, pp. I–511.

[4] J. Li and Y. Zhang, "Learning surf cascade for fast and accurate object detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3468–3475.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*, 2006, pp. 404–417.

[6] M. N. Dailey, C. Joyce, M. J. Lyons, M. Kamachi, H. Ishi, J. Gyoba, and G. W. Cottrell, "Evidence and a computational explanation of cultural differences in facial expression recognition." *Emotion*, vol. 10, no. 6, p. 874, 2010.

[7] M. Kamachi, M. Lyons, and J. Gyoba, "The japanese female facial expression (jaffe) database," *URL http://www. kasrl. org/jaffe. html*, vol. 21, 1998.

[8] M. Bartlett, G. Littlewort, I. Fasel, and J. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction." in *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*, vol. 5. IEEE, 2003, pp. 53–53.

[9] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[10] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 1, pp. 96–105, 2006.

[11] J. Chen, Y. Ariki, and T. Takiguchi, "Robust facial expressions recognition using 3d average face and ameliorated adaboost," in *Proceedings of the 21st ACM International Conference on Multimedia*, ser. MM '13. New York, NY, USA: ACM, 2013, pp. 661–664.

[12] J. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, 2005, pp. 886–893 vol. 1.

[15] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[16] J. Platt, "Probabilities for sv machines," in *Advances in Large Margin Classifiers*, A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA: MIT Press, 2000, pp. 61–74.

[17] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on platts probabilistic outputs for support vector machines," *Machine learning*, vol. 68, no. 3, pp. 267–276, 2007.

[18] Z. Sun, N. Ampornpunt, M. Varma, and S. Vishwanathan, "Multiple kernel learning and the smo algorithm," in *Advances in Neural Information Processing Systems 23*, J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, Eds., 2010, pp. 2361–2369.

[19] N. Alldrin, A. Smith, and D. Turnbull, "Classifying facial expression with radial basis function networks, using gradient descent and k-means," *CSE253*, 2003.