Facial Expression Recognition with PCA and LBP Features Extracting from Active Facial Patches

Yanpeng Liu[a], Yuwen Cao[a], Yibin Li[a], Ming Liu, Rui Song[a]

Yafang Wang, Zhigang Xu , Xin Ma[a]†

Abstract—**Facial expression recognition is an important part of Natural User Interface (NUI). Feature extraction is one important step which could contribute to fast and accurate expression recognition. In order to extract more effective features from the static images, this paper proposes an algorithm based on the combination of gray pixel value and Local Binary Patterns (LBP) features. Principal component analysis (PCA) is used to reduce dimensions of the features which are combined by the gray pixel value and Local Binary Patterns (LBP) features. All the features are extracted from the active facial patches. The active facial patches are these face regions which undergo a major change during different expressions. Softmax regression classifier is used to classify the six basic facial expressions, the experimental results on extended Cohn-Kanade (CK+) database gain an average recognition rate of 96.3% under leave-one-out cross validation method which validates every subject in the database.**

I. INTRODUCTION

Facial expression plays an important role in our communication with other people in our daily life. For the continuous progress and development of intelligent robot, emotional interactions between these robots and people also are the foundational function of these intelligent robots. Facial expression is a significant part of emotion, so it is worth to do research on this field.

The universal expressions which are mentioned in the papers are usually anger, disgust, fear, happiness, sadness and surprise [1] while some researchers add neutral as the seventh expression [2, 9]. The information of [3,4] show that most expressions are invoked by the facial muscles around the mouth while in Facial Action Coding System (FACS) [5], each muscle movement is considered as an Action Unit (AU), these AUs' combination contributes the different facial expressions. Due to the difficult to detect the AUs, it is a big challenge to recognize the facial expressions using the AUs [6]. Therefore, researchers applied the active patches [6,7,8] or used the whole face to extract features for classification. There are two main branches which are using handcrafted features for recognition [6,7,8,9] and using neural networks or deep learning network frame[2,10,14]. Applying machine learning methods for classification on handcrafted features has a continuous progress and a great development, there are many feature extraction methods and classifiers developed in the past years. For facial expression recognition, SVM [7] classifier is the most common machine learning classifier while some researchers chose to use KNNs. In the recent years, deep learning frameworks also achieve better grade than the traditional machine learning methods. Though the deep learning methods could gain better grade, more memory and better computers are required in the experiment [15]. Also the deep learning frameworks are data driven methods. In order to use the combination of the traditional machine learning method and the neural frameworks, we chose the combination features coming from handcraft features and the gray value, and use the softmax regression classifier to recognize the six facial expressions. Softmax regression classifier is hardly used to classify the traditional features. Among the features LBP [9], Gabor [13], HOG [12,17] and their combination [6] are most popular for representing the faces. To be honest, though using one of the methods could get a decent grade, the combination features also could achieve a better result.

In this paper we choose the combination features for classification. We propose an automated and simply equipped facial expression recognition framework which could combine the LBP features and gray values to classify the facial expressions. Then we apply the softmax regression [18] method in the classification step. Comparing with the normal recognition methods, our algorithm combines the original image data with the LBP features which could use less computational cost to handle more information of the images. Moreover, we use the gray values and the LBP features which extract from the active patches of the faces to classify the different expressions while some researchers use the whole faces. The papers which use the whole face can be categorized into two types. Nearly all the deep neural frames [2,11,25] use the whole face to recognize and these works also could gain

*This work was supported in part by the National High Technology Research and Development Program 863, China (2015AA042307), Shandong Province Science and Technology Major Projects, China (2015ZDXX0801A02), and Independent Innovation & Achievement Transformation Special Foundation, China (2014ZZCX04302, 2014ZZCX04303), Shandong Provincial Scientific and Technological Development Foundation, China (2014GGX103038), and the Fundamental Research Funds of Shandong University (2015JC027, 2015JC051).

[a] The authors are all with School of Control Science and Engineering, Shandong University, Jinan, China, 250061.

Ming Liu is with Department of Mechanical and Biomedical Engineering City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.

Zhigang Xu is with Institute of Developmental Biology, School of Life Sciences, Shandong University Jinan, China, 250100.

Yafang Wang is with School of Computer Science and Technology, Shandong University, Jinan, China, 250100.

† The corresponding author. E-mail addresse : maxin@sdu.edu.cn.

higher recognition accuracy. The papers using handcrafted features are usually proposed in early days [20]. We have proposed an applied algorithm framework to recognize facial expressions. The appliance of PCA could reduce the dimensions and make it possible to run the program on the common computers which have ordinary configuration.

This paper is organized as follows: Section II describes the framework applied for facial expression recognition in this paper. Section III explains the results obtained in the juxtaposed experiments, followed by conclusion and future work in Section IV.

II. METHODOLOGY

In this paper, we have proposed a method to combine LBP features and gray values to classify the facial expressions. The gray values and the LBP features are all extracted from the actives patches. Active patches are these face regions which undergo a major change during different expressions. The active patches usually lie around eyes, nose and mouth. So by choosing these active parts we could gain the key information of different expressions. In order to gain the active patches we use the automated facial landmark detection method to mark the faces. Using these landmarks we could gain patches. By the application of PCA we could reduce the dimensions of the features.

The algorithm we have applied needs less memory and computational cost and gains a good grade which is close the state-of-the-art recognition accuracy which uses the deep learning framework [25]. The proposed framework in our paper has 7 stages which are shown in Fig. 1. In the next of this section, we will explain the 7 stages in details.

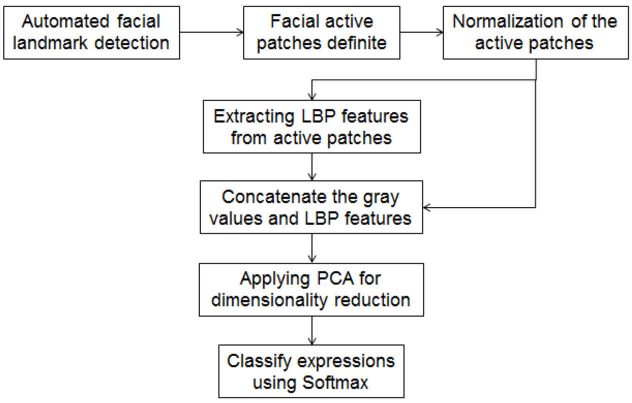

Figure 1. The proposed algorithm framework

A. Automated Facial Landmark Location

Feature extraction is one important step for expression recognition, which could contribute to fast and accurate expression classification. For the expression recognition, extract features from the whole faces may collect enormous useless data which could cost much time for calculating, also these useless data may make it hard to classify expressions. For these reasons, we chose to use the active patches to extract the key information.

Automated facial landmark detection is the first step in our method. Facial landmark detection is an important basement for facial expression classification. There are many methods to detect landmarks from the faces. Paper [23] has proposed a Fast-SIC method for fitting AAMs to detect marks on the faces. We use the method which is applied in the paper [24]. In the paper the author use a model based on mixture of trees with a shared pool marks 68 landmarks on the face. For our method, more landmarks are better for the patches' location, because more landmarks mean more accurate location. Comparing with the faces marked with fewer landmarks, we choose to use the method which locates 68 landmarks on the face. These landmarks mark the shape of eyebrows, eyes, nose, mouth and the whole face, we could use the landmark to cut the active patches. The image which is marked 68 marks is shown in Fig. 2.

Figure 2. 68 landmarks on the face

B. Facial Active Patches Definition and Normalization

In the second step we define the active facial patches which are follow the paper [6]. In the paper the writers use multitask sparse learning to confirm the common part of a face. Therefore, in our paper we choose three parts in the face for classification. The chosen result is illustrated in the Fig. 3, and we call the patch in the red, green, blue rectangles forehead patch, cheek patch and mouth patch. As these faces have different sizes, the patches get from the faces have various dimensions, it is important to apply normalization on the patches.

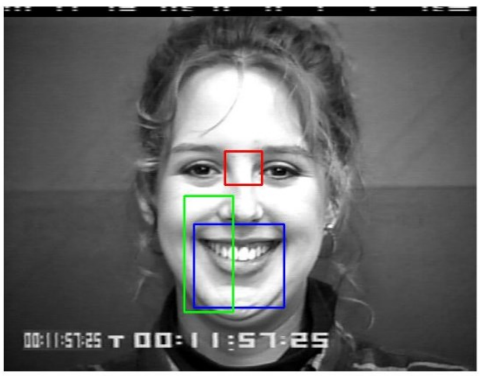

Figure 3. Facial active patches definite

In this paper, we use the landmarks which are found from the landmarks detection method to find the patches, then using the proportion of the lateral edge with the normal size to normalize the other edge. Normalization is an important step to get the information from the images, after the normalization we could draw features in the same dimensions. Also by applying the normalization method we could gain more information from the images, this could help us to gain better grade using the small benchmarks.

C. Local Binary Patterns (LBP) Features

Texture information is an important descriptor for the pattern analysis of image, LBP [16,20,21] was presented to get the texture information from the images. The Fig. 4 shows the calculation progress of the LBP value. The feature vector is processed using the softmax regression classifier. A useful extension to the original operator is the so-called uniform pattern [22], which can be used to reduce the length of the feature vector and implement a simple rotation invariant descriptor.

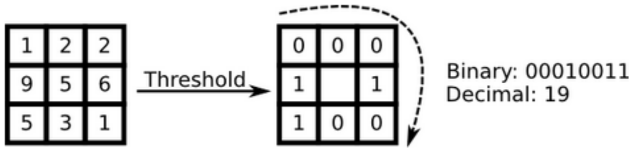

Figure 4. Calculation of LBP value

In our paper we use the uniform pattern LBP to gain features from the patches, the patches are all separated to small patches. Using uniform patterns, the length of the feature vector for a single cell reduces from 256 to 59. For example, the size of the mouth patch is 40*60 and the small patches' size is 10*15, so the mouth patch is divided to 16 patches. The uniform LBP features are extracted from each small patch and mapped to a 59-dimensional histogram. In the Fig. 5 we separate the mouth to 16 small patches.

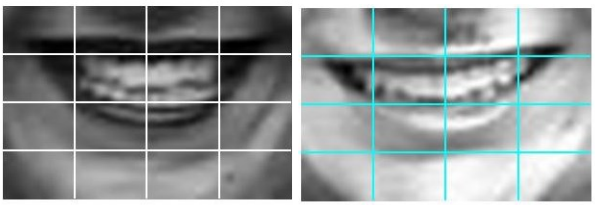

Figure 5. Calculation of LBP value

D. Concatenation and Principal Component Analysis(PCA)

The gray value of the patches contain the fundamental information of the face, some researchers only use the PCA feature of the gray value and could get decent recognition accuracy. Also the LBP features after PCA could get a good grade for facial expression recognition. In our paper we propose a method which combines the gray value and LBP feature.

For the active patches of faces the LBP feature gained from the small gray value patch has number with big variance while the gray value has smaller variance than the LBP feature. If we use the concatenation data for classification, the small number in the LBP feature will have little effect. To use more complete information of the vectors, we propose a method to reduce the big variance in the vectors and retain the different of the data in the vectors. In order to achieve this goal, we use square root to deal with the vectors of LBP features. The process of the concatenation method will show next.

(1) Normalize the gray value of the active patches.

$$Gray\_value = Gray\_value / 255 \qquad (1)$$

(2) Normalize the LBP feature vectors. First normalize the vectors to (0-1). Second solve the n square root of LBP feature vectors. In our paper to find the key information of LBP features we chose n equal to 24.

$$L\_vec = L\_vec / \max(L\_vec) \qquad (2)$$

$$L\_vec = nthroot(L\_vec, n) \qquad (3)$$

(3) Concatenate the features and use PCA to reduce the dimensions.

PCA has done an important job in this method. The result in next section shows the effect of combination the data. For the dimensions of the fusion features, as we only have 618 images we could chose 300 dimensions for most of the principal component will be preserved. But while we change the dimensions of the fusion features, we gain the result which is shown in the Fig. 6. So in our paper we chose the dimension number is 450. In the figure we change the dimensions by adding 50 from 300 to 850.

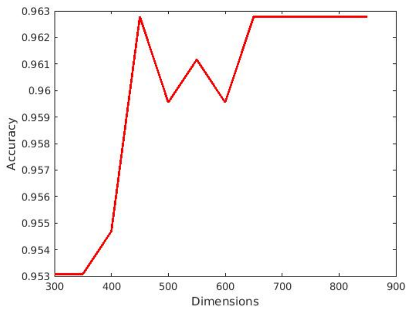

Figure 6. Relationship between accuracy and dimensions

E. Apply Softmax Regression for classification

Softmax regression classifier always is the last layer which is after a fully-connected layer in the deep learning network [25]. In our paper we have not applied the neural network, but we also use the softmax layer since the good effective of it. The softmax regression model generalizes logistic regression to classification problems where the class label can take on more than two possible values. Softmax regression is a supervised learning algorithm [18].

Compare to some traditional methods which are proposed in the papers [6,7], the softmax could classify more than two classes at once. For the facial expressions classification job we have six expressions to be recognized, after applying the softmax we could only use one classification to sort the six expressions rather than design 15 classifiers [6] to solve the problem.

III. EXPERIMENTAL RESULTS

We use extended Cohn-Kanade (CK+) database to evaluate the proposed framework [8,19]. CK+ database is a standard database which consists of 100 university students aged from 18 to 30 years old, of which 65% were female, 15% were African-American and 3% were Asian or Latino. The CK+ database contains 327 expression-labeled image sequences, each of which has one of 7 expressions, i.e., anger, contempt, disgust, fear, happiness, sadness, and surprise activated.

In our paper we only use six basic expressions as those papers classify the six expressions. In order to get a good train result, we chose the last two peak images which contribute to 618 images. To be honest, these expressions have different number with each other, in order to use all the images in the database we chose to use leave-one-out validation method to validate the classification effect.

A. Use the Gray Value for Classification

For the purpose to know the influence of the part patch, such as the mouth patch and the cheek patch, we design two experiments. One is using only the data of mouth patch for classification, and the other uses the gray value of all the three patches. The progress is shown in Fig. 7.

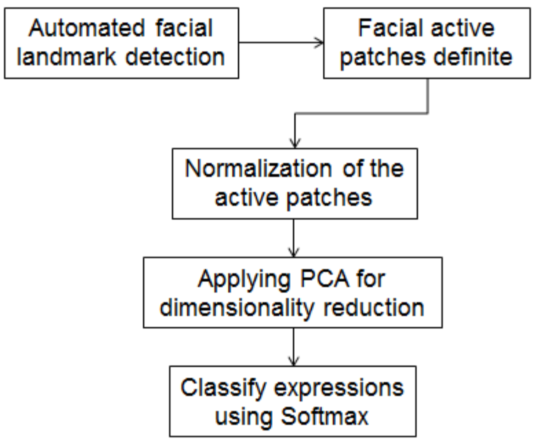

Figure 7. Process of using only gray value

This idea of using the active part comes from the paper [6], in the paper the authors using multi-task learning algorithm to distinguish active and inactive patches. In this paper the authors show the active patches in the image, also the authors distinguish the active patches to common and specific patches. The common patches in the paper are these face regions which change most during different expressions. The specific patches are these patches which different expressions have different display.

Different from the method in the paper [6], we chose only the mouth patches to classify the different expressions to do the first experiment and we apply all the active patches to recognize the expressions as the second experiment. This is different from these methods which use the common patches to distinguish all the six expressions and apply the specific patches to sort the special expressions which have different performs in these regions. The results are shown in Fig. 8 and Fig. 9.

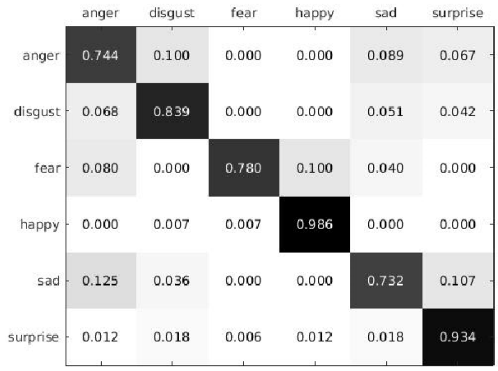

Figure 8. Confusion matrix of using mouth patch's gray value

Learning from the results of the experiments we could draw a conclusion that more effective data means higher accuracy. Besides, our framework also could gain a good result with these data.

Figure 9. Confusion matrix of using three patches' gray value

B. Apply the LBP Feature for Classification

We also design one experiment to evaluate the LBP feature's effectiveness. We extract the LBP features from the mouth patch, cheek patch and forehead patch. In the second section we divide the mouth patch into 16 blocks, then we gain 59 dimensions feature from every small block. So only from the mouth patch we can extract 59*16=944 dimensions feature.

In the third experiment we use the LBP features of the all patches and we get a better result than the gray value. The result is shown in the Fig. 10. In the next part we will compare all the results from the experiments.

C. Use the Concatenation of the Patches' Gray Value and LBP Features for Classification

In the second part of the paper, we propose a method to concatenate the gray value and the LBP features. We do as the forth part in the second section, normalize all the gray value, then fusion the gray value of all the patches and concatenate the LBP features extracted from all the patches and use the PCA to reduce the dimensions.

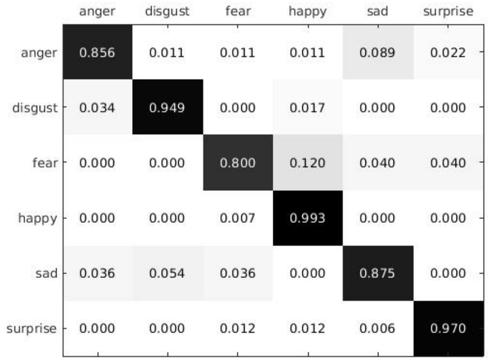

Figure 10. Confusion matrix of using LBP features

The Fig. 11 shows the result of the classification. From the figure we could know that using the combination features could gain more accurate recognition rate. Our method proposed in the paper could gain better result than the gray value and LBP features.

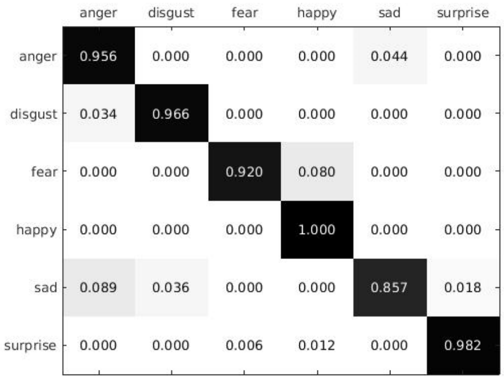

Figure 11. Confusion matrix of using combination feature

The table below gives the recognition result of all the four experiments. Compare these details in the Table. I, we could find that the accuracy of the expressions all have promotion by using the all patches' features. On one hand, this shows that more information means higher accuracy in our algorithm. On the other hand, by using the fusion features we could gain better grades of all the expressions. This means that our algorithm could improve the recognition accuracy of facial expression.

As for the computational cost, paper [15] took about 8 days to complete the overall training for 6 expressions in an 8-fold experimental setup on a 6-core 2.4GHZ PC using Matlab implementation. For our algorithm, we take about 5 seconds to train for 6 expressions in leave-one-out validation method. The experiment is running on a 4-core 3.2GHZ PC using Matlab. Though we use less time than the algorithm in the paper, we also could get a grade similar to their result.

IV. CONCLUSION

In this paper, we have proposed an automated and simply equipped facial expression recognition framework which uses combined LBP features and gray values to classify the facial expressions. Besides we use the active patches of the face instead of the whole face, by using these active patches we could use the key information of the different expressions. The combination of the LBP features and gray values has better classify result than only one part, so we could merge these features to do other recognition work. In this paper, we apply PCA method to reduce dimensions which could reduce computational cost and memory cost. Our algorithm could gain super result from the static images, but recognize these basic expressions from the videos would be a big challenge for the researchers. In the future, we will apply our algorithm on the videos.

TABLE I. RESULT OF CONTRAST EXPRIMENTS

Facial expressions	Classification Result			
	Exp. one	Exp. two	Exp. three	Exp. four
Anger	0.744	0.844	0.856	0.956
Disgust	0.839	0.932	0.949	0.966
Fear	0.780	0.840	0.800	0.920
Happy	0.986	0.986	0.993	1.000
Sad	0.732	0.696	0.875	0.857
Surprise	0.934	0.940	0.970	0.982
Average	0.869	0.905	0.932	0.963

ACKNOWLEDGMENT

The authors would like to thank Prof. Jeffery Cohn for providing the Cohn–Kanade database and extending the database.

REFERENCES

[1] C. E. Izard. The face of emotion. New York: Appleton-Century-Crofts, 1, 1971.
[2] M. Ali, C. David, H. Mohammad. Going Deeper in Facial Expression Recognition using Deep Neural Networks. IEEE Winter Conference on Applications of ComputerVision (WACV), 2016.
[3] G. Duchenne. Mecanisme de la Physionomie Humaine.1862.
[4] http://en.wikipedia.org/wiki/Facial_expression.
[5] P. Ekman, W. V. Friesen and J. C. Hager, "FACS Manual," Salt Lake City, UT: A Human Face, May 2002.
[6] L. Zhong , Q. Liuz , P. Yangy , B. Liuy , J. Huangx and D. N. Metaxasy. Learning active facial patches for expression analysis. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 2562-2569, 2012.
[7] SL H, A. Routray. Robust Facial Expression Classification Using Shape and Appearance Features. International Conference on Advances in Pattern Recognition, 2015.
[8] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete expression dataset for action unit and emotion-specified expression. In CVPR Workshops, pages 94–101, 2010.
[9] T.H.H. Zavaschi, A.S. Britto, L.E.S. Oliveira, A.L. Koerich. Fusion of feature sets and classifiers for facial expression recognition. Expert Systems with Applications, 40 (2013), pp. 646–655.
[10] M. Liu, S. Li, S. Shan, and X. Chen. Au-inspired deep networks for facial expression feature learning. Neurocomputing, 159:126–136, 2015.
[11] Z, Yu, C. Zhang. Image based Static Facial Expression Recognition with Multiple Deep Network Learning. International Conference on Multimodal Interaction, 2015.

[12] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon. Emotion recognition using PHOG and LPQ features. in Proc. IEEE Int. Conf.Autom. Face Gesture Recog. Workshops, 2011, pp. 878–883.
[13] S. Bashyal, and G. Venayagamoorthy. Recognition of Facial Expressions using Gabor Wavelets and Learning Vector Quantization. Engineering Applications of Artificial Intelligence, Vol. 21, No. 7, pp. 1056-1064, 2008.
[14] H. Jung, S. Lee, S. Park, I. Lee, C. Ahn, and J. Kim. Deep temporal appearance-geometry network for facial expression recognition. arXiv preprint arXiv:1503.01532, 2015.
[15] P. Liu, S. Han, Z. Meng, and Y. Tong. Facial expression recognition via a boosted deep belief network. In CVPR, pages 1805–1812, 2014.
[16] T. Ojala, Pietikainen, M. Harwood. A comparative study of texture measures with classification based on featured distribution. Pattern Recognition. 29(1), 51–59, 1996.
[17] N. Dalal, B. Triggs. Histograms of Oriented Gradients for Human Detection. CVPR, 2005.
[18] http://ufldl.stanford.edu/wiki/index.php/Softmax_Regression.
[19] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In FG, pages 46–53, 2000.
[20] T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585.
[21] https://en.wikipedia.org/wiki/Local_binary_patterns
[22] M. Heikkilä, M. Pietikäinen. A texture-based method for modeling the background and detecting moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(4):657-662, 2006.
[23] G. Tzimiropoulos, and M. Pantic, "Optimization problems for fast AAM fitting in-the-wild," ICCV 2013.
[24] X. Zhu, D. Ramanan. Face detection, pose estimation and landmark localization in the wild. Computer Vision and Pattern Recognition (CVPR) Providence, Rhode Island, June 2012.
[25] P. Khorrami, T.L. Paine, T.S. Huang. Do Deep Neural Networks Learn Facial Action Units When Doing Expression Recognition? Preprint at http://arxiv.org/abs/1510.02969(2015).