Automated classification of facial expressions using bag of visual words and texture-based features

Nouzha Harrati*†, Imed Bouchrika†, Abdelkamel Tari* and Ammar Ladjailia†‡
†*Faculty of Science and Technology, University of Souk Ahras, Algeria*
**Department of Computer Science, University of Bejaia, Algeria*
‡*Department of Computer Science, University of Annaba, Algeria*
n.harrathi@univ-soukahras.dz

***Abstract*—As facial expression plays undoubtedly a key role in conveying human emotions and feelings, research into how people react to the world and communicate with each other still stands as one of the most scientific challenges to be addressed. Recent research has shown that facial expressions can be a potential medium for various applications. In this research paper, we explore the use of texture-based facial features obtained using the Local Binary Patterns operator. The facial expression signature is constructed via encoding the textural information using the bag of features. Features are trained to robustly distinguish different seven facial emotions including: happiness, anger, disgust, fear, surprise, sadness as well as the neutral case. Based on a gallery dataset containing 76 images, a classification rate of 93.4% is achieved using the Support Vector Machine classifier. The attained results assert that automated classification of facial expression using an appearance-based approach is feasible with an acceptable accuracy.**

***Keywords*-Facial Expressions, LBP, Bag of Features.**

I. Introduction

The human face is a source of rich and wealthy information that are more essential to social interaction as well as communication. Such information can be either static such as identity, age, gender and ethnicity. Alternatively, facial data can be dynamic such as the emotional state of the person. As facial expression plays undoubtedly a key role in conveying human emotions and feelings, research into how human beings react to the world and communicate with each other still stands as one of the most scientific challenges to be addressed. The first research and analysis for facial emotion analysis dates back to the nineteenth century as Charles Darwin [1] is considered the pioneer to conduct the initial study on emotion indication from animal and human faces. Ekman and Friesen [2] categorized the human emotions into six basic emotions that each owns its particular content, appearance and distinctive facial expression. The classified basic emotions are considered to be universal across different ethnicities and cultures. The six basic emotions are : happiness, sadness, fear, disgust, surprise and anger. The computer vision community showed unprecedented interest for research related to automated classification of emotions based on facial expressions.

Automated classification of facial expressions has been a subject of intense investigation for the last two decades due to the various number of applications with pioneer work dating back to 1991 which was presented by Mase and Pentland [3] about automated lipreading using optical-flow analysis. As facial expression plays a vital role in conveying human emotions, an automated facial expression system can provide a non-intrusive way to apprehend the emotional state or activity of a person. It is becoming an important part of affective computing whose main objectives are to recognize and respond to users' emotions. Facial expressions can be deployed in various applications including emotion analysis, automated tutoring systems, indexing and retrieval of rich multimedia databases as well as a human computer interaction [4], [13]. Furthermore, facial expression is regarded as a valuable non-verbal feedback for evaluative application where the opinion of the user or customer is analyzed based on their emotion. Such applications enjoy the potency to comprehend and understand the individual's cognitive appraisals in addition to the social intentions which are to an extent related to the emotional experience of the user [5].

As a person can undertake little efforts to interpret and to understand faces of other people so robustly and under most difficult conditions from a single still image, it is still a challenging computer vision problem. Although substantial progress has been made by many researchers [6], [7], [8], [9], achieving a high accuracy is still a challenging problem due to the complicated variations of the facial dynamics. In addition, a number of factors impact the complexity further that can be related to personal or environmental conditions such as occlusion, makeup, imagery resolution, camera movement and recording viewpoints. Moreover, the process of how to extract and represent the dynamic facial features is a key issue that is still to be further investigated.

As recent research has proven that facial expressions can be a potential medium for various applications together with the fact there still remains a number of obstacles to resolve and large number of avenues to explore, we investigate in this research study a textural approach based on marker-less extraction of facial features using Local Binary Patterns (LBP) operator combined with the use of bag-of-features for encoding a signature histogram for each facial expression based on texture information. Initially, the face is being detected using the method proposed by Viola and Jones [10] which is applied to a sequence of images taken from the MUG dataset [11]. A histogram-based feature vector is being constructed using the detection of the LBP operator. K-means clustering

978-1-4673-9234-1/15/$31.00 ©2015 IEEE

is employed to construct the codebook used for the the bag-of-features. Features are trained to robustly distinguish different seven facial emotions including: happiness, anger, disgust, fear, surprise, sadness as well as the neutral case. The classification process is based on the multi-class support vector machine to recognize the different classes of facial expressions.

The remainder of this paper is organized as follows. The next section outlines the previous related studies for the different methods being proposed for the automated detection of facial expression. The theoretical description of the presented framework for extracting an LBP-based features for recognizing the basic facial expression is discussed in Section 3. Section 4 introduces the experimental results and analysis applied.

II. Related Work

There are a large number of studies devoted recently to automated analysis and classification of facial expressions. The approaches can be categorized into two main streams based on the extracted features being used for the detection [5]. The features can be classified into either geometric-based or appearance-based. For the first category, features describe the shape, dimensions or relative distances of the different facial parts as the eyes, mouth and nose. Geometric features are usually defined in terms of distances and angles between facial points through a set of fiducial facial landmarks. The main drawback of using geometric-based method is their dependence on accurate facial point localisation which is not guaranteed for real case scenarios. On the other hand, appearance-based features depict the state of the face when expressing an emotion such as wrinkles and furrows. Therefore, the facial emotions become a case of texture classification and recognition. Both of such approaches have their own benefits and setbacks and there are interestingly recent hybrid methods that harness both types of geometric and appearance-based features for the classification process to attain the highest accuracy.

For the geometric-based features, various approaches have been outlined in the literature. The majority of methods use Active Appearance Model (AAM) or derivatives of such technique to extract and track a dense set of facial points, which are used to construct the geometric properties of the face. Asthana *et al* [12] proposed an AAM-based method which is evaluated against the Cohn-Kanade (CK) database reporting a recognition rate of 93% for the classification of the basic six emotions. Similarly, a correct classification rate of 93% is achieved by Sebe *et al* [14] for emotions detection using Piecewise Bezier volume deformation tracking along with the k-nearest neighbor classifier (knn) on the same dataset. Facial landmark points are manually labeled. The knn classifier is surprisingly reported as the best classifier to yield the best classification rate compared to a large number of machine learning methods experimented on the same dataset. For the detection of Action Units (AU) for facial expressions, Valstar [15] extracted 20 facial points automatically and temporally tracked using particle filtering. An average AU recognition rate of 95.3% is reported on the CK dataset with an achieved facial expressions rate of 72% when tested on spontaneous expressions.

For the appearance based features, a number of research studies have been proposed for the last decade. The Local Binary Operator (LBP) [16] which was initially proposed for texture analysis; has been successfully adopted for facial expressions due to its low computational cost, efficiency and promising results for both still images and video sequences. Shan *et al* [17] constructed a classifier using LBP and Support Vector Machine (SVM) which was applied on low resolution imageries from the Cohn-Kanade database. A high recognition rate of 92.1% is being reported for a sequence of 320 different image sequences. Recently, Zhi et al [18] introduced a method called the Graph Preserving Sparse NMF(GSNMF) to solve the problem of the six basic emotions. The method has shown potential to work for both supervised and unsupervised manner via transforming high dimensional images into a locality preserving feature space with sparse representation. Zhi reported a high recognition of 94.3% for the CK database. There are many other methods based on the appearance features such as: Local Gabor Binary Pattern, Local Phase Quantization and histogram of oriented gradients.

For using both appearance and geometric features, Simon *et al* [19] proposed a k segment-based SVM approach where face-based features are extracted and tracked via a person-specific AAM along with extracting SIFT features. Liu [20] proposed a deep boosted deep belief machine learning algorithm for the classification of facial expressions through appearance and geometric changes that are derived from images or video sequences. The classification process is usually carried out sequentially in three stages: feature generation, feature selection, and classifier construction. In their research study, a high classification value of 96.7% is obtained using the CK database.

III. Proposed Approach

Figure (1) shows the different stages being performed from the detection of the face to the formulation of the feature vector using a histogram construction based on the local binary operator and bag of visual features. For the initial phase, the face of a person is detected from a single static image using the Viola and Jones algorithm. The implementation is provided within the computer vision Matlab toolbox. To further refine and tune the detection accuracy, symmetrical analysis of face using color intensity is performed so that non-prominent features of the face such as hair are suppressed. Subsequently, the local binary pattern operator is applied on the face image divided into a grid of cells. For the classification stage, a histogram is generated for the LBP features that can be used as input for the SVM classifier.

A. Local Binary Patterns

The Local Binary Pattern (LBP) operator was first introduced for texture analysis by Ojala et al. [21] in 1996

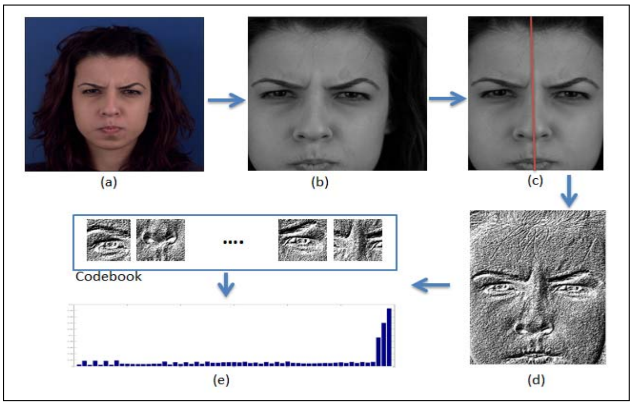
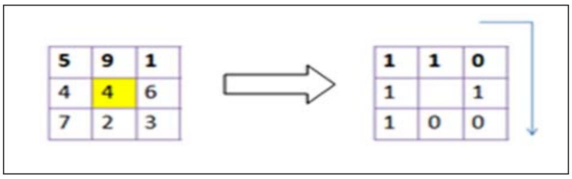


Figure 1. Face Detection & Feature Vector: (a) Input Image. (b) Face detection using Viola & Jones Algorithm. (b) Symmetry Analysis to tune face detection. (c) LBP (d) Bag of Visual Features.

in their editorial “a comparative study of texture measures with classification based on featured distribution”. The LBP can be efficiently and swiftly computed in a single image scan offering facial recognition capabilities even for lower resolution images. The operator sets the pixels of a given image by thresholding each number of the neighboring pixels against the centre pixel within a 3x3 matrix and therefore, resulting a series of values of consecutive 1 or 0 as shown in Figure (2). By reading in the same direction of the arrow, a binary number is formulated which is converted to a decimal number i.e. a label where the binary number: 11010011 is converted to 211. The 256-bin histogram of the resulting labels is computed and employed as a texture descriptor for facial-based applications.



5	9	1
4	4	6
7	2	3

1	1	0
1		1
1	0	0

Figure 2. The basic Local Binary Pattern Operator (LBP)

The main drawback of the basic local binary operator is its small neighborhood area of (3*3) whereby it may ignore or disregard prominent features for larger structures. An extended version of the LBP operator is outlined in recent research studies by Ojala et al. [21] to use neighborhoods of different sizes. The extended LBP operator is represented by a circular neighborhood area written as (P, R) where P is the number of pixel points in the circular neighborhood whilst R is the radius of the circular area as shown in Figure (3). The value of the LBP for pixel point having the coordinate (x_c, y_c) is computed as shown in Equation (1):

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c)2^i \qquad (1)$$

Where g_i is the grayscale value of the pixel point i. c is the centre pixel. The function $s(a)$ is a thresholding function returning 1 for the case of $a \geq 0$ and 0 otherwise.

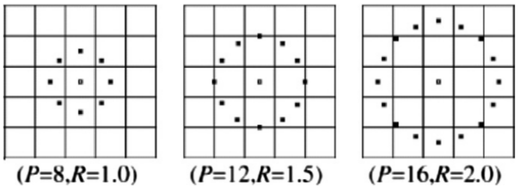


Figure 3. The Extended Local Binary Pattern Operator [17].

An $LBP(P, R)$ can produce 2^p different values according to the 2^p different patterns made by P which is the number of points in the chosen circular neighborhood. Research studies have shown that some patterns contain more discriminative information than the others. Further extension of the LBP has been introduced to take into account only uniform patterns which are defined as the patterns containing at most two bitwise transitions between 0 to 1 and vice versa, i.e.: the number of times that a digit chnages from 0 to 1 or vice versa. As an example, the following binary numbers 00000000 and 00111000 and 11100001 are uniform patterns. An LBP operator written as (LBP^{U2_P}, R) meaning that this LBP operator is using the (P, R) circular neighborhood area with only uniform patterns being considered. The histogram with an n-bin for the Local Binary Pattern operator is derived from a labeled image f_l as defined in Equation (2):

$$H_i = \sum_{x,y} b(f_i(x, y) == i) \qquad i = 0, 1, , , , n - 1 \qquad (2)$$

Where b is a Boolean function returning 1 for true cases and 0 for false conditions. In [17], an improved spatially-based histogram is described which divides the image into m smaller regions R for the aim to retain spatial features where the histogram is computed as set in Equation (3):

$$H_{i,j} = \sum_{x,y} b(f_i(x, y) == i)b((x, y) \in R_j) \qquad (3)$$

where R_j is the j^{th} region of an image divided into an m region.

B. Classification method

The bag-of-features (BoF) method is a well-known classification method for object recognition in computer vision applications demonstrating greater performance. It was inspired from the bag of words method for handling textual document based on word frequencies. For the bag of features, every image is defined as a set of orderless or unstructured features locally described. Two major concepts form the basis for the approach: Local features and Codebook representation. Local features consist in extraction of global image descriptors, then to represent this image as set of characteristics aggregated from a series of smaller images which are called patches. To obtain patches many techniques are used among them the local binary pattern. Once the patches are obtained, they are mapped in clustered vectors called visual words through the use of unsupervised K-means clustering. The resulting codewords are regrouped to form the codebook or dictionary. Subsequently each image or instance is described by the frequency histogram indicating the occurrence of visual words for the given image. The similarity d of images can be measured by comparing between the two BoF histograms A and B as shown in the following Equation:

$$d(A,B) = 1 - \sum_{i=1}^{n} min(a_i, b_i) \tag{4}$$

such that a_i, b_i denote the frequencies of the i^{th} visual words in the image A and B respectively.

Support Vector machine is used as the main classifier for the automated classification of facial expressions. Given the training set of labeled data $D = \{(f_i, c_i), i = 1, .N\}$ where f_i is a multi-dimensional feature vector and $c_i \in \{-1, 1\}$ is a binary class label for the candidate f_i. Given a test data x, it is classified as :

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i c_i K(f_i, x) + b) \tag{5}$$

such that α_i is the Lagrange multipliers for the dual optimization case, $K(f_i, x)$ is the kernel function whilst b is the parameter for the optimal separating hyperplane. The support vector machine deduces a linear hyperplane which maximizes the separation margins between the different clusters in the training data. Multi-class classification is achieved through a cascade manner with binary SVM classifiers via a voting scheme.

IV. Experimental Results

In order to evaluate the use of textual facial features taken using the extended local binary operator for facial expression detection, a gallery of 280 still images is collected from the MUG dataset [11]. The selected set contains 10 different individuals. The seven different scenarios of facial emotion are being taken for every subject including: happiness, sadness, anger, disgust, fear, surprise and the neutral case with 4 sequences for every individual per emotion. For the MUG facial expression dataset, the candidates were sitting in a chair in front of one camera. The background was a blue screen. Two light sources of 300W each, mounted on stands at a height of 130cm approximately were used. On each stand one umbrella was fixed in order to diffuse light and avoid shadows. The camera was able to capture images at a rate of 19 frames per second. Each image is saved in jpg format with a resolution of 896x896 pixels and a disk size ranging from 240 to 340 KB. Sample of the imagery set being used for the evaluation during this research studies are being shown in Figure (4).

Initially, the histogram from the visual bag of features is derived for every still image after running the local binary pattern for the detected faces. To assess the classification merit of the textual features derived using the local binary pattern operator, the multi-class support vector machine classifier is employed with the Leave-one-out cross-validation rule. In the leave-one-out cross validation, every sample from the dataset is employed for testing in such a way it is validated against all the remaining instances in the dataset. This is repeated for all other instances in the same fashion. The recognition rate is estimated as the average of all cross-validations. In addition, the K-nearest neighbor (KNN) classifier is applied separately at the classification phase because of its fast computation and simplicity besides the convenience to compare against existing studies.

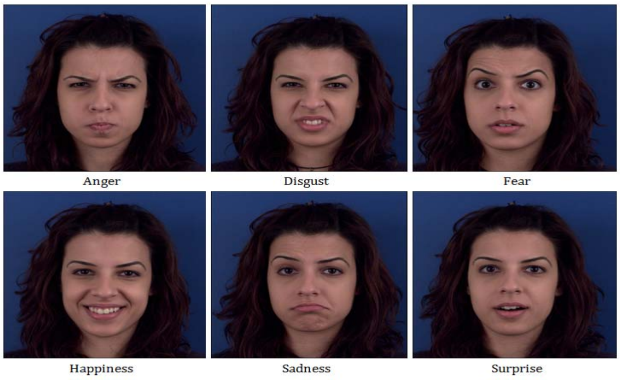


Figure 4. The MUG dataset

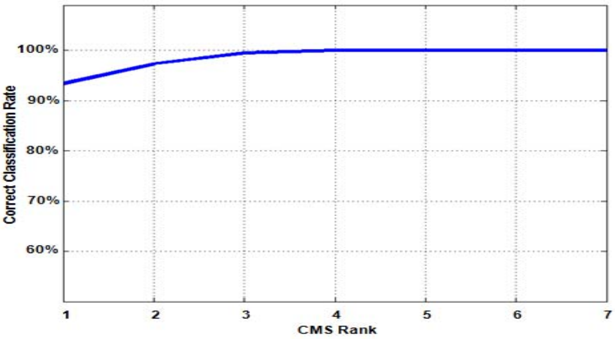


Figure 5. Cumulative Match Score for Facial Expressions

In this experiment, the Cumulative Match Score evaluation method is estimated which was first described during the FERET protocol for face biometrics, a correct classification rate of 93.4% is reported for the 7 basic emotions at rank $R = 1$ and 100% at rank $R = 4$ using SVM meanwhile the following rates 82.7% and 95.1% are achieved using the Knn rule for the same ranks respectively. Figure (5) shows the CMS curve for the classification process. The achieved results are promising because the recognition is based purely on local textural information and this can be boosted through adding geometric features of the face. Table (1) shows comparative results for different methods for automated classification of facial expression on the same MUG dataset. The obtained results inspire to certain potentials in addressing the intricate issue of automated recognition of expressions.

V. Conclusions

In this study, an automated framework for the classification of facial expressions using the Local binary

Method	**CCR**
Our method : LBP + BoF+SVM	93.4 %
Our method : LBP + BoF+Knn	82.7 %
LBP + Self Organizing Maps (SOM) [22]	86.5%
Projection in the Encrypted Domain [23]	95.2%

Table I
COMPARATIVE RESULTS FOR THE MUG DATASET

Operator combined with a weight-based feature selection algorithm is described. The expressions include the basic six emotions defined by Ekmen which are: happiness, sadness, disgust , fear, anger, surprise as well as the neutral cases which is considered in the classification phase. We have shown that the facial dynamic features embed most of the discriminatory traits for the detection of human emotions with an obtained classification rate of 93.4% using a probe dataset taken from the MUG dataset.

REFERENCES

[1] C. Darwin, P. Ekman, and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998.

[2] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion.” *Journal of personality and social psychology*, vol. 17, no. 2, p. 124, 1971.

[3] K. Mase and A. Pentland, “Automatic lipreading by optical-flow analysis,” *Systems and Computers in Japan*, vol. 22, no. 6, pp. 67–76, 1991.

[4] N. Harrati, I. Bouchrika, A. Tari, and A. Ladjailia, “Automating the evaluation of usability remotely for web applications via a model-based approach,” in *International Conference on New Technologies of Information and Communication (NTIC)*, 2015.

[5] M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer, “Meta-analysis of the first facial expression recognition challenge,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 4, 2012.

[6] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, “The first facial expression recognition and analysis challenge,” in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 921–926.

[7] Y. Tian, T. Kanade, and J. F. Cohn, “Facial expression recognition,” in *Handbook of Face Recognition*. Springer, 2011, pp. 487–519.

[8] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, “The computer expression recognition toolbox (cert),” in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 298–305.

[9] J. Whitehill, G. Littlewort, I. Fasel, M. Bartlett, and J. Movellan, “Toward practical smile detection,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 2106–2111, 2009.

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

[11] N. Aifanti, C. Papachristou, and A. Delopoulos, “The mug facial expression database,” in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*. IEEE, 2010, pp. 1–4.

[12] A. Asthana, J. Saragih, M. Wagner, and R. Goecke, “Evaluating aam fitting methods for facial expression recognition,” in *Affective Computing and Intelligent Interaction and Workshops, 3rd Conference on*, 2009, pp. 1–8.

[13] A. Ladjailia, I. Bouchrika, H. F. Merouani, and N. Harrati, “On the use of local motion information for human action recognition via feature selection,” in *4th IEEE International Conference on Electrical Engineering (ICEE)*, 2015.

[14] N. Sebe, M. S. Lew, Y. Sun, I. Cohen, T. Gevers, and T. S. Huang, “Authentic facial expression analysis,” *Image and Vision Computing*, vol. 25, no. 12, pp. 1856–1863, 2007.

[15] M. F. Valstar and M. Pantic, “Fully automatic recognition of the temporal phases of facial actions,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 1, pp. 28–43, 2012.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, 2002.

[17] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[18] R. Zhi, M. Flierl, Q. Ruan, and W. Kleijn, “Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 1, pp. 38–52, 2011.

[19] T. Simon, M. H. Nguyen, F. De La Torre, and J. F. Cohn, “Action unit detection with segment-based svms,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2737–2744.

[20] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial expression recognition via a boosted deep belief network,” in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1805–1812.

[21] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[22] S. Agarwal and D. P. Mukherjee, “Decoding mixed emotions from expression map of face images,” in *Automatic Face and Gesture Recognition (FG),10th IEEE International Conference and Workshops on*, 2013, pp. 1–6.

[23] Y. Rahulamathavan, R. C.-W. Phan, J. Chambers, D. J. Parish *et al.*, “Facial expression recognition in the encrypted domain based on local fisher discriminant analysis,” *Affective Computing, IEEE Transactions on*, vol. 4, no. 1, pp. 83–92, 2013.