Graph Spectral Domain Feature Representation for in-Air Drawn Number Recognition

Basheer Alwaely* , Charith Abhayaratne†
Department of Electronic and Electrical Engineering
The University of Sheffield
Sheffield, S1 3JD, United Kingdom
Email: *b.alwaely@sheffield.ac.uk, †c.abhayaratne@sheffield.ac.uk

***Abstract*—The emerging field of graph signal processing has brought new scope in understanding the spectral properties of arbitrary structures. This paper proposes a novel graph spectral domain feature representation scheme for recognising in-air drawn numbers. It provides the solution by forming the hand's path as a graph and extracting its features based on the spectral domain representation by computing the graph spectral transform. A novel graph generation model is proposed to form the topology of the shapes of numbers. The experiments show that the proposed features are flip and rotation- invariant which makes insensitive to changes in the rotation angle of the drawn numbers. The proposed solution achieves a high level of accuracy of nearly 98% for in-air hand drawn number recognition.**

I. Introduction

Irregular data and complex structures are rapidly increasing number of data due to the vast improvement in the technology. As a result, there is a serious need to understand and characterise complex shapes which can be applied in real-world applications like sign language. The new emerging field of graph signal processing (GSP) has attracted a lot of attention in evaluating and understanding the irregular data because the classical discrete signal processing (DSP) does not apply for unstructured data. Conversely, DSP deals mostly with structured data or grid data in continuous domain.

The proposed method falls into two areas, which are complex shapes matching field and in-air hand gesture recognition. The use of the graphs for complex structure matching is explored in terms of graph matching to find a perfect map between two sets of nodes that align the relationships of these sets. For example, Horaud and Sossa [1], Carcassoni and Hancook[2], Umeyama [3] , Cour *et al.*[4] and Zhou and Torre [5] applied an approximate method to explore the maximum probability of correspondence mapping between two patterns through the weight matrix based on polynomial characterization, centre of clusters, eigen domain, spectral relaxation and Kronecker product of the graph incident matrix respectively. Also, many problems are formed as bipartite matching. For example, Shamaie and Sutherland [6], Leordeanu and Hebert [7], Riba et al. [8] performed a bipartite graph matching based on shortest edges, largest eigenvalue and convex path inside text images. However, the massive time requirement for implementation makes these approaches unsuitable for real-time applications. To address these limitations, this paper proposes feature-to-feature assignments, which is based on the geometric structure properties of the objects. Our method is totally different than the existing works in the fields of graph matching and hand gestures recognition.

In-air human hand movement is an obvious example of forming irregular structures and the available work for in air hand writing recognition can be classified into two groups: image representation and node representation. Image representation means that the numbers are saved as an image. In this case, the image has two types of pixels: number path pixels and background pixels. Thus, a great deal of data is required to represent the numbers, which is the limitation of this approach. For instance, [9] [10] [11] normalized number data into a binary table where logic 1 refers to the hand writing path and 0 to the background. The node representation method refers to the path of hand movement. For example, [12], [13] used skeletal coordinates of the user's hand, then features were obtained based on a pixel domain.

Surprisingly, the use of graph in the hand gesture recognition field is very limited and restricted by identifying hand poses only, which makes this paper the first to apply the spectral graph feature for in-air hand drawn numbers recognition. The available studies of hand gesture recognition based on graphs have been restricted to static hand gestures rather than dynamic hand gestures [14] [15]. They also focused on employing graph theory rather than investigating the spectral graph domain [6] [16]. For example, [14] [17] [18][19] implemented static gesture recognition by using an edge detector to extract the hand features. The key point of these works is how to allocate the graph nodes over the hand based on features. However, these features are rotation and flip variants that make the feature sensitive to a change of rotation angle. Usually a great deal of data is required to achieve the optimal level of accuracy, which leads to a lack of real-time implementation.

The main contributions of this work are:

- A novel graph generation model to form the topology of hand movements for in-air drawing shapes.
- A novel set of spectral domain features for in-air drawn numbers by exploring the graph spectral transform.
- Incorporating the rotation and flip-invariance in feature modelling enabling high robustness for various rotation angles of the hand drawn numbers.
- Real-time recognition with very low computational time

ISBN 978-0-9928626-7-1 © EURASIP 2017

The rest of this paper is organized as follows: Section 2, illustrates the proposed graph spectral domain features. Section 3, presents in air drawn number recognition as the application. Section 4, shows the results and discussions. Finally, the work will conclude in section 5.

II. Proposed Graph Spectral domain features

A. Graph concepts

Initially, the graph concepts will be define as it will be used throughout the paper. Graph (G) contains of nodes (N) or vertices (v) which are connected by edges (e). These edges take the form of any relation between vertices in the graph adjacency matrix (A) such as Euclidean distance, pixel value or sensor measurement. A is obtained as in Eq. (1)

$$A = \begin{cases} e_{(i,j)}, & \text{if i and j are connected;} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

D is a diagonal matrix equals to the sum of edge weights connected to the vertex $_i$ which calculated by Eq. (2).

$$D_{(i,i)} = \sum_{j=0}^{N-1} A_{(i,j)}. \tag{2}$$

Then, the non-normalized graph Laplacian matrix (L) is computed:

$$L = D - A, \tag{3}$$

and the normalized version ($\mathcal{L}$) of the graph Laplacian matrix is obtained :

$$\mathcal{L} = D^{-1/2} L D^{-1/2}. \tag{4}$$

Eigenvalues λ and Eigenvectors χ of the Laplacian matrix are computed to extract the features. Graphs are useful form to describe the geometric structures of data based on the connectivity. Also, it is important to highlight the fact that this connectivity is rotation, flip and mirror invariant because it depends on relative measurements between the nodes. In other words, the graph connectivity description of the object does not change by moving or rotation the object in any direction unless the object structure will change. This concept leads to investigate matching samples in different direction based on the connectivity. To do so, the second eigenvector, which is known as Fiedler value [20] or the algebraic connectivity (Fig. 2) is analysed for in-air handwritten digits (0-9) (Fig. 1). For example, number zero has an equal distribution of nodes except in the first and last nodes, which are either close or far in the distance compared to other nodes. Also, number one has a uniform distribution of nodes. Thus, the mid node has a strong connectivity than other nodes and it will decrease by moving away from the mid node. At number four, we can see clearly that highest and lowest point occurred at after starting and before finishing nodes because they cross each other and the distance between them will be close. Among all the connectivity representation of numbers, two and five are the most similar in non-normalize version because they have almost the same structure; except that number two has smooth change at the beginning and fluctuated at the end,

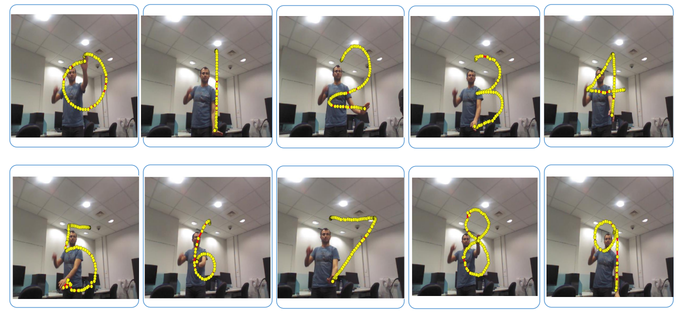

Fig. 1. In-air hand writing digits (0-9).

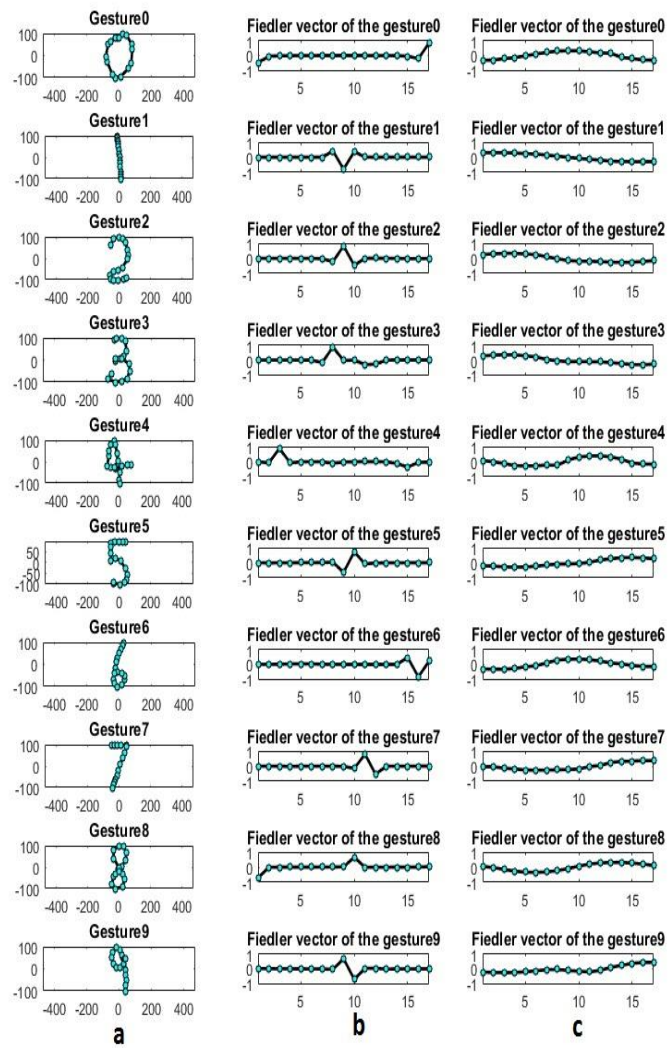

Fig. 2. (a): Handwritten number digits, (b): the second eigenvector of their non-normalized Laplacian matrixEq. (3), (c): the second eigenvector of their normalized Laplacian matrix Eq. (4).

ISBN 978-0-9928626-7-1 © EURASIP 2017

whereas number five is vice versa. Also, there is a similarity between number one and three as both numbers have a strong connection in the middle. In normalized version, one, three and seven have a high degree of similarity in their second eigenvector.

B. Graph model and feature extraction

This paper proposes undirected graph (G). G is built over the nodes which represent the X,Y coordinates of the hand's path. G is a fully connected graph which means each node has (N-1) connections as shown in (Fig. 3) and these edges represent the Euclidean distance between nodes.

This paper suggests utilizing two types of feature as follows:

1) First moment components (M_i)
   Although there is a similarity in the second eigenvector vector between samples, a small difference details is shown at the beginning or at the end of the Fiedler vector such as number two and five. Therefore, to overcome this issue, the first moment components of the second eigenvector is proposed to highlight these differences as shown in Eq. (5)

$$M_i = \chi_{(j,i)} * (i+1), \qquad (for \;\; i = 0, 1, .., N-1.) \quad (5)$$

   where $(j = 0, 1, ..., N-1.)$ refers to the index of the eigenvector.
2) Zero crossings (S).
   The number of zero crossings at each eigenvector is computed. We ignore the first eigenvector because its values have the same sign in all cases which means zero crossing always equal to zero. S provides the information about the positions of high connectivity areas. Also, it is important to mention that the order of the nodes is important to get an optimal S value because S will be varied for the same structure if and only if the nodes order strategy is different between patterns.

Regarding to the length of feature,
From (M_i): N
From (S): N-1
In total, the length is equal to $2N - 1$.

C. Classification

For the classification process, several types of classifiers are trained as will be shown in Section IV. Quadratic Discriminant Analysis classifier with 'Pseudo' as a kernel function shows better performance compared to the other classifiers.

III. In air drawn number recognition

This paper proposes recognising digits from (0-9) (Fig. 1) as an application to test the spectral graph feature for in-air hand writing recognition because there is a similarity (*i.e.*, digits 2 and 5) and variety (*i.e.*, digits 4 and 8) in terms of the structure. To do so, the Kinect sensor is used to provide fast hand tracking based on the skeleton coordinate. Users have to stand in front of Kinect around (1-3) meters [21] to draw handwritten numbers. Right hand is used to draw numbers in the air. Then, users can raise their left hand higher than the shoulder to finish the drawing. After the hand coordinates are acquired from the Kinect, a pre-processing steps are implemented in order to remove outlier nodes and keep all the samples in the same scale. Therefore, four steps are performed to do so:

1) Outlier remover:
   Once the hand's path is detected, a fixed threshold (t) is used to remove the nodes that are outliers to the accepted path. These nodes, considered as noise, are removed by considering the distance higher than a threshold (Fig. 3).
2) Node selection:
   The prior threshold adjusted the distance between the nodes. N_{th} nodes are selected from the hand's path, as shown in (Fig. 3). Let the S is the original node path with length (L), a space vector (V) is generated with length (N) from 1 to L by increment equal to $\left(\frac{L-1}{N-1}\right)$. After Outlier remover and node selection steps, a balance distribution of nodes along hand's path will be adjusted.
3) Image resizing:
   Numbers have different sizes in terms of image size. Therefore, to have a uniform structure of numbers in the same scale, image is resized into (K $\times$ K) image pixels as shown in Fig. 3 in order to get the same scale for all the samples.

After the pre-processing steps, G is constructed over the hand's path. Then, its features are extracted for classification.

IV. Results and discussions

In order to evaluate the proposed algorithm, a new and large dataset is created. The dataset was implemented by 13 volunteers performing 40 samples per each number. Volunteers follow the same direction rules to write the samples in the air. The dataset consists of 5200 samples, 520 samples for each class (0 - 9). In this work, we try to minimize the number of nodes to form the samples as much as possible, which leads to reduce the number of features and to improve the time performance of the system. We evaluate three different lengths of each sample which are 9,13 and 17 nodes with fixed threshold value equal to 20 and image size (K= 100) pixels which are determined experimentally. For the classification, 50% of random samples are used to train the classifier and the remaining samples were used for testing.

Initially, the proposed method is evaluated based on different classifiers. To do so, the database is randomly partitioned into 50%(training) and 50%(testing) four times to identify the better classifier. As can be seen in Table I, Nearest Neighbour (KNN) and Naive Bayes (NB) have more false prediction and they record relatively lower degree of accuracy than Quadratic Discriminant Analysis (QDA), Quadratic Support Vector Machine (QSVM), Classification Tree (CT) and Neural Network (NN). For the neural network (NN) classifier, a two-layer feed-forward network, with sigmoid hidden and softmax

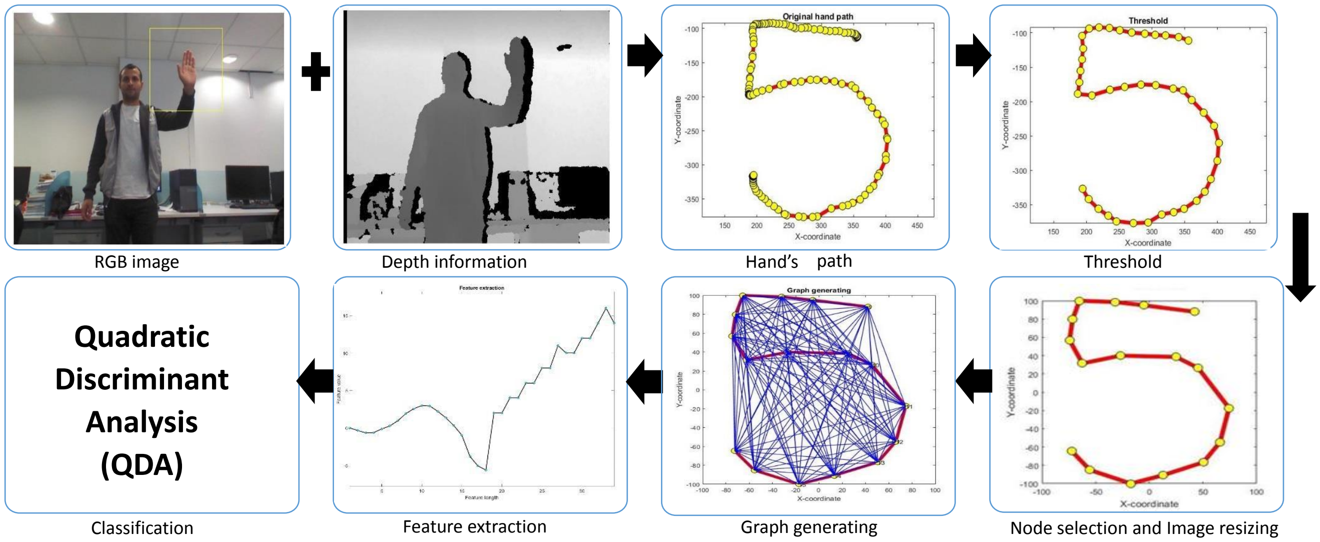

Fig. 3. Steps of the proposed method.

TABLE I
DIFFERENT CLASSIFIERS TEST(%) BASED ON 20 EXPERIMENTS.

	QDA	NN	QSVM	CT	KNN	NB
Mean	97.20	97.24	96.38	94.19	88.81	88.38
Standard deviation	0.39	0.19	0.39	0.45	0.49	2.63

TABLE II
PERFORMANCE EVALUATION OF THE PROPOSED METHOD.

Proposed method	9 nodes	13 nodes	17 nodes
Non-normalized	94.92	94.84	96.11
Normalized	97.23	97.46	97.53

output neurons, with 40 neurons in its hidden layer. Although NN has the maximum level of accuracy, this ratio is change based on the number of neurons in its hidden layer. Therefore, we prefer to utilized QDA for the rest of experiments.

To set up the optimal criteria of the spectral graph features, the evaluation is implemented using both the normalized and non-normalized graph Laplacian matrices for three different lengths. The proposed features achieve a highest level of accuracy by (97.53%) using normalized graph Laplacian matrix as shown in Table II.

This case is evaluated and the confusion matrix (Fig. 4) shows that, number one, five and seven manifest the maximum error recognition rate. This is because they have a high connectivity ratio in the middle as in class three. As a consequence, the error samples are misclassified as a classes three. In addition, Table III shows that the graph spectral features achieve the highest level of accuracy compared to the available works on in-air handwritten number recognition based on Kinect.

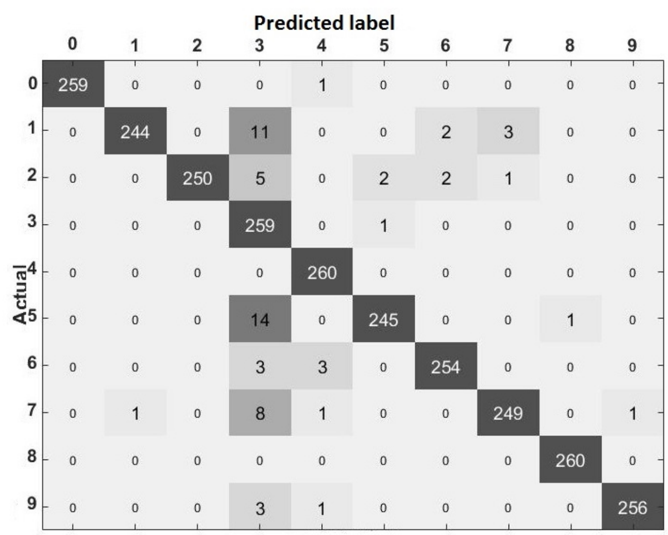

Fig. 4. Confusion matrix rate (%) for individual class.

TABLE III
AVERAGE RECOGNITION RATE (%) FOR EACH CLASS.

	Accuracy of [9]	Accuracy of [10]	Accuracy of [11]	Proposed Solution	Time (ms)
Zero	92.61	97.83	97.83	99.62	834
One	76.09	94.78	91.3	93.85	468
Two	86.96	95.65	96.09	96.15	834
Three	86.96	93.48	97.39	99.62	856
Four	91.74	96.09	97.39	100	750
Five	75.96	89.13	98.26	94.23	804
Six	86.96	95.65	97.83	97.69	662
Seven	91.3	94.35	97.83	95.38	573
Eight	87.83	93.91	95.22	100	1050
Nine	89.57	95.22	99.13	98.46	816
Average	**90.8**	**94.6**	**96.8**	**97.53**	**765**

ISBN 978-0-9928626-7-1 © EURASIP 2017

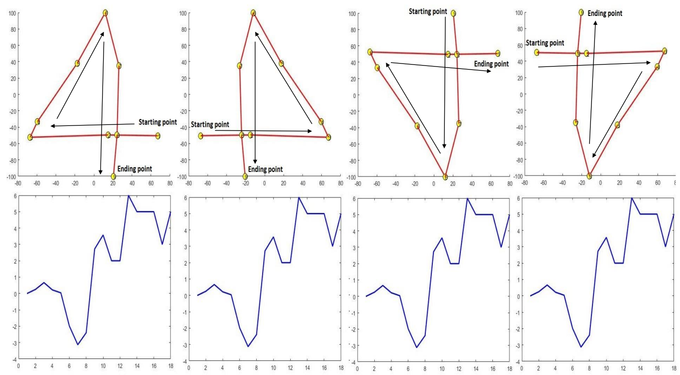

Fig. 5. Hand drawing in different angle with its feature.

The real-time system was implemented using Matlab R2015b Intel processor, CPU@3.6GHz and RAM 16GB. Average speed of the recognition 0-9 digits is achieved by 765 ms, which is a suitable range for the real-life applications.

Finally, since the graph adjacency matrix was defined based on the connectivity, it is insensitive to the rotation and flip changing. In other words, it does not matter what is the angle or direction of writing unless it follows the rules of starting and ending point. For example, all the cases shown in (Fig. 5) are detected as number four.

Note that the stroke order of drawing numbers affects the graph formation and the detection results. For example, if we write number five by starting from bottom to the top, the system will detect it as a number two (*i.e.*, flip 2). The only difference between gesture 2 and 5 is that the short connectivities happens at the beginning in 5 and at the end in 2. Despite the angle of drawing, stroke order is an essential concept for the detection.

For further illustration, swiping the hand to the left, right, top or bottom are classified as a single category. This is a significant property for many applications that require detection of samples at different angles. However, it may cause limitations for applications that require distinguishing samples at different angles.

V. Conclusions

Graph is a new mechanism to deal with unstructured data that provides an optimal description of the geometric structure. In this paper, we have proposed a new algorithm for in air hand writing recognition based on spectral graph feature. Matching problem was solved by forming the hand's path as a graph and extract its features. The proposed feature includes the first moment components of the graph eigenvector and the number of zero crossing in each eigenvector. The proposed method achieves up to 97.53% with low complexity and real time implementation. Finally, the proposed method has shown that the spectral graph feature provides an optimal rotation and flip invariant features for in-air number recognition.

References

[1] R. Horaud and H. Sossa, "Polyhedral object recognition by indexing," *Pattern recognition*, vol. 28, no. 12, pp. 1855–1870, 1995.

[2] M. Carcassoni and E. R. Hancock, "Alignment using spectral clusters." in *BMVC*, 2002, pp. 1–10.

[3] S. Umeyama, "An eigendecomposition approach to weighted graph matching problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 695–703, Sept 1988.

[4] T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching," *Advances in Neural Information Processing Systems*, vol. 19, p. 313, 2007.

[5] F. Zhou and F. de la Torre, "Factorized graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2015.

[6] A. Shamaie and A. Sutherland, "Graph-based matching of occluded hand gestures," in *Applied Imagery Pattern Recognition Workshop, AIPR 2001 30th*, Oct 2001, pp. 67–73.

[7] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, Oct 2005, pp. 1482–1489 Vol. 2.

[8] P. Riba, J. LladÃčs, and A. FornÃl's, "Handwritten word spotting by inexact matching of grapheme graphs," in *13th International Conference on Document Analysis and Recognition (ICDAR)*, Aug 2015, pp. 781–785.

[9] T. T. Chu and C. Y. Su, "A Kinect-based handwritten digit recognition for TV remote controller," in *International Symposium onIntelligent Signal Processing and Communications Systems (ISPACS)*, Nov 2012, pp. 414–419.

[10] F. A. Huang, C. Y. Su, and T. T. Chu, "Kinect-based mid-air handwritten digit recognition using multiple segments and scaled coding," in *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, Nov 2013, pp. 694–697.

[11] C. Wang, C. Y. Su, and C. L. Lin, "A novel recognition system for digits writing in the air using coordinated path ordering," in *International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, Nov 2015, pp. 244–249.

[12] J.-S. Sheu and Y.-L. Huang, "Implementation of an interactive TV interface via gesture and handwritten numeral recognition," *Multimedia Tools and Applications*, pp. 1–22, 2015.

[13] T. Murata and J. Shin, "Hand gesture and character recognition based on Kinect sensor," *International Journal of Distributed Sensor Networks*, vol. 1, pp. 1–6, July 2014.

[14] J. Triesch and C. von der Malsburg, "Robust classification of hand postures against complex backgrounds," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Oct 1996, pp. 170–175.

[15] M. Avraam, "Static gesture recognition combining graph and appearance features," *International Journal of Advanced Research in Artificial Intelligence (IJARAI)*, vol. 3, no. 2, 2014.

[16] H. Hamer, K. Schindler, E. Koller-Meier, and L. V. Gool, "Tracking a hand manipulating an object," in *IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 1475–1482.

[17] P. P. Kumar, P. Vadakkepat, and L. A. Poh, "Graph matching based hand posture recognition using neuro-biologically inspired features," in *11th International Conference on Control Automation Robotics Vision (ICARCV)*, Dec 2010, pp. 1151–1156.

[18] Y.-T. Li and J. P. Wachs, "Recognizing hand gestures using the weighted elastic graph matching (WEGM) method," *Image and Vision Computing*, vol. 31, no. 9, pp. 649–657, 2013.

[19] ——, "Hegm: A hierarchical elastic graph matching for hand gesture recognition," *Pattern Recognition*, vol. 47, no. 1, pp. 80–88, 2014.

[20] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 619–633, 1975.

[21] K. Khoshelham, "Accuracy analysis of kinect depth data," in *ISPRS workshop laser scanning*, vol. 38, no. 5, 2011, p. W12.

ISBN 978-0-9928626-7-1 © EURASIP 2017