HAL open science

A novel machine learning approach in Image Pattern Recognition under invariance constraints

Dragos Nastasiu, Angela Digulescu, Cornel Ioana, Maxime Bernier, Frederic Garet, Alexandru Serbanescu

To cite this version:

Dragos Nastasiu, Angela Digulescu, Cornel Ioana, Maxime Bernier, Frederic Garet, et al.. A novel machine learning approach in Image Pattern Recognition under invariance constraints. EUSIPCO 2022 - 30th European Signal Processing Conference, Aug 2022, Belgrade, Serbia. pp.558-562, 10.23919/EUSIPCO55093.2022.9909907 . hal-03980830

HAL Id: hal-03980830

https://hal.science/hal-03980830v1

Submitted on 9 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A novel machine learning approach in Image Pattern Recogniton under invariance constraints

Nastasiu Dragoș
IMEP-LAHC, UMR 5130 CNRS
GIPSA-Lab, UMR 5216 CNRS
Univeristé Grenoble-Alpes
Grenoble, France
dragos.nastasiu@mta.ro

Digulescu Angela
Department of Communications and Information Technology
Military Technical Academy
"Ferdinand I"
Bucharest, Romania
angela.digulescu@mta.ro

Ioana Cornel
GIPSA-Lab, UMR 5216 CNRS
Univeristé Grenoble-Alpes
Grenoble, France
cornel.ioana@gipsa-lab.fr

Bernier Maxime
IMEP-LAHC, UMR 5130 CNRS
Univeristé Savoie Mont-Blanc
Chambéry, France
maxime.bernier@univ-smb.fr

Garet Fréderic
IMEP-LAHC, UMR 5130 CNRS
Univeristé Savoie Mont-Blanc
Chambéry, France
frederic.garet@univ-smb.fr

Șerbănescu Alexandru
Department of Communications and Information Technology
Military Technical Academy
"Ferdinand I"
Bucharest, Romania
alexandru.serbanescu@mta.ro

***Abstract*—Two of the main challenges of image recognition in radar, acoustic or T-ray imaging regard the view-point variation of the pattern and the feature extraction techniques that must retrieve the most discriminative information about different classes. In this paper, we focus on feature extraction and image classification techniques by using a Rotation Invariant Wavelet Packet Decomposition and a novel entropy-based feature extraction technique to characterize an image. The entropy-based characterization described in the paper offers an extended analysis compared to usual approaches such as the energy of the wavelet sub bands. The computed features will be further used to train a Graph Neural Network adapted to a quad-tree decomposition which has the powerful advantage of considering the structural information of the rotation-invariant decomposition. We successfully classified the images with an accuracy of 99.3%. The results are compared to other classic feature extraction techniques such as k-NN, SVM and WPD, proving the increased capability of our method.**

Keywords—rotation invariant wavelet packet decomposition, directional entropy, feature extraction, graph neural networks, pattern recognition, image classification

I. Introduction

Pattern recognition is a field that studies the characteristics of data by means of computer algorithms in order to classify it into different categories [1]. The image recognition task involves four main components that are usually addressed in the following order: sensing or image acquisition, image preprocessing, image characterization or feature extraction and finally, image classification. The applications of pattern recognition are vast and include computer-aided diagnosis such as ultrasound, imaging, X-rays and ECG; industrial automation; speech and character recognition; industrial fingerprinting and entity authentication [2]. Most of the methodologies used in these applications are based on a hybrid approach, meaning that they use complementary information, tools and paradigms to create an accurate classification system. In the same framework, our paper aims to develop a new approach of image analysis and in-depth feature extraction based on entropy, invariant wavelet decomposition and "best basis" selection.

One of the popular choices in the field of image processing is Wavelet Packet Decomposition (WPD). WPD is the generalization of the Wavelet Transform (WT) that offers a detailed decomposition of an image into multiple frequency sub-bands [3], similar to how the human's visual system perceives reality. However, the overcomplete WPD is computational expensive and therefore, it motivates the selection of the most representative sub-bands usually based on entropy cost functions [4]. Additionally, from a human perspective, an object can be identified regardless of the point of view. But from computer vision perspective, the extracted features from the rotated image are different with respect to the ones coming out from the original image. This will certainly lead to a wrong classification. In order to overcome this, the WPD is improved to Rotation-Invariant WPD (RI-WPD), making the decomposition identical for images with the same pattern, but which is rotated.

In most classification applications, the widely used features are the statistical properties [5], energy [6] or the entropy [7] of the wavelet sub-bands. The aforementioned approaches wield good results in many applications, but lack an in-depth analysis of the wavelet sub-bands. In opposition, in this paper, the wavelet features are in fact the N-directional entropy of each wavelet sub-band. This novel feature emphasizes the distribution of entropy in N directions for each sub-band and offers an optimal analysis of the image. The idea behind using the entropy concept as a means of characterization comes from the fact that the wavelet sub-band selection is also based on an entropy cost function. Moreover, in image processing, the entropy can be regarded as the quantization of the structural information contained in an image: the lower the entropy of an image, the more structured the information are, while the higher the entropy, the more the image is closer to noise.

Recently, Graph Neural Networks (GNN) have received more attention because of the great expressive power of graphs [8]. They are used in various domains, including traffic forecasting, recommender systems, medical diagnosis, etc. A GNN operates directly on non-Euclidean data structures and can be used to predict node labels, links, edges or even the entire graph. The major advantage of GNN stems from considering the hierarchal information of data, as opposed to the classical neural networks which do not.

The contributions of this paper include the characterization of images in terms of entropy distribution among wavelet sub-band images and finally, the use of GNN adapted to RI-WPD to classify a set of images accordingly.

The paper is organized as follows. Section II briefly describes the classic WPD, "best basis" selection and the additional property of rotation invariance. Section III details

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

the novel N-directional entropy characterization of the RI-WPD. Section IV describes the GNN and the way we use it in our study. Section V presents the results of the classification and Section VI closes the paper with the conclusions.

II. Wavelet Packet Decomposition

WPD is an extension of WT in order to obtain an over-complete time-frequency analysis of a signal. In 2-D WPD, an image is decomposed into one approximation and three detail images. These images are further decomposed into other four images, and the process is repeated. The classic 2-D WPD can be implemented using the multiresolution, filter-bank and pyramidal image decomposition principles [9]. The 2-D WPD of an $M \times M$ image up to a depth of $D+1$, where $D \le \log_2 N$, is defined as follows:

$$C^{d+1}_{4k,(i,j)} = \sum_m \sum_m h(m) h(n) C^d_{k(m+2i,m+2j)} \,. \tag{1}$$

$$C^{d+1}_{4k+1,(i,j)} = \sum_m \sum_m h(m) g(n) C^d_{k(m+2i,m+2j)} \tag{2}$$

$$C^{d+1}_{4k+2,(i,j)} = \sum_m \sum_m g(m) h(n) C^d_{k(m+2i,m+2j)} \tag{3}$$

$$C^{d+1}_{4k+3,(i,j)} = \sum_m \sum_m g(m) g(n) C^d_{k(m+2i,m+2j)} \tag{4}$$

where C^0_0 is the initial image, k is the node's index in the wavelet packet tree, representing each sub-band, h and g are a pair of quadrature mirror filters' impulse responses. Recursively, the image C^d_k is filtered into four images at 2 times lower resolution, C^{d+1}_{4k}, C^{d+1}_{4k+1}, C^{d+1}_{4k+2}, C^{d+1}_{4k+3}. In Fig. 1 we find the graph representation of an over-complete wavelet analysis.

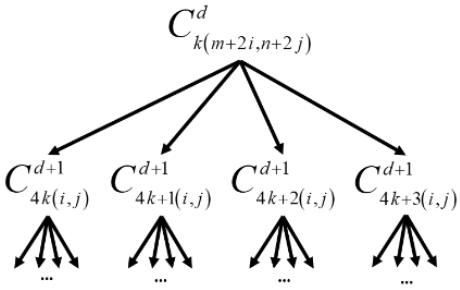

Fig. 1. Full WPD analysis graph

A. "Best basis" selection

The number of possible decompositions in a WPD is often large and thus it is expensive to analyze all the possible options. There are efficient algorithms for finding optimal decompositions which use classical entropy and energy cost functions [10]. In our study, the "best" decomposition from the overcomplete 2-D WPD is based on the minimization of the nonnormalized Shannon entropy function for an image, $f(x,y)$:

$$E\left[f(x,y) \right] = \sum_x \sum_y f^2(x,y) \ln\left[f^2(x,y) \right]. \tag{5}$$

The algorithm of selecting a subspace of minimum entropy distribution is synthesized in the following. Starting with the initial image and proceeding level by level to the lower resolution images:

- Compute the Shannon entropy for each node η_P (parent entropy) and the entropy of its four children nodes denoted as η_A (approximation entropy), η_H (horizontal details entropy), η_V (vertical detail entropy), η_D (diagonal detail entropy).
- If the summed entropy of the children nodes is less than the entropy of the parent, keep the children nodes in the optimal decomposition tree. Otherwise, keep only the parent node.

After finishing the algorithms' iterations, we find the optimal decomposition as it is depicted in Fig. 2, where only the most representative time-frequency bands are included.

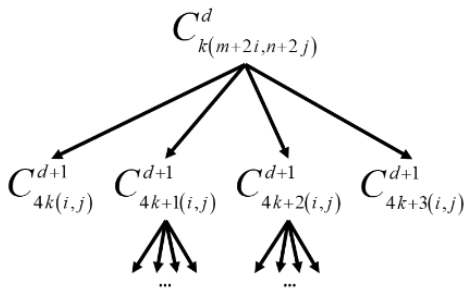

Fig. 2. "Best basis" or the optimal decomposition graph

B. Rotation invariancy

The WPD can be modified accordingly in order to achieve rotation invariance [11]. Fig. 3 (a) presents the strategy to obtain a rotation invariant decomposition. Firstly, the image is represented in polar coordinates. The polar representation transposes the complex problem of rotation in digital images to a more simpler translation problem. A translation in the polar domain represents a rotation in the cartesian grid. The mapping from an image $I(x,y)$ to a polar representation, $PR(\rho,\theta)$ is defined by the following parameters relations:

$$\rho = \sqrt{(x-x_c)^2 + (y-y_c)^2} \tag{6}$$

$$\theta = \tan^{-1} \frac{y-y_c}{x-x_c} \tag{7}$$

where (x_c, y_c) is the center of the image, (x,y) denotes the sampling pixel in the cartesian grid and (ρ,θ) is the radius and angular position in the polar representation.

Rotation-invariance comes from an additional degree of freedom, generated at the decomposition stage and incorporated into the "best basis" selection algorithm as it is shown in Fig. 3 (b), (c). In accordance with Mallat's pyramidal decomposition principle [9], at each node, we generate the subspace of all wavelet packet coefficients and their row-shifted versions using the analysis operator, here denoted A. Row shifting indicates that the translation operator (T) can only translate the image in horizontal direction, thus the accepted translations are $(0,0)$ and $(1,0)$, where the first argument expresses the translation on the Ox axis (horizontal), respectively, the Oy axis (vertical) of the polar image.

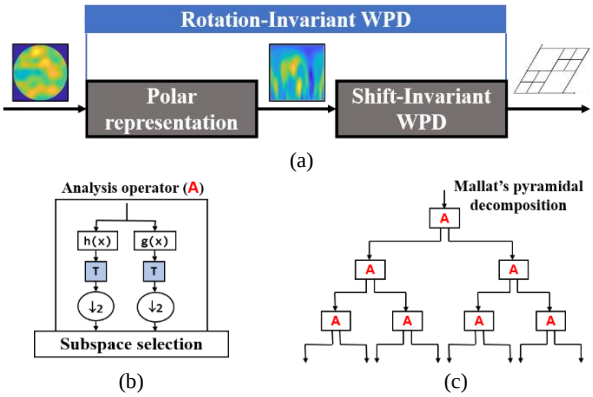

Fig. 3. Presentation of the: (a) conceptual scheme of RI-WPD; (b) RI-WPD analysis operator A – high- and low-pass filtering followed by one sample delay (T) and subsequently a 2:1 downsampling; (c) Mallat's pyramidal decomposition with the presented analysis operator, A.

With respect to the row-shifts, where $T \in \{0,1\}$, the RI-WPD can be defined as follows:

$$C^{d+1,(T,0)}_{4k,(i,j)} = \sum_m \sum_n h(m)h(n)C^d_{k(m+2i+T,n+2j)} \tag{8}$$

$$C^{d+1,(T,0)}_{4k+1,(i,j)} = \sum_m \sum_n h(m)h(n)C^d_{k(m+2i+T,n+2j)} \tag{9}$$

$$C^{d+1,(T,0)}_{4k+2,(i,j)} = \sum_m \sum_n g(m)h(n)C^d_{k(m+2i+T,n+2j)} \tag{10}$$

$$C^{d+1,(T,0)}_{4k+3,(i,j)} = \sum_m \sum_n g(m)g(n)C^d_{k(m+2i+T,n+2j)} \tag{11}$$

Equations (8)-(11) represent all the coefficients that appear if the original analyzed image is translated individually by $(0,0)$ and then by $(1,0)$, before the filtering procedure of each case. We select the "best" time-frequency band from the generated subspace based on the entropy cost function. By repeating the procedure, at each node, we select a space of minimum entropy distribution, invariant to rotations defined in the polar representation.

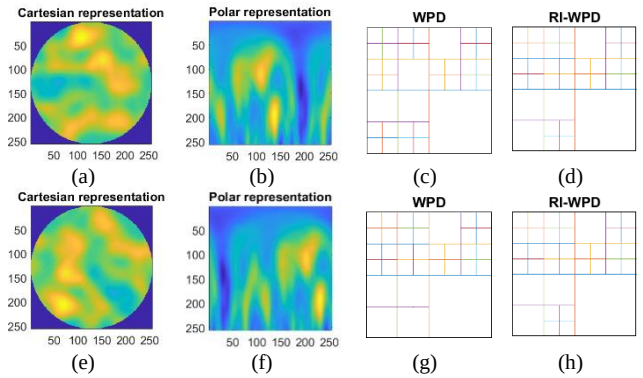

Fig. 4. A comparison between WPD and RI-WPD for: (a) an image and (e) its rotated version in cartesian grid; (b), (f) same images in polar coordinates; (c), (g) WPD "best basis"; and (d), (h) RI-WPD "best basis".

Figure 4 (a), (b), (e), (f) shows an image and its rotated version in both cartesian and polar coordinates. We provide the corresponding WPD and RI-WPD analysis for a depth $D = 3$ in Fig. 4 (c), (d), (g) and (h), respectively. We observe that the RI-WPD "best basis" is the same regardless the orientation of the initial image, while in the classic WPD the basis changes. This feature of RI-WPD allows us to provide the same decomposition of a pattern regardless of its orientation.

III. N-DIRECTIONAL ENTROPY

After an invariant decomposition of the image, the next step is to introduce a novel method of feature extraction based on the Shannon entropy. The entropy measures the randomness of an image. For a $M \times M$ dyadic image, the entropy is 0 when the image is of constant intensity and is maximum, $M \log_2 M$, when the probabilities of a pixel value to occur are uniformly distributed. It is important to mention that lower values of the entropy are strictly correlated to structural information appearing in the image.

Based on this concept, we describe the distribution of information in N directions or regions of interest (ROI). The choice of N depends on the information present in the image. As a rule of thumb, the maximum value should be a power of 2 and lower than the smallest dimension of an image, in our case, M. In this manner we assure no overlapping pixels between the regions. If N does not respect the imposed constraints, the analysis should still provide a good description, but with redundant information and at a higher computation cost.

Fig. 5 (a), (b) provides two examples of a directional entropy analysis, where $N = 8$ and thus, forming 8 regions to compute the entropy. The computed entropies are presented in a polar plot as it provides an adequate way to visualize the distribution inside the analyzed image. The cases from Fig. 5 represent two binary images, one being the noised version of the other. The two lobes present in each polar plot show that in specific directions there is a structure of information, in our case, a square and an ellipse. It is worth mentioning that this method is not dependent on the geometric structures, but rather on the scattering of information in the image. Moreover, the additive white noise has a low impact on the results, as we can see in Fig. 5 (b). The lobes keep their aspect, but their entropy is slightly increased due to the randomness introduced by the additive white noise in the whole image.

The N-directional entropy characterization has high potential in being a powerful tool to describe not only time-frequency bands from wavelet decompositions, but even raw patterns and motives present in images.

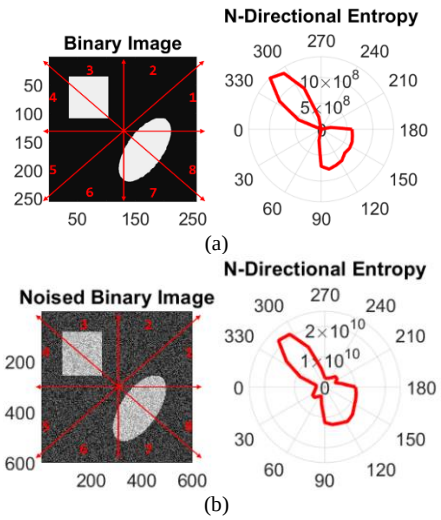

Fig. 5. N-directional entropy characterization of binary image: (a) ideal case; (b) noisy case.

IV. Graph Neural Networks

A GNN is a type of neural network which directly operates on a graph structure. The main advantage of GNN regards the opportunity to exploit not only the features from all edges and nodes, but also the structural information of the graph, which is omitted by the classic neural network architectures [12]. A GNN is an optimizable transformation of all attributes of a graph (nodes or edges) that offers an optimal graph representation used to predict other node values, edges or even classify the entire graph.

The motivation of using GNNs comes from the general hierarchical structure of the RI-WPD. In other words, the RI-WPD decomposes an image into a wavelet graph, where each node represents a sub band image and each edge represents the connection between an image and one of its children. The general graph convolutional network (GCN) used in our study follows the message passing paradigm [13] and is defined as follows:

$$h_i^{(l+1)} = \sigma\left(b^{(l)} + \sum_{j \in V(i)} \frac{e_{ji}}{c_{ji}} h_j^{(l)} W^{(l)} \right) \tag{12}$$

where V is the RI-WPD "best basis", $V(i)$ is the set of neighbors of node i, c_{ji} is the product of the square root of node degrees, e_{ji} is the scalar weight of the edge from node j to node i, and σ is the non-linear activation function ReLU. By using the notations W and b we denote the weights and bias vector, respectively. The proposed network has 3 rounds of graph convolution with 128 neurons. The next step is graph readout or aggregation by averaging over all node features:

$$h_{agg} = \frac{1}{|V|} \sum_{j \in V} h_j \tag{13}$$

The final layer of the proposed GNN is a SoftMax classifier which offers a probabilistic interpretation of the output. The optimization technique used is Adam [14] with a learning rate of $\eta = 0.001$. In our study, this type of neural network is adapted to work on a wavelet graph. Fig. 6 presents an image characterization and classification example using the neural architecture mentioned in the previous paragraphs. Each of the "best basis" nodes are described using the N-directional entropy. The edges are assumed to be of the same priority or weight, that is $e_{ji} = 1$, regardless of i, j values.

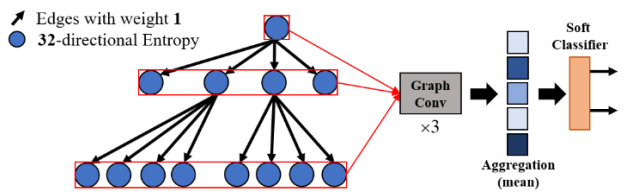

Fig. 6. Image feature extraction and GNN classification from RI-WPD "best basis"

V. Database Creation and Results

To demonstrate our proposed approach in pattern recognition we used enhanced THz images [15] of a sample composed from paper substrate (95%) and 3*mm* metallic fibers (5%) printed on it with a random distribution. Fig. 7 (a) shows the optical image of the sample and the corresponding THz image acquired with a THz imaging system, TeraPulse Ltd. from Teraview. The THz image is reconstructed using the maximum-peak value of the time-domain pulses. The spatial resolution is $300\mu m$ and the sample is $4.5cm$ by $4.5cm$. The spatial resolution is limited by the refraction limit and the wavelength of the THz waves. Therefore, because of the small carbon fibers, the THz image does not resemble the optical image.

To generate a database of images, we divide the initial 150×150 image into 9 classes or patterns as showed in Fig. 7 (a). Each pattern is interpolated to extract 256×256 images from inside of it. The training dataset creation is depicted in Fig. 7 (b), where for each interpolated image we apply a circular mask. The idea is to assure the independence of the images generated for a class. Inside the circular mask, we can freely translate and rotate a 256×256 mask without the risk of getting inter-class pixel information. In this manner, we generate 400 images for each category. To further validate our concept, we divide the database. The percentages from the initial dataset of 3600 images corresponding to the training/validation/testing sets are 60/20/20.

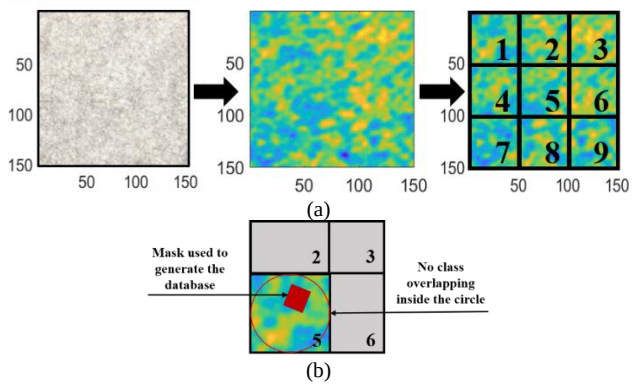

Fig. 7. Presentation of the: (a) optical sample, THz image and the classes (noted from 1 to 9); (b) The means of class generation with a square mask inside a circle to avoid inter-class overlapping.

The GNN architecture described in Section IV is trained until the validation accuracy is not greatly improving for 5 consecutive iterations. The evolution of the learning metrics (training/validation loss and accuracy) can be seen in Fig. 8. The curves have a steady and a low-variance that correspond to an adequate learning procedure. The validation accuracy goes up to 99.8% and the validation loss converges to a minimum, solidifying the performance of the classifier.

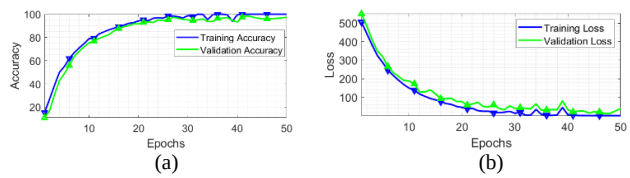

Fig. 8. Monitoring of learning metrics: (a) training/validation accuracy; (b) training/validation loss.

The confusion matrix provides an insight of the classifier's capabilities on the testing set. The rows correspond to the predicted class and the columns correspond to the target class. On the first diagonal, the numbers represent the correctly classified images, while the off-diagonal cells are the misclassified ones. The far-right column corresponds to the precision and false discovery rate (FDR), while the bottom

row shows the recall and false negative rate (FNR). The cell in the bottom right shows the overall accuracy. As we observe, the resulted accuracy is 99.3%. Only a few images are misinterpreted by the proposed GNN, the FDR achieving a maximum of 2.5% for the third class.

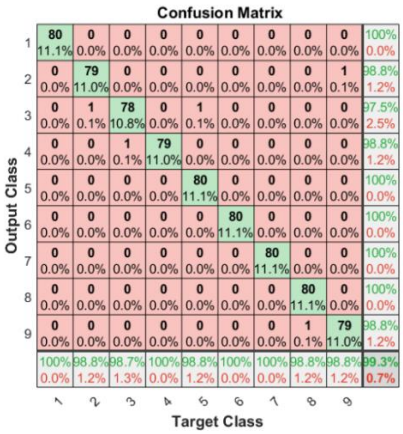

Fig. 9. Confusion matrix

The bar graphic in Fig. 10 presents a comparison between our proposed method and other classic image classification techniques such as k-Nearest Neighbors (k-NN), Support Vector Machine (SVM) and WPD. In this comparison, we used the same image database with the same proportions for the training/validation and testing set. For WPD, the entropy features extracted are provided to train a two-layer neural classifier for 50 epochs. As for the k-NN, we analyzed multiple number of neighbors and found that $k = 5$ offered the best outcome. As we can observe, the performance of our proposed method is superior compared to the other techniques by a large margin. Second and third to our method are the classic k-NN and WPD.

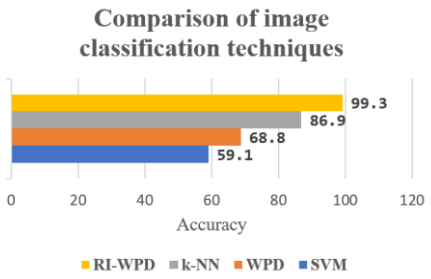

Fig. 10. Performance analysis of different techniques of image classification

VI. Conclusions

The paper presents a novel approach in image pattern recognition. The method is based on the rotation invariant version of the WPD and the selection of a ”best basis” using the mimization of the entropy cost function. In this manner, we provide a precise decomposition, invariant to rotational variations of the observed pattern.

Each time-frequency band is individually analyzed to emphasize the distribution of entropy in N directions or regions. Thus, a pattern analysis using our method will result in graph-structured features representing the most concise information extracted from the image in terms of information organization and structures.

The graph structure imposes the use of GNN to classify the images, as this type of neural network works directly on the graph-structured data, improving the generalization power of the classifier.

Our study showed that the proposed method can considerably outperform classic techniques of pattern recognition. Moreover, using GNNs allows to use multiple stacked features of the analyzed pattern to increase the performance of the classifier. Further studies include the possibility to extract complementary features from both a rotation and translation invariant WPD to increase the performance and robustness of the classifier.

Acknowledgment

This work is supported by the French Agence Nationale de la Recherche (AUSTRALE project – grant ANR-18-CE39-0002) and Région AURA (AUTHANTIC project – grant C2019-086)

References

[1] S. Asht and R. Dass, “Pattern Recognition Techniques: A Review,” *Int. J. Comput. Sci. Telecommun.*, vol. 3, no. 8, 2012.

[2] V. Kober, T. Choi, V. Diaz-Ramírez, and P. Aguilar-González, “Pattern Recognition: Recent Advances and Applications,” *Math. Probl. Eng.*, vol. 2018, Nov. 2018, doi: 10.1155/2018/8510319.

[3] R. C. Gonzalez and R. E. Woods, *4TH EDITION Digital image processing*. 2018.

[4] R. R. Coifman and M. V. Wickerhauser, “Entropy-based algorithms for best basis selection,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, 1992, doi: 10.1109/18.119732.

[5] G. G. Yen, “Wavelet packet feature extraction for vibration monitoring,” *IEEE Trans. Ind. Electron.*, vol. 47, no. 3, pp. 650–667, 2000, doi: 10.1109/41.847906.

[6] W. Ting, Y. Guo-zheng, Y. Bang-hua, and S. Hong, “EEG feature extraction based on wavelet packet decomposition for brain computer interface,” *Measurement*, vol. 41, no. 6, pp. 618–625, Jul. 2008, doi: 10.1016/J.MEASUREMENT.2007.07.007.

[7] D. Wang, D. Miao, and C. Xie, “Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection,” *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14314–14320, Oct. 2011, doi: 10.1016/J.ESWA.2011.05.096.

[8] J. Zhou *et al.*, “Graph Neural Networks: A Review of Methods and Applications,” *AI Open*, vol. 1, pp. 57–81, Dec. 2018, doi: 10.1016/j.aiopen.2021.01.001.

[9] S. G. Stéphane G. Mallat and G. Peyré, “A wavelet tour of signal processing : the sparse way,” p. 805, 2009.

[10] K. Huang and S. Aviyente, “Choosing best basis in wavelet packets for fingerprint matching,” *Proc. - Int. Conf. Image Process. ICIP*, vol. 5, pp. 1249–1252, 2004, doi: 10.1109/ICIP.2004.1419724.

[11] C.-M. Pun, “Efficient and Adaptive Rotation Invariant Wavelet Transform,” *http://dx.doi.org/10.1142/S0219691303000219*, vol. 01, no. 03, pp. 353–372, Jan. 2012, doi: 10.1142/S0219691303000219.

[12] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A Comprehensive Survey on Graph Neural Networks,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2019, doi: 10.1109/TNNLS.2020.2978386.

[13] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural Message Passing for Quantum Chemistry,” *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 3, pp. 2053–2070, Apr. 2017, Accessed: Feb. 18, 2022. [Online]. Available: https://arxiv.org/abs/1704.01212v2.

[14] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014, Accessed: Feb. 18, 2022. [Online]. Available: https://arxiv.org/abs/1412.6980v9.

[15] D. Nastasiu, M. Bernier, C. Ioana, C. Trehoult, L. Lyannaz, and F. Garet, “Phase diagram method for efficient THz images reconstructing,” pp. 1–2, Oct. 2021, doi: 10.1109/IRMMW-THZ50926.2021.9566864.