# Image pattern classification under invariant constraints

Dragoş Nastasiu[1,2,3], Maxime Bernier[2], Angela Digulescu[2], Frederic Garet[2], Cornel Ioana[1], Alexandru Şerbănescu[3]

[1]GIPSA-Lab, UMR 5216 CNRS, Université Grenoble–Alpes, 38402 Grenoble, France
[2]IMEP-LAHC, UMR 5130 CNRS, Université Savoie Mont-Blanc, 73376 Le Bourget du Lac, France
[3]Department of Communications and Information Technology, Military Technical Academy "Ferdinand I", Bucharest, Romania

[1,2,3]Dragos.Nastasiu@mta.ro

Résumé - Deux des principaux défis de la reconnaissance d'images dans l'imagerie radar, à rayons X ou à rayons T sont la variation du motif selon le point de vue et les techniques d'extraction de caractéristiques qui doivent extraire les informations les plus discriminantes des données analysées. Dans cet article, nous décrivons une nouvelle technique d'extraction de caractéristiques et de classification d'images qui atteint une précision presque maximale et surpasse les approches classiques, telles que k-NN et SVM avec les vecteurs de caractéristiques habituels d'entropie. Dans notre algorithme, les images sont décomposées en utilisant une décomposition en paquets d'ondelettes invariante en rotation. Ensuite, nous calculons la nouvelle technique d'extraction de caractéristiques, l'entropie N-directionnelle, pour chaque sous-bande d'ondelettes à partir de la "meilleure base". L'arbre de caractéristiques résultant est soumis à un classificateur à réseau de neurones graphiques, adapté pour fonctionner avec des décompositions en ondelettes.

Abstract - Two of the main challenges of image recognition in radar, X-ray or T-ray (Terahertz radiation) imaging are the view-point variation of the pattern and the feature extraction techniques that must retrieve the most discriminative information from the analyzed data. In this paper, we aim to describe a novel feature extraction and image classification technique that achieves an almost maximum accuracy and surpasses the classical approaches, such as k-NN and SVM with usual entropy feature vectors. In our algorithm, the images are decomposed in using a Rotation Invariant Wavelet Packet Decomposition. Afterwards, we compute the novel feature extraction technique, the N-directional entropy, for each wavelet sub-band from the "best basis". The resulted feature tree is fed to a Graph Neural Network classifier, adapted to operate on wavelet decompositions.

## 1 Introduction

The main objective of image pattern recognition is to study data characteristics in order to classify it [1]. Pattern recognition is widely used in industrial automation; speech and character recognition; computer-aided diagnosis such as ultrasound imaging, THz (T-rays) and X-rays; entity fingerprinting and authentication [2]. Our paper proposes to develop a new approach of pattern analysis and in-depth feature extraction based on entropy and rotation invariant wavelet decomposition.

Wavelet Packet Decomposition (WPD) is a popular tool in the field of image processing. WPD is the extension of the Wavelet Transform (WT) that provides an over-complete analysis of an image in terms of time-frequency sub-bands. Due to the exhaustive possibilities of wavelet basis, a selection algorithm is used to determine the best wavelet decomposition under the entropy minimization constraint. Our objective is to tackle the rotation variance problems in image processing. As human beings, we can identify an object from many perspectives or view-points, but from a computer vision perspective the extracted features from the rotated image are distinct compared to the ones extracted from the original image. This directly impacts the robustness of a classifier, certainly leading to image misclassification.

In some classification problems, the feature vector used to train a classifier is created from the entropy or the energy of each wavelet sub-band [3], [4]. Their main issue regards the lack of an in-depth analysis of each sub-band. We propose a novel feature extraction method which emphasizes the distribution of entropy in N directions for each wavelet sub-band and therefore, providing an optimal analysis of an 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. Graph Neural Networks (GNN) are becoming popular due to the great expressive power of graphs [5]. Their main characteristic, is the ability to directly operate on non-Euclidean data structures such as wavelet decomposition trees. Non-Euclidean data refer to data with an underlying structure that do not obey the Euclidean distance as a metric between points. They consider the hierarchical information of graph-structured data, opposed to classic neural network which omit it.

The paper is organized as follows. Section 2 and 3 describes the WPD and the rotation invariance. Section 4 details the novel N-directional entropy characterization of the RI-WPD. Section 5 describes the GNN and the way we use it in our study. Section 6 presents the results of the classification and Section 7 closes the paper with the conclusions.

## 2 Wavelet Packet Decomposition

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 until it reaches a specific depth level. The classic 2-D WPD can be implemented using the multiresolution, filter-bank and pyramidal image decomposition principles [6]. The 2-D WPD of an $M \times M$ image up to a depth of $D+1$, where $D \leq \log_2 M$, is defined as follows:

$$C_{4k,(i,j)}^{d+1} = \sum_m \sum_m h(m)h(n)C_{k(m+2i,m+2j)}^d \quad (1)$$

$$C_{4k+1,(i,j)}^{d+1} = \sum_m \sum_m h(m)g(n)C_{k(m+2i,m+2j)}^d \quad (2)$$

$$C_{4k+2,(i,j)}^{d+1} = \sum_m \sum_m g(m)h(n)C_{k(m+2i,m+2j)}^d \quad (3)$$

$$C_{4k+3,(i,j)}^{d+1} = \sum_m \sum_m g(m)g(n)C_{k(m+2i,m+2j)}^d \quad (4)$$

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 $C_k^d$ is filtered into four images at 2 times lower resolution, $C_{4k}^{d+1}$, $C_{4k+1}^{d+1}$, $C_{4k+2}^{d+1}$, $C_{4k+3}^{d+1}$ The number of possible decompositions in a WPD is often large and thus it is expensive to analyze all the possible options. In our study, the "best" decomposition from the overcomplete 2-D WPD is based on the minimization of the nonnormalized Shannon entropy [6] function for an image, $f(x,y)$:

$$E\left[f(x,y)\right] = \sum_x \sum_y f(x,y)\ln\left[f(x,y)\right] \quad (5)$$

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

- Compute the Shannon entropy for each node $\eta_P$ (parent entropy) and the entropy of its four children nodes denoted as $\eta_A$ (approximation entropy), $\eta_H$ (horizontal details entropy), $\eta_V$ (vertical detail entropy), $\eta_D$ (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.

## 3 Rotation-Invariant Wavelet Packet Decomposition

Rotation invariance can be achieved by combining the polar representation of an image and a row-shift invariant WPD. 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 relations:

$$\rho = \sqrt{(x-x_c)^2 + (y-y_c)^2}; \ \theta = \tan^{-1}\frac{y-y_c}{x-x_c} \quad (6)$$

where $(x_c, y_c)$ is the center of the image, $(x,y)$ denotes the sampling pixel in the cartesian grid and $(\rho,\theta)$ is the radius and angular position in the polar representation. Fig. 1. presents the block diagram of RI-WPD.
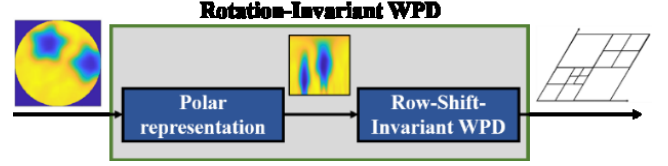


**Fig. 1. Block diagram of RI-WPD.**

In order to achieve a rotation-invariant decomposition, an additional degree of freedom is considered and generated at the decomposition stage and then, incorporated into the "best basis" selection algorithm, as it is shown in Fig. 2. Iteratively, at each node, we generate a subspace of all wavelet packet coefficients and their row-shifted versions, using the analysis operator, here denoted A. A row-shift 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. It is important to note that in our case, horizontal translations in polar domain are in fact rotations in the cartesian grid.
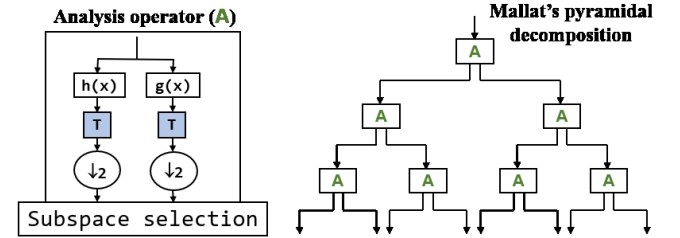


**Fig. 2. Presentation of RI-WPD analysis operator – high- and low-pass filtering, followed by one sample delay (T) and subsequently a 2:1 down sampling; and its implementation using Mallat's pyramidal decomposition principle.**

The following RI-WPD definition represents all the coefficients that appear if the decomposed image is firstly translated and filtered individually for each case of row-shifts, where $T \in \{0,1\}$. Considering the aforementioned, RI-WPD definition is stated as follows:

$$C_{4k,(i,j)}^{d+1,(T,0)} = \sum_m \sum_n h(m)h(n)C_{k(m+2i+T,n+2j)}^d \quad (7)$$

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

$$C_{4k+2,(i,j)}^{d+1,(T,0)} = \sum_m \sum_n g(m)h(n)C_{k(m+2i+T,n+2j)}^d \quad (9)$$

$$C_{4k+3,(i,j)}^{d+1,(T,0)} = \sum_m \sum_n g(m)g(n)C_{k(m+2i+T,n+2j)}^d \quad (10)$$

Fig. 3 presents a pattern and its rotated version in both cartesian and polar grid. We note that the WPD "best basis" is different in each case, while the RI-WPD provides an identical decomposition.
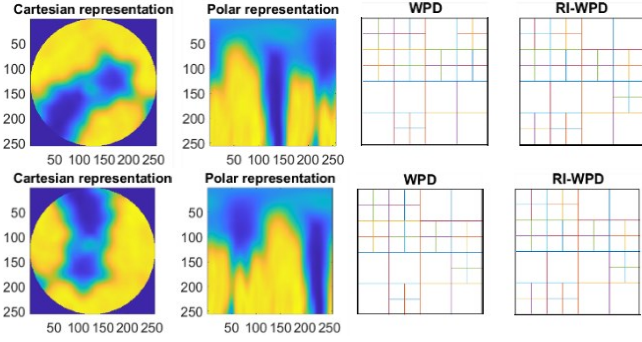
**Fig. 3. Comparison between WPD and RI-WPD for an image and it rotated version**

## 4 N-directional entropy characterization

The RI-WPD is further exploited by extracting features based on the Shannon entropy. This measure quantizes the informational randomness of an image. Considering a dyadic image of width and height, $M$, its entropy is $0$ when the image has a constant intensity and is maximum, $M \log_2 M$, when the pixels' probabilities are uniformly distributed. This novel feature extraction method is based on the concept that lower values of the entropy are correlated to geometric information appearing in the image. Based on this idea, our method offers a description of the structural information in $N$ directions or regions.

In Fig.4 (a) we present the manner in which we compute the directional entropy for an image. In Fig. 4 (b), (c) we show an example of directional entropy characterization in $N = 16$ directions. The entropies are presented in a polar grid as this representation is adequate to visualize the distribution inside the analyzed image.
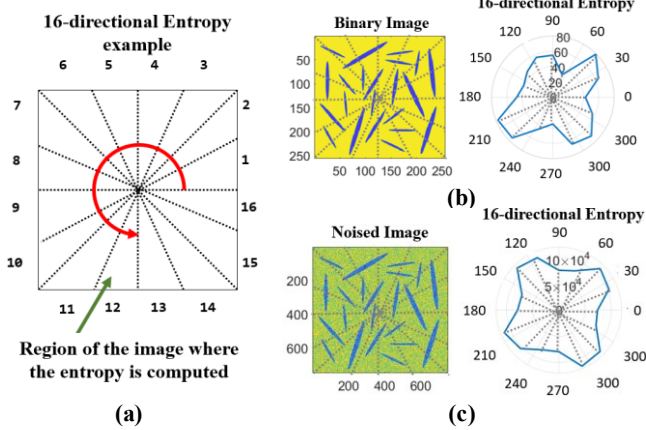


**Fig. 4. (a) N-directional entropy algorithm; (b) 16-directional characterization of binary image and (c) its noised version.**

The two cases from Fig. 4 (b), (c) represent two binary images, one being the noised version of the other. The polar plot describes the directions in which there is structured information in the image. It is worth mentioning that this characterization is dependent only on the scattering of information and not on the geometric structures. The impact of the additive noise is observed in an increase of entropy due to the randomness introduced by the additive noise in the image. This novel tool proves to be powerful enough to describe not only time-frequency bands from wavelet decompositions, but even plain patterns present in images.

## 5 Graph Neural Networks

The motivation of using GNNs comes from the 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 [7] and is defined as follows:

$$h_i^{(l+1)} = \sigma\left(b^{(l)} + \sum_{j \in V(i)} \frac{e_{ji}}{c_{ji}} h_j^{(l)} W^{(l)}\right) \quad (11)$$

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 $\sigma$ 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 64 neurons. The next step is graph readout or aggregation by averaging over all node features:

$$h_{agg} = \frac{1}{|V|} \sum_{j \in V} h_j \quad (12)$$

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 with a learning rate of $\eta = 0.001$. Fig. 5 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 $i, j$.
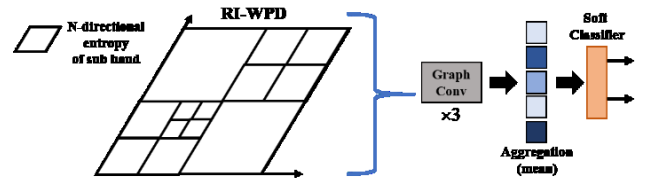


**Fig. 5. Image decomposition, feature extraction and GNN classification**

## 6 Results

To demonstrate our proposed approach in pattern recognition we used enhanced THz images [8] of a sample composed from bars ($10mm \times 100\mu m$) printed with metallic-ink on a polyethylene substrate. Fig. 7 (a) shows the 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 $3cm$ by $3cm$.

To generate a database of images, we divide the initial $100 \times 100$ image into 4 classes or patterns as showed in Fig. 6. Inside these 4 regions, we use a square mask to extract images corresponding to each class and thus generating the database as it is depicted in Fig. 7. The

percentages from the initial dataset of 1600 (400 per class) images corresponding to the training/validation/testing sets are 70/15/15.
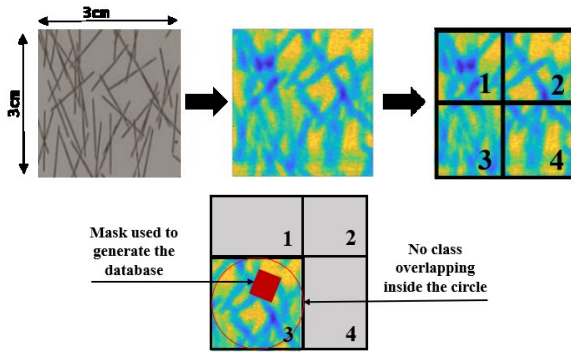


**Fig. 7. Database creation from the optical image to the corresponding THz images.**

The GNN architecture described in Section IV is trained until the validation accuracy is not improving for 5 consecutive iterations. The confusion matrix from Table 1 provides an insight of the classifier's capabilities on the testing set. As we observe, only one image is misinterpreted by the proposed GNN. The resulting accuracy is 99.58%, verified on the testing set.

| 240 | Cls 1 | Cls 2 | Cls 3 | Cls 4 |
|-----|-------|-------|-------|-------|
| Cls 1 | 60 | 0 | 0 | 0 |
| Cls 2 | 0 | 59 | 0 | 1 |
| Cls 3 | 0 | 0 | 60 | 0 |
| Cls 4 | 0 | 0 | 0 | 60 |

**Table 1. Confusion matrix**

The bar graphic in Fig. 9 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 a separate testing set. In every tested approach we used the entropy feature vectors, only in our approach we exploited the N-directional entropy. For WPD, the 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.
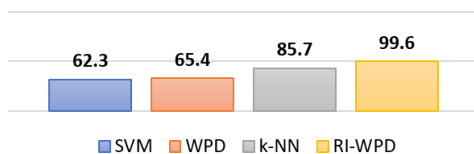
## Comparison of image classification techniques



**Fig. 9, Performance analysis of different techniques of image classification**

## 7    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 view-point variations of the analyzed pattern.

Each sub-band is analyzed to emphasize the distribution of entropy in $N$ directions. 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 structural information. 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 use a complementary feature such as N-directional energy.

## 8    References

[1]    X. Y. Zhang, C. L. Liu, and C. Y. Suen, "Towards Robust Pattern Recognition: A Review," *Proc. IEEE*, vol. 108, no. 6, pp. 894–922, Jun. 2020, doi: 10.48550/arxiv.2006.06976.

[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]    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.

[4]    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.

[5]    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.

[6]    S. Mallat, "A wavelet tour of signal processing: the sparse way 3rd Edition", Academic Press, p. 805, 2009.

[7]    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.

[8]    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.