Evaluation of Similarity Measures for Shift-Invariant Image Motif Discovery

Sahar Torkamani and Volker Lohweg
inIT – Institute Industrial IT
Ostwestfalen-Lippe University of Applied Sciences
Liebigstr. 87, D-32657 Lemgo, Germany
Email: {sahar.torkamani, volker.lohweg}@hs-owl.de

***Abstract*—The rapid growth of optical imaging technologies increased the access and collection of data, which boosts the demand of data and knowledge discovery. This is a fast growing topic in several industry and research areas. Nowadays, a large number of images and signals must be analysed in order to gain and learn proper knowledge. Detecting images with similar contents without specifying an image, recently attracts the researches in image processing domain. Motif discovery in image processing aims to tackle the problem of deriving structures or detecting regularities in image databases. Most of the motif discovery methods solve this problem by converting images into one dimensional time series in a pre-processing step and then applying a motif discovery on these one dimensional time series for image motifs detection. Nevertheless, this conversion might lead to information loss and also the problem of inability to discover shifted and multi-scale image motifs of different size. Contrary to other approaches, here, a method is proposed to find image motifs of different size in image data sets by employing images in original dimension (2D) without converting them to one dimensional time series. The proposed approach consists of three steps: Mapping or transformation, feature extraction and measuring similarities. First, images are inspected by the Complex Quad Tree Wavelet Packet transform, which provides broad frequency analysis of an image in various scales. Next, statistical features are extracted from the wavelet coefficients. Finally, image motifs are detected by measuring the similarity of the features applying various similarity measures. Here, the performance of six similarity measures are benchmarked in details. Moreover, the efficiency of the proposed method is demonstrated on a data set with images from diverse applications such as hand gesture, text recognition, leaf and plant identification, etc. Additionally, the robustness of this method is examined with the image data overlaying with distortions such as noise and blur.**

Keywords–Motif discovery; Image processing; Wavelet transformation.

I. INTRODUCTION

The accelerated growth of digital computation, telecommunication and imaging technologies results in a flood of information and data. These data are obtained in various forms such as text, graphics, pictures, videos or integrated multimedia. Such data are valuable if efficient information can be acquired from them. This issue is addressed by data mining and machine learning tasks. These tasks can be categorised into clustering, classification, anomaly detection and *motif discovery* [1].

Information such as number of clusters or classes, prototype patterns/images for each class or providing an image query to find, is necessary for such tasks [2]. The problems of clustering or classifying images as well as finding a query image in an image database are fairly known problems, which have been investigated during last decades [3]–[5]. The problem of deriving structures or detecting regularities in image databases is rather new topic and investigated by researchers [6]. This new topic is called motif discovery and aims to detect frequently repeated unknown images in a database without any prior information. The term motif has its roots in genetics and DNA sequences. A sequence motif in a DNA is a widespread amino-acid sequence pattern, which shows a biological significance [7]. In time series data mining, the term motif was first triggered by Patel et al. [8].

Motif discovery recently applied in image processing applications with various image databases. The aim of the image motif discovery is also to detect similar images and shapes within an image database without prior information. Such images are called *image motifs*. Fig. 1 aims to enhance the role of image motifs by given examples of some petroglyphs that are gathered in the USA [9]. The study of such petroglyphs is important for anthropologists, since these images show the spread of cultures and people. Therefore, detecting similar images that captured in different locations are in concerns for anthropologists. As depicted in Fig. 1, the images (a) and (c) captured in Capitol reef are similar to (b) and (d) that are obtained in Nine Mile Canyon [10]. Consequently, anthropologists are interested to discover such images (image motif) in a petroglyph image data set [9].

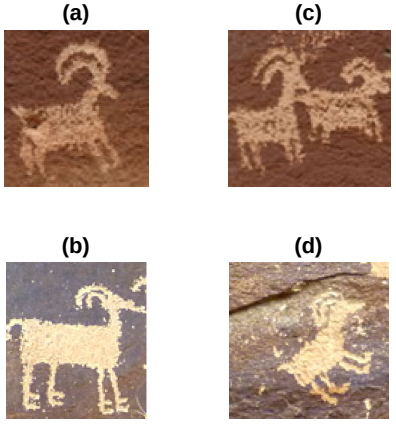

Figure 1. Examples of petroglyphs from Capitol reef and Nine Mile Canyon in Utah, USA [10]. Images (a) and (c) are from Capitol reef, and images (b) and (d) are captured in Nine Mile Canyon.

Detecting motifs add valuable insights about the problem under investigation to the user. Huge research effort has been performed on this topic [6], [11]. However, most of the image

motif discovery methods detect motifs by converting images into one-dimensional time series and then attempt to find motifs in such data by operating a motif discovery algorithm. This converting might lead to information loss and also the problem of inability to detect shifted and multi-scale motifs of different size [9]. Correspondingly, a method is proposed to find shifted and multi-scale motifs of different size in image data sets by applying the images in original dimension without converting them to one dimensional time series [1], [9].

This contribution is the extend version of the article published in [1]. Detailed information about the approach and comprehensive results are provided in this work. The proposed approach is benchmarked also with the distorted test cases in order to obtain the robustness of this method. This paper is structured as follows: the related work in motif discovery for image data type is described in Section II. Section III explains the proposed approach. The evaluation and the obtained results are illustrated in Section IV. At the end, a conclusion and the future work are indicated in Section V.

II. Related Work

Over the past decades, image and shape analysis have attracted several researchers and been a matter for discussion. Huge amount of research has been performed in several image processing tasks such as clustering, classification, query by content, segmentation, etc. [4], [12]–[15]. Recently, a new topic namely motif discovery in image and shape analysis is added to this research area. Motif discovery has evoked the interest in several researches, who aimed to link time series data mining tasks and issues to the image and shape analysis domain [6], [9], [16]. For instance, Barone et al. [17] studied the problem of classifying ordered sequences of digital images.

The first approach in image motif discovery is proposed by Xi et al. [9]. The authors detected image motifs in image data sets by representing an image or a shape in a one dimensional time series. This method extracts a time series from the contour of an image. The main problem of such an approach is that transforming a two dimensional data to a one dimensional might lead to information loss. Moreover, the image should be segmented in order to obtain the shapes in it.

The same procedure as in [9] is applied by Chi et al. [18] in order to detect image motifs in face image data sets. The term shapelet was introduced by Ye and Keogh [16]. Shapelets are a discriminative subsequence of the time series, which is considered instead of analysing the whole time series. Ye and Keogh [9] as well as Grabocka et al. [19] extended the proposed approach in [9]. After transforming an image to a one dimensional representation, shapelets are analysed to detect motifs. The performance of these methods is promising, but these approaches transform the data to a one dimensional time series. Caballero and Aranda [20] proposed an effective shape-based image retrieval system for leaf images. This contour descriptor reduces the number of points for the shape representation considerably.

Rakthanmanon and his colleagues [21] handled this problem by detecting motifs in images without representing them into a one dimensional signal. They, first, segmented the tested images using a sliding window of a fixed size, then the similarity between these segments are measured by the generalised Hough transform [22]. The fixed size of the sliding window is one of the disadvantages of this method. Since, a fixed size sliding window results in inability of detecting motifs with various proportions. En et al. [23] followed a similar approach, nevertheless they employed sliding windows with varying sizes of 20, 40, 80, and 160 pixels.

In our first approach [24], motifs in an image data base are discovered in their original dimension without converting them to time series. Images are decomposed into several frequency scales by the dual tree complex wavelet transform (DTCWT) [25], next features are extracted from the wavelet coefficients and finally motif images are found by measuring the similarity of their features. However, further experiments showed that the DTCWT is shift tolerance and not shift invariant [26]. For this reason, in this work, an approach is proposed, which is based on a shift-invariant feature extraction method for motif discovery (SIMD), given in [26]. This method is applied as core in our approach and explained in the following section. Additionally, this contribution is an extended version of the paper presented in [1] with comprehensive experiments.

III. Proposed Approach

The proposed motif discovery algorithm combines two research areas: pattern recognition and motif discovery. Motif discovery algorithms mainly consist of a representation and a similarity measure step. In this contribution, feature extraction step, which mostly applies in pattern recognition tasks, is added to the procedure of the approach depicted in Fig. 2.

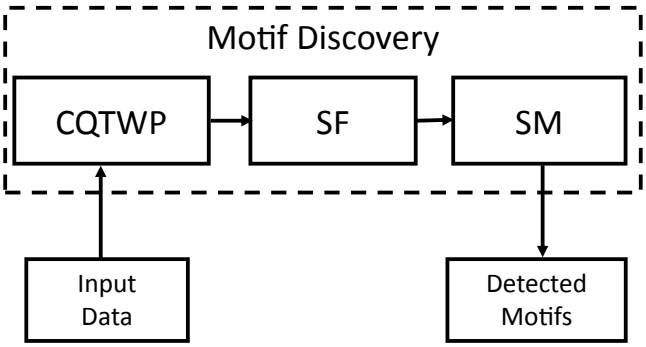

Figure 2. The proposed approach; CQTWP is the Complex Quad Tree Wavelet Packet; SF is the statistical features and SM represents similarity measures.

First, images are transformed by the *Complex Quad Tree Wavelet Packet* (CQTWP) into a broad frequency scales. Wavelets have several properties such as: ability to analyse data into different frequency scales, flexible time-frequency resolution and prefect reconstruction. Wavelet transformations proved their performance in signal and image processing applications [27]–[29]. In the second step, features are extracted from the normalised wavelet coefficients. At last, motifs are discovered by measuring the similarity between features using various distance measures. Before explaining these steps in details, some notations and useful definitions used in this paper are described in the following.

A. Definitions and Notations

Definition 1 (Image). *A digital image $\boldsymbol{X}_{m,n}$ is represented in a 2D discrete space as a $m \times n,\ m,n \in \mathbb{N}$ matrix:*

$$\boldsymbol{X}_{m,n} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1} & x_{m,2} & \cdots & x_{m,n} \end{pmatrix}.$$

Images can vary in their size and the applications they are captured from.

Definition 2 (Image Motif). *An image motif in an image data base is a pair of images $(\boldsymbol{X}_{m,n}, \boldsymbol{Y}_{p,q})$, where $m,p \in \mathbb{N}$ are the number of rows and $n,q \in \mathbb{N}$ are the number of columns, so that $distance(\boldsymbol{X},\boldsymbol{Y})$ is the smallest among all possible pairs* [9].

Function $distance(\mathbf{X},\mathbf{Y})$ is a distance similarity measure.

Definition 3 (1st-Image Motifs). *Given an image data base $D = \boldsymbol{X}^i, i = 1,2,...,N,\ N \in \mathbb{N}$, the most significant image motif in D is the image $\boldsymbol{X}^j$ that has the highest amount of matches. This image motif is called the 1st-Image Motif.*

Definition 4 (K-Image Motifs). *The K-th most significant image motif in D is the image $\boldsymbol{X}^k$ with the k^{th} highest amount of image matches.*

B. Complex Quad Tree Wavelet Packet Transform

1) 1D-CQTWP: The CQTWP is proposed to overcome the drawbacks of the DTCWT. It is an extended version of the DTCWT [25] and it consists of two wavelet packet trees (WPT) working parallel to each other. "WPT A" represents the real part and "WPT B" provides the imaginary part of the signal. A graphical representation of the "*1D-WPT A*" is given in Fig. 3, where $\downarrow 2^e$ and $\downarrow 2^o$ depict the even and odd down-sampling. The low and high-pass filters are denoted by ${}^s g_a$ and ${}^s h_a$, for $s \in \mathbb{N}$. Parameter s represents the scale of the decomposition. The wavelet coefficients are given by ${}^s c_i$ for $i \in [0, 4^s]$.

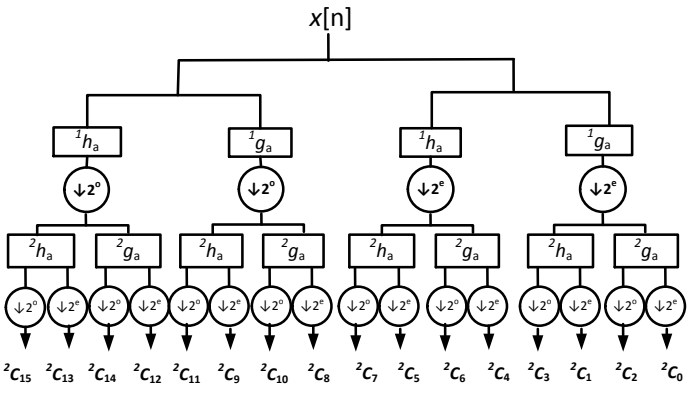

Figure 3. First wavelet packet filter bank of a two scale CQTWP. The second wavelet packet is obtained by replacing the filters ${}^1 g_a$ and ${}^1 h_a$ with ${}^1 g_b$ and ${}^1 h_b$ for the first scale and ${}^2 g_a$ and ${}^2 h_a$ with ${}^2 g_b$ and ${}^2 h_b$ for the second scale.

The low and high-pass filters applied in CQTWP are similar to the filters of DTCWT. The filters of the DTCWT satisfy the conditions required for having an analytic and complex wavelet transforms [25]. An analytic representation of a signal is achieved if and only if the filters of the CQTWP form a Hilbert pair [25], [26].

Definition 5. *Wavelets ψ_a and ψ_b with the following property*

$$\Psi_a(j\omega) = \begin{cases} -j\Psi_b(j\omega), & \omega > 0, \\ j\Psi_b(j\omega), & \omega < 0, \end{cases}$$

are called the Hilbert pair, *where $\Psi(j\omega)$ is the Fourier transform of $\psi(t)$.*

Consequently, the response of each branch of the "WPT A" and the corresponding branch of the "WPT B" forms a Hilbert pair and therefore, the CQTWP is approximately analytic in each sub band. Besides obtaining complex wavelet coefficients, the analytic representation has advantages such as reduction of aliasing.

To accomplish wavelets with Hilbert form, they must be designed by the following theorem:

Theorem 1 (Half-sample delay [30]). *Wavelets ψ_a and ψ_b form a Hilbert pair, if the filters ${}^s g_a$ and ${}^s g_b$ satisfy the condition,*

$${}^s G_a(e^{j\omega}) = {}^s G_b(e^{j\omega}) e^{-j\frac{\omega}{2}}. \tag{1}$$

Eq. (1) can be presented in terms of the magnitude and phase functions:

$$|{}^s G_a(e^{j\omega})| = |{}^s G_b(e^{j\omega})|, \quad \angle {}^s G_a(e^{j\omega}) = \angle {}^s G_b(e^{j\omega}) - \frac{1}{2}\omega, \tag{2}$$

which is the so-called "half-sample delay" *condition between two low-pass filters ${}^s g_a,\ {}^s g_b$.*

Proof. Proof is represented by Selsnick in [30]. □

Based on the half-sample delay theorem, the scaling low-pass filters must be offset from one another by a half sample. This is the necessary and sufficient condition for two wavelets to form a Hilbert transform pair, proved by Yu and Ozkaramanli [31].

Definition 6 (q-shift filters [32]). *Kingsbury's solution for design such suitable filters is called* "q-shift"*, which satisfies the "half-sample delay" condition given in Theorem* 1*, where the low-pass filters are set as*

$${}^s g_a[n] = {}^s g_b[M-1-n]. \tag{3}$$

Here, $M \in \mathbb{N}^+$ is the even length of filter ${}^s g_b$, which is supported on $0 \le n \le M-1$.

In order to achieve the half-sample delay theorem, at each scale the filters of WPT A translated by 2^s must be fall midway between the translated filters of WPT B. However, this condition leads to have filters in the first scale that have one sample delay difference. All the filters are real, orthonormal and are obtained by the design given by Abdelnour [33] and Kingsbury [32]. In the first scale, the filters have the even-length of 10 [33] and in the scale greater than one, filters have the even-length of 14 [32].

The wavelet and scaling functions of the CQTWP are defined as:

Definition 7. *Let* $\psi_{a,2J+1}(t)$,$\psi_{a,2J+3}(t)$, $\psi_{b,2J+1}(t)$, $\psi_{b,2J+3}(t)$ *and* $\phi_{a,2J}(t)$,$\phi_{a,2J+2}(t)$, $\phi_{b,2J}(t)$, $\phi_{b,2J+2}(t)$ *be the wavelet and scaling functions of the CQTWP. The wavelet and scaling functions in "WPT A",* $\forall n \in \mathbb{N}$ *are given by*

$$
\begin{aligned}
{}^{s+1}\psi_{a,2J+1}(t) &= \sqrt{2}\sum_{n=0}^{M} {}^{s}h_a[n]\ {}^{s}\phi_{a,2J}(2t-n),\\
{}^{s+1}\psi_{a,2J+3}(t) &= \sqrt{2}\sum_{n=0}^{M} {}^{s}h_a[n]\ {}^{s}\phi_{a,2J+2}(2t-n+1),\\
{}^{s+1}\phi_{a,2J}(t) &= \sqrt{2}\sum_{n=0}^{M} {}^{s}g_a[n]\ {}^{s}\phi_{a,2J}(2t-n),\\
{}^{s+1}\phi_{a,2J+2}(t) &= \sqrt{2}\sum_{n=0}^{M} {}^{s}g_a[n]\ {}^{s}\phi_{a,2J+2}(2t-n+1).
\end{aligned}
$$

Parameter $J = 2j$ *where* $0 \leq j < 2^s \cdot (s-1)$*, and* $s \in \mathbb{N}$ *is number of scales, and* $M \in \mathbb{N}^+$ *is the length of the filters.*

For "WPT B" the wavelet and scaling functions are defined in the same manner, but the high-pass filter ${}^{s}h_a$ *and the low-pass filter* ${}^{s}g_a$ *are replaced by* ${}^{s}h_b$ *and* ${}^{s}g_b$ *respectively. All filters are causal so* ${}^{s}h_{a,b}[n] = 0$ *and* ${}^{s}g_{a,b}[n] = 0$ *for* $n < 0$*.*

The wavelet and scaling coefficients of the CQTWP for the "*WPT A*" are defined in Def. 8.

Definition 8. *Coefficients of the CQTWP for the "WPT A" are given by* ${}^{s}C[n] = \{\ {}^{s+1}C_{2J}[n],\ {}^{s+1}C_{2J+1}[n],\ {}^{s+1}C_{2J+2}[n],\ {}^{s+1}C_{2J+3}[n]\}$ *and obtained by*

$$
\begin{aligned}
{}^{s+1}C_{2J}[n] &= \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C_j[2n-k],\\
{}^{s+1}C_{2J+1}[n] &= \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C_j[2n-k],\\
{}^{s+1}C_{2J+2}[n] &= \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C_j[2n+1-k],\\
{}^{s+1}C_{2J+3}[n] &= \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C_j[2n+1-k].
\end{aligned}
\tag{4}
$$

where $Len = length({}^{s}C_j)$*,* $J = 2j$*, and* $0 \leq j < 2^s \cdot (s-1)$*. Similarly, the wavelet and scaling coefficients of the "WPT B" are obtained by replacing the high and low-pass filters* ${}^{s}h_a$ *and* ${}^{s}g_a$ *to* ${}^{s}h_b$ *and* ${}^{s}g_b$*. These coefficients are depicted by* ${}^{s}D[n] = \{\ {}^{s+1}D_{2J}[n],\ {}^{s+1}D_{2J+1}[n],\ {}^{s+1}D_{2J+2}[n],\ {}^{s+1}D_{2J+3}[n]\}$.

Beside comprehensive frequency analysis, the CQTWP has another advantage of being shift-invariant [26].

Definition 9 (shift-invariant)**.** *The* shift-invariant *is defined by studying the wavelet coefficients of every scale* $s \in \mathbb{N}$ *from both the original and translated signal. This means if* $x[n]$ *and* $x[n-S]$ *are respectively the original and translated signal shifted by* $S \in \mathbb{Z}$*, then the corresponding wavelet coefficients are given by* ${}^{s}C[n]$ *and* ${}^{s}C[n,S]$*. Let the wavelet transformation be presented by* $x[n] \mapsto {}^{s}C[n]$*, then this transformation must satisfy* $x[n-S] \mapsto {}^{s}C[n,S]$*, where* ${}^{s}C[n,S] = {}^{s}C[n-S]$.

The shift-invariant property is obtained by decomposing a non-shifted and a shifted version of the input signal in each scale. Thus, the wavelet and scaling functions of the CQTWP select both even and odd samples of the signal in order to detect the occurred shift. The results of this property is identical wavelet coefficients for both the original signal and its shifted versions. The shift invariance property is proved by the following corollary:

Corollary 2. *Assume* $x[n]$ *is a discrete signal and let* $S_{e/o} \in \mathbb{Z}$ *be shifts occurred on signal* $x[n]$*, where* $S_{e/o}$ *can be even or odd. The CQTWP wavelet coefficients of* $x[n-S_e]$ *and* $x[n-S_o]$ *from "WPT A" in scales s, are depicted by* ${}^{s}C'_e[n,S_e]$ *and* ${}^{s}C'_o[n,S_o]$*, given by:*

$$
\begin{aligned}
{}^{s}C'_{e/o}[n,S_{e/o}] = \{\ &{}^{s+1}C'_{2J}[n,S_{e/o}],\ {}^{s+1}C'_{2J+1}[n,S_{e/o}],\\
&{}^{s+1}C'_{2J+2}[n,S_{e/o}],{}^{s+1}\,C'_{2J+3}[n,S_{e/o}]\},
\end{aligned}
$$

and for "WPT B" are provided by

$$
\begin{aligned}
{}^{s}D'_{e/o}[n,S_{e/o}] = \{\ &{}^{s+1}D'_{2J}[n,S_{e/o}],\ {}^{s+1}D'_{2J+1}[n,S_{e/o}],\\
&{}^{s+1}D'_{2J+2}[n,S_{e/o}],{}^{s+1}\,D'_{2J+3}[n,S_{e/o}]\},
\end{aligned}
$$

Then, the following equations hold

$$
\begin{cases}
\forall\ x[n-S_e], & \begin{cases} {}^{s}C[n] = {}^{s}C'_e[n-\lfloor\frac{S_e}{2^s}\rfloor],\\ {}^{s}D[n] = {}^{s}D'_e[n-\lfloor\frac{S_e}{2^s}\rfloor]. \end{cases}\\[2ex]
\forall\ x[n-S_o], & \begin{cases} {}^{s}C[n] = {}^{s}C'_o[n-\lfloor\frac{S_o}{2^s}\rfloor],\\ {}^{s}D[n] = {}^{s}D'_o[n-\lfloor\frac{S_e}{2^s}\rfloor]. \end{cases}
\end{cases}
\tag{5}
$$

Proof. Proof is given in Appendix A. □

For simplicity, the odd and even wavelet and scaling functions of "WPT A" are denoted by $\psi_{a,e}(t) = \psi_{a,2J+1}(t)$ and $\psi_{a,o} = \psi_{a,2J+3}(t)$; and $\phi_{a,e} = \phi_{a,2J}(t)$, $\phi_{a,o} = \phi_{a,2J+2}(t)$. The functions of "WPT B" are represented in the same manner.

2) 2D-CQTWP: It is able to expand the CQTWP to a higher dimension. The 2D-CQTWP analyses an image into various frequency bands. The structure of two scales decomposition of the "*2D-WPT A*" is depicted in Fig. 4(b), where both low and high-pass filtered sub bands decomposed further. This property results in a more flexible and broad frequency decomposition of the images.

The first scale of the 2D-CQTWP is similar to the 2D-discrete wavelet transform [34], where an image is decomposed into four sub bands namely LL_1, LH_1, HL_1 and HH_1, cf., Fig. 4(a). However, in the first scale, the 2D-CQTWP has two LL, two LH, two HL and two HH sub bands obtained from both "*2D-WPT A*" and "*2D-WPT B*".

The product of the low-pass function $\phi_a()$ along the first dimension (row) and the low-pass function $\phi_a()$ along the second dimension (column) results in LL_1. LH_1 is the product of the low-pass function $\phi_a()$ along the first dimension and the high-pass function $\psi_a()$ along the second dimension. Similarly, the HL_1 and HH_1 are labelled, and the index 1 determines the decomposed scale. The same procedure is performed on each sub band in order to obtain the second scale coefficients.

The wavelet and scaling functions of the 2D-CQTWP are defined as:

Definition 10. *The "2D-WPT A"of the 2D-CQTWP is characterised by twelve wavelets and four scaling functions.*

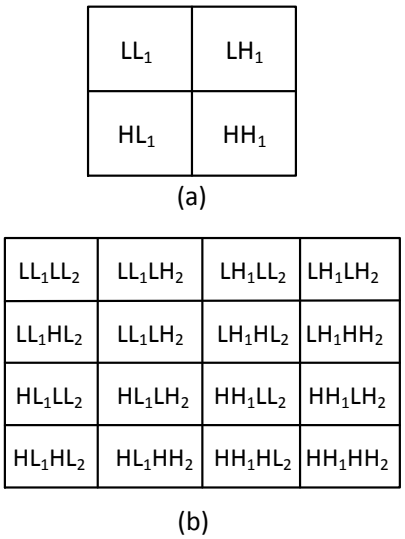

Figure 4. Structure of two scales decomposition of the "*2D-WPT A*": (a) the first scale decomposition, (b) the second scale decomposition.

The 2D-wavelet $\psi(x,y) = \psi(x)\psi(y)$ is associated with the row-column implementation of the wavelet transform. The wavelet functions for the wavelet packet tree A are given by

$$\psi_{a,1}(x,y) = \phi_{a,e}(x)\psi_{a,e}(y), \qquad \psi_{a,4}(x,y) = \phi_{a,e}(x)\psi_{a,o}(y),$$
$$\psi_{a,2}(x,y) = \psi_{a,e}(x)\phi_{a,e}(y), \qquad \psi_{a,5}(x,y) = \psi_{a,e}(x)\phi_{a,o}(y),$$
$$\psi_{a,3}(x,y) = \psi_{a,e}(x)\psi_{a,e}(y), \qquad \psi_{a,6}(x,y) = \psi_{a,e}(x)\psi_{a,o}(y).$$

The rest of the wavelet functions are obtained similarly. The scaling functions are defined as

$$\phi_{a,1}(x,y) = \phi_{a,e}(x)\phi_{a,e}(y), \qquad \phi_{a,2}(x,y) = \phi_{a,e}(x)\phi_{a,o}(y),$$
$$\phi_{a,3}(x,y) = \phi_{a,o}(x)\phi_{a,e}(y), \qquad \phi_{a,4}(x,y) = \phi_{a,o}(x)\phi_{a,o}(y).$$

The wavelet and scaling functions of the "2D-WPT B" are given accordingly.

The wavelet and scaling coefficients of the 2D-CQTWP for the "2D-WPT A" are given by

Definition 11. *Coefficients of the 2D-CQTWP for the "2D-WPT A" are given by ${}^sC[x,y] = \{ {}^{s+1}C_{2J}[x,y], {}^{s+1}C_{2J+1}[x,y], ..., {}^{s+1}C_{2J+11}[x,y]\}$ and obtained by*

$$\begin{aligned}
{}^{s+1}C_{2J}[x,y] &= {}^sC_j[x,y] * {}^sg_a[2x]\ {}^sg_a[2y],\\
{}^{s+1}C_{2J+1}[x,y] &= {}^sC_j[x,y] * {}^sg_a[2x+1]\ {}^sg_a[2y],\\
&\vdots\\
{}^{s+1}C_{2J+5}[x,y] &= {}^sC_j[x,y] * {}^sg_a[2x]\ {}^sh_a[2y],\\
{}^{s+1}C_{2J+6}[x,y] &= {}^sC_j[x,y] * {}^sg_a[2x]\ {}^sh_a[2y+1],\\
&\vdots\\
{}^{s+1}C_{2J+14}[x,y] &= {}^sC_j[x,y] * {}^sh_a[2x]\ {}^sh_a[2y],\\
{}^{s+1}C_{2J+15}[x,y] &= {}^sC_j[x,y] * {}^sg_a[2x+1]\ {}^sg_a[2y+1].
\end{aligned}$$

Parameter $s \in \mathbb{N}$ is the number of scales and parameter $J = 8j$ is the index of the coefficient nodes, whereby for $s = 1$, $j = 0$ and for $s > 1$, $0 \leq j < 4^s$. The wavelet coefficients for the "2D-WPT B" are computed similarly and denoted by ${}^sD[x,y] = \{ {}^{s+1}D_{2J}[x,y], {}^{s+1}D_{2J+1}[x,y], ..., {}^{s+1}D_{2J+15}[x,y]\}$.

The 2D-CQTWP has the same properties of the one dimensional CQTWP.

3) Selection of the Best Nodes: Decomposing the data in each scale leads to the number of nodes which grows exponentially in each wavelet packet tree. Therefore, selecting the nodes with the most information content reduces the amount of redundant and unnecessary information. Every node of the wavelet packet tree A and B has a potential to be chosen as a proper node, which provides meaningful information for feature extraction. In order to select the best nodes, a method is applied, which is based on the algorithm introduced in [35] for the discrete wavelet packet and its concept is established by an additive cost function.

Definition 12 (Cost function [35])**.** *A cost function CF that maps the sequences $\{x_i\}_{i=1}^N$ to real numbers considers as additive, if $CF(\{x_i\}) = \sum_{i=1}^N g(x_i)$ for some $g : \mathbb{R} \to \mathbb{R}$ and for all $\{x_i\}_{i=1}^N$.*

An entropy-based cost function is considered here.

Definition 13 (Entropy-based cost function)**.** *The entropy-based cost function for the wavelet packet "WPT A" is denoted by $W_H({}^sC[n])$ and obtained by*

$$W_H({}^sC[n]) = -\sum_{n=1}^{N} E_c[n]\log(E_c[n]),$$

where $({}^sC[n])$ is the wavelet coefficients defined in Def. 8, and the normalized energy is given by $E_c[n] = \frac{({}^sC[n])^2}{\sum_n ({}^sC[n])^2}$. The entropy-based cost function for the wavelet packet "WPT B" is obtained by replacing ${}^sC[n]$ with ${}^sD[n]$.

The normalized energy of the wavelet coefficients applied in the above definition allows to adjust and compare coefficients from different scales. The algorithm for detection the best node, Algorithm 1, has the following steps:

Algorithm 1 Best Nodes Selection
Input: Entropy-based cost function $W_H({}^sC[n])$
Output: Best nodes

```
for s = s − 1 : 1 do
    for j = 0 : 2^s do
        J = 2j
        if W_H(^sC_j[n]) < W_H(^{s+1}C_{2J}[n]) + ... + W_H(^{s+1}C_{2J+7}[n]) then
            ^sBN_j = (^sC_j[n]) is selected as best node.
        else
            W_H(^sC_j[n]) = W_H(^{s+1}C_{2J}[n]) + ... + W_H(^{s+1}C_{2J+7}[n])
        end if
    end for
end for
```

The best nodes selection algorithm computes the entropy-based cost function for each coefficients node upwards from the scale s to the first scale. The same approach applies for the 2D-CQTWP transform.

C. Feature Extraction

Feature extraction plays an important role in pattern recognition applications, and it helps to reduce the size of the data. Since, features present the special characters of the data, it is important that they are detectable under changes in proportion, location or even under noise circumstances. A proper feature extraction method must be able to generalise over differences within a class (intra-class) and determine the variations between various classes (inter-class).

In the second step, from wavelet coefficients features must be extracted. But before extracting features, it is necessary to normalise the coefficients of each scale. The normalisation is performed because the proposed method is able to analyse images of various size, therefore the wavelet coefficients have also different size. Thus, normalisation allows to rescale all the coefficients in order to compare them. The normalised histogram of the wavelet coefficients is denoted by $H(p)$ and is given by

$$H(p) = \frac{1}{v \cdot u} \cdot h(p),$$

where $u, v \in \mathbb{N}$ determine the size of the matrix coefficients and parameter p is number of the histogram bins. The rate in each bin is presented by $h(p)$.

The first four statistical moments [36], namely, mean value, variance, skewness and kurtosis are extracted from the wavelet coefficients in both wavelet packet trees. As 2D-CQTWP is shift-invariant, then these features have identical values even in the case of shift occurrence in the data.

Additionally, the energy of the wavelet coefficients is considered as another feature. Since the CQTWP is shift-invariant, the energy of the wavelet coefficients and their shifted ones are similar. Moreover, according to the Parseval's theorem the energy of the signal or image is preserved in the coefficients and as described in Section III-B1, the scaling and wavelet functions of the CQTWP are orthonormal, which satisfy the Parseval's theorem.

D. Similarity Measures

In order to detect image motifs, the similarity between their features must be measured. In general, similarity measures can be divided into four groups: shape-based, edit-based, model-based and feature-based methods [6].

Shape-based distance similarity measures compare the total shape of the signals or images. Members of the Minkowski distance family [37], and Dynamic Time Warping (DTW) [37] belong to this group of measures. Here, the two members of the Minkowski distance or L_p-distance namely, Euclidean distance (ED) and Canberra distance (CD) are applied. Both of these measures have linear computational time complexity $O(n)$, and are metric. The Euclidean distance is obtained by setting $p = 2$ in L_p-distance. This measure is also known as L_2-distance. Besides the advantages of the Euclidean distance, results of this similarity measure are not promising, when performing directly on the data, in the case of outliers. The Canberra distance is actually a weighted version of Manhattan distance or L_1-distance, and is useful in the case of ranking lists or results. DTW matches various sections of a signal by warping of the time axis, or finding the proper alignment. This similarity measure is more flexible than Euclidean or Canberra distance although its time-complexity is $O(n^2)$. Apart from its quadratic computational time complexity, still DTW is one of the most popular approaches for measuring similarity/dissimilarity.

Edit-based similarity measures compare two signals according to the minimum number of operations needed to transform one signal or feature vector into another one. Such operations are insertion, deletion, and substitution. These methods are also known under Levebshtein distances [38]. Examples of these similarity measures are Edit Distance [38], and the Longest Common SubSequence(LCSS) [39]. The Edit distance method is usually applied on the string data sets. This can be seen as one disadvantage for this method. If $s_1 =$'Hello' and $s_2 =$'Have' are two strings, then the Editdistance$(s_1, s_2) = 4$. Since 4 operations must be done: replace(e,a), replace(l,v), replace(l,e) and delete(o).

LCSS aims to detect the characteristic segment between two time series by looping over all possible Edit distances.

Definition 14 (Longest Common SubSequence)**.** *The LCSS of two time series* $x[n] = (x_1, x_2, ..., x_N)^T$ *and* $y[n] = (y_1, y_2, ..., y_M)^T$ *of lengths* $N, M \in \mathbb{N}$ *is denoted by* $LCSS(x, y)$ *and computed by* [39]

$$LCSS = \begin{cases} 0 & \text{if } N = 0 \text{ or } M = 0, \\ LCSS(rest(x), rest(y)) + 1 & \text{if } dist(x_1, y_1) \leq \epsilon, \\ \max(LCSS(rest(x), y), LCSS(x, rest(y))) & \text{else}, \end{cases}$$

where the threshold $0 < \epsilon < 1$ *should be defined in advance, in order to show if two elements match. The* $dist()$ *function is defined by* $dist(x_1, y_1) = |x_1 - y_1|$ *and* $rest(x)$ *defines the remaining sequence of* x*.*

The main problem of the LCSS is being sensitive to noise. Similar to the Euclidean and Canberra distance, the time-complexity of the Edit distance is $O(n)$. LCSS for $n \in \mathbb{N}$ number of time series or sequences performs in $O(2^n)$.

Typically model-based methods use prior knowledge about the model that generated the data sets. These methods compute the similarity between data sets by first modelling one data set and then examine the likelihood that other data sets are also generated by the same model. Methods such as Hidden Markov Models (HMM) [40] and Autoregressive Moving Average model (ARMA) [41] belong to this group. Since these methods need prior knowledge about the data, they are not applied in this work.

Feature-based methods measure the similarity between different data sets based on the obtained sets of features. In these methods, first features are derived from the data and then distance measures are applied to capture patterns. Likelihood ratio [42] is a measure belongs to the feature-based methods.

Definition 15 (Likelihood ratio LR)**.** *Given the two time series* $x[n] = (x_1, x_2, ..., x_N)^T$ *and* $y[n] = (y_1, y_2, ..., y_N)^T$ *with periodograms* a_i *and* b_i *respectively, the likelihood ratio between them is determined by* [42]

$$LR(X(\omega), Y(\omega)) = 4 \sum_{i=1}^{k} \{2 \log(a_i + b_i) - \log a_i - \log b_i\},$$

where $X(\omega)$ *and* $Y(\omega)$ *are the DFT of the time series* $x[n]$ *and* $y[n]$*. Periodogram* a_i *is obtained by* $a_i = p_i^2 + q_i^2$*, where* (p_i, q_i) *are Fourier coefficients of the time series* $x[n]$*.*

The class of shape-based similarity measures usually is considered as another candidate for feature-based similarity measures. The edit-based measures can be also utilized as feature-based similarity measures. Since, feature extraction belongs to the process of our approach, the performance of feature-based similarity measures are tested in this work.

IV. Experiments and Results

In this section, the results of the proposed method are described. All the tests are executed on Windows 10 with a AMD Ryzen 5 1600 core processor and 16GB RAM. The codes are performed by MATLAB R2017a [43].

A test case image data base with different images is considered, which is sent as input data to the proposed approach. The experiments followed the procedure given in Fig. 2. To evaluate the performance of the proposed approach, different validation principles are performed, given in Section IV-A. After that, the captured results of image motif discovery are presented in Section IV-C. The experiments are executed in two parts: the first part is performed on the test case images without any added distortions. In the second part of experiments, two types of noise are added to the data. Finally, the test images are distorted by effects such as blurring.

A. Validation Principles

As described in Section III-C, different features are extracted from the normalised histogram of the wavelet coefficients. The quality of the selected features is measured by the linear discriminant analysis (LDA) algorithm [2], [44].

There are various validation methods to analyse the performance of a method. Here, the outcome of the investigations is benchmarked by the quality measures explained in the following section.

1) Linear Discriminant Analysis: Linear Discriminant Analysis (LDA) is a supervised method, which projects the features from the samples of the two or more classes onto a lower dimensional space with good class separability in order to avoid over-fitting and computational costs reduction. This method projects a data set into a lower dimensional space with good class separability.

Given samples from two motif groups, C_1 and C_2, LDA's aim is to find the direction $W = (w_1, w_2, ..., w_N)$ such that when the data are projected onto W, the motif examples of each group are as perfectly separated as possible:

$$F_{prj} = W^T F,$$

where $F = (f_1, f_2, ..., f_N)$ is the vector of the objects and F_{prj} is a scalar that samples in F are projected onto.

To be able to obtain a good projection vector, a measure for separation between the projections must be defined. The arithmetic mean value of the vector F and its projected one F_{prj} are given by [2]

$$\mu = \frac{1}{N}\sum_{i=1}^{N} f_i, \qquad \widetilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} W^T f_i.$$

One possibility is to consider the distance between the projected means of each motif group, but this option is not a proper measure since it does not consider the standard deviation within each motif group.

Fisher proposed a solution to maximise a function that represents the difference between the means, normalised by a measure of the within-class (group or cluster) scatter. For each motif group, the scatter is defined as [2]

$$\widetilde{\sigma_i} = \sum_{F_{prj} \in C_i} (F_{prj} - \widetilde{\mu_i})^2,$$

where parameter $i \in \mathbb{N}$ is the number of motif groups (here $i = 2$). The Fisher linear discriminant is determined by [2]

$$J(W) = \frac{|\widetilde{\mu_1} - \widetilde{\mu_2}|^2}{\widetilde{\sigma_1^2} + \widetilde{\sigma_2^2}}.$$

Thus, LDA searches for a projection where the motifs belonging to the same group are very close to each other, and the motifs of various groups are as farther apart as possible [2]. Therefore, to estimate the efficiency of the extracted features, the classification error by LDA is considered here. This error is denoted by e where $0 \leq e \leq 1$. The less the error, the better is merit of the features. If the data can be separated linearly and correctly, the error will be 0, and if the whole data cannot be classified linearly and correctly, then the error has its maximum amount of 1.

2) Quality Measures: An image motif which matches all the images in the target class and no other images out of that class, is considered as a perfect motif. To qualify a motif matching an image, four possibilities of the confusion matrix are available; namely, true positive rate (TP), false negative rate (FN), true negative rate (TN), and false positive rate (FP). Parameter (TP) represents a positive example that is also predicted positive. A positive example with a false prediction shows by (FP). (TN) depicts a negative example when the prediction is also negative. Finally, (FN) is a result of having a positive prediction for a negative example [2].

The results of the proposed algorithm are evaluated by the following quality measures [2]: Correct motif discovery rate CR, Sensitivity Sn, Precision Pr and F-Measure $F - M$.

Definition 16 (Correct motif discovery rate)**.** *This rate expresses the performance of the algorithm. It is given by*

$$CR = \frac{n^+}{N},$$

where $N \in \mathbb{N}$ is number of all motifs and n^+ is number of correctly detected motifs.

Definition 17 (Sensitivity)**.** *Sensitivity measures the capacity of images of the target class correctly matched by the motif. This measure is also denoted by recall.*

$$Sn = \frac{TP}{TP + FN},$$

where $Sn \in [0,\ 1]$ and the optimal case is $Sn = 1$.

Definition 18 (Precision)**.** *This measure provides the fraction of images of the target class that are matched by the motif and the images that are not correctly matched by the motif.*

$$Pr = \frac{TP}{TP + FP},$$

where $Pr \in [0,\ 1]$. In other words, Pr relates the number of correct detected motifs to all positive determined motifs with the optimal case of $Pr = 1$.

Definition 19 (F-Measure)**.** *F-Measure considers both precision and sensitivity and is determined by*

$$F - M = 2 \cdot (\frac{Pr \cdot Sn}{Pr + Sn}).$$

The best value for F-Measure is 1 and the worst is 0.

B. Test Case

The test image data base consists of images from diverse applications and domains like hand gesture, leaf identification [45], [46], and text and object recognition. Fig. 5 represents some images of four groups. All the images have various size and scale, to analyse the performance of the proposed method. Since, both images of fixed and variable size can be analysed in this work.

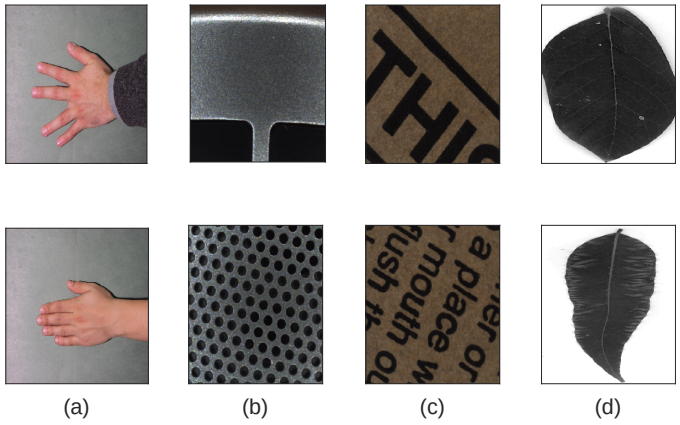

Figure 5. Data set of different images captured from various applications.

Top inserted image motifs or the most occurred images are the pictures of hands and leaves, which are depicted in Fig. 6. In order to demonstrate the shift-invariant property of the 2D-CQTWP in feature space, images such as given in Fig. 6 (a-d) are considered. These images are the shifted version of the image (a), and image (e) is the rotated version of image (a). Images (f-j) are different leaf types with various size and shape. The number of test images is increased from 280 to 2202 images. From these figures, 400 images are the inserted motif images.

C. Results and Evaluation

The proposed method starts with a pre-processing step, where all the images are converted in grey-scale, since the colour information is not required.

Next, all the images are sent to the main part of the method namely to the 2D-CQTWP transform. As explained, the 2D-CQTWP is able to decompose the images into various signals (up to $s = \log_2^{(m \times n)}$). In this work, the wavelet coefficients of the second scale are selected, since the amount of noise is usually reduced in the second scale for the noisy data. The best nodes with the highest information content are selected from these scales, according to the algorithm 1.

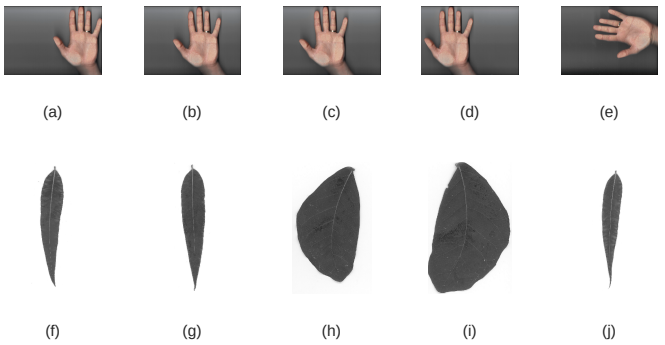

Figure 6. Images of hands and leaves. Images (b-d) are the shifted version of image (a); Image (e) is the rotated version of image (a). Images (f-j) are various sorts of leaves.

As the test case consists of images of various size, the wavelet coefficients have also different size. Therefore, the normalised histograms of the selected wavelet coefficients are calculated. Fig. 7 is the graphical representation of the normalised histogram of the HL sub band coefficients of "2D-WPT A" from the images depicted in Fig. 5.

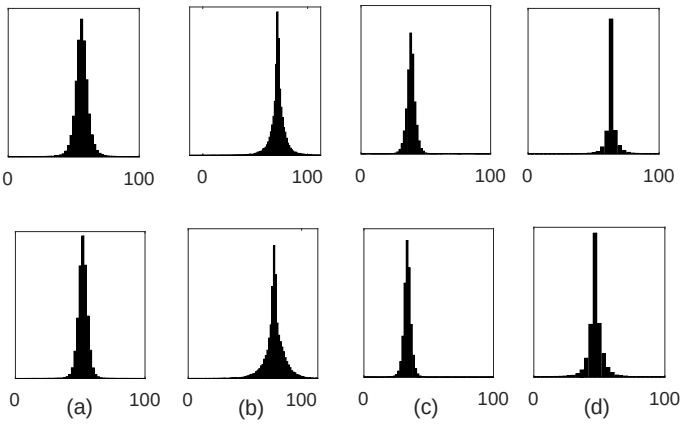

Figure 7. Normalised histogram of the HL sub band coefficients, obtained from the corresponding images from Fig. 5 (a-d).

According to Fig. 7, the histograms of wavelet coefficients from the two depicted images in each group (a), (b), (c) and (d) are similar to each other but different to the histograms of other groups. This helps to determine the variations between various motif classes (inter-class).

In order to represent the shift-invariant property of the 2D-CQTWP, the hand images in upper subfigures of Fig. 8 are considered. The position of the hand is shifted in these images. Based on the 2D-CQTWP transform the wavelet coefficients and therefore, the normalised histograms of these images must be identical to each other. As illustrated, the normalised histograms are all identical to each other, which shows the shift-invariant property of the 2D-CQTWP. The normalised histograms depicted in Fig. 8 are obtained from the HL sub band coefficients of the hand images in Fig. 6 (a)-(d).

After determining the normalised histograms from the wavelet coefficients, the five stated features are extracted from

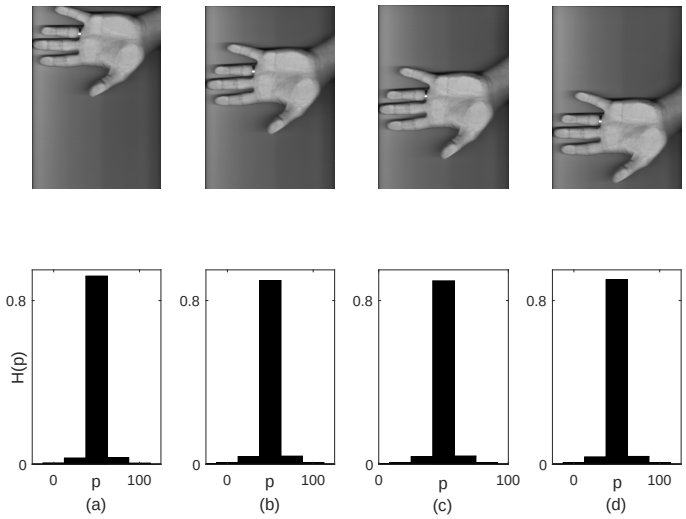

Figure 8. (a) a hand pattern; (b) shifted version of image in (a); (c) and (d) represent the normalised histograms of the HL sub band coefficients from images (a) and (b).

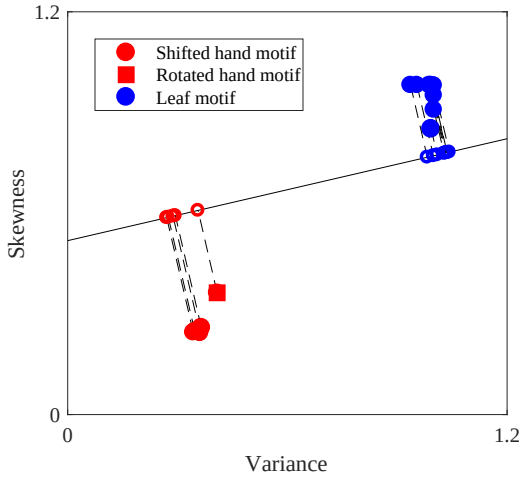

Figure 9. LDA projection of the two features from some of the hand and leaf image motifs; the distance between features within an image motif group is as minimum as possible and the distance between features of different image motif groups is large enough. Red circle markers represent the shifted images of the hand where the red square marker depicts the rotated image of the hand. Blue circle markers demonstrate the leaves images.

them. The efficiency of these features are investigated by the linear discriminant analysis (LDA) algorithm [2], [44]. The experiments show that for most of the tested features the minimum error is $0 \leq e \leq 0.01$. Furthermore, the distance between feature clusters is as great as possible, which facilitates the grouping.

The result of the LDA projection of the two extracted features (skewness and variance) from the image motifs in Fig. 6 is given in Fig. 9. As demonstrated, the distance between the two groups is large enough in order to separate them correctly. Moreover, the distance between features belonging to the same image motif group (represented on the projection line) is minimised.

The features of the first four hand images are as close as possible to each other and their projection on the projection line is at the same position. This also depicts a graphical representation of the shift-invariant property of the 2D-CQTWP transform. These images are depicted by the circle red marker. Nevertheless, the projection of the features extracted from the rotated image (illustrated by the square red marker) is not at the same position of other hand images. This illustrates that the 2D-CQTWP is nearly rotation invariant, but still we are able to detect this image motif and separate it from other image motifs.

Another example is presented in Fig. 10, where similar to the Fig. 9 the features (energy and kurtosis) of the first four hand images are as close as possible to each other and their projection on the projection line is at the same position. On the contrary to Fig. 9, in Fig. 10 the projection of the features extracted from the rotated image (illustrated by the square red marker) is closer to the position of other hand images.

In the last step, the similarity between feature values is measured by the Euclidean, Canberra and Edit distance, and also by the Dynamic Time Warping, the Longest Common SubSequence measures and the Likelihood ratio.

The results of these measures the are given Table I. The best performance is obtained by the Canberra distance and

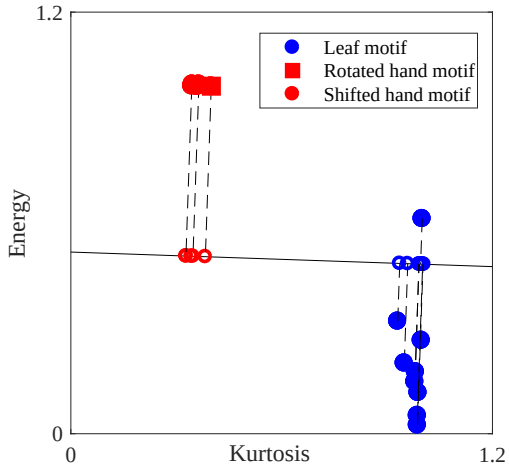

Figure 10. LDA projection of the kurtosis and energy features from some of the hand and leaf image motifs; the images of both groups can be easily separated. Red circle and square markers represent the shifted and rotated image of the hand. Blue circle markers demonstrate the leaves images.

the LCSS provided the inadequate results (under 0.5). The LR measure is applied only on the wavelet coefficients of the images with the fixed size, since this measure is unable to compare images of various size and also the periodogram of the extracted features does not provide any useful information.

From 400 image motifs, Euclidean distance and DTW are able to detect respectively 327 and 322 motif images. The Edit distance and LCSS distinguished 244 and 139 motif images. Number of 154 image motifs are identified by the LR measure. It should be noticed that for the LR measure only 250 inserted motifs are tested, as these images have the same size. The

TABLE I. Results of detected motifs considering the best selected wavelet nodes, CR: Correct motif discovery rate, F-M: F-Measure, Sn: Sensitivity, Pr: Precision; ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence, and LR: Likelihood ratio.

Similarity Measure	CR	Sn	Pr	F-M
ED	0.812	0.812	0.820	0.816
DTW	0.807	0.807	0.794	0.801
CD	0.927	0.927	0.913	0.920
Edit	0.617	0.617	0.601	0.609
LCSS	0.344	0.344	0.339	0.341
LR	0.625	0.625	0.615	0.627

maximum amount of 370 image motifs is detected by the CD similarity measure.

As mentioned, the most repeated image motif is the hand images. All the tested similarity measures are able to detect this image motif as the 1st-Image motif. The highest amount F-Measure is obtained by the Canberra and the Euclidean distance measures, which confirms the accuracy of these measures.

In the experiments given in the last contribution [1], the best motif discovery rate was achieved by the Euclidean distance (correct motif discovery = 0.861), and the Canberra and DTW measures performed in the same manner. Here, by increasing the size of the data set the performance of the Canberra distance improves since this measure involves some standardisation across the two observations being compared rather than simply adding the distance differences. The DTW achieves the lower results than the Euclidean distance by increasing the size of the data. This issue is also mentioned in [47], where the authors showed that by increasing the size of the data, the Euclidean distance outperforms the DTW measure. Moreover, by selecting the nodes with the best information content the performance of these similarity measures is increased.

In order to test the robustness of the proposed method, the same experiments are performed by adding noise to the test images. The test case is overlaid with two different types of noise namely Gaussian and Salt & Pepper [48], cf., Fig. 11. The Gaussian noise is the most occurring noise in images. It has the same discrete probability density function as the normal distribution.

$$p[X] = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(X-\mu)}{2\sigma^2}},$$

where $X(m,n)$ is the original image (grey-scaled) and μ and σ are the mean value and the standard deviation. Thus, the values of the noise are Gaussian-distributed.

The Salt & Pepper noise does not corrupt the whole image, instead some pixel values of the image are changed. The damaged image by Salt & Pepper looks like that several black and white dots scattered on the image. The Salt & Pepper noise can be simply modelled by

$$\begin{aligned} p[X = X_N] &= 1 - \alpha, \\ p[X_N = max] &= \alpha/2, \\ p[X_N = min] &= \alpha/2, \end{aligned}$$

where $X(m,n)$ is the original image and $X_N(m,n)$ is the image altered by the Salt & Pepper noise. The *min* and *max* are the minimum and maximum image values (for 8-bit images $min = 0$ and $max = 255$), and $0 \leq \alpha \leq 1$ is the probability that a pixel is corrupted. By the discrete probability density equals to $1-\alpha$, the pixels stay unchanged and with probability $\alpha/2$, the pixels are changed to the largest or smallest values [48]. The added Gaussian and Salt & Pepper noise to the images are respectively 20dB and 13dB.

The performance of the proposed motif discovery algorithm under the influence of noise and applying the above similarity measures is given in Tables II-III.

TABLE II. Detected motifs from test images overlaid with the Gaussian noise, CR: Correct motif discovery rate, F-M: F-Measure, Sn: Sensitivity, Pr: Precision; ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence, and LR: Likelihood ratio.

Similarity Measure	CR	Sn	Pr	F-M
ED	0.781	0.781	0.769	0.775
DTW	0.781	0.781	0.769	0.775
CD	0.835	0.835	0.822	0.829
Edit	0.601	0.601	0.592	0.596
LCSS	0.329	0.329	0.324	0.326
LR	0.590	0.590	0.584	0.586

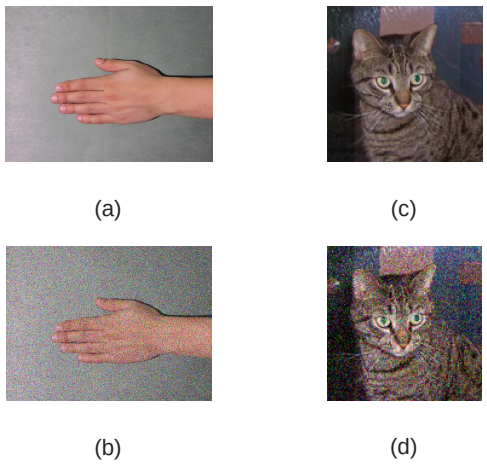

Figure 11. Example of the tested images overlaid with the Gaussian and Salt & Pepper noise. Images (a) and (c) are original images, and images (b) and (d) are images overlaid respectively with the Salt & Pepper and the Gaussian noise.

As stated, the performance of the LCSS is very poor under the noise circumstances, and it provides under 50% correct motif discovery rate. The best outcome is obtained by the Canberra distance in all three cases. The rest of the similarity measures provide the similar performance. As the CQTWP transform reduces the amount of noise, the performance of most of these similarity measures stays alike, but in general the correct motif discovery of the noisy test data is lower than the original test data (without noise). The Euclidean and Canberra distances and the DTW measure are more robust to noise than the Edit distance and LCSS measure. The outcomes of the LR measures is only obtained from images of the equal size.

TABLE III. Evaluating results of detected motifs under Salt & Pepper noise, CR: Correct motif discovery rate, F-M: F-Measure, Sn: Sensitivity, Pr: Precision; ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence, and LR: Likelihood ratio.

Similarity Measure	CR	Sn	Pr	F-M
ED	0.732	0.732	0.721	0.727
DTW	0.748	0.748	0.736	0.742
CD	0.832	0.832	0.820	0.826
Edit	0.565	0.565	0.557	0.551
LCSS	0.326	0.326	0.354	0.340
LR	0.527	0.527	0.519	0.523

Image blurring [49] is another distortion occurs by an optical system. Fig. 12 is a graphical representation of this effect. The same experiments are carried out on the test case with blurred images and the outcome is given in Table IV.

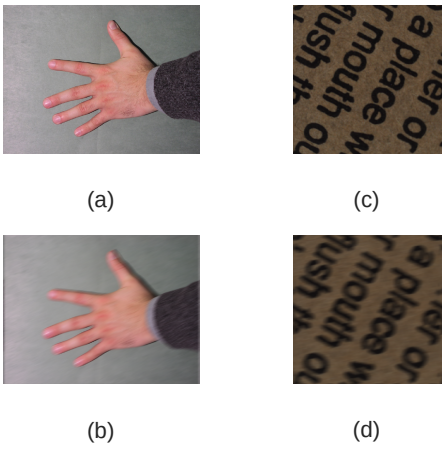

Figure 12. Image blurring effect. Subfigures (b) and (d) are the blurred images, figures (a) and (c) are the original images.

TABLE IV. Evaluating results of detected motifs under image blurring effect, CR: Correct motif discovery rate, F-M: F-Measure, Sn: Sensitivity, Pr: Precision; ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence, and LR: Likelihood ratio.

Similarity Measure	CR	Sn	Pr	F-M
ED	0.794	0.794	0.792	0.793
DTW	0.807	0.807	0.794	0.801
CD	0.786	0.786	0.774	0.780
Edit	0.421	0.421	0.415	0.418
LCSS	0.331	0.331	0.326	0.329
LR	0.606	0.606	0.619	0.613

The highest correct motif discovery rate is obtained by the DTW measure (CR=0.807). The Euclidean distance outperforms the Canberra distance, since the Canberra distance is very sensitive to the values close to zero.

The performance time for each of these similarity measures that took to detect image motifs is given in Table V and Fig. 13, where the number of the image motifs is increased up to 2000 images.

TABLE V. Evaluating the performance time took by the applied similarity measures ;ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence,, and LR: Likelihood Ratio.

Measures	ED	DTW	CD	Edit	LCSS	LR
Run-Time(s)	0.09	3.28	0.05	10.28	29.15	3.94

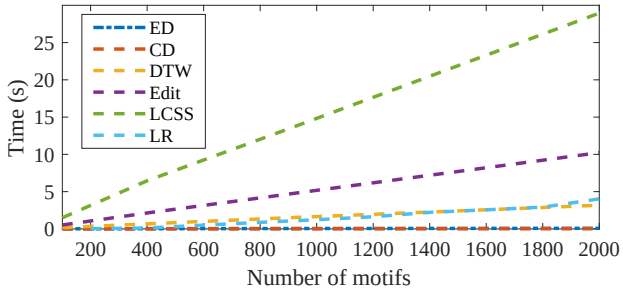

Figure 13. Performance time of the proposed method applying different similarity measures while increasing the size of the data. ED: Euclidean distance, DTW: Dynamic Time Warping, CD: Canberra distance, Edit: Edit distance, LCSS: Longest common subsequence, and LR: Likelihood Ratio.

As the size of the data increases, the performance takes longer. Among these similarity measures, Canberra distance was the fastest one with 0.05 s and the LCSS was the slowest measure with 29.15 s. The DTW and LR have similar execution time.

V. Conclusion and Outlook

In order to handle the drawbacks of existing methods, in this contribution an approach for detecting image motifs is proposed. This method overcomes the existing limitation by considering both fixed and variable size. Moreover, the tested images are not transformed to a one dimensional representation form, thus no information is lost.

This image motif discovery method is performed within three steps: In the first step, the Complex Quad Tree Wavelet Packet transform (CQTWP) analyses the images in various frequency scales. In this work, images are decomposed up to the second scale, since the amount of noise is mostly decreased at this scale. The nodes with the highest information are chosen in order to reduce the amount of redundant information and increase the execution time. The CQTWP consists of two wavelet packets working parallel to each other. Besides the advantages of wavelet transformations, the CQTWP transform has an efficient property of being shift-invariant. Also, its ability for approximately analytic representation is helpful in order to reduce aliasing.

In the second step, features are extracted from the normalised histograms obtained from the wavelet coefficients. These features are the first four statistical moments, and the energy of the wavelet coefficients. Since motif discovery is an unsupervised task, there is no information about the tested images. Consequently, the statistical features are applied in

this work, but depending on the task it is possible to employ other types of features. The efficiency of these features is benchmarked with the linear discriminant analysis (LDA) algorithm [2].

In the last step, motifs are detected by measuring the similarity between their feature values. Six different similarity measures are applied here. The performance of the proposed method and these similarity measures are evaluated by different quality measures. The highest amount of correct motif discovery rate is achieved by the Canberra distance. The Euclidean distance and DTW provide the second best correct motif discovery. By increasing the size of the data, the performance of the Canberra distance improves while the performance of the DTW decreases. The Euclidean distance provides better results than the DTW in the case of larger test cases. The Canberra distance includes the standardisation of the differences between various test data and therefore, it provides the higher correct motif discovery rate compared with the Euclidean distance.

All the experiments carried out with the test cases that are overlaid with noise and blurring effects. These distortions are added to the data to measure the robustness of the proposed method. The best outcome is obtained by the Canberra distance and the LCSS provided the lowest result. The correct motif discovery of the noisy test data is lower than the original test data (without noise). However, as the CQTWP transform decreases the amount of noise, the results obtained from these similarity measures in the case of noisy data are still proper. In case of image blurring, the Euclidean distance executed robuster than the Canberra distance, since the Canberra distance is sensible to the values near zero.

From these similarity measures, the Canberra and Euclidean distance were the fastest one, and the Longest Common SubSequence has the lowest execution time among all.

In the future approach, our aim is to examine other cost functions or approaches to detect the proper nodes of the 2D-CQTWP with the best information content. The approach applied here, is based on the entropy-based cost function, nevertheless other cost functions such as energy or variance can be applied as well. Investigation in effects of image rotation for image motif discovery is another concept, which has to be regarded in the future work. Finally, discovery of motifs within various images without segmenting these images, is the last issue that must be considered in outlook.

ACKNOWLEDGMENT

This research is partly supported by the International Graduate School of Intelligent Systems in Automation Technology (ISA), which is run by scientists of the Faculty of Computer Science, Electrical Engineering and Mathematics and the Faculty of Mechanical Engineering of the University of Paderborn and the Institute of Industrial Information Technologies (inIT) of the Ostwestfalen-Lippe University of Applied Sciences.

APPENDIX A

Proof: The coefficients of signal $x[n-S_{e/o}]$ for both odd and even shifts are given in following:

1. Even Shifts. If $x[n-S_e]$ where the shift $S_e = 2m, \quad m \in \mathbb{Z}$ then CQTWP's coefficients ${}^{s+1}C'_{2J}[n,S_e]$ and ${}^{s+1}C'_{2J+1}[n,S_e]$ are able to detect the shift. Thus,

$$
\begin{aligned}
{}^{s+1}C'_{2J}[n,S_e] &\overset{S_e=2m}{=} \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C'_j[2n-2m-k] = \\
&= \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C'_j[2(n-m)-k] = {}^{s+1}C_{2J}[n-m], \\
{}^{s+1}C'_{2J+1}[n,S_e] &\overset{S_e=2m}{=} \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C'_j[2n-2m-k] = \\
&= \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C'_j[2(n-m)-k] = {}^{s+1}C_{2J+1}[n-m],
\end{aligned}
\tag{6}
$$

where $Len = length({}^{s}C_j)$ and $M \in \mathbb{N}^+$ is the length of the filters.

2. Odd Shifts. If $x[n-S_o]$ where the shift $S_o = 2m+1, \quad m \in \mathbb{Z}$ then CQTWP's coefficients ${}^{s+1}C'_{2J+2}[n,S_o]$ and ${}^{s+1}C'_{2J+3}[n,S_o]$ are able to detect the shift. Thus,

$$
\begin{aligned}
&{}^{s+1}C'_{2J+2}[n,S_o] \overset{S_o=2m+1}{=} \\
&= \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C'_j[2n+1-2m-1-k] = \\
&= \sum_{k=0}^{M+Len-1} {}^{s}g_a[k]\ {}^{s}C'_j[2(n-m)-k] = {}^{s+1}C_{2J}[n-m], \\
&{}^{s+1}C'_{2J+3}[n,S_o] \overset{S_o=2m+1}{=} \\
&= \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C'_j[2n+1-2m-1-k] = \\
&= \sum_{k=0}^{M+Len-1} {}^{s}h_a[k]\ {}^{s}C'_j[2(n-m)-k] = {}^{s+1}C_{2J+1}[n-m].
\end{aligned}
\tag{7}
$$

Similarly, the coefficients for the second wavelet packet "*WPT B*" are obtained. ■

REFERENCES

[1] S. Torkamani and V. Lohweg, "Shift-invariant motif discovery in image processing," in The Seventh International Conference on Performance, Safety and Robustness in Complex Systems and Applications; Special track MAIS: Machine Learning Algorithms in Image and Signal Processing. IARIA, 2017.

[2] E. Alpaydın, Introduction to Machine Learning, 2nd ed. Cambridge: The MIT Press, 2010.

[3] N. M. A. S. Khan, S. A. and N. Riaz, "Gender classification using image processing techniques: A survey," in 2011 IEEE 14th International Multitopic Conference. IEEE, 2011, pp. 25–30.

[4] M. G. K. J. C. S. Nath, S. S. and N. Dey, "A survey of image classification methods and techniques," in International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). IEEE, 2014, pp. 554–557.

[5] S. Gangwar and R. P. Chauhan, "Survey of clustering techniques enhancing image segmentation process," in 2015 Second International Conference on Advances in Computing and Communication Engineering. IEEE, 2015, pp. 34–39.

[6] P. Esling and C. Agon, "Time-series data mining," vol. 45. ACM, 2012, pp. 1–34.

[7] M. K. Das and H.-K. Dai, "A survey of dna motif finding algorithms," BMC bioinformatics, vol. 8 Suppl 7, 2007, p. 21.

[8] K. E. L. J. Patel, P. and S. Lonardi, "Mining motifs in massive time series databases," in Proceedings IEEE International Conference on Data Mining. IEEE, 2002, pp. 370–377.

[9] K. E. W. L. Xi, X. and A. Mafra-Neto, "Finding motifs in a database of shapes," in Proceedings of the 2007 SIAM international conference on data mining. SIAM, 2007, pp. 249–260.

[10] D. Crowther and M. Thompson, "Geology.com-petroglyph photo gallery," 2005-2017. [Online]. Available: http://geology.com/articles/petroglyphs/more-petroglyphs.shtml

[11] S. Torkamani and V. Lohweg, "Survey on time series motif discovery," in Journal: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Article ID: WIDM1199, 2017.

[12] H. R. X. N. L. H. Hu, W. and S. Maybank, "Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering," IEEE Transactions on Image Processing, vol. 23, no. 4, 2014, pp. 1513–1526.

[13] N. M. Zaitoun and M. J. Aqel, "Survey on image segmentation techniques," Procedia Computer Science, vol. 65, 2015, pp. 797–806, Elsevier.

[14] G. Azzopardi and N. Petkov, Eds., Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015 Proceedings, Part I. Springer International Publishing, 2015.

[15] B. K. Gayathri and P. Raajan, "A survey of breast cancer detection based on image segmentation techniques," in 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16). IEEE, 2016, pp. 1–5.

[16] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '09. ACM, 2009, pp. 947–956.

[17] C. M. F. Barone, P. and R. March, "Segmentation, classification and denoising of a time series field by a variational method," Journal of Mathematical Imaging and Vision, vol. 34, no. 2, Jun 2009, pp. 152–164.

[18] F. Y. C. H. Chi, L. and Y. Huang, "Face image recognition based on time series motif discovery," in IEEE International Conference on Granular Computing. IEEE, 2012, pp. 72–77.

[19] S. N. W. M. Grabocka, J. and L. Schmidt-Thieme, "Learning time-series shapelets," in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '14. ACM, 2014, pp. 392–401.

[20] C. Caballero and M. C. Aranda, "Wapsi: Web application for plant species identification using fuzzy image retrieval," Advances on Computational Intelligence, 2012, pp. 250–259, Springer.

[21] Z. Q. Rakthanmanon, T. and E. Keogh, "Mining historical documents for near-duplicate figures," in IEEE 11th International Conference on Data Mining. IEEE, 2011, pp. 557–566.

[22] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," Pattern recognition, vol. 13, no. 2, 1981, pp. 111–122, elsevier.

[23] P. C. N. S. En, S. and L. Heutte, "Segmentation-free pattern spotting in historical document images," in 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015, pp. 606–610.

[24] D. H. Torkamani, S. and V. Lohweg, "Multi-scale motif discovery in image processing," in Workshop on Probabilistic Graphical Models, Heidelberg, Germany, Oct 2015.

[25] B. R. G. Selesnick, I. W. and N. G. Kingsbury, "The dual-tree complex wavelet transform," Signal Processing Magazine, IEEE, vol. 22, no. 6, 2005, pp. 123–151.

[26] S. Torkamani and V. Lohweg, "Shift-invariant feature extraction for time-series motif discovery," in 25. Workshop Computational Intelligence, ser. Schriftenreihe des Instituts für Angewandte Informatik - Automatisierungstechnik am Karlsruher Institut für Technologie, vol. 54. KIT Scientific Publishing, 2015, pp. 23–45.

[27] L. Q. Z. S. Li, T. and M. Ogihara, "A survey on wavelet applications in data mining," ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, 2002, pp. 49–68, ACM.

[28] G. A. K. G. Chaovalit, P. and Z. Chen, "Discrete wavelet transform-based time series analysis and mining," ACM Computing Surveys (CSUR), vol. 43, no. 2, 2011, p. 6, ACM.

[29] S. A. T. S. M. L. Y. Y. C. S. C. I. S. S. Y. J. S. Meng, T. and P. Iyengar, "Wavelet analysis in current cancer genome research: A survey," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 10, no. 6, 2013, pp. 1442–1459, ACM.

[30] I. W. Selesnick, "Hilbert transform pairs of wavelet bases," IEEE Signal Processing Letters, vol. 8, no. 6, 2001, pp. 170–173.

[31] R. Yu and H. Ozkaramanli, "Hilbert transform pairs of orthogonal wavelet bases: Necessary and sufficient conditions," IEEE Transactions on Signal Processing, vol. 53, no. 12, 2005, pp. 4723–4725, IEEE.

[32] N. Kingsbury, "Design of q-shift complex wavelets for image processing using frequency domain energy minimization," in Proceedings International Conference on Image Processing, vol. 1. IEEE, 2003, pp. I–1013.

[33] A. F. Abdelnour and I. W. Selesnick, "Symmetric nearly shift-invariant tight frame wavelets," IEEE Transactions on Signal Processing, vol. 53, no. 1, 2005, pp. 231–239, IEEE.

[34] C. S. Burrus, R. A. Gopinath, and H. Guo, Introduction to wavelets and wavelet transforms: A primer. Upper Saddle River and NJ: Prentice Hall, 1998.

[35] M. V. Wickerhauser, "Lectures on wavelet packet algorithms," in Lecture notes, INRIA. Citeseer, 1991, pp. 31–99.

[36] H. Niemann, Pattern analysis and understanding. Springer Science & Business Media, 2013, vol. 4.

[37] M. M. Deza and E. Deza, Encyclopedia of distances. Springer, 2009.

[38] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," in Soviet physics doklady, vol. 10, 1966, p. 707.

[39] M. Paterson and V. Dancik, Longest common subsequences. Springer Berlin Heidelberg, 1994, vol. 841.

[40] Z. Guoqing and D. Wei, "An HMM-based hierarchical clustering method for gene expression time series data," in IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA). IEEE, 2010, pp. 219–222.

[41] Y. Xiong and D.-Y. Yeung, "Time series clustering with ARMA mixtures," Pattern Recognition, vol. 37, no. 8, 2004, pp. 1675–1689, Elsevier.

[42] B. A. J. Janacek, G. J. and M. Powell, "A likelihood ratio distance measure for the similarity between the Fourier transform of time series," in Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam. Springer Berlin Heidelberg, 2005, pp. 737–743.

[43] MathWorks, "MATLAB," 2017, last access: 31.08.17. [Online]. Available: https://de.mathworks.com/products/matlab.html

[44] C. Bayer, M. Bator, U. Mönks, A. Dicks, O. Enge-Rosenblatt, and V. Lohweg, "Sensorless drive diagnosis using automated feature extraction, significance ranking and reduction," in 18th IEEE Int. Conf. on Emerging Technologies and Factory Automation (ETFA 2013). IEEE, 2013, pp. 1–4.

[45] M. B. Stegmann and D. D. Gomez, "A brief introduction to statistical shape analysis," 2002, informatics and Mathematical Modelling, Technical University of Denmark, DTU.

[46] M. A. R. S. Silva, P. F. B. and da Silva R. A., "UCI machine learning repository: Leaf data set," 2014. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Leaf

[47] M. A. D. H. T. G. S. P. Wang, X. and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," 2013, pp. 1–35, data Mining and Knowledge Discovery.

[48] C. Boncelet, "Image noise models," in The Essential Guide to Image Processing. Elsevier, 2009, pp. 143–167.

[49] W. Burger and M. J. Burge, Principles of digital image processing. Springer, 2009.