Optimizing Feature Extraction from Sequential Sensor Data Using Neural Network-Based Image Transformation

Chang-Hyun Kim
Advanced Mobility System Group
Korea Institute of Industrial
Technology Daegu, Republic of Korea
limition@kitech.re.kr

Seung-Hwan Choi
Advanced Mobility System Group
Korea Institute of Industrial
Technology Daegu, Republic of Korea
csw1496@kitech.re.kr

Hyoeun Kwon
Advanced Mobility System Group
Korea Institute of Industrial
Technology Daegu, Republic of Korea
hekwon525@kitech.re.kr

Hiroaki Kawamoto
Institute of System and Information
Engineering University of Tsukuba
Tsukuba, Japan
kawamoto@iit.tsukuba.ac.jp

Sanghun Choi
Department of Mechanical
Engineering, Kyungpook National
University, Daegu, Republic of Korea
s-choi@knu.ac.kr

Suwoong Lee
Advanced Mobility System Group
Korea Institute of Industrial
Technology Daegu, Republic of Korea
lee@kitech.re.kr

***Abstract*—The transformation of time-series data into images and the subsequent application of Convolutional Neural Networks (CNNs) has recently emerged as a promising approach for sensor signal analysis, owing to CNNs' superior capability in extracting spatially encoded patterns and complex hierarchical features. While prior studies have shown that CNN-based learning on image-transformed time-series data can surpass the performance of traditional direct learning from raw time-series data, this advantage is not consistently observed in all cases. We investigate key issues, including potential quantization error and loss of information during the transformation process, CNN's inherent limitations in capturing temporal dependencies and long-term correlations, the mismatch between transformation methods and specific signal patterns, and the risk of overfitting due to the complexity imbalance between the data and model architecture. The experimental results demonstrate that image-based learning achieves higher accuracy and robustness, especially in complex environments with noise and overlapping signals.**

Keywords—gas sensor, image encoded, time-series, neural network

I. Introduction

Advancements in high-performance sensor technology have significantly enhanced the accuracy and efficiency of real-time monitoring and control across various domains, including environmental monitoring, industrial automation, and healthcare. Sensors typically generate continuous output over time, resulting in time-series signals characterized by non-stationary behavior, varying levels of noise, and complex cross-sensitivity effects. Effective analysis of such signals is crucial for identifying patterns, detecting anomalies, and generating accurate predictions in diverse application settings [1], [2].

Traditional sensor signal analysis often relies on Long Short-Term Memory (LSTM) networks. However, LSTM models struggle to capture complex signal patterns and long-term dependencies in noisy environments, limiting classification accuracy. LSTM models are particularly Effective at for capturing long-term dependencies due to their ability to maintain hidden state information over extended time steps, thereby mitigating the vanishing gradient problem associated with standard Recurrent Neural Networks (RNNs) [3], [4]. However, the complexity and non-linearity of sensor signal data often pose significant challenges for direct time-series learning, especially in the presence of noise and long-term dependencies. Traditional methods for sensor signal analysis involve direct processing of time-series data using LSTM and other sequential models. However, direct learning from raw time-series data often fails to capture complex patterns and long-term dependencies effectively.

An emerging alternative involves transforming time-series data into images and processing them using Convolutional Neural Networks (CNNs). CNNs, renowned for their exceptional performance in image-based spatial pattern recognition tasks, are capable of extracting hierarchical spatial features from complex input data [5].

The Recurrence Plot (RP) encodes recurrence dynamics of a signal over time by calculating pairwise distances between time points and applying a threshold to determine recurrence states. This approach is particularly effective in identifying cyclic patterns and state similarity. The Gramian Angular Field (GAF) captures the angular relationship between time-series values by transforming them into polar coordinates. It preserves both the magnitude and directional information of the signal, allowing CNN models to detect fine-grained temporal structures. The Markov Transition Field (MTF) represents state transition probabilities of a time-series as a structured matrix. This approach enables CNN models to learn dynamic transitions and capture complex state changes in the signal [6].

These image-based approaches leverage CNN's local receptive fields and weight-sharing capabilities to uncover complex patterns that may not be readily apparent in the raw time-series format. As a result, CNN-based models trained on image-transformed data have demonstrated improved

This work was supported by the Industrial Technology Innovation Program(RS-2024-00422721, Development of safe and reliable XAI-based robotic workcell safety sensors and control modules) funded by the Ministry of Trade, Industry Energy (MOTIE,Korea). And this work was supported by the Industrial Technology Innovation Program(RS-2024- 00424974, Robot and robot service digital twinization framework technology) funded by the Ministry of Trade, Industry Energy (MOTIE,Korea).

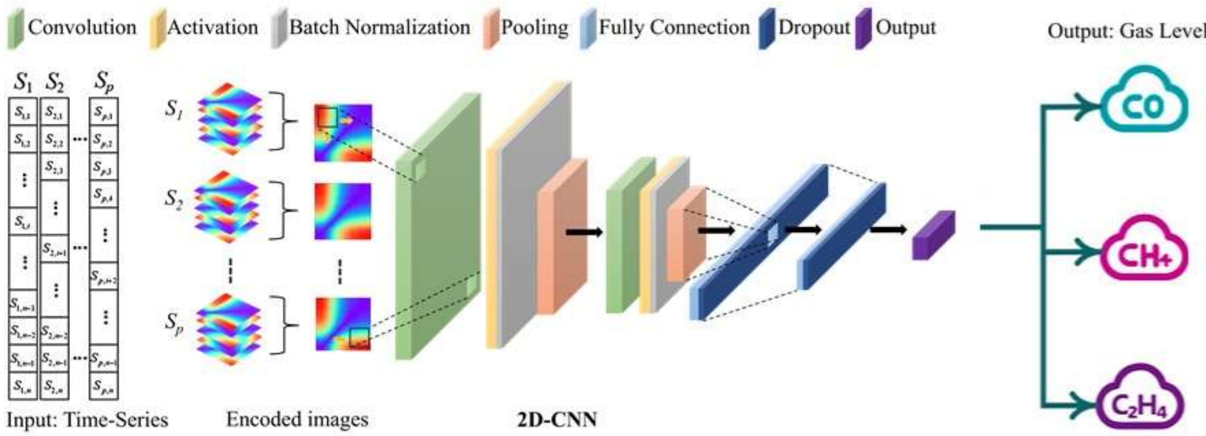

Fig. 1. The schematic of the 2D-CNNs with image encoding

performance in certain sensor signal classification tasks compared to direct LSTM-based learning [6-8].

Despite the encouraging performance gains associated with CNN-based learning on image-transformed data, this approach is not universally superior. In certain scenarios, the transformation process itself may introduce artifacts, degrade temporal information, or suppress key features, thereby diminishing classification accuracy. Furthermore, CNN's inherent limitation in modeling sequential dependencies and long-term correlations may lead to suboptimal performance when handling time-series data with strong temporal structures.

This study aims to investigate the performance and limitations of CNN-based learning on image-transformed time-series sensor data. Section II provides an overview of the dataset and the preprocessing steps involved. Section III introduces the transformation techniques used to convert time-series data into images, including RP, GAF, and MTF. Section IV describes the experimental setup and the CNN-based model architecture. Section V presents a detailed analysis of the classification performance, including a comparison with direct LSTM-based learning. Finally, Section VI discusses the limitations and future research directions to enhance model performance and generalization across different sensor types.

II. METHOD

In this study, we employed three widely used time-series transformation techniques to convert sensor signals into images: RP, GAF, and MTF. Figure 1 illustrates the overall process of transforming time-series sensor data into image representations using RP, GAF, and MTF. The RP encodes cyclic patterns within the signal by computing pairwise distances between time points and applying a threshold to determine recurrence. The GAF preserves angular relationships between data points, capturing both the magnitude and direction of the signal. The MTF represents state transition probabilities, which allows the model to learn dynamic transitions within the signal. These transformed images serve as input to the CNN-based model, enabling the extraction of hierarchical spatial patterns from the data.

Figure 2 shows the architecture of the CNN model used for classification. The model consists of three convolutional layers followed by batch normalization and max-pooling layers. A fully connected layer with 128 units is used before the final softmax layer for classification. The CNN model is designed to leverage local spatial patterns and hierarchical features extracted from the image representations of time-series signals. This architecture allows the model to effectively identify complex signal variations that are not easily captured by direct learning from raw time-series data.

A. Dataset

Publicly available sensor signal time-series datasets were used for this study. The data consists of continuous resistance measurements over time from different types of sensors exposed to various gases and environmental conditions. The goal is to classify the sensor response accurately based on the signal pattern.

We use the dataset introduced in [9], consisting of 180 time-series measurements from a chemo-resistive gas detection platform composed of 8 MOX gas sensors. Each measurement lasted 300 seconds with a sampling rate of 20 ms. The dataset includes sensor readings, temperature, and humidity. The experiments were conducted in a 2.5 m × 1.2 m × 0.4 m wind tunnel with two independent gas sources. The sensors were exposed to mixtures of Ethylene with Methane or Carbon Monoxide at four different flow rates (zero, low,

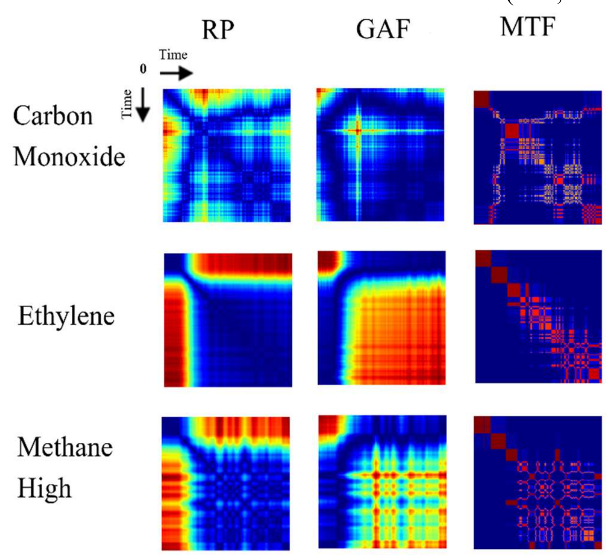

Fig. 2. Example representations of time-series data using RP, GAF, and MTF for different gas types. Each row shows the transformation results for Carbon Monoxide, Ethylene, and Methane High concentration levels. RP highlights cyclic patterns, GAF captures angular relationships, and MTF reflects state transition probabilities.

medium, high), resulting in 30 mixture configurations. Each configuration was repeated 6 times, generating a total of 180 measurements. The data encode the gas type and concentration levels. Each data contains 11 columns: time, temperature, humidity, and the 8 sensor readings. Sensor readings can be converted to resistance using the formula: Rs(KΩ) = 10 × (3110 − A) / A. The dataset has no missing values and provides both raw and down-sampled time series at 100 ms.

We preprocess the dataset consisting of 180 time-series measurements from 8 MOX gas sensors, each sampled at 20 ms for 300 seconds. First, the sensor readings are converted to resistance values for consistency. To reduce noise and computational load, we apply uniform down sampling with three factors: DF = 2 (40 ms), DF = 4 (80 ms), and DF = 8 (160 ms), ensuring the key signal patterns are preserved. The sensor data, along with temperature and humidity, are normalized using min-max scaling to ensure consistency across different ranges. The data are then labeled based on gas type and concentration and organized by down sampling factor.

B. Time-Series to Image Conversion Techniques

In this study, we employed three widely used time-series transformation techniques to convert sensor signals into images:

1) Recurrence Plot (RP)

A Recurrence Plot visualizes the similarity between different time points in a time-series. It transforms the temporal relationship into a spatial pattern by computing pairwise distances between time points and applying a threshold to determine recurrence.

$$R_{i,j} = \Theta\left(\epsilon - \| x_i - x_j \|\right) \tag{1}$$

where $R_{i,j}$ represents the recurrence state between time points i and j, ϵ is the threshold distance, and Θ is the Heaviside step function.

RP is effective for identifying cyclic behavior, repeated patterns, and chaotic structures in time-series signals, especially under stable periodic conditions [10].

2) Gramian Angular Field (GAF)

GAF transforms a time-series into polar coordinates, encoding the angular relationships between data points into a matrix. It captures both the magnitude and direction of signal changes [6].

a) Normalize the data to [−1,1][-1, 1]:

$$\tilde{x}_t = \frac{x_t - \mathrm{mi}\ (X)}{\max(X) - \min(X)} \times 2 - 1 \tag{2}$$

b) Convert to angular coordinates:

$$\phi_t = \arccos\left(\tilde{x}_t\right) \tag{3}$$

c) Compute the Gramian matrix:

$$G_{i,j} = \cos\left(\phi_i + \phi_j\right) \tag{4}$$

where $\phi_i = \arccos(x_i)$ represents the angular encoding of the signal.

GAF highlights the temporal structure of the signal and allows CNNs to learn patterns based on angular relationships.

3) Markov Transition Field (MTF)

MTF captures the state transition probabilities of a time-series as a matrix. It encodes the likelihood of transitioning from one state to another.

a) Quantize the time-series into discrete states.

b) Compute the state transition probability matrix:

$$M_{i,j} = P\left(s_j \mid s_i\right) \tag{5}$$

where $P\left(s_j \mid s_i\right)$ denotes the transition probability from state s_i to state s_j.

MTF effectively represents the dynamic behavior and transition patterns within the signal [6].

C. Machine Learning Models

Two types of models were trained for comparison. The LSTM (Direct Learning) model was trained directly on raw time-series data to capture temporal dependencies and long-term patterns. In contrast, the CNN (Image-Based Learning) model was trained on converted images generated from RP, GAF, and MTF transformations. This allowed the CNN model to extract spatial patterns and hierarchical features from the image-based representations of the time-series data.

III. Limitations of Image-Based Time-Series Transformation

A. Information Loss During Transformation

Transforming time-series data into images often results in information loss due to quantization, resolution reduction, and normalization processes.

1) Quantization and Resolution Loss

Transforming time-series data into images often results in information loss due to quantization, resolution reduction, and normalization processes. In the quantization process, small variations in the original signal are often rounded off or eliminated, which can reduce the overall sensitivity of the model. Additionally, downsampling the data to reduce computational complexity may lead to the loss of high-frequency components, making it difficult for the model to recognize detailed signal patterns.

For example, in GAF transformation:

a) Time-series data is normalized to [−1,1][-1, 1]:

$$\tilde{x}_t = \frac{x_t - \min(X)}{\max(X) - \min(X)} \times 2 - 1 \tag{6}$$

b) Small variations are reduced to the same angular value, causing information loss.

Consider the following time-series data:

$$[0.1, 0.15, 0.2, 0.25, 0.3] \tag{7}$$

After GAF transformation and quantization, this may result in:

$$[0.0, 0.0, 0.1, 0.1, 0.1] \tag{8}$$

Small variations in the original data are lost due to quantization, resulting in a loss of distinctive features that could aid in classification.

2) Clipping and Saturation

Normalization during the transformation process can lead to clipping or saturation, where extreme values are forced into a narrow range. This compression effect results in the loss of information at the tails of the distribution, reducing the model's ability to detect outlier patterns or extreme signal variations. As a result, important variations in signal amplitude may be ignored during the learning process, which can degrade classification accuracy, especially for signals with high dynamic range.

B. CNN's Limited Capacity for Capturing Temporal Dependencies

CNN models are primarily designed for spatial pattern recognition, making them less effective in capturing long-term temporal dependencies within non-stationary sensor signals. The convolutional filters in CNN models extract local features from 2D images, which allows them to identify spatial patterns efficiently. However, this local feature extraction process makes it difficult for CNN models to capture the sequential nature and long-term dependencies inherent in time-series data. In contrast, LSTM networks are explicitly designed to capture long-term dependencies by maintaining hidden state information over time:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \tag{9}$$

where:

- h_t = hidden state at time step t
- x_t = input at time step t
- W_{hh}= weight matrix for hidden state
- W_{xh}= weight matrix for input state

The vanishing gradient problem in RNNs and LSTMs arises from repeated multiplication of gradients over time:

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h_t} \cdot \prod_{i=1}^{t} W_{hh} \tag{11}$$

CNN does not face this problem but also does not preserve temporal dependencies as effectively as LSTM networks. When the signal contains long-term dependencies or cyclic patterns, LSTM-based models tend to outperform CNN-based models.

C. Inappropriate Fit of Transformation Techniques

Each transformation technique is suitable for specific types of time-series patterns. The RP is effective for capturing cyclic and repetitive patterns. However, when the sensor signal lacks cyclic behavior, RP tends to generate noisy or random-looking plots, which can degrade classification accuracy. The GAF is well-suited for smooth, continuous changes, as it encodes angular relationships between time-series values. However, when the signal contains abrupt state changes, GAF tends to smooth out those changes, thereby reducing the model's ability to detect sudden transitions. The MTF effectively represents state transitions and discrete events. However, when the transition probability is low, MTF generates sparse matrices, which limits the CNN's ability to effectively learn patterns from the data.

D. Data Complexity vs. Model Complexity

CNN models typically require large amounts of data due to their complex architecture. When the dataset is small, CNN-based models tend to overfit, leading to poor generalization on unseen data. The number of parameters in a CNN model is calculated using the equation:

$$parameter = (f_w \cdot f_h \cdot c_{in} + 1) \cdot c_{out} \tag{12}$$

where f_w and f_h represent the filter width and height, and c_{in} and c_{out} represent the input and output channels, respectively. A higher number of parameters increases the model's complexity, which can lead to overfitting when the dataset size is limited. This tendency is particularly problematic for small datasets, where the large number of parameters may cause the model to memorize training data rather than generalize to new data.

E. Noise Sensitivity

While CNN models are relatively robust to noise, the transformation process itself can introduce additional noise. The RP may amplify noise in cyclic patterns, making it difficult to distinguish meaningful patterns from artifacts. The GAF may exaggerate small variations during the angular transformation process, leading to distorted signal representation. Similarly, the MTF may misrepresent state transitions when the input signal is noisy, which can reduce the model's ability to learn accurate transition patterns.

F. Experimental Setup

We conducted experiments using sensor signal data from publicly available sources. The direct learning approach involved training LSTM models on raw time-series data to capture temporal dependencies. In contrast, the image-based learning approach involved transforming the time-series data using RP, GAF, and MTF and training CNN models on the resulting images. This allowed CNN models to leverage spatial patterns extracted from the transformed images.

Hyperparameters for CNN and LSTM models were optimized using grid search. Each experiment was repeated five times, and the average performance was reported.

IV. Results and Discussion

The detailed classification performance of each model is shown in Table I. Figure 3 illustrates the confusion matrix for the GAF-based CNN model. The matrix shows the true and predicted labels for various gas types and concentration levels. The GAF-based CNN model demonstrates strong classification accuracy across different gas types with minimal misclassification, indicating that the model effectively captures key signal characteristics.

A. Performance Comparison

Experimental results show that CNN-based learning on image-transformed data outperforms direct LSTM-based learning under certain conditions, particularly when the signal exhibits well-defined cyclic patterns or state transitions. However, in cases involving non-stationary behavior, high-frequency noise, or cross-sensitivity effects, direct LSTM-based learning achieved higher classification accuracy and improved robustness.

Table I shows the detailed performance metrics for each model. The GAF-based CNN model achieved the highest

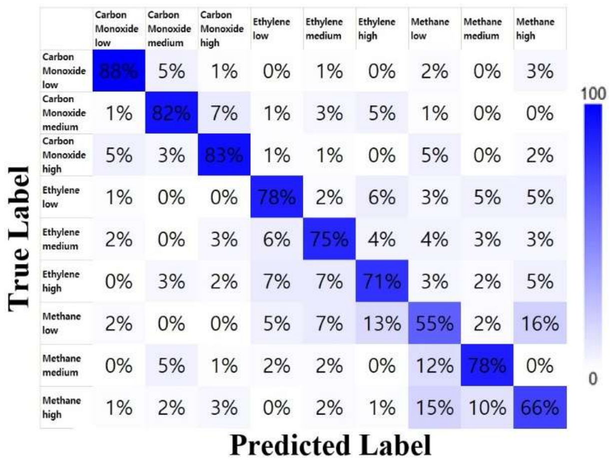

Fig. 3. Gas types summarized in a confusion matrix. Ethylene (low: 31 ppm, midium: 46 ppm, high: 96 ppm), Carbon Monooxide (low: 270 ppm, midium: 397 ppm, high: 460 ppm), Methane (low: 51 ppm, midium: 115 ppm, high: 131 ppm)

classification accuracy of 82.5 percent, outperforming direct LSTM-based learning by 5.4 percent. The CNN-based models demonstrated stronger overall precision and recall compared to the LSTM-based model, particularly in signals exhibiting cyclic or state-transition patterns.

TABLE I. PERFORMANCE EVALUATION

Method	Accuracy (%)	F1-Score	Precision	Recall
LSTM (Direct)	77.1	0.75	0.76	0.74
CNN (RP)	80.3	0.78	0.79	0.78
CNN (GAF)	82.5	0.81	0.83	0.79
CNN (MTF)	78.5	0.77	0.78	0.76

The GAF-based CNN model showed balanced performance across precision, recall, and F1-score, suggesting that it effectively captured both the magnitude and directional structure of the signal. The RP-based model performed well for cyclic patterns but underperformed when the signal was less repetitive. The MTF-based model demonstrated strength in recognizing discrete state transitions but showed reduced accuracy when transition probabilities were low.

B. *Why CNN Outperformed LSTM*

The CNN model demonstrated superior performance in recognizing spatial patterns in the transformed data, especially when the signal exhibited cyclic or repeating structures. The RP effectively captured cyclic behavior by representing the similarity between time points. This allowed the CNN model to identify repetitive patterns more effectively. The GAF preserved angular relationships and encoded both magnitude and direction, enabling CNN models to detect fine-grained temporal variations. The MTF provided valuable information about state transitions, which improved the model's ability to identify structured changes within the signal.

In contrast, the LSTM model struggled with high-frequency noise and non-stationary behavior. LSTM models are designed to capture long-term dependencies, but the presence of noise and rapid state transitions reduced their ability to maintain consistent hidden state information. The CNN model's reliance on local patterns and spatial features allowed it to adapt more effectively to these complex signals.

C. *Misclassification Analysis*

Misclassifications primarily occurred in signals with overlapping patterns or high-frequency noise. The RP-based model showed reduced accuracy in signals lacking cyclic behavior, while the GAF-based model struggled with abrupt state changes. The MTF-based model misclassified low-probability state transitions due to data sparsity. A detailed analysis of the confusion matrix revealed that misclassifications often occurred in cases with similar signal patterns or overlapping feature sets.

V. FUTURE WORK

Although the proposed approach demonstrates enhanced performance for sensor signal classification using CNN-based image learning, several areas remain for future exploration. Future research could investigate hybrid architectures combining CNN and LSTM models to leverage both spatial and temporal pattern recognition capabilities. This approach may address the limitations of CNN in capturing long-term dependencies.

Extending the model to work with sensor signals from different domains, such as biomedical signals and industrial sensor data, could validate the generalizability of the proposed approach. Future work could also focus on optimizing the computational efficiency of the model to enable real-time processing in practical applications. Furthermore, exploring adaptive transformation techniques that adjust dynamically to signal characteristics could enhance the model's ability to capture complex patterns.

References

[1] Yaqoob, Usman, and Mohammad I. Younis. "Chemical gas sensors: Recent developments, challenges, and the potential of machine learning—A review." Sensors 21.8 (2021): 2877.

[2] Nazemi, Haleh, et al. "Advanced micro-and nano-gas sensor technology: A review." Sensors 19.6 (2019): 1285.

[3] Hochreiter, Sepp, et al. "Gradient flow in recurrent nets: the difficulty of learning long-term dependencies." (2001).

[4] Hu, Yuhuang, et al. "Overcoming the vanishing gradient problem in plain recurrent networks." arXiv preprint arXiv:1801.06105 (2018).

[5] Chen, Long, et al. "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[6] Wang, Zhiguang, and Tim Oates. "Imaging time-series to improve classification and imputation." arXiv preprint arXiv:1506.00327 (2015).

[7] Ismail Fawaz, Hassan, et al. "Deep learning for time series classification: a review." Data mining and knowledge discovery 33.4 (2019): 917-963.

[8] Hatami, Nima, Yann Gavet, and Johan Debayle. "Classification of time-series images using deep convolutional neural networks." Tenth international conference on machine vision (ICMV 2017). Vol. 10696. SPIE, 2018.

[9] Tian, Fengchun, Leilei Zhao, and Siyuan Deng. "Gas sensor array low-concentration ." UCI Machine Learning Repository, 2023, https://doi.org/10.24432/C5CK6F.

[10] Eckmann, J-P., S. Oliffson Kamphorst, and David Ruelle. "Recurrence plots of dynamical systems." Turbulence, Strange Attractors and Chaos. 1995. 441-445.