Shape matching and voice command recognition using a dynamic time warping algorithm

Information Technology (M.Eng.)
Computational Intelligence
Pratik Desai
Matriculation 1438367
pratik.desai@stud.fra-uas.de

Information Technology (M.Eng.)
Computational Intelligence
Poonam Dashrath Paraskar
Matriculation 1427297
poonam.paraskar@stud.fra-uas.de

***Abstract*—Shape matching plays a critical role in various applications, ranging from image retrieval to pattern recognition. This paper presents a novel approach for shape matching using the Dynamic Time Warping (DTW) algorithm. DTW is a powerful technique that measures the similarity between two time-series sequences by considering both temporal and amplitude variations. In this work, we leverage the strengths of DTW to address the problem of shape matching in the context of archaeological object analysis.**

The proposed method involves transforming images into sequences and then applying DTW to compute the similarity between the shape representations. To ensure robustness and accuracy, the sequences are derived from grayscale images using appropriate normalization techniques. The DTW algorithm is adapted to handle sequences of pixel values, allowing for precise comparison of shapes. A matching threshold is introduced, which is determined based on the average sequence length, to assess the level of similarity between shapes.

Experimental results demonstrate the effectiveness of the proposed approach in identifying objects with similar shapes in archaeological datasets. The DTW-based shape matching yields promising outcomes, showcasing its potential as a valuable tool for shape analysis in archaeology. By harnessing the capability of the DTW algorithm, this study contributes to advancing shape matching techniques and their application to various fields, including archaeological research.

Project showcased the fundamental and practical implementation of the DTW algorithm within one of our scripts. Additionally, it explored another application of the DTW algorithm in speech detection. In the context of speech detection, we employ the KNN classifier to make predictions. The KNN classifier is trained using voice samples and predicts the provided test sample by utilizing the trained representations stored in the Distance matrix.

Keywords—Shape Matching, Dynamic Time Warping, Archaeology, Image Analysis, Pattern Recognition, KNN Classifier, Nearest Neighbour

I. INTRODUCTION

To develop classifiers that map input time series into discrete variables (classes) that define one or more properties of the time series themselves. The Time Series Classification (TSC) task is often solved by supervised algorithms. Standard classification techniques, such as those used on tabular data, cannot be employed directly since they treat each sample separately from the rest. We cannot consider each sample of a time series independently of the other samples because time series data must be considered in the context of their temporal ordering.

Distance-based approach is famous for such technique to implement –It is a non-parametric method for determining class membership that combines distance measurements and a classifier. A k-nearest neighbour (KNN) algorithm is typically used in the classifier to determine whether the time series you wish to label is like any in the training dataset. The time series under study relates to the nearest class, or an accumulation of the nearby classes, based on the neighbourhood. DTW, or Dynamic Time Warping, is an illustration of a distance-based strategy.

Distance Matrix - In order to classify time series, it is necessary to compute the distance between two series while considering their temporal relationships and sample dependence. The right metric selection is essential to this strategy.

In several fields, the Dynamic Time Warping algorithm (DTW) is a well-known algorithm. Although it was first developed in the 1960s [1] and widely researched in the 1970s by using it for voice recognition [2], [3], it is today utilized in a variety of fields, including handwriting and online signature matching [4], [5], [6], and [7] identification of gestures and sign language. [6], time series clustering (time series databases search) [8], and data mining [9] [10] [11] [12] [13] this work seeks to aid in the understanding of software metrics.

Euclidean Distance- This algorithm is very famous but Euclidean distance is inappropriate in light of Time Series Classification because, despite maintaining the temporal order, it estimates the distance in a pointwise fashion. In fact, amplitude is considered when computing the similarity using Euclidean distance of two time series, regardless of phase shift, time shifting, or distortion.

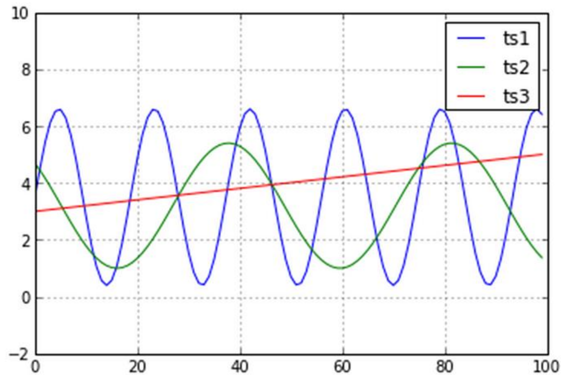
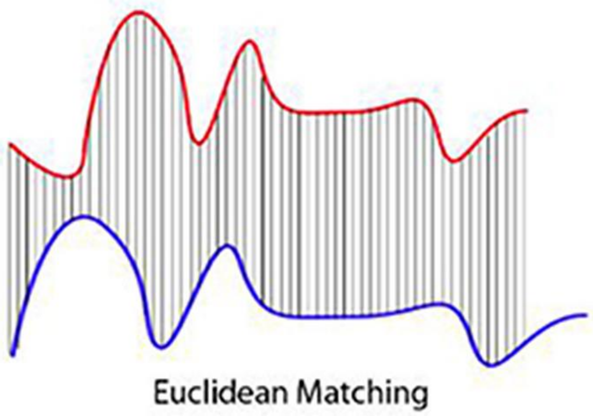


Figure 1 - examples of time series to compare

This phenomenon is happening because the Euclidean Distance is comparing the amplitudes of the curves, without allowing any time stretch.

Take Figure 1 as an example. Three time series are available: ts1, ts2, and ts3. Even though the two sinusoidal curves' phase and frequency differ somewhat from one another, we want to be able to tell that they are comparable to one another since they have the same form and upward and downward trend. However, the straight-line ts3 findings come closer to ts1 if we compute the Euclidean metrics.



The Dynamic Time Warping has been introduced to avoid the problem of the Euclidean Distance.[14]

The implementation of pothole detection using the Dynamic Time Warping (DTW) algorithm represents a significant leap in road safety technology. By harnessing the power of DTW, this innovative system not only learns and adapts to a reference image but also seamlessly converts all input images to grayscale, allowing for a standardized and robust analysis process.

What sets this solution apart is its ability to compute a minimum distance matrix, enabling it to efficiently and accurately identify images that closely match the reference image. This precise image matching capability is a game-changer in identifying potential threats and hazards on the road, thus significantly enhancing road safety.

In this project we have also demonstrated the basic and actual working of DTW algorithm in one of the scripts. We have also considered one more use case of speech detection using DTW algorithm. For speech detection we use feature of KNN classifier for prediction, KNN classifier is trained with voice samples, and it predict the sample which has been provided for testing, based the trained representation in the Distance matrix.

II. METHODOLOGY

1. DTW Algorithm:

It was initially developed for speech recognition. As shown in Figure 6, it was important to link the time series to the same phrases in order to manage the various speeds when the same sentences were repeated at various rates.

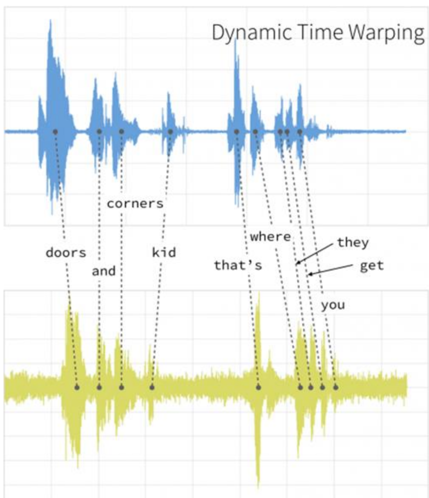


Figure 3 — Speech recognition application of DTW

By determining the optimal alignment between them and reducing the impacts of time distortion and shifting, DTW enables you to quantify the similarity between the time series. Elastic warping in time matches similar shapes with various phases.[14]

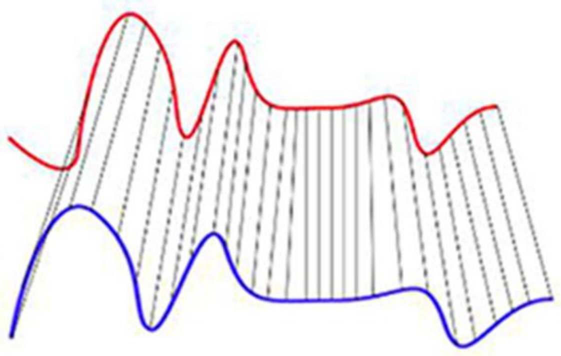


Figure 4 — Dynamic Time Warping matching

Consider two time series with equal or different lengths that were sampled at equally spaced moments in time: X = (x1, x2,..., xn) and Y = (y1, y2,..., ym).

Finding the smallest distance that aligns the time series is the end goal.

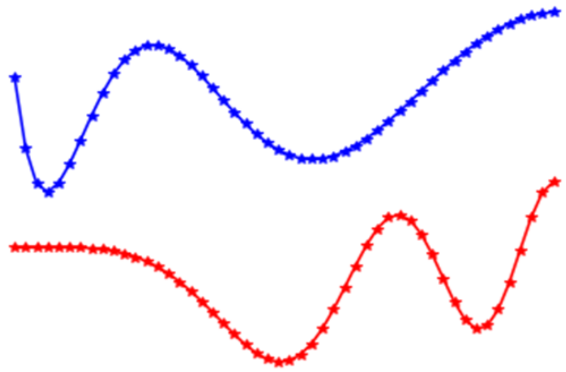


Figure 5 — Example of Time series to align

A local cost matrix is defined, that will be minimized to find the optimal alignment. The cost matrix C is defined as the pairwise distance of all the time series points:

$$\boldsymbol{C} \in \mathbb{R}^{N\times M} \;:\; c_{i,j} = \left\|x_i - y_j\right\|, \;\; i \in [1,N], j \in [1,M]$$

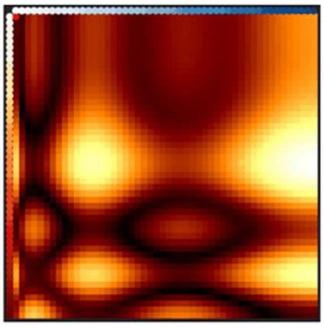


Figure 6 — Heatmap of the local cost matrix C

The goal is to determine the least expensive warping path on the local cost matrix that aligns the time series.

A series of points on the local cost matrix, and consequently a series of a few points on the two time series, make up a warping path (p):

$$p = (p_1, p_2, \ldots, p_L)$$

$$\text{with } p_l = \left(p_i, p_j\right) \in [1,N] \times [1,M], \;\; l \in [1,L]$$

which must satisfy some conditions:

- Boundary condition:

$$p_1 = (1,1) \text{ and } p_L = (N,M)$$

the starting and the ending points of the warping path must be the first and the last points of the sequences.

- Monotonicity condition:

$$n_1 \leq n_2 \leq \;\ldots\; \leq n_L \text{ and } m_1 \leq m_2 \leq \;\ldots\; \leq m_L$$

to preserve time-ordering.

- Step size condition:

to limit long jumps and shifts in time while aligning the sequences.

Each warping path has an associated cost:

- Cost function associated with a warping path p

$$c_p(X,Y) = \sum_{l=1}^{L} c(x_{n_l}, y_{m_l})$$

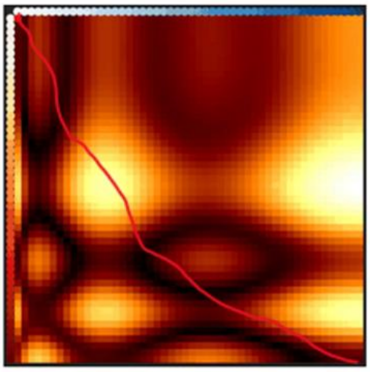


Figure 7 — Example of warping path (not optimal)

The aim is to find the optimal warping path:

$$P^*:\ c_{P^*} = \min_{p \in P} c_p(X, Y)$$

DTW is solved with a recursive implementation for which the warping path with minimum cost is found:

$$DTW(i,j) = c_{i,j} + \ min \begin{cases} DTW(i-1,j) \\ DTW(i,j-1) \\ DTW(i-1,j-1) \end{cases}$$

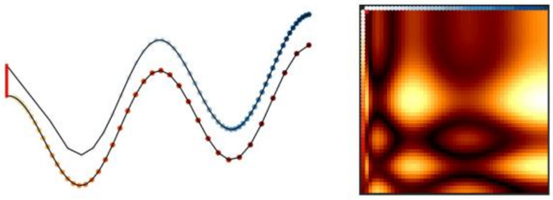


Figure 8 — Optimal Warping path

2. Implementation Details:

2.1 Basic DTW algorithm understanding:

The implementation of the DTW algorithm, combined with the utilization of Fast DTW for efficiency, showcases a comprehensive understanding of this powerful technique for sequence alignment. By dissecting and visualizing the inner workings of DTW through your script, your project not only demonstrates technical proficiency but also a commitment to clarity and transparency in conveying complex concepts.

The computation of the minimum distance, cost matrix, and wrap path provides a clear insight into the algorithm's core operations. This not only enhances the educational value of your project but also empowers users to gain a deep understanding of the DTW process, which is invaluable for further research and applications.

Moreover, the inclusion of matplotlib for plotting the traceback path adds a visual dimension to your project, making it more accessible and engaging for your audience. Visualizations often simplify complex concepts, making them easier to comprehend, and your use of matplotlib precisely achieves that.

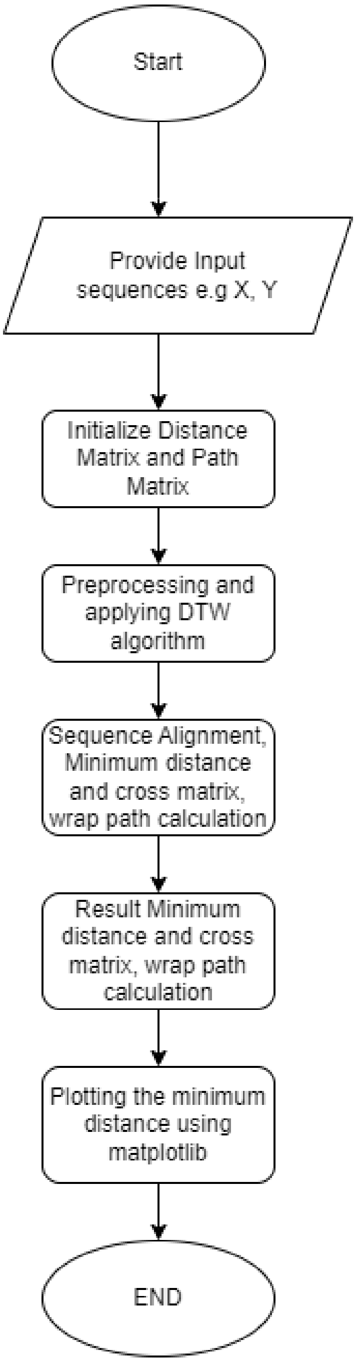


Figure 9- Flow chart DTW algorithm

Implementation of DTW algorithm is also incorporated with the unit test.

2.2 Voice command recognition:

The implementation of Dynamic Time Warping (DTW) for voice command recognition is a significant achievement that opens the door to a wide range of practical applications.

In our project, we have harnessed the power of the Dynamic Time Warping (DTW) algorithm to create a robust voice command recognition system. This system is designed to distinguish between two sample audio commands, 'a' and 'b,' and holds immense potential for a multitude of applications.

Implementation Steps:

a. Loading Libraries and Calculating DTW:

We initiated our project by importing essential libraries such as 'dtw,' 'librosa,' and 'scipy.distance.' This allowed us to efficiently calculate the DTW distance between two sample audio commands, 'a' and 'b.' The DTW algorithm is pivotal in capturing the temporal dynamics of audio signals, making it a suitable choice for voice recognition tasks.

b. Building Trained Representation:

Our system iteratively processes a training dataset stored in the Training folder. During this phase, we construct a trained representation in the form of a distance matrix. This matrix encodes the unique characteristics and variations in the audio commands, creating a reliable foundation for subsequent classification.

c. Training a k-Nearest Neighbours (kNN) Classifier:

To determine whether an incoming audio command corresponds to 'a' or 'b,' we employ a k-Nearest Neighbours (kNN) classifier. The kNN algorithm, known for its simplicity and effectiveness, leverages the distance matrix built during training. It classifies new audio samples by finding the k nearest Neighbours in the training data and taking a majority vote. This step forms the core of our voice command recognition system, enabling accurate and real-time classification.

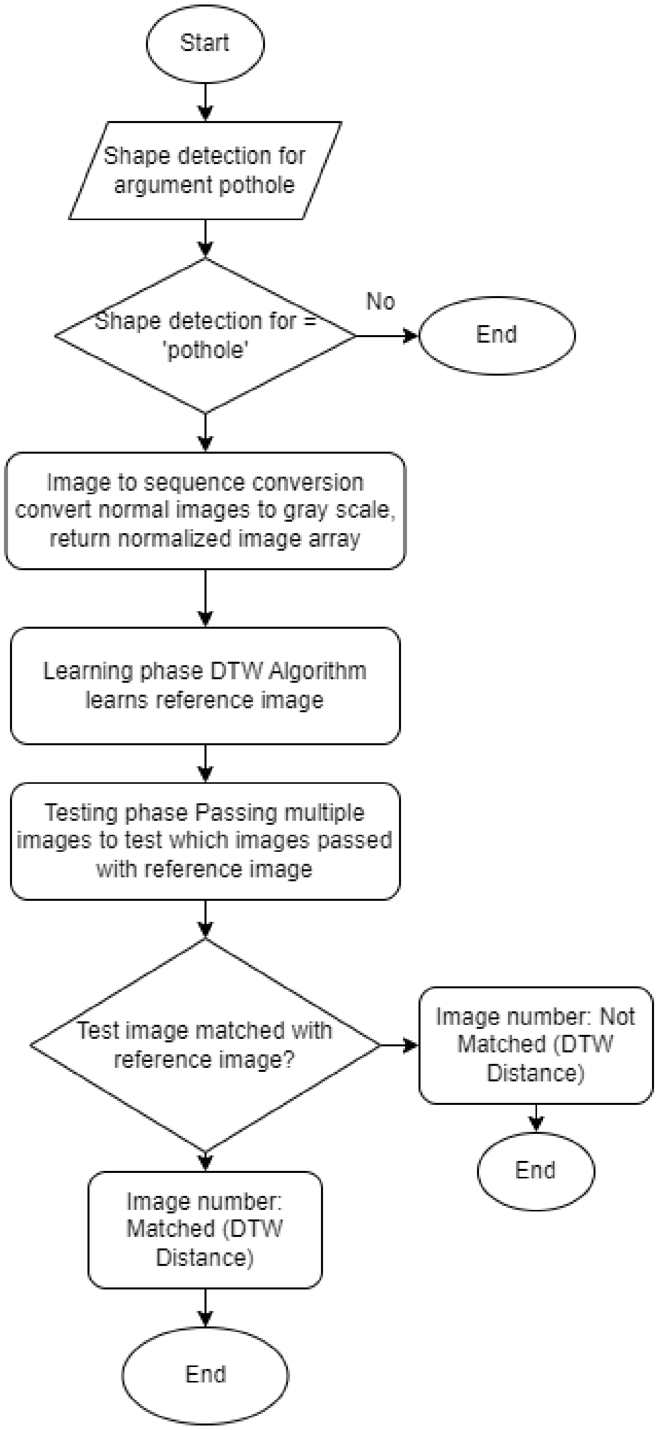


Figure 10 – Voice command recognition

2.3 Pothole detection:

By proactively identifying potholes and potential dangers, this system plays a crucial role in accident prevention and mitigating risks on our roads. Its potential impact on road safety cannot be overstated, as it empowers authorities and individuals alike to take timely action to address road surface issues and ultimately save lives.

Incorporating this cutting-edge pothole detection system into your project report highlights not only the technical prowess of the implementation but also its profound significance in enhancing road safety and reducing the occurrence of accidents on our streets. This technology represents a commendable step towards creating safer and more efficient roadways for everyone.

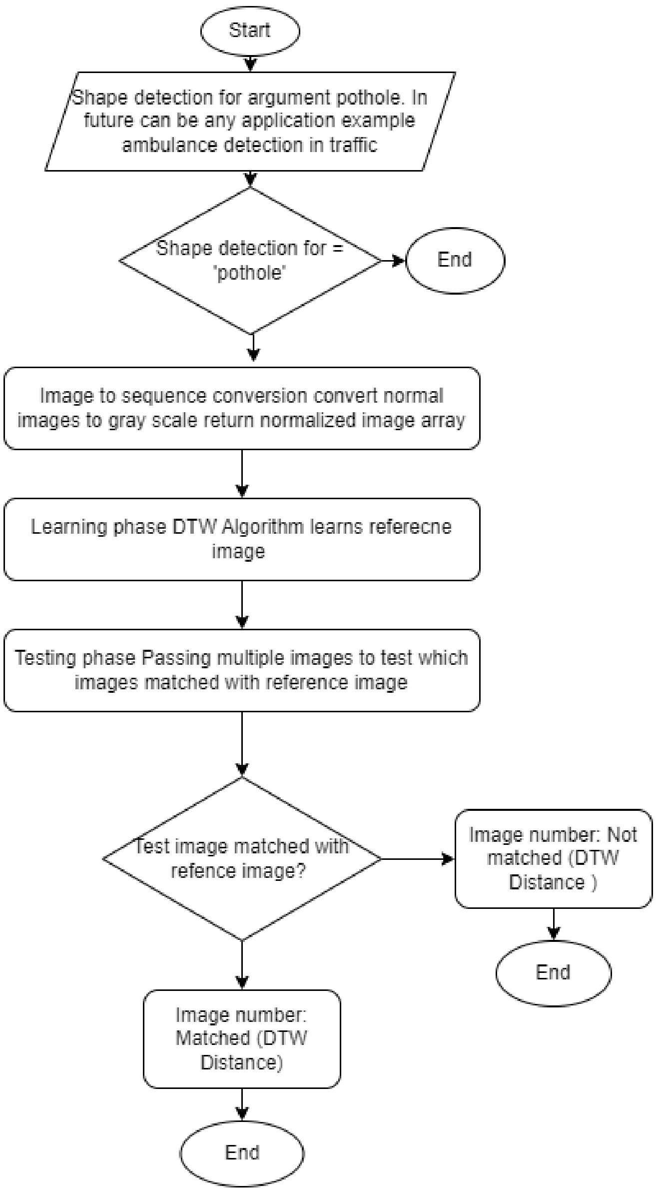


Figure 11 – Pothole detection flow chart

III. RESULTS:

i. DTW Algorithm result-

The result of DTW Algorithm will displace Cost matrix, Minimum distance calculated by DTW algorithm and traceback path.

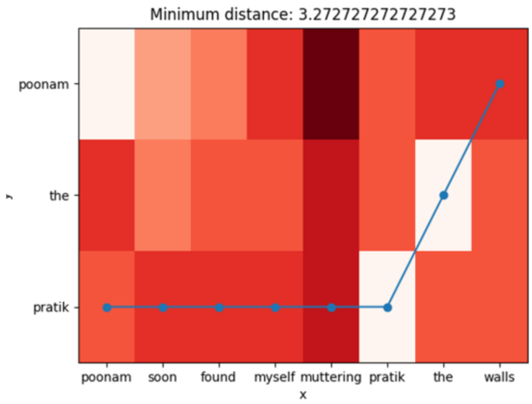


Figure 12- Result of traceback route.

ii. Speech command prediction:

Voice command detection program will need a test argument to identify if user pass sample audio ‘a’ path the respective trained KNN classifier will Predict audio ‘a’ and vice-versa for audio prediction ‘b’ will work.

In below result audio ‘b’ is predicted by algorithm as user passed the same argument for testing.

```
C:\Users\admin\AppData\Local\Programs\Python\Python311\python.exe D:\F
Normalized distance between the two sounds: 141.96766490936278
Time used: 7.359142699977383s
1.0
Predict audio is: 'b'

Process finished with exit code 0
```

Figure 13- Result of voice command prediction

iii. Pothole detection result-

Pothole detection logic will display the result for all the images which matched and not matched with the reference pot hole image.

In the result below we can see that result in the form of image number – Matched/No matched.

Algorithm will also calculate the DTW Distance and can be seen in the result image.

```
C:\Users\admin\AppData\Local\Programs\Python\Python311\pyth
Image 1: Matched (DTW Distance: 7.6627)
Image 2: Not Matched (DTW Distance: 14.1412)
Image 3: Matched (DTW Distance: 9.4000)
Image 4: Matched (DTW Distance: 11.0824)
Image 5: Not Matched (DTW Distance: 15.5647)
Image 6: Not Matched (DTW Distance: 18.3412)
Image 7: Matched (DTW Distance: 13.1412)
Image 8: Not Matched (DTW Distance: 17.2941)
Image 9: Not Matched (DTW Distance: 18.4902)
Image 10: Matched (DTW Distance: 1.5804)
Image 11: Matched (DTW Distance: 11.2000)
Image 12: Matched (DTW Distance: 12.3294)
Image 13: Matched (DTW Distance: 8.1725)
Image 14: Matched (DTW Distance: 13.7176)
Image 15: Matched (DTW Distance: 10.3765)
Image 16: Not Matched (DTW Distance: 14.3961)
Image 17: Matched (DTW Distance: 0.0431)
Image 18: Matched (DTW Distance: 9.6157)
Image 19: Matched (DTW Distance: 12.5882)
Image 20: Matched (DTW Distance: 8.9137)
Image 21: Matched (DTW Distance: 12.4980)

Process finished with exit code 0
```

Figure 14- Result of pothole detection.

IV. CONCLUSION AND FUTURE SCOPE:

We have described and demonstrated the working of dynamic time warping algorithm and worked on two use cases namely shape (pothole detection), voice command recognition at low level. In the future we hope to extend the technique to higher level representations such as Voice-Activated Smart Devices, Security Systems as Voice recognition can enhance the security of access control systems, ensuring that only authorized individuals gain entry to restricted areas.

REFERENCES

[1] R. Bellman and R. Kalaba, “On adaptive control processes,”Automatic Control, IRE Transactions on, vol. 4, no. 2, pp. 1–9,1959. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=1104847

[2] C. Myers, L. Rabiner, and A. Rosenberg, “Performance tradeoffsin dynamic time warping algorithms for isolated word recognition,”Acoustics, Speech, and Signal Processing [see also IEEE Transactions onSignal Processing], IEEE Transactions on, vol. 28, no. 6, pp. 623–635,1980. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=1163491

[3] H. Sakoe and S. Chiba, “Dynamic programming algorithm optimizationfor spoken word recognition,” Acoustics, Speech and Signal Processing,19 IEEE Transactions on, vol. 26, no. 1, pp. 43–49, 1978. [Online]. Avail-able: http://ieeexplore.ieee.org/xpls/abs all.jsp?arnumber=1163055

[4] A. Efrat, Q. Fan, and S. Venkatasubramanian, “Curve matching, timewarping, and light fields: New algorithms for computing similaritybetween curves,” J. Math. Imaging Vis., vol. 27, no. 3, pp. 203–216, April2007. [Online]. Available: http://dx.doi.org/10.1007/s10851-006-0647-0

[5] C. C. Tappert, C. Y. Suen, and T. Wakahara, “The state of theart in online handwriting recognition,” Pattern Analysis and MachineIntelligence, IEEE Transactions on, vol. 12, no. 8, pp. 787–808, 1990.[Online]. Available: http://dx.doi.org/10.1109/34.57669

[6] A. Kuzmanic and V. Zanchi, “Hand shape classification using dtwand lcss as similarity measures for vision-based gesture recognitionsystem,” in EUROCON, 2007. The International Conference on”Computer as a Tool”, 2007, pp. 264–269. [Online]. Available:http://dx.doi.org/10.1109/EURCON.2007.4400350

[7] A. Corradini, “Dynamic time warping for off-line recognition ofa small gesture vocabulary,” in RATFG-RTS ’01: Proceedings ofthe IEEE ICCV Workshop on Recognition, Analysis, and Trackingof Faces and Gestures in Real-Time Systems (RATFG-RTS’01).Washington, DC, USA: IEEE Computer Society, 2001. [Online].Available: http://portal.acm.org/citation.cfm?id=882476.883586

[8] V. Niennattrakul and C. A. Ratanamahatana, “On clusteringmultimedia time series data using k-means and dynamic timewarping,” in Multimedia and Ubiquitous Engineering, 2007. MUE ’07.International Conference on, 2007, pp. 733–738. [Online].

Available:http://dx.doi.org/10.1109/MUE.2007.165 20
[9] J. Gu and X. Jin, “A simple approximation for dynamic time warpingsearch in large time series database,” 2006, pp. 841–848. [Online].Available: http://dx.doi.org/10.1007/11875581 101
[10] C. Bahlmann and H. Burkhardt, “The writer independent onlinehandwriting recognition system frog on hand and cluster generativestatistical dynamic time warping,” IEEE Trans. Pattern Anal. Mach.Intell., vol. 26, no. 3, pp. 299–310, 2004. [Online]. Available:http://dx.doi.org/10.1109/TPAMI.2004.1 262308
[11] T. Kahveci and A. Singh, “Variable length queries for time seriesdata,” in Data Engineering, 2001. Proceedings. 17th InternationalConference on, 2001, pp. 273–282. [Online]. Available: http://dx.doi.org/10.1109/ICDE.2001.914838
[12] T. Kahveci, A. Singh, and A. Gurel, “Similarity searching for multi-attribute sequences,” in Scientific and Statistical Database Management,2002. Proceedings. 14th International Conference on, 2002, pp. 175–184.[Online]. Available: http://dx.doi.org/10.1109/SSDM.2002.1029718
[13] W. Euachongprasit and C. Ratanamahatana, “Efficient multimediatime series data retrieval under uniform scaling and normalisation,”2008, pp. 506–513. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-78646-7 49
[14] Tizzo, A. (2022, November 9). search medium. Retrieved from Dynamic Time Warping for Time Series Classification: Dynamic Time Warping for Time Series Classification | Eni digiTALKS (medium.com)