# Ownership Protection of Shape datasets with Distance Based Preservation

**Kalaiselvi. N, Meenatchi. A, Priyadharsini. T, Ponnudurai. N**

***Abstract*— Security is the one of the reasonable property of this topic with important technological and legal facets. We provide mechanisms for establishing the ownership of a dataset consisting of multiple objects. The algorithms also preserve important properties of the dataset, which are important for mining operations, and so guarantee both right protection and utility preservation. We consider a right-protection scheme based on watermarking. Watermarking may distort the original distance graph. Our watermarking methodology preserves important distance relationships, such as: the Nearest Neighbors (NN) of each object and the Minimum Spanning Tree (MST) of the original dataset. This leads to preservation of any mining operation that depends on the ordering of distances between objects, such as NN-search and classification, as well as many visualization techniques. We prove fundamental lower and upper bounds on the distance between objects post-watermarking. In particular, we establish a *restricted isometry property*, i.e., tight bounds on the contraction/expansion of the original distances. We use this analysis to design fast algorithms for NN-preserving and MST-preserving watermarking that drastically prune the vast search space. We observe two orders of magnitude speedup over the exhaustive schemes, without any sacrifice in NN or MST preservation.**

***Index Terms*— Minimum Spanning Tree (MST), Nearest Neighbors (NN), Restricted Isometry Property (RIP), watermarking.**

## I. INTRODUCTION

The protection of intellectual property rights became recently a pressing need especially with the rapid growth of transmission techniques. A watermarking technique based on the image content. We choose edges for watermark insertion. The image is decomposed into a set of sub bands and each one is individually marked. Watermarking may distort the original distance graph. Under the development of transmission techniques, data became, volatile and easily processed. Our watermarking methodology preserves important distance relationships .Duplication of any digital data, easier than before, made the definition of protection protocols an important issue to prevent illegal use of such content. Consequently, many watermarking techniques have been developed. The digital watermarking defined as an approach to insert an invisible and robust mark in a medium, is the most efficient technique for intellectual property rights protection.

The concept of digital watermarking is defined as embedding a digital signature or a digital watermark in a document to assert its ownership. The insertion step may be performed in spatial [2,3], frequency [4,5,6] or multi resolution domains. The watermark can then be extracted from the watermarked media to identify the owner. This step can be developed in two different ways: using the original image or not. Each digital watermarking technique must satisfy three essential requirements: perceptual transparency, robustness to attack and capacity of the embedded signature. In fact, the quality of watermarked data must be preserved so that no distortion of the original data should be visible and the perceptual quality of the host media must be the same as the original one. The human visual system may be exploited to improve the mark invisibility. The robustness of the signature is also very important. The embedded signature must be still recoverable and recognizable by the user after many operations including compression and filtering. Moreover, good capacity is very important to prove that the mark should be able to contain significant information.

Watermarking allows the user to hide innocuous pieces of information inside the data. The value of watermarking becomes increasingly important because of the proliferation of digital content, and because of the ease of data sharing particularly through data clouds.

## II. RELATED WORK

Most watermarking methods for images and video have been proposed are based on ideas from spread spectrum radio communications, namely additive embedding of a pseudo-noise watermark pattern, and watermark recovery by correlation [1]. Even methods that are not presented as spread spectrum methods often build on these principles. We review proposed attacks on spread spectrum watermarks are systematically. Further, modifications for watermark embedding and extraction are presented to avoid and counterattack these attacks.

A digital watermark is information that is imperceptibly and robustly embedded in the host data such that it cannot be removed [2]. A watermark typically contains information about the origin, status, or recipient of the host data. The basic concepts of watermarking systems are outlined and illustrated with proposed watermarking methods for images, video, audio, text documents, and other media. Finally, a few remarks are made about the state of the art and possible future developments in watermarking technology.

**Kalaiselvi.N** undergoing her Master of Computer Science and Engineering in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu,India.

**Meenatchi.A** undergoing her Master of Computer Science and Engineering in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.

**Priyadharshini.T** obtained her M.E., M.B.A., degree from various recognized universities. She is currently working as a associate professor in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.

**Ponnudurai.N** obtained his M.E., degree from various recognized universities. He is currently working as a associate professor in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.

An effective watermarking technique geared for relational data [3]. This technique ensures that some bit positions of some of the attributes of some of the tuples contain specific values. The tuples attributes within a tuple, bit positions in an attribute, and specific bit values are all algorithmically determined under the control of a private key known only to the owner of the data. This bit pattern constitutes the watermark. Only if one has access to the private key can the watermark be detected with high probability. The watermark can be detected even in a small subset of a watermarked relation as long as the sample contains some of the marks.

A new and flexible approach for privacy-preserving data mining that does not require new problem-specific algorithms, since it maps the original data set into a new anonymized data set [4]. These anonymized data closely match the characteristics of the original data including the correlations among the different dimensions. We will show how to extend the method to the case of data streams. We present empirical results illustrating the effectiveness of the method. We also show the efficiency of the method for data streams. The Nearest Neighbor of a given sample in approximately constant average time complexity (i.e. independent of the data set size) [5]. The algorithm does not assume the data to be structured into any vector space, and only makes use of the metric properties of the given distance, thus being of general use in many present applications of Pattern Recognition.

In most watermarking applications, the marked data is likely to be processed in some way before it reaches the watermark receiver. An embedded watermark may unintentionally or inadvertently be impaired by such a processing. Other types of processing may be applied with the explicit goal of hindering watermark reception. In watermarking terminology, an attack is any processing that may impair detection of the watermark or communication of the information conveyed by the watermark.

## III. WATERMARKING ATTACKS

### A. Geometric Attack

This attack do not actually remove the embedded watermark itself, but intend to distort the watermark detector synchronization with the embedded information. The detector could recover the embedded watermark information when perfect synchronization is regained. However, the complexity of the required synchronization process might be too great to be practical.

### B. Cryptographic attack

The cryptographic attack aim at cracking the security methods in watermarking schemes and thus finding a way to remove the embedded watermark information or to embed misleading watermarks. One such technique is the *brute-force* search for the embedded secret information. Another attack is *Oracle attack*, which can be used to create a non-watermarked signal when a watermark detector device is available. Practically, application of these attacks is restricted due to their high computational complexity.

## IV. EXISTING ALGORITHM

### A. N-Preservation

The NN-P problem can thus be solved by find- ing the solution of a system of quadratic inequalities, for which we provide the NN-Preservation algorithm. The main idea is that by performing all pairwise violation checks, we compute a set of feasible powers (i.e., powers that preserve the NN property) for each object by successive updates. We can then compute the largest power that belongs in the set of feasible powers of at least a fraction of $1-\tau$ of the objects in D.In the sequel, let solve $f(p) \leq g(p) \mid S$ return the subset of watermark embedding powers in S, such that $f(p) \leq g(p)$,

where $f, g$ are quadratic functions of $p$ in $S$.

```
INPUTS: $D, W, p_{min}, p_{max}, \tau$
OUTPUT: $p^*$
NN(D) = find 1-Nearest Neighbors of D
for all $x \in D$ do
    feasible_powers(x) = $[p_{min}, p_{max}]$
    for all $y \in D, y = x, y = NN(x)$ do
  feasible powers(x) = $D^2$
end if
end for
end for
$(x, NN(x)) \leq D_p(x, y) \mid$ feasible_powers(x)
```

### B.MST-Preservation

A similar rationale applies for the MST preservation algorithm. The algorithm progressively removes infeasible powers, under which the MST properties are violated. Let $T(D, E)$ be a Minimum Spanning Tree of the distance graph of dataset D, where E is the set of $|D| - 1$ edges composing the tree. If we remove an edge $e = (x, y) \in$, e original tree into connected components $U_e$and $V_{e.}$ Since T is a Minimum Spanning Tree, such edge = $(x, y)$ has the property of being a shortest edge that connects $U_e$ with $V_{e.}$ If for edge e = (x, y) we use D(e) to denote the Euclidean distance $D(x, y)$, for every edge $e \in$ Eit holds that

$$D(e) \leq D(u, v) \quad \forall u \in U_e, \forall v \in V_e$$

This defining property of an edge e of the MST is preserved after the watermark embedding with watermark W and power p if and only if

$$D_p(e) \leq D_p(u, v) \forall u \in U_e, \forall v \in V_e$$

The MST-P Watermarking Problem can be solved again via a system of quadratic inequalities. Algorithm 2 provides the pseudocode for this process.

```
INPUTS: $D, W, p_{min}, p_{max}, \tau$
OUTPUT: $p^*$
T (D, E) = find MST of D (using Kruskal's algorithm)
for all $e \in E$ do
    feasible_powers (e) = $[p_{min}, p_{max}]$
    for all $u \in U_e$ do
            feasible_powers (e)
  $D_p(x, y) \mid$ feasible_powers(x)
        end for
    end for
end for
```

## V. PROPOSED ALGORITHM

### A. Fast NN-Preservation

We state and prove a sufficient condition for preservation of the Nearest Neighbor of an object x. We show that if the ratio of the Euclidean distance between x and some other object y in the original dataset over the distance D(x, NN(x)) is greater than or equal to a threshold depending solely on $p_{max}$, then y does not violate the NN of x after the watermark embedding, regardless of the details of the dataset or the watermark embedding. In such case, y can be safely removed from the violation checks, without computing the quadratic $D^2(x, y)$.

```
INPUTS: $D, W \in W(D), p_{min}, p_{max}, \tau$
OUTPUT: $p^*$
NN(D) = find 1-Nearest Neighbors of D
for all $x \in D$ do
    feasible_powers(x) = $[p_{min}, p_{max}]$
    for all $y \in D, y = x, y = NN(x)$ do
$D(x,NN(x)) < 1 - p_{max}$ then

end if
end for
end for
```

### B. Fast MST-Preservation

A similar sufficient condition holds for MST preservation. For an edge e in an MST of the distance graph of D, and objects $u \in U_e, v \in V_e$, if

$$D(u, v) \leq 1 + p_{max}$$

$$D(e) \leq 1 - p_{max}$$

then edge (u, v) does not violate the MST at edge e after the watermark embedding, for all watermarks $W \in W(D)$ and embedding powers $p \in [p_{min}, p_{max}]$.

The proof is similar to the proof of Corollary 1 and is therefore omitted. Algorithm 4 uses Corollary 2 to prune the search space of the naive MST-Preservation algorithm.

```
INPUTS: D, $W \in W(D)$, pmin, pmax, $\tau$
OUTPUT: $p*$
T (|D|, E) = find Euclidean MST of D (using Kruskal's
algorithm)
for all $e \in E$ do
feasible_powers (e) = [pmin, pmax]
for all $u \in Ue$ do
for all $v \in Ve$ do
feasible_powers(x) = solve $D^2(e) \leq D^2(u, v)$ |
feasible_powers(x)
end if
end for
end for
end for
$p* = \max \{p: \{e: p \notin$ feasible_powers (e)$\} \leq \tau \cdot (|D|
- 1)\}$
```

## VI. MODULE DESCRIPTION

### A. Implementing Data publisher portal

In this module we are going to create data owner publishing portal. In this portal help of uploading data's from owner and sharing to other users in this portal via. Data owner uploading various kinds of data's and various types of formats. Normally data owner publishing data from our other users without any constraints. In this portal help of identifying user information, Data sharing information, data view and accessing information and recommended sites information and etc.

### B. Construction of datasets using MST

In this module help of identifying data objects and similarity of data objects help of datasets. Dataset is a temporary storage and consist of multiple in disciplinary data objects. Our Dataset construction based on the minimum spanning tree approach. In this technique help of classifying various objects (labeled and unlabeled) and constructing tree structure help of learning algorithms.

### C. Creation of watermarking

In this module we are going to create watermark shapes help of datasets. Watermarking is the one type of security technique it helps of improving data confidentiality and data integrity. A digital watermark is information that is imperceptibly and robustly embedded in the host data such that it cannot be moved. We have uses a spread-spectrum approach this embeds the watermark across multiple frequencies of each object and across multiple objects of the dataset. As such, it renders the removal of the watermark particularly difficult without substantially compromising the data utility.

### D. Analyzing distance and alteration due to watermarking

In this module, we are going to analyze calculation of distortion of object distances due to watermarking. We has provided closed-form expressions for the Euclidean distance between two objects before and after the watermark embedding. Then, we derive tight lower and upper bounds on the distortion of distances due to watermarking, which hold uniformly for any watermark compatible with the given dataset. Our analysis gives crisp insight about the effect of watermarking on object distances and forms the basis for the fast algorithms for NN and MST preservation proposed scheme.

### E. Performance Evaluation

In this module we are going to perform data owner right-protection scheme based on additive watermarking preserving the NN structure. The Present work presents a multiplicative watermarking framework; therefore the algorithm analysis is quite different. More importantly, here we examine fundamental properties of distance distortion due to multiplicative watermarking. The current work represents an extended version we augment the original publication by presenting a comprehensive theoretical analysis of the distance distortion.

## VII. CONCLUSION

Watermarking techniques have been used to right protect databases. There is also relevant work on watermarking for streaming time series. They examine watermarking on a single numerical sequence, as opposed considering a collection of sequences and aiming at maintaining the original pair wise relationships. Also they do not consider resilience to geometric data transformations. Our setting poses additional challenges compared to traditional watermarking or privacy-preservation techniques. Not only do we work on the perturbed data, but more importantly, we provide guarantees on preservation of distance properties. Although the focus in this work is on preservation of the NN and the MST, our formulation is applicable on any mining operation that depends on the order of distances between objects. This makes our approach relevant for a wide range of distance-based learning, search, and mining algorithms. A right-protection scheme based on additive watermarking preserving the NN structure was presented by the present work presents a multiplicative watermarking framework therefore the algorithm analysis is quite different. More importantly, here we examine fundamental properties of distance distortion due to multiplicative watermarking. The current work represents an extended version of we augment the original publication by resenting a comprehensive theoretical analysis of the distance distortion. We leverage this analysis to provide fast versions of the exhaustive algorithms presented previously.

## REFERENCES

[1]F. Hartung, J. Su, and B. Girod., “Spread spectrum watermarking: Malicious attacks and counterattacks,” in Proc. SPIE Security Watermarking Multimedia Contents, vol. 3657, San Jose, CA, USA, 1999.

[2]S. J. Shyu, Y. T. Tsai, and R. C. T. Lee, “The minimal spanning tree preservation approaches for DNA multiple sequence alignment and evolutionary tree construction,” J. Combinat. Optim., vol. 8, no. 4, pp. 453–468, 2004.

[3]N. Paivinen, “Clustering with a minimum spanning tree of scale-free-like structure,” Pattern Recognit. Lett., vol. 26, no. 7, pp. 921–930, 2005.

[4] Y. Xu, V. Olman, and D. Xu, “Minimum spanning trees for gene expression data clustering,” Genome Inform., vol. 12,pp. 24–33, 2001.

[5]J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” Sci.,vol. 290, no. 5500, pp. 2319–2323, 2000.

[6]M. Vlachos, C. Lucchese, D. Rajan, and P. S. Yu, “Ownership protection of shape datasets with geodesic distance preservation,” inProc. 11th Int. Conf. EDBT, Nantes, France, 2008, pp. 276–286.

[7] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoo, “Secure spread spectrum watermarking for multimedia,” IEEE Trans.Image Process., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.

[8] J.-P. M. G. Linnartz and M. van Dijk, “Analysis of the sensitivity attack against electronic watermarks in images,” in Proc. 2nd Int.Workshop IH, Portland, OR, USA, 1998, pp. 258–272.

[9] D. Aha, D. Kibler, and M. Albert, “Instance based learning algorithms,” Mach. Learn., vol. 6, no. 1, pp. 37–66, 1991.

[10]V. Solachidis and I. Pitas, “Watermarking polygonal lines using Fourier descriptors,” IEEE Comput. Graph. Appl., vol. 24, no. 3,pp. 44–51, May/Jun. 2004.

[11]P. Das, N. R. Chakraborti, and P. K. Chaudhuri, “Spherical minimax location problem,” Comput. Optim. Appl., vol. 18, no. 3, pp. 311–326, 2001.

[12] G. Economou, V. Pothos, and A. Ifantis, “Geodesic distance and MST-based image segmentation,” in Proc. 12th EUSIPCO, Vienna,Austria, 2004, pp. 941–944.

**Kalaiselvi.N** undergoing her Master of Computer Science and Engineering in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu,India.

**Meenatchi.A** received B.E., computer science and engineering degree from Arasu engineering College, Kumbakonam in 2012.Now currently pursuing doing undergoing her Master of Computer Science and Engineering in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.

**Priyadharsini.T** obtained her M.E., M.B.A., degree from various recognized universities. She is currently working as a associate professor in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.

**Ponnudurai.N** obtained his M.E., degree from various recognized universities. He is currently working as a associate professor in Meenakshi Ramaswamy Engineering College, Thathanur, Tamil Nadu, India.