IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 November 2023, accepted 11 December 2023, date of publication 14 December 2023, date of current version 20 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343190

RESEARCH ARTICLE

Sampling Fingerprints From Multimedia Content Resource Clusters

UMER RASHID[1], SAMRA NASEER[1], ABDUR REHMAN KHAN[2], MUAZZAM A. KHAN[1], GAUHAR ALI[3], NAVEED AHMAD[3], AND YASIR JAVED[3], (Member, IEEE)

[1]Department of Computer Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2]Department of Computer Science, National University of Modern Languages, Lahore 54000, Pakistan
[3]College of Computer and Information Sciences, Prince Sultan University, Riyadh 12435, Saudi Arabia

Corresponding author: Umer Rashid (umerrashid@qau.edu.pk)

This work was supported by Prince Sultan University.

ABSTRACT Nowadays, the growth of multimedia content over the web is exponential. The fingerprints are inconspicuously embedded in multimedia content. The fingerprints can be exploited to trace divergent information from multimedia resources. Sampling fingerprints, particularly from multimedia resources, is challenging since they are complex, heterogeneous, and diverse. This research proposed an approach to sample fingerprints from multimedia resources. Our approach partitions the multimedia content space into converged clusters using variations of Canberra distance and identifies the most diverged samples using Kullback-Leibler (KL) divergence. The resultant clusters represent the information belonging to particular concepts and the diverged samples within the clusters represent multimedia fingerprints. The fingerprint sampling process is leveraged using unsupervised learning algorithms, instantiated across various multimedia descriptors, and tested over standard multimedia datasets. The average results obtained over various standard visual and acoustic datasets reveal 80%, 77%, and 78% accuracy, precision, and recall, respectively, surpassing most of the existing baseline clustering methods such as K-Means, Mean-Shift, and DBSCAN. Furthermore, the rigorousness of the proposed algorithm clustering is evaluated using the internal clustering stability silhouette coefficient and the fingerprint diversity scores. The results unveil a maximum of 94% diversity score. The proposed variation of Canberra distance and KL divergence provides the most stable performance (SD=0.02) and creates promising implications in future multimedia retrieval, summarization, and exploration activities.

INDEX TERMS Algorithms, convergence, clustering, divergence, fingerprints, multimedia, unsupervised.

I. INTRODUCTION

Nowadays, exponential growth in the online production of multimedia content has been observed [1], [2]. The multimedia content in different media formats, i.e., text, audio, image, video objects, etc., collectively accumulated over massive multimedia resources [3]. Multimedia content has associated textual, acoustic, and visual information modalities [4]. Approximately 2.6 exabytes of multimedia content are consumed, replicated, and explored over the online multimedia resources [5]. Almost 82% of the global data traffic over the web is multimedia-based [6]. The contents in different media formats with multiple modalities are archived, retrieved, and interacted with by the web users in everyday exploration activities via search applications [7].

Web users become overwhelmed with multimedia content, causing information overload, which hinders multimedia content exploration and access [8], [9]. Synthesizing vast multimedia resources with an abundance of different media formats and multiple information modalities via computing technologies is challenging [10], [11]. Ensuring users can access specific content from multimedia resources is a challenging endeavor [12]. Additionally, retrieving relevant information from immense piles of multimedia resources over the web becomes cumbersome. The retrieved multimedia content may include irrelevant, redundant, and insignificant content, leading to a partial satisfaction of information needs [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Geng-Ming Jiang.

© 2023 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

The massive amount of information in multimedia resources is undoubtedly invaluable in various user domains and retrieval scenarios [13], [14]. The techniques to access vast multimedia resources are becoming integral to user interaction and exploration scenarios [15], [16]. The exploration scenarios require fingerprint sampling from multimedia resources in the user's exploration activities [17]. Fingerprinting is about extracting information subsets from a divergent information resource that may represent a concept as a whole [18]. The fingerprints are inconspicuously embedded in multimedia content resources and are used to trace precise information from divergent resources [19], [20].

The existing literature broadly defines the multimedia fingerprinting concept in the context of audio content and copyright protection [21], [22]. The former is to provide the effective matching of audio clips, and the latter is to preserve the copyright of the multimedia content. However, the main objective of multimedia fingerprinting is to facilitate the precise identification of massive content via signature matching [23]. In this research, we extended the idea of fingerprinting to solve the problem of multimedia content accessibility in retrieval and exploration contexts. We will generalize the fingerprinting concept to identify the samples from multimedia resources. The samples are fingerprints, which may give a holistic representation of multimedia resources.

This research proposes an approach to sample fingerprints from multimedia resources containing audio-visual content. Our approach initially clusters multimedia content instances into a dynamic number of the most converged clusters. The key representative samples are finally extracted from the most diverged samples as fingerprints based on their convergence in perspective clusters. We employed Canberra distance and Kullback-Leibler divergence measures in clustering and fingerprinting, respectively. The former is to distribute multimedia resources into clusters and later to identify fingerprints from them. We also proposed clustering and fingerprint identification algorithms that employ the variations of Canberra distance and Kullback-Leibler divergence measure, respectively.

Our proposed approach provides a baseline to extract fingerprints from multimedia resources. To our knowledge, we are the first to employ multimedia fingerprinting to ease multimedia content accessibility and exploration. The proposed approach was instantiated over diverse audio-visual standard multimedia datasets. We extracted a variety of audio-visual descriptors from the multimedia contents and employed them in instantiation. The performance of our proposed approach in terms of precision, recall, and accuracy measures was revealed. We also used Mean-Shift, K-Means, and DBSCAN as baseline algorithms in a comparative evaluation. Our approach outperforms other baseline methods. We found that our proposed approach is more accurate and precision-oriented. The silhouette coefficients analysis highlights cluster stability across the different datasets and extracted descriptors. Our proposed approach is generic and effective since the approach is equally applicable across multiple datasets and descriptors.

The rest of the discussion is organized as follows. Section II provides a literature review. Section III discusses the proposed approach. Section IV provides approach instantiation details. Section VI explains the experimental details and results. Section VII provides a comparative discussion. Finally, section VIII concludes the discussion and highlights future research directions.

II. LITERATURE REVIEW

A. MULTIMEDIA RESOURCES

In recent years, multimedia resources have converged over the web due to the emergence and proliferation of advanced computer and communication technologies [24]. The web has become a vast distributed multimedia resource. The multiple media objects have been accumulated over the web as massive multimedia resources that enabled the exploration of several different media types via advanced computing applications, i.e., digital libraries, social media platforms, knowledge-based systems, etc. [25], [26], [27]. The multimedia information resources enable access and interaction with multiple media objects [28]. For example, Google[1] provides users interaction with more than 30 Trillion web pages containing textual content; Flickr[2] enables social interaction with more than 10 billion images; SoundCloud[3] contains 50 million tracks of audio content; YouTube[4] contains more than 800 million videos clips of variable length.

B. FINGERPRINTS

1) FINGERPRINTING: BASIC CONCEPT

Traditionally, fingerprinting involves bio-metric of people's unique physical or biological characteristics required to identify them, e.g., thumb lines, retina, ears, etc. [29], [30]. The fingerprinting concept was first conceptualized from the theory of uniqueness [31]. However, in recent years, fingerprinting has been further employed in source identification, duplicate detection, copyright prevention, etc., in different domains [32], [33], [34], [35]. The research concerning multimedia fingerprints has recently gained the attention of researchers with a significant focus on the audio domain [36], [37]. The same idea of the theory of uniqueness in fingerprinting is also adopted in the context of fingerprinting of multimedia content [38]. The fingerprinting mainly distinguishes perceptually different artifacts on the uniqueness basis from multimedia resources [31]. Fingerprint identification from multimedia resources can be determined as extracting a subset of information as representatives of a multimedia resource [39].

[1] https://www.google.com/
[2] https://www.flickr.com/
[3] https://soundcloud.com/
[4] https://www.youtube.com/

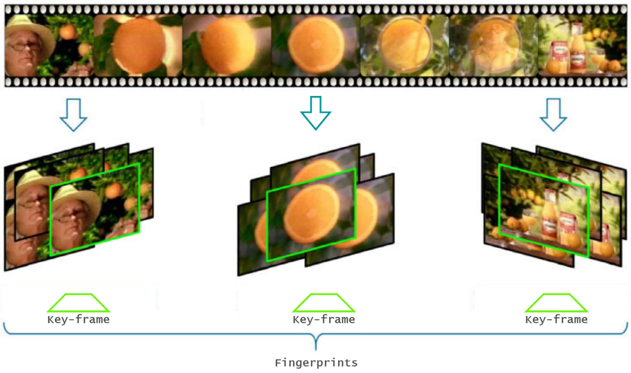


FIGURE 1. **Analogy of multimedia fingerprints.**

2) FINGERPRINTING: ANALOGY

The fingerprinting approach mainly identifies a set of samples from the large resource sets [40]. In a way, fingerprinting comprises selective representative examples from the original datasets or resources [41]. Figure 1 illustrates the Analogy of fingerprint identification. As it emerges from Figure 1 that fingerprinting involves the selection of representative samples from information resources that can further be used in comparison, analysis, and management. The fingerprinting captures vital information from the original dataset more efficiently than any random sampling technique [13]. Fingerprints dependably and effectively portray the whole dataset and address the vital issues in scientific data analysis, having diverse utilization in artificial intelligence, signal processing, data recovery domains, etc. [18].

C. MULTIMEDIA FINGERPRINTS

The proliferation of multimedia data creates challenges in accessing, interacting, and exploring massive multimedia resources [11], [42], [43]. However, fingerprinting is a non-trivial task that enhances human understanding of information resources via a smaller set of representatives identified as samples [44], [45]. Multimedia content available on the web is large and highly redundant and could be represented by a relatively small subset [36], [45], [46], [47]. The relevant and representative subset that demonstrates the global view of the entire resource can be nominated as fingerprints [13]. The identified fingerprints can be further used in processing since they precisely indicate the possible attributes of a collection. In multimedia resources, the fingerprints exist as a condensed content-based mark that synopsis content and provides evidence of uniqueness [48]. In multimedia resources, fingerprinting can be categorized into acoustic and visual.

1) ACOUSTIC FINGERPRINTS

Audio fingerprints have become popular because they permit the detection of audio self-reliant from its structure. However, it may not include the meta-data requirements [49]. An acoustic fingerprint is a digital summary generated from an audio clip. The objective is to locate an audio clip or similar from the audio database [50]. Anguera et al. computed masks near the spectral peaks in the spectrogram for robust audio fingerprinting [51]. Yu et al. proposed hybrid high-performance data structures for indexing massive amounts of audio fingerprinting data for efficient search [52]. Ouali et al. quantized spectrogram regions into a series of horizontal and vertical slices, which are then represented as 48-dimensional fingerprints [53]. Malekesmaeili et al. computed scale-invariant features from two-dimensional time-chroma representations of spectrogram patches [54]. Saravanos et al. proposed a novel audio fingerprinting technique based on the expression of audio signals by establishing a dictionary [55]. Li et al. proposed a compact representation for audio fingerprints executed from local linear embedding that is further utilized in the retrieval task [36].

2) VISUAL FINGERPRINTS

In visual fingerprinting, most work is done in either the context of prototype selection from an image dataset or key-frame extraction from a series of video frames [46], [56]. Traditionally visual fingerprints are employed to verify human identities; the objective is to improve security and safety against impersonal attacks [57]. The concept can be generalized to identify the sample visuals from a diverse set of video objects. Pandya et al. suggested the identification of fingerprints from the visual content by employing texture features, histogram equalization, Gabar filters, and deep learning approaches [58]. Li et al. proposed a fingerprinting method for video retrieval and copy detection by considering convolution neural networks, quantization coding, and feature extraction method [59]. Tseytlina et al. proposed a video fingerprinting plan for content-based video retrieval. The approach was based on Fourier Mellin, features, and compaction [60]. Mandelli et al. dealt with stabilizing video from the recording devices, particularly the method that involves the identification of images or video clips as fingerprints [61].

D. FINGERPRINTING MECHANISMS

Ye et al. proposed multimedia content fingerprinting by employing Cellular Automata (CA), Social Network Analysis (SNA), and Discrete Wavelet Transform (DWT). Mainly the fingerprinting code is produced via SNA [62]. Pinto et al. suggested a novel technique to extract time-spectral descriptors as low-level features from the visuals. They constructed a visual codebook to drive mid-level feature descriptors as fingerprints [63]. Egorova et al. devised identifiable parent property (IPP) coding mechanism to detect the unauthorized distribution of multimedia content. They theoretically generated IPP signatures from multimedia content [64]. Ouali et al. suggested an approach to extract fingerprints from the visual contents by encoding the positions of salient features from the gray-scale transformed images of video objects [65]. Phan et al. targeted minimizing the Sensor Pattern Noise to effectively identify the image fingerprints using Large Scale Sparse Subspace Clustering. The technique produces many clusters from unclustered images [66]. Chen et al. introduced Deep Marks, a framework to retrieve authorship information and unique users from multimedia content as fingerprints. The framework provided the design of a unique codebook and encoding scheme to extract fingerprints from multimedia content [67]. Fan et al. investigated signature codes using the weighted binary adder channel and collusion-resistant to extract the multimedia fingerprinting. They theoretically experimented and generated adversarial traceability fingerprints [18]. Panday et al. devised fingerprint Singular Value Decomposition (SVD) to generate the image fingerprints. The notion was to construct the fingerprint regardless of the rotation of the image [68]. Sharma et al. employed Local Adaptive Binary Patterns (LABP) and Uniform Local Binary Patterns (ULBP) along with Support Vector Machine (SVM) to learn LABP and ULBP features as fingerprints [69]. Ye et al. proposed a novel fingerprinting that decomposes the image fingerprint code via structure fingerprint embedding. The objective was to use a unique image fingerprint to encrypt the images [70].

E. ISSUES AND MOTIVATION

In the present era, multimedia resources are growing exponentially. Contrarily, individuals have constrained resources due to limitations in their manual comprehension. The existing fingerprinting approaches provide the identification of fingerprints from audio-visual content. However, they exploit the low-level representation of multimedia content, such as binary encoding and signal manipulations [18], [63], [68], [70]. Moreover, the prime purpose of the existing fingerprinting approach is to uniquely identify multimedia content for the prevention of unauthorized distribution [64]. Therefore, fingerprinting in the context of multimedia content identification is the least discussed in the literature. Most of the fingerprinting work has been leveraged in the context of source identification, duplicate Selection, similarity-based retrieval, inverted index management, etc. However, almost all of the fingerprinting techniques are for particular domains. The research needs perceptual divergence to provide a comprehensive fingerprint identification mechanism for heterogeneous multimedia content resources. Hence, in this research, we are interested in exploring a generic multimedia fingerprinting approach based on state-of-the-art descriptors that provide representative samples to help aid immense multimedia data exploration.

III. FINGERPRINTING APPROACH

In this research, we extended the fingerprinting analogy to address the issues in identifying fingerprints from multimedia resources. The objective is to suggest a generic approach that locates the most desired samples of multimedia resources as fingerprints. Notably, we extended the multimedia fingerprinting idea to sampling the most diverged fingerprints that may provide the sample-based coverage of the entire multimedia resource via the clusters with the most similar multimedia content. We aim to improve the performance of our generic algorithms and compare them with standard benchmarks that are applicable regardless of domain knowledge. We have proposed a novel approach to identify fingerprints from audio-visual resources. Primarily, we employed an unsupervised approach and developed an algorithmic fingerprint selection strategy from multimedia resources.

We hypothesized that the most convergent samples within clusters might have the most diverged characteristics within an entire multimedia resource. The clusters individually represent the unique concepts within an entire multimedia resource since a multimedia resource is a collection of diverse clusters. The most converged sample within a cluster shows maximum similarity with the other samples of the cluster. In this way, (i) a unique sample as a fingerprint from a cluster can be identified, (ii) the fingerprints can be sampled from the clusters as representative of the entire multimedia resource, and (iii) the sample representations of the entire resource can be recognized as fingerprints of the multimedia resource. We identified the most converged items as multimedia fingerprints from the most diverged clusters. In the following section, We will discuss the approach overview, preliminaries, distance measure, and algorithms employed to sample fingerprints from the multimedia resources.

A. APPROACH OVERVIEW

Our approach sampled the most diverged components from the most converged multimedia clusters, where Components are media objects belonging to a particular multimedia resource type, i.e., text, image, audio, video, etc. We introduced new variations of Canberra distance to identify L_{Most} converged clusters. Alternatively, the proposed variations of Kullback-Leibler divergence identify M_{Most} diverged components from the components of L_{Most} converged clusters. M_{Most} and L_{Most} represent the dynamic number of clusters and sample fingerprints, respectively. Figure 2 demonstrates a schematic overview of our proposed fingerprinting approach.

Our proposed approach dynamically samples the fingerprints from the clusters by accommodating media objects belonging to a particular media type in separate media object spaces (Figure 2 (a)). The components are loaded into media set space (Figure 2 (b)). The media set space is converged into the most relevant components in separate partitions called clusters (Figure 2 (c)). The divergence process is applied to the entire sets of clusters to identify divergent samples (Figure 2 (d)). Amongst the divergent samples, the proposed approach identifies the fingerprints, which are the most discriminating and maximally correlated components in a media object space and clusters (Figure 2 (e)). Finally, the results of the obtained fingerprints are obtained empirically and compared with existing state-of-the-art algorithms (Figure 2 (f)).

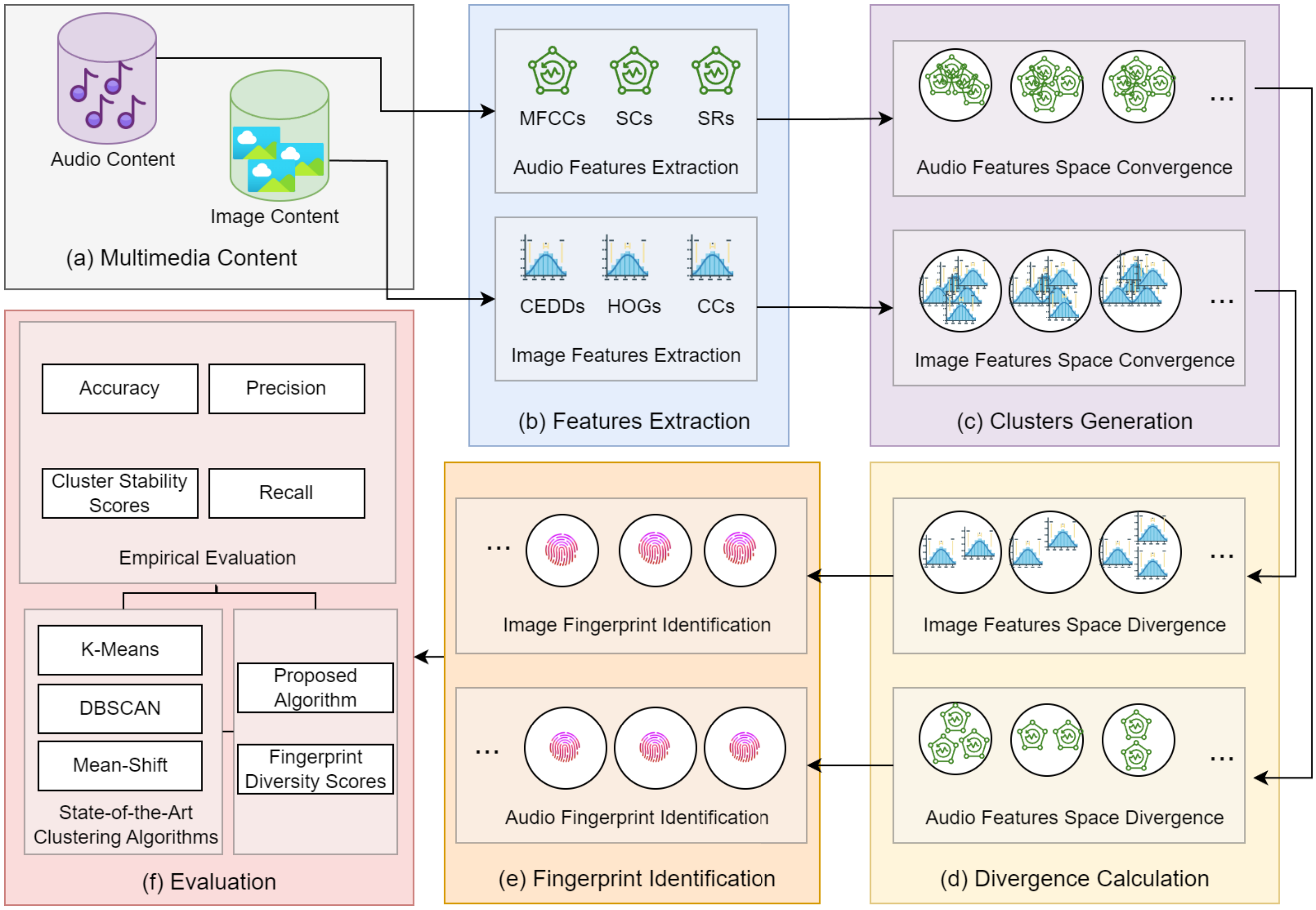

Audio Content
Image Content
(a) Multimedia Content
MFCCs
SCs
SRs
Audio Features Extraction
CEDDs
HOGs
CCs
Image Features Extraction
(b) Features Extraction
Audio Features Space Convergence
Image Features Space Convergence
(c) Clusters Generation
Accuracy
Precision
Cluster Stability Scores
Recall
Empirical Evaluation
K-Means
DBSCAN
Mean-Shift
State-of-the-Art Clustering Algorithms
Proposed Algorithm
Fingerprint Diversity Scores
(f) Evaluation
Image Fingerprint Identification
Audio Fingerprint Identification
(e) Fingerprint Identification
Image Features Space Divergence
Audio Features Space Divergence
(d) Divergence Calculation

FIGURE 2. The overview of the fingerprint sampling approach comprising media (a) object space accommodation, (b) features extraction, (c) cluster generation, (d) sample extraction, (e) fingerprint identification, and (f) approach evaluation.

B. APPROACH FORMALIZATION

Let $T = \{T_i, T_2, T_3, .., T_n\}$ is a media object space containing media objects (components) belonging to a particular type, i.e., either text (α), audio (β), image (γ), or video (δ). The α, β, γ, and δ are disjoint sets. $\forall T_i \in T$ represent a unique components in T. $R = \{R_1, R_2, R_3, \ldots, R_n\}$ is a feature set space extracted $\forall T_i \in T$ and $T_i \cong R_i$. $C = \{C_1, C_2, C_3, \ldots, C_n\}$ is a set of clusters and T_{is} are converged in distinct clusters C_{is} by considering R_{is} similarities in T_{is} and T_{js}, where $C_i \cap C_j = \phi$ and $i \neq C_j$. $C_i = \{T_{i1}, T_{i2}, T_{i3}, \ldots, T_{in}\}$ is a cluster set containing most converged components, where $T = \{C_1 \cup C_2 \cup C_3 \cup, \ldots, C_n\}$. $F = \{F_1, F_2, F_3, \ldots, F_n\}$ is a sample fingerprints set extracted from C, where $F_i \in C_i$, each F_i is unique within a C_i, and $F_{sim} > (T_{sim}) \rightarrow C$ since $F \subseteq T$ and F_{is} clusters are most diverged components of T.

C. DISTANCE MEASURES

1) CANBERRA DISTANCE

Our approach employs basic Canberra distance (d_{cn}) to split $\forall T_i \in T$ into most converged $\forall C_i \in C$. The d_{cn} in a pair of components T_i and T_j is computed as:

$$d_{cn}(T_i, T_j) = \sum_{k=1}^{n} \frac{|R_i - R_j|}{|R_i| + |R_j|} \qquad (1)$$

Equation 1 can not give the component-wise mean Canberra distance (C_oCn_{mean}) of a component T_i concerning all other components in set S, where S are non-clustered components and $\forall S_k \in T$. The C_oCn_{mean} is computed as:

$$C_oCn_{mean}(T_i, S_j) = \sum_{k=i+1}^{n} d_{cn}(T_i, S_j)/|S_j|$$
$$\therefore S_j = \{T_{i+1}, T_{i+2}, \ldots, T_n\} \qquad (2)$$

Equation 2 associate individual mean Cn distance $\forall T_i \in T$. In fact, C_oCn_{mean} computes the degree of uniqueness $\forall T_i \in T$. However, the uniqueness of individual components is not normalized; it varies in components, hence only utilized to find a convergence of randomly distributed components into the dynamic number of clusters. We proposed normalization of C_oCn_{mean} to compute a normalized factor, which can be used as threshold values to decide the inclusion of a T_i in a particular C_i. The normalized component-wise mean Canberra distance ($NE_oCan_{mean}T$) $\forall T_i \in T$ is computed as:

$$N_oCn_{mean}(S_{N_{com}}) = Max(C_oCn_{mean}(C_{N_{com}})) - Min(C_oCn_{mean}(S_{N_{com}})) \qquad (3)$$

The $NC_oCan_{mean}T$ is derived by taking the difference between maximum and minimum non-zero values of

$C_oCan_{mean}T_i$ and $\forall T_i \in T$. The equation 10 represents a set of all components that are not converged in any $\forall C_i \in C$ and their $C_oCan_{mean}T_i > 0$. Equation 4 represents the components excluded in C.

$$S_{N_T} = \{S_T\} - \{S_{1_{C_T}} \cup S_{2_{C_T}} \cup S_{3_{C_T}} \ldots \cup S_{L_{C_T}}\} \\ - \{T_{0-N_oCan_{std}}\} \tag{4}$$

In equation 4, S_{N_T} is a set of not-clustered components T_i, where $T_i \in T$ and $T_i \notin C$. $S_{L_{C_T}}$ is a set of all clustered components T_i, where $T_i \in \{T, C\}$; and $T_{0-N_oCan_{std}}$ is a set of all components T_i, where $T_i \in T$ and T_i C_oCn_{mean} distance with respect $T_j \in T = 0$, where $i \neq j$.

2) KULLBACK-LEIBLER DIVERGENCE

We propose variations of Kullback-Leibler divergence as object-wise Kullback-Leibler divergence and normalized object-wise Kullback-Leibler divergence. These variations are used to sample M-Most divergent objects from the objects of L-Most convergent clusters. Kullback-Leibler divergence calculates the degree of dissimilarity between two objects. It can be used to compute the divergence between objects. Kullback-Leibler divergence can be measured between the objects of vectors E_i of objects M_i and M_j as:

$$d_{KL}(M_i, M_j) = \sum_{k=1}^{n} \left(M_jE_k \log \left(\frac{M_jE_k}{M_iE_k} \right) \right) \\ + \sum_{k=1}^{n} \left(M_iE_k \log \left(\frac{M_iE_k}{M_jE_k} \right) \right) \tag{5}$$

The range of the Kullback-Leibler divergence measure is [0,∞]. The lower and upper bound will represent the degree of convergence between the pair of objects (M_i, M_j). The Individual divergence measure of any two individual objects can be calculated using equation 5. This measure can not calculate the Kullback-Leibler divergence of an object concerning all other remaining objects. Kullback-Leibler divergence measure is used to calculate this measure. Equation 6 represents object Wise Mean Kullback-Leibler Divergence of an object ($E_0KL - Divergence_{mean}M_i$) for all other objects.

$$E_oKL - Divergence_{mean}(M_i, E_k) = \sum_{j=1}^{n} d_{KL}(M_i, E_k)/|E_k| \tag{6}$$

($E_0KL - Divergence_{mean}M_i$) can be calculated by dividing the sum of all the Kullback-Leibler divergence of an object M_i with all the objects M_j with the cardinality of cluster object set E_k^k, where j≠i, $E_0KL - Divergence_{mean}M_i$ represent individual divergence of each object for all other cluster objects. $E_0KL - D_{mean}M_i$ represents its uniqueness in the set of objects in a cluster E_k. The proposed variation of Kullback-Leibler divergence only calculates the individual $E_0KL - D_{mean}M_i$ in the set of media objects. It represents only the uniqueness of an object for all other objects in the cluster. The uniqueness of individual objects in the clusters is not normalized; it varies from object to object. It can not only be utilized to calculate the normalized divergence in cluster objects. Normalized object-wise Kullback-Leibler Mean Deviation($NE_0KL - Divergence_{mean}M$) is also proposed. The normalized factor can be used as a threshold value to decide the inclusion of an object in the set of candidates. ($NE_0KL - Divergence_{mean}M$) can be calculated as:

$$N_oKL - D_{mean}(E_{N-obj}) = Max(E_oKL - Dmean(E_{N-obj})) \\ + Min(E_oKL - Dstd(C_{N-obj}))/2 \tag{7}$$

The normalization factor can be calculated by taking an average of the maximum and minimum ($E_0KL - Divergence_{mean}M_i$) for all other objects in the cluster.

D. ALGORITHMS

We developed four novel algorithms to sample the most diverged media objects (instances) as sample fingerprints from the most converged clusters. The algorithms compute instance-wise mean Canberra distance of all the non-clustered instances, instance-wise Kullback-Leibler (KL) standard deviation for all instances in a cluster, instances into a dynamic number of clusters, and sample most divergent instances from the clusters as sample instances. The *Algorithm* 1 computes instance-wise mean Canberra distance as $E_oCan_{mean}M_i$ of all non-clustered instances. It provides a threshold value as a normalization factor during the clustering of the instances. The threshold value is updated dynamically. It is re-computed for the objects remaining in a set after instance inclusion in a cluster. The threshold values are calculated dynamically until the inclusion of all objects in their corresponding most convergent cluster.

Algorithm 1 Canberra Distance Computation

```
Data: Feature Object Space
Result: Uniqueness of Media object and Threshold
        value
$S_{n-objects} \leftarrow \{S_{objects}\} - \{S_{Cluster_{objects}}\}$;
$k \leftarrow 0$;
while $|S_{n-Objects}| > 1$ do
    $M_i \leftarrow S_{n-object}[1]$;
    $\{S_{n-object}\} \leftarrow \{S_{n-objects}\} - \{M_i\}$;
    $d_{Can} \leftarrow 0$;
    foreach ($M_j in S_{n-objects}$) do
        $d_{can} \leftarrow d_{can} + d_{can}(M_i, M_j)$
    end
    $E_oCab_{mean}M_j \leftarrow d_{can}/|S_{n-objects}|$;
end
$NE_oCan_{mean}M \leftarrow max(Cab_{MEAN_j}) - min(Cab_{MEAN_j})$;
```

The *Algorithm* 2 computed Kullback-Leibler standard deviation of all instances in a cluster. The algorithm performs normalization on the calculated instance-wise Kullback-Leibler standard deviations. *Algorithm* 2 provides a threshold value in sampling the pair of instances. The threshold factor is a normalization factor. The threshold value is calculated dynamically for each cluster. The *Algorithm* 1 and *Algorithm* 2 are further exploited to group instances in a dynamic number of converged clusters and sample most diverged instances from the clusters as sample fingerprints in *Algorithm* 3 and *Algorithm* 4, respectively.

Algorithm 2 Kullback-Leibler Normalization

```
Data: Feature object set contained in Clusters
Result: Uniqueness of Media object and Threshold
        value
$C_{n-objects} \leftarrow \{Cluster - objects\}$;
$k \leftarrow 1$;
$while - loop - size = |C_{n-objects}|$;
while ($k! = while - loop - size$) do
    $M_i \leftarrow E_{n-objects}[k]$;
    $d_{KL} \leftarrow 0$;
    $\{C_{n-objects}\} \leftarrow \{C_{n-objects}\} - \{M_i\}$;
    foreach ($M_j in S_{n-objects}$ do
        $d_{KL} \leftarrow d_{KL} + dKL(M_i, M_j)$;
    end
    $\{C_{n-objects}\} \leftarrow \{C_{n-objects}\} U \{M_i\}$;
    $E_o KL - Divergence_{mean} M_j \leftarrow d_{can}/|C_{n-objects}|$;
end
$NE_o KL - Divergence_{mean} \leftarrow$
  $max(Cab_{MEAN_j}) + min(Cab_{MEAN_j})/2$;
```

Algorithm 3 initially takes the first instance of the media object space as cluster centroid. The normalization factor for the centroid is dynamically calculated using the pseudo-code mentioned in algorithm-1. The objects from the set are included in the cluster and excluded from the object set if their Canberra distance for the centroid is less than $NE_o Can_{mean} M$ (computed via *Algorithm* 1). The procedure continues until all the objects are clustered into disjoint sets, and the cardinality of the media object set becomes zero. The *Algorithm* 3 creates the number of clusters dynamically.

Algorithm 4 selects Each cluster object will be chosen individually, and its Kullback-Leibler divergence for all other objects is computed. An object is considered divergent and sampled if its Kullback-Leibler divergence concerning all other objects is more significant than that of the KL-Divergence threshold. A pair of objects were selected from each cluster as a sample candidate. The Algorithm sample an object from the pair of objects with the least mean KL-Divergence for all other cluster objects. This procedure eliminates boundary objects from the candidate samples. This procedure continues cluster by cluster for all the objects until the sampling of all the M-Most divergent objects. The workflow is defined in *Algorithm* 4. The complexity of this algorithm is O(nk) as it will compute all the distances, and from each cluster, the fingerprint will be selected.

Algorithm 3 Centroid Initialization

```
Data: Media objects
Result: Clusters
$\{S_{n-objects}\} \leftarrow \{objects_{objects}\}$;
$l \leftarrow 0$;
$\{C_{L-objects}\} \leftarrow \{Empty\}$;
while ($|S_{n-objects}| \neq 0$) do
    $M_i \leftarrow S_{n-object}[1]$;
    $\{C_l\} = M_i$;
    $\{S_{n-objects}\} \leftarrow \{S_{n-objects}\} - \{M_i\}$;
    foreach ($M_j in S_{n-objects}$) do
        $d_{can} \leftarrow d_{can} + d_{can}(M_i, M_j)$;
        if ($d_{can} < NE_o Can_{mean} M_i$) then
            $\{C_l\} = \{C_l\} U \{M_j\}$;
            $\{S_{n-objects}\} \leftarrow \{S_{n-objects}\} - \{M_i\}$;
        end
    end
end
```

Algorithm 4 Kullback-Leibler Divergence Calculations

```
Data: Clusters
Result: Fingerprints
$\{N_{cluster-sets}\} \leftarrow \{\{C_1\}, \{C_2\}, \{C_3\}, \ldots, \{C_m\}\}$;
$\{S_{samples}\} \leftarrow \{empty\}$;
$l \leftarrow 0$;
$\{C_{L-objects}\} \leftarrow \{Empty\}$;
while ($|N_{cluster-sets}| \neq 0$) do
    $\{C_l\} \leftarrow \{N_{cluster-sets}\}$;
    $M_i = C_l$;
    $\{S_{c-samples}\} = \{\}$;
    foreach ($M_j in C_{ls}$) do
        if ($M_i \neq M_j$) then
            $d_{KL} \leftarrow d_{KL}(M_i, M_j)$;
            if ($d_{KL} > NE_o CKLD_{mean} M_i$) then
                $\{S_{c-samples}\} = \{S_{c-samples}\} U \{C_l\}$;
            end
        end
        $D_n = avg(d_{KL}(S_{c-samples}[1]), \{S_{c-samples}\})$;
        $D_m = avg(d_{KL}(S_{c-samples}[2]), \{S_{c-samples}\})$;
        if ($D_n < D_m$) then
            $\{S_{samples}\} = \{S_{samples}\} U \{S_{c-samples}[1]\}$;
        end
        $\{S_{samples}\} = \{S_{samples}\} U \{S_{c-samples}[2]\}$
    end
end
```

IV. INSTANTIATION

Our proposed approach is instantiated and executed on a publicly available dataset. It also defines the implementation of various measures and approaches. The following subsection briefly overviews fingerprinting instantiation

TABLE 1. Datasets employed in the instantiation of the proposed fingerprinting research.

Media Type	Datasets	# of Elements	Type
Image	I-Search	10k	General Images
	Oxford-IIIT Pet	7k	Pet
Audio	I-Search	10k	General sounds
	audioMNIST	30k	Spoken digits

details, including the dataset, implementation details, experimental setup, and baseline algorithms.

A. MULTIMEDIA DATASETS

We instantiated our approach on different publicly available widely used datasets. The details of the datasets are given in Table1. We have instantiated our approach on image datasets named I-Search[5] and Oxford-IIIT Pet.[6] The I-Search dataset contains 10305 images. The I-search dataset is divided into 51 categories, and each type has approximately 200 images. Similarly, the Oxford-IIIT dataset consists of 37 category pet datasets with approximately 200 images for each class totaling around 7349 images. The I-Search audio dataset consists of 637 audio files classified into 43 categories. Finally, the audioMNIST[7] dataset consists of 30000 audio samples of spoken digits (0-9) of 60 different speakers. These datasets contain the ground truth value which facilitates the calculation of accuracy, precision, and recall measures.

B. DESCRIPTORS

The visual features extracted via routines are mainly implemented in C# and MATLAB. We extracted features from image objects that include the Color and Edge Directivity Descriptor (CEDD), Color-Correlogram (CC), and Histogram of Oriented Gradients (HoG) features. These features were extracted via openCV library.[8] The resultant fingerprints from the CEDD, CC, and HoG are shown in Figure 3. In the case of audio datasets, Spectral Roll-off (SR), Spectral Centroids (SC), and Mel Frequency Cepstral Coefficients (MFCC) are extracted via Librosa[9] library. The *librosa.display* routine is used to display the audio files in different formats, such as wave plots, spectrograms, or color maps. Amplitude and frequency are important parameters of the sound and are unique for each audio, for which *librosa.display.waveplot* routine is used. Figure 4 shows the fingerprints obtained via acoustic descriptors. The information contained in image and audio objects is extracted as vectors and matrices, respectively. These vectors and matrices are finally stored in text files comprising numeric values in corresponding matrices and vectors.

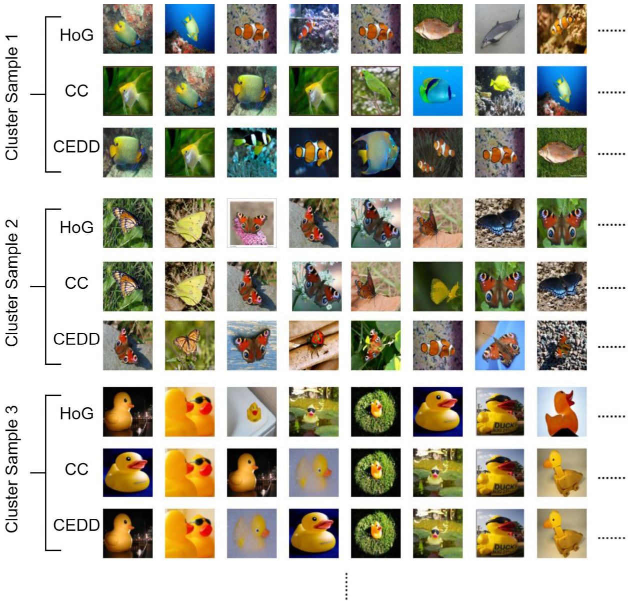


FIGURE 3. The clustering samples obtained for each feature set.

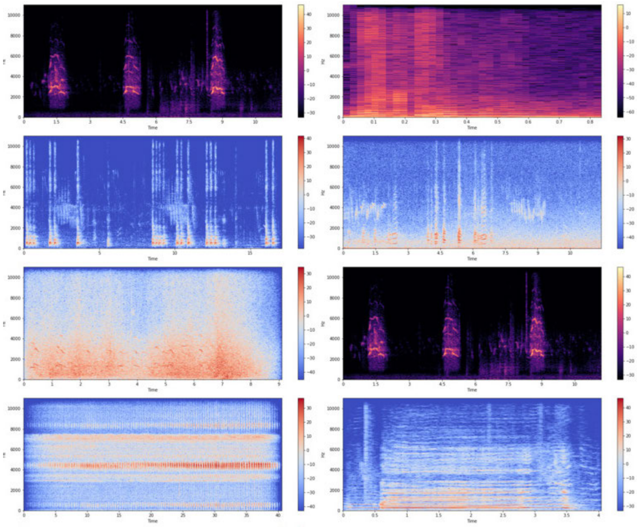


FIGURE 4. The fingerprints obtained via acoustic descriptors.

C. IMPLEMENTATION

Concretely, each media object M_i in the original set M is initially considered a single object, denoted as $M_1, M_2, M_3, .., M_n$. Each object has n features. Initially, the objects are deemed non-clustered objects. The $NE_oCan_{mean}M$ of all the objects in the set are calculated using the previously mentioned $NE_oCan_{mean}M$ - Algorithm. The $NE_oCan_{mean}M$ can be calculated using the $NE_oCan_{mean}M$ - Algorithm. An object is randomly selected from the set of non-clustered objects as a cluster centroid.

The remaining objects in the set whose Canberra distance is less than the normalization factor are deemed cluster elements. The clustering procedure continues until the normalization factor of the last created clusters is not less than the normalization factor of the non-clustered objects. The

[5] https://vcl.iti.gr/dataset/i-search-multimodal-dataset/
[6] https://www.robots.ox.ac.uk/∼vgg/data/pets/
[7] https://www.kaggle.com/datasets/sripaadsrinivasan/audio-mnist
[8] https://opencv.org/
[9] https://librosa.org/

objects in the non-clustered set, whose normalization factor is less than the last made cluster, are included in a new cluster. The procedure automatically stops until the partition of the set of objects into a dynamic number of clusters. It is revealed from the simulation that our proposed clustering approach distinguishes the effective results (Figure 5).

The media objects with the highest average similarity to the other objects will offer the highest content coverage in the set. The working sampling algorithm samples the most similar objects from the cluster created by the clustering algorithm. The algorithm calculates the $E_oCKL - D_{mean}M$ of all the objects in the first cluster. Points with maximum KL Divergence from the cluster are sampled as candidate samples. An object from the candidate samples with maximum KL-divergence for all other objects in the cluster is considered a sample from the cluster. This procedure continues until the objects are sampled from all clusters. Figure 5 demonstrated the fingerprints extracted from the clusters.

V. EVALUATION

A. EXPERIMENTAL SETUP

We have applied our approach on a quad-core Intel (R) Core (TM) i7-6700 @ 3.4 GHz desktop computer with 8GB DDR3 RAM. All methods were implemented in the Python 3.8 version of the Spyder[10] environment with the 64-bit interpreter. Pandas,[11] Sci-Kit,[12] Flask,[13] Keras,[14] and OpenCV[15] libraries.

B. BASELINE ALGORITHMS

We have proposed a new and novel method for clustering and provided an unsupervised approach that only requires prior information like the number of clusters or initial value. However, We have compared the performance of our algorithm with standard benchmarks to determine the efficiency of our algorithm. The algorithms such as Mean-Shift, K-Means, and DBSCAN were utilized to test the effectiveness of our algorithms.

C. EVALUATION MEASURES

The results are evaluated in terms of the quality and performance of clusters. The results are also compared with traditional clustering methods such as K-Means, DBSCAN, and Mean-Shift. The details are discussed in the subsequent subsections. An information-theoretic approach has been conducted for clustering to view it as a series of decisions. To evaluate the performance of clustering, a contingency matrix has been measured as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP is those decisions when similar elements are assigned to the same cluster, whereas TN decisions give dissimilar elements to a different cluster. In that case, two types of errors can be committed. FP and FN. An FP decision assigns dissimilar elements to the same cluster, and an FN decision refers to those decisions when similar items are assigned to different clusters.

According to the literature, precision reveals the effectiveness of clusters. It illustrates the fraction of relevant results among the retrieved results [71]. The appropriate score is divided by a total score to measure the precision. The precision can be measured as $P_c = TP/(TP + FP)$, Where the average precision can be calculated as $AP_c = \sum_{i=1}^{n} P_c/n$. Recall can be discussed as the completeness of outcomes, which can be defined as the fraction of relevant results retrieved over the total number of relevant results. In mathematics, we can define recall as $R_c = TP/(TP + FN)$. Similarly, the average recall rate can be calculated as $AR_c = \sum_{i=1}^{n} R_c/n$. Another measure that we can use to check the performance of clusters is accuracy which tells us how correctly our elements are clustered. Accuracy can be defined as $A_c = (TP+TN)/(TP+TN+FP+FN)$, Where the average accuracy can be defined as $AA_c = \sum_{i=1}^{n} A_c/n$.

VI. EXPERIMENTAL RESULTS

A. BASELINE RESULTS

The evaluation was performed on visual and acoustic datasets. For the former, we used the I-Search dataset and the Oxford-IIIT Pet dataset. For the latter, we used the AudioMNIST and I-Search datasets. The results were obtained on the existing state-of-the-art clustering algorithms (K-Means, Mean-Shift, and DBSCAN) and the proposed algorithm. For the I-Search image dataset, the proposed algorithm achieved the average highest accuracy and recall of 84.39% and 80%, respectively, gained from CEDD embedding. Meanwhile, the highest precision is recorded at 89% in the case of K-Means and CEDD embedding. The detailed results obtained are summarized in Figure 6. Hence, the proposed algorithm outperforms in accuracy and recall over all of the existing baselines in CEDD embedding. Amongst the baselines, the K-Means was observed as the close competitor. However, K-Means only surpassed in case of the precision while the proposed approach was able to outperform the accuracy and recall.

For the Oxford-IIIT Pet image dataset, the CEDD embedding again yielded the best overall accuracy, precision, and recall scores of 87%, 79%, and 77%, respectively, when compared to all existing baselines. The highest recall reported also belonged to the proposed system reported at 82%. Similarly, amongst the baseline algorithms, only the K-Means was able to achieve the best results, with accuracy, precision, and recall reported at 79%, 78%, and 75%, respectively, for the CEDD embedding. Holistically, the proposed algorithm surpassed all the existing baselines for all the other feature sets e.g., HoG and CC. Hence, the proposed algorithm presents a new promising baseline.

[10] https://www.spyder-ide.org/
[11] https://pandas.pydata.org/
[12] https://scikit-learn.org/stable/
[13] https://flask.palletsprojects.com/en/2.2.x/
[14] https://keras.io/
[15] https://opencv.org/

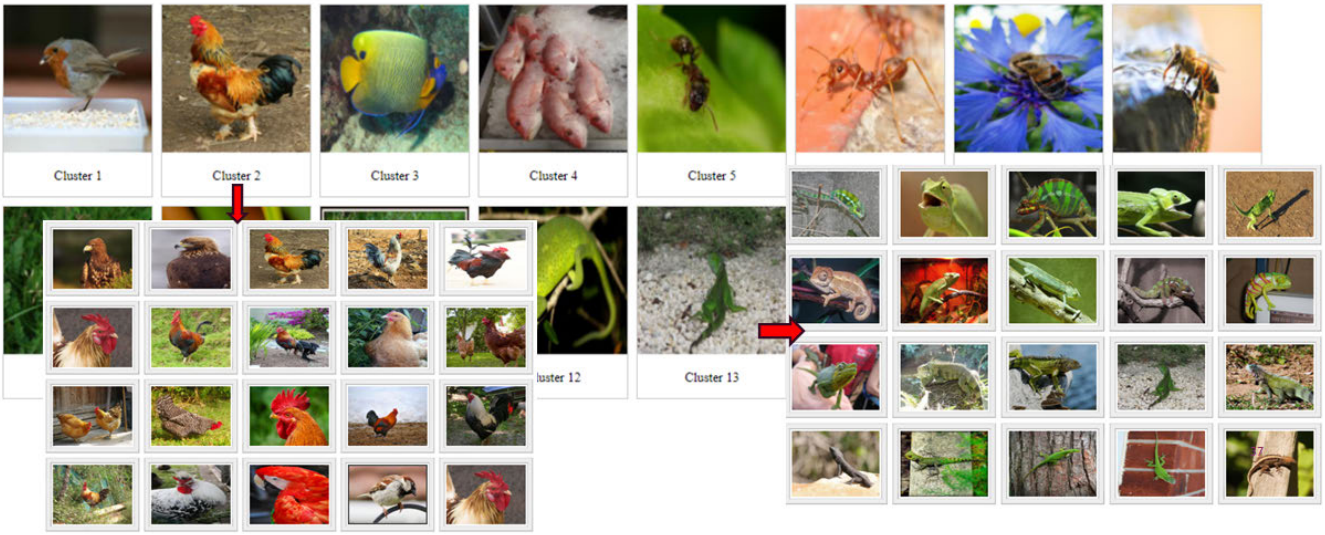


FIGURE 5. Fingerprints-based presentation of clusters.

We also evaluated the proposed approach using the AudioMNIST and I-Search acoustic datasets. For the I-Search audio dataset, the proposed approach and the Mean-Shift clustering algorithms performed the best, achieving 85% accuracy scores on the SC and MFCC feature sets, respectively. For the precision, the K-Means and the Mean-Shift performed marginally (2%) better than the proposed. The recall was the highest in the K-Means clustering algorithm reported at 85%. However, the proposed approach was able to outperform all the existing baselines in the accuracy and recall of the SC feature set. Holistically, the proposed approach performs nearly as well as the existing baselines in the I-Search acoustic dataset. The detailed results are presented in Figure 8.

For the AudioMNIST dataset, the DBSCAN outperforms existing baselines by achieving accuracy, precision, and recall rates of 88%, 88%, and 87%, respectively for the MFCC feature set. The Mean-Shift algorithm closely follows up with a margin of 1% in the accuracy. The proposed and the Mean-Shift performs nearly as well with a difference of 1% recall margin. The proposed approach outperforms the baselines in SC recall by achieving a recall of 79%. The detailed results are shown in Figure 9.

B. APPROACH RESULTS

The proposed approach outperformed the image datasets' results in terms of AA by achieving a maximum of 83%. The AR was also the best amongst all the baselines by achieving a maximum score of 79%. The best average precision was reported in the Oxford-IIIT image dataset of 78%. However, the proposed algorithm stayed marginally behind the K-Kmeans algorithm in the I-Search image dataset. The proposed approach achieved stable performance across the audio datasets. The AA remained the highest in the I-Search audio dataset (81%). For the same dataset, the AR was the second best by 1% margin. The rest of the baseline algorithms demonstrated a variable performance for each instance of the dataset. The averaged results for the image and audio datasets are presented in Figure 10 and Figure 11,

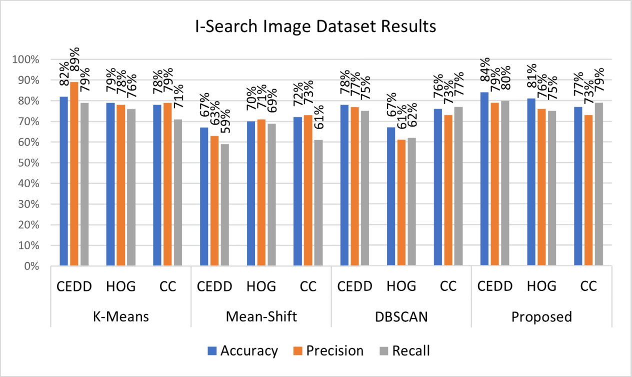


FIGURE 6. Comparison of clustering results on I-Search image dataset.

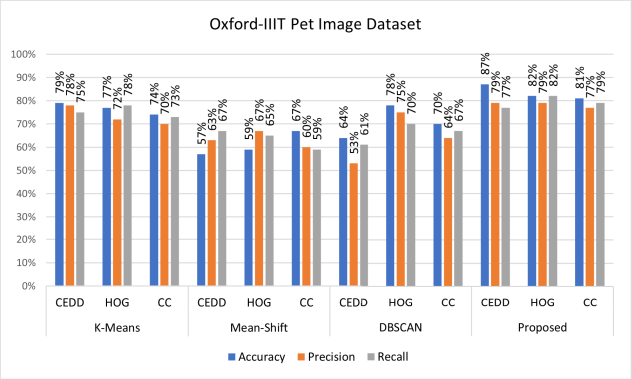


FIGURE 7. Comparison of clustering results on Oxford-IIIT pet image dataset.

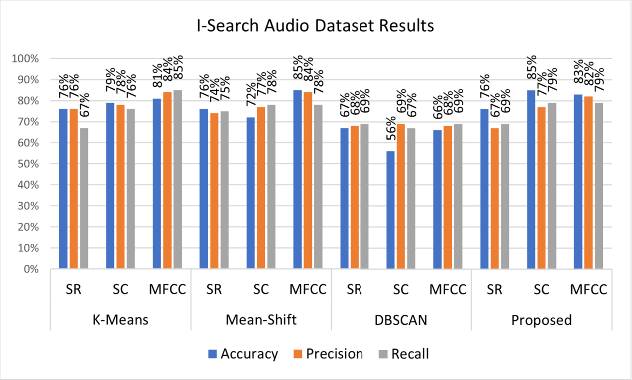


FIGURE 8. Comparison of clustering results on I-Search audio dataset.

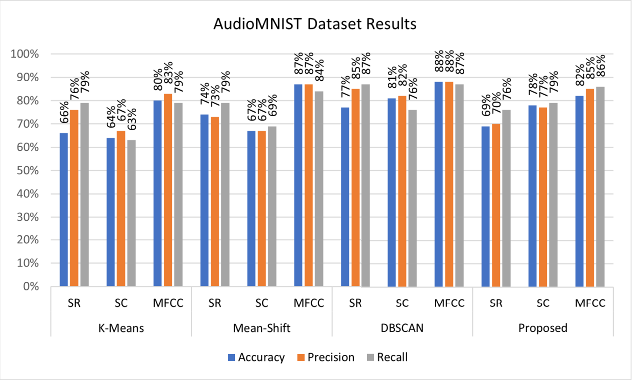


FIGURE 9. Comparison of clustering results on AudioMNIST pet image dataset.

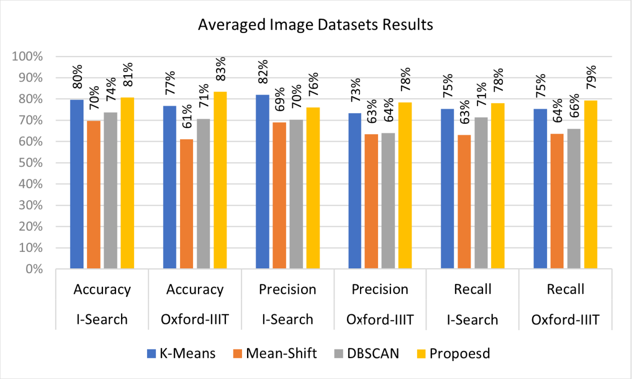


FIGURE 10. Averaged feature set image datasets results of the proposed with state-of-the-art baselines.

respectively. Hence, the proposed approach was able to achieve the best performance for the image datasets with stable results (SD=0.02). Similarly, the proposed approach remained stable with acceptable performance (SD=0.02). The detailed results are provided in Table 2.

C. CLUSTERING ANALYSIS

We also measured cluster performance with Silhouette analysis, which is used to compute the stability of clusters.

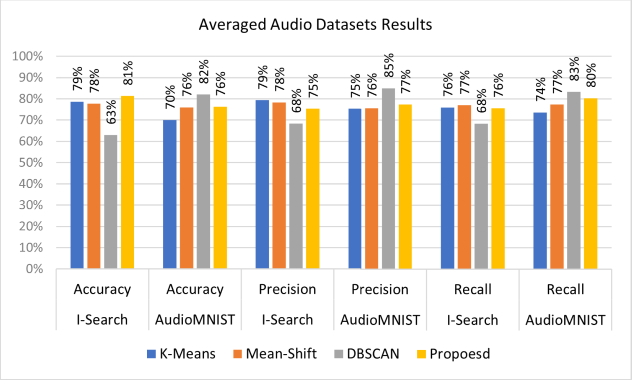


FIGURE 11. Averaged feature set audio datasets results of the proposed with state-of-the-art baselines.

Silhouette analysis is also utilized to measure the interruption distance between clusters. The investigation is done by generating a plot that illustrates the assessment of cluster numbers visually. Mathematically, these can be calculated as $S = (b - a)/\max(a, b)$, Where the term “a” represent the mean distance among all points in the similar cluster and a sample. In contrast, “b” represents the mean distance between all points in the next closest cluster and sample. The scores are in the range of -1 and $+1$. As the value reaches $+1$, it demonstrates precise clustering, whereas the value zero reveals the overlapping of clustering. A higher score defines the stability of clusters.

The silhouette coefficients have also been extracted to test the stability of Clusters. The silhouette analysis is employed to select an optimal standard for n-clusters [72], [73]. It also illustrates the stability of clusters. Figure 12 shows the silhouette plot of the CEDD features set, which presents that the n-cluster value for K-Means of 30, 70, and 90 are poor choices for the given multimedia objects because of the occurrence of clusters with lower average silhouette scores. It also presents that these n-cluster numbers are appalling because of the wide variations in the size of silhouette plots. However, this plot is more indecisive in choosing an n-cluster number between 10 and 50. Moreover, the results illustrate that the choice of 50 is quite beneficial as it has a high score. At the same time, the Mean-Shift algorithm for CEDD features demonstrates that 10 and 50 clusters are not providing promising results. The n-clusters of 30, 70, and 90 indicate a good number of clusters. The result shows that n-clusters of 70 are more practical for evaluating mean shifts. However, the mean shift lacks satisfactory accuracy, precision, and recall results.

The results in Figure 12 demonstrate that the n-cluster values for DBSCAN of 10, 30, 50, and 90 are not a good choice due to the below-average silhouette scores. The extensive fluctuation in the range renders it a poor choice. However, the n-cluster of 70 shows the stability of DBSCAN clustering. The results show that our proposed approach obeys the

TABLE 2. Detailed experimental clustering results for each algorithm and corresponding features evaluation (CEDD/HOG/CC for visual datasets and SR/SC/MFCC for acoustic datasets), where the highest obtained results are bolded.

Algo.	Dataset	I-Search Image			Oxford-IIIT			I-Search Audio			AudioMNIST		
	Features	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
K-Means	CEDD/SR	82%	**89%**	79%	79%	78%	75%	**76%**	**76%**	67%	66%	76%	79%
	HoG/SC	79%	**78%**	**76%**	77%	72%	78%	79%	**78%**	76%	64%	67%	63%
	CC/MFCC	**78%**	**79%**	71%	74%	70%	73%	81%	**84%**	**85%**	80%	83%	79%
	Avg.	80%	**82%**	75%	77%	73%	75%	79%	**79%**	76%	70%	75%	74%
Mean-Shift	CEDD/SR	67%	63%	59%	57%	63%	67%	**76%**	74%	**75%**	74%	73%	79%
	HoG/SC	70%	71%	69%	59%	67%	65%	72%	77%	78%	67%	67%	69%
	CC/MFCC	72%	73%	61%	67%	60%	59%	**85%**	**84%**	78%	87%	87%	84%
	Avg.	70%	69%	63%	61%	63%	64%	78%	78%	**77%**	76%	76%	77%
DBSCAN	CEDD/SR	78%	77%	75%	64%	53%	61%	67%	68%	69%	**77%**	**85%**	**87%**
	HoG/SC	67%	61%	62%	78%	75%	70%	56%	69%	67%	**81%**	**82%**	76%
	CC/MFCC	76%	73%	77%	70%	64%	67%	66%	68%	69%	**88%**	**88%**	**87%**
	Avg.	74%	70%	71%	71%	64%	66%	63%	68%	68%	**82%**	**85%**	**83%**
Propoesd	CEDD/SR	**84%**	79%	**80%**	**87%**	**79%**	**77%**	**76%**	67%	69%	69%	70%	76%
	HoG/SC	**81%**	76%	75%	**82%**	**79%**	**82%**	**85%**	77%	**79%**	78%	77%	**79%**
	CC/MFCC	77%	73%	**79%**	**81%**	**77%**	**79%**	83%	82%	79%	82%	85%	86%
	Avg.	**81%**	76%	**78%**	**83%**	**78%**	**79%**	**81%**	75%	76%	76%	77%	80%

natural portioning as the scores are above average for 10, 30, 50, and 70. However, the outcome of silhouette scores illustrates that the drastic increase in the change of clusters does not give promising results. It also demonstrates the stability of n-clusters between 30 and 50, as they both give favorable scores.

Figure 13 illustrates the silhouette coefficients plot resultant from the CC feature set. The plot demonstrates that the n-cluster value for K-Means of 30 and 90 is not a good choice because of its low score of silhouette coefficients. However, the analysis is more cautious in determining between 10, 50, and 70. Similarly, n-clusters of 70 and 90 are poor choices for the Mean-Shift algorithm. Whereas the n-clusters between 30 and 50 provide more balanced results. Similarly, DBSCAN presents that n-clusters between 50 and 70 provide promising stability of Clusters. Our approach offers stable results for the CC feature set when the value of the n-cluster is between 30 and 50. This feature set also assures that our algorithm does not support drastic expansion in several clusters.

Figure 14 presents the average silhouette coefficients plot for the HoG feature set. The results demonstrate that the n-cluster for K-Means offers the stability of the n-cluster value between 10 and 50. The silhouette analysis is more ambivalent in determining between 10 and 50. However, the 30, 70, and 90 are rejected because of below-average scores. Similarly, the DBSCAN offers stability on the n-cluster value of 70. The Mean-Shift algorithm shows that n-cluster values of 10, 70, and 90 are insufficient. The stability is indicated between 30 and 50 as they demonstrate good scores. Our proposed approach presents that the partition of media objects gives promising results.

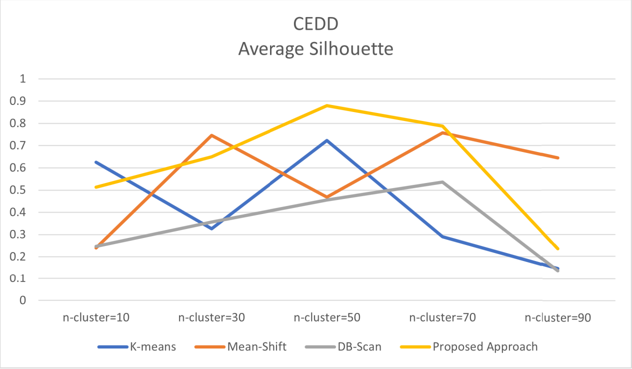


FIGURE 12. Comparison of silhouette coefficients results based on CEDD embeddings on the I-Search image dataset.

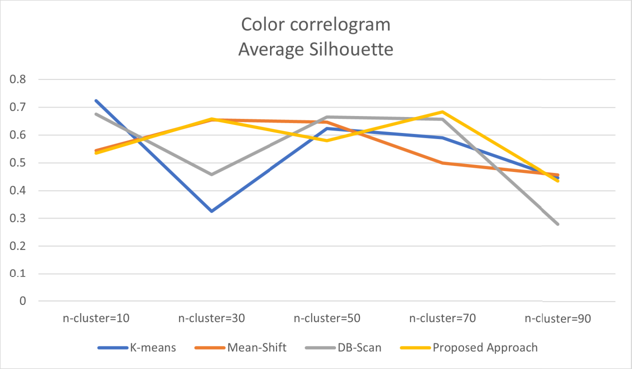


FIGURE 13. Comparison of silhouette coefficients results based on CC embeddings on the I-Search image dataset.

D. FINGERPRINTS ANALYSIS

The core idea of fingerprinting is to capture the divergent elements of a dataset. The fingerprint diversity scores

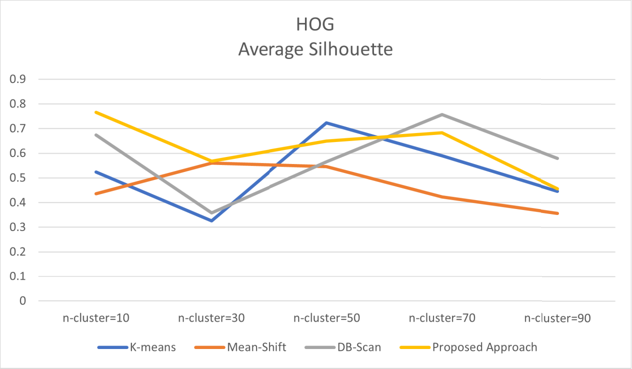


FIGURE 14. Comparison of silhouette coefficients results based on HoG embeddings on the I-Search image dataset.

are calculated based on the variability of the multimedia fingerprints within a cluster. This metric is suitable for assessing the diversity of sampled fingerprints, especially when using techniques like Canberra distance and KL divergence. In contrast, traditional clustering methods like K-Means, DBSCAN, and Mean-Shift do not inherently produce or focus on multimedia fingerprints. Instead, they aim to cluster data points based on their feature vectors. They aim to find natural clusters in the data based on the chosen distance metrics. The proposed approach focused on identifying multimedia fingerprints as a distinct task. The verification of the audio and image fingerprint sampling is therefore leveraged via fingerprint diversity score. The diversity of the result set can be measured based on the distance between the visual and acoustic features. The features such as HoG, CEDD, and CC were extracted from the result set, and their diversity was calculated. Similarly, the acoustic features, i.e., SR, SC, and MFCC, were extracted from the result set, and their diversity was calculated against each feature set. We assume that media objects are encoded in the R feature vector of E-dimensions. The diversity of a result set F with N elements can be formalized as:

$$Diversity(F) = \sum_{i=1}^{E}\sum_{j=1}^{E} var(i)xvar(j)x\delta(i,j) \tag{8}$$

where var(i) and var(j) are the various modules computed as the standard deviation of the feature vector of all M media objects in F. Similarly, $\delta(i,j)$ have been computed as a distance function between i_{th} and j_{th} dimension feature. Here, the $\delta(i,j)$ has been calculated as a reciprocal of similarity among features as $\delta(i,j) = \frac{1}{Q(i,j)}$ which has been computed as:

$$Q(i,j) = \frac{\sum N(r_i, r_j)}{\sqrt{\sum N(r_i^2)}\sqrt{\sum N(r_j^2)}} \tag{9}$$

The similarity is calculated by cosine distance. In this context, we have taken a media object feature vectors as r_i & r_j as the $i_{th}\,j_{th}$ elements. In particular, $Q(i,j)$ can be considered as the probability of i_{th} and j_{th} feature vectors as elements that coincide in all media objects.

TABLE 3. Diversity scores of the proposed fingerprinting approach.

Modality	Dataset	FeatureSet	Diversity Scores
Image	I-Search	HoG	0.86
		CEDD	0.83
		CC	0.77
	Oxford IIIT-Pet	HoG	0.76
		CEDD	0.80
		CC	0.78
Audio	I-Search	SR	0.91
		SC	0.84
		MFCC	0.94
	AudioMNIST	SR	0.83
		SC	0.83
		MFCC	0.86

$$\begin{aligned} P(r_i = i, r_2 = j) &= \sum_F P(r_1 = i, r_2 = j|M)P(F) \\ &= \sum_F P(i|F)P(j|F)P(F) \end{aligned} \tag{10}$$

$P(i|F)$ and $P(j|F)$ indicate the conditional probability of a feature in result set F whereas P(F) comprises the prior probability. if N elements are enclosed in the result set then it is equal to $\frac{1}{N}$ elements.

The diversity scores of the results set are obtained on different audio and image datasets, as shown in Table 3. These diversity scores reveal the dissimilarity of the images contained in the result set. In our experiments, we take the result set as M and extracted features as HoG, CEDD, and CC in the image dataset. For the audio dataset, we extracted MFCC. The normalized score [0,1] informs about the diversity as the "1" score shows the complete diverse set of results. However "0" score exposes highly redundant data. We achieved a maximum of 94% fingerprint diversity scores. The detailed scores are also provided in Table 3 which reveals that our fingerprints are diverse in nature.

VII. DISCUSSION

Our proposed approach identifies the relevant samples as fingerprints, which may demonstrate the multimedia resources. Our approach provides a diversified representation of a multimedia resource. It enhances the user's information-seeking journey to find relevant content from multimedia resources. The more diversity as a whole accommodates the richer information that is accessible to the system, and the higher performance is estimated via our proposed approach. In the fingerprints selection problem, we aim to pick a few representative samples that capture distinguished characteristics of an entire multimedia resource. We proposed a generic approach to seek fingerprints that proportionally reflect specified characteristics exemplified in a target population.

A. CONTRIBUTIONS

The proposed approach outperformed clustering results with the additional advantage of diversity in the final results. The proposed approach consisted of three distinctive phases.

TABLE 4. Statistical significance of the overall obtained clustering results.

Algorithm	Mean	Std. Deviation	Std. Error Mean	Sig. (2-tailed)
K-Means	0.76	0.03	0.01	0.000
Mean-Shift	0.71	0.07	0.02	0.000
DBSCAN	0.72	0.07	0.02	0.000
Proposed	0.78	0.02	0.01	0.000

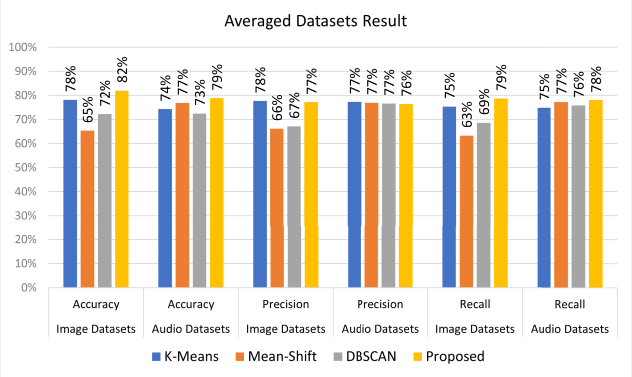


FIGURE 15. Averaged audio and image datasets summarized result.

Firstly, the convergence of the distinct media set space was calculated using a novel variation of Canberra distance. Afterwards, using a variation of Kullback-Leibler divergence, the most distinct samples within each media set space were identified. Finally, the proposed approach was evaluated in terms of clustering and fingerprint identification perspectives. According to the obtained results, the proposed algorithm outperformed the existing state-of-the-art clustering algorithms for image datasets in terms of commutative accuracy (82%) and recall (79%). The precision was also on par with the existing baselines (79%), marginally behind the K-Means by 1%. Similar results were obtained for the audio datasets where the accuracy (79%) and recall (78%) surpassed all the existing baseline results. The precision (76%) was marginally behind (1%) compared to the existing baselines (77%). The proposed approach was able to demonstrate stable performance (SD=0.02) across various feature spaces and datasets, as shown in Table 4. The summarized averaged results obtained for the image and audio datasets are shown in Figure 15, respectively. The averaged fingerprint diversity scores obtained for the image and audio datasets were 80% and 87%, respectively. To the best of our knowledge, no previous fingerprinting approach was introduced that processed diverse multimedia datasets and surpassed the existing baseline algorithms.

B. FINGERPRINTS

This research proposed a unique approach to sampling fingerprints from multimedia resources using a combination of Canberra distance and Kullback-Leibler (KL) divergence to identify the most diverged samples within multimedia content clusters. The proposed approach is different from traditional fingerprinting methods that may rely on other techniques or algorithms for feature extraction and clustering where the aim is to find natural clusters in the data based on the chosen distance metrics. In contrast, the proposed approach focused on identifying multimedia fingerprints as a distinct task. This research leveraged unsupervised learning algorithms to create clusters of multimedia content based on their fingerprints which is not limited to a specific algorithm but is instantiated across various multimedia descriptors, representing flexibility in adapting to different modalities and datasets.

The proposed research was evaluated against performance metrics such as accuracy, precision, and recall, which are commonly used to evaluate the effectiveness of fingerprinting methods. The reported high values (80%, 77%, and 78%) for these metrics indicate the effectiveness of the proposed approach, surpassing existing baseline clustering methods like K-Means, Mean-Shift, and DBSCAN.

Furthermore, the clustering stability was measured using the silhouette coefficient, which represented how well-defined the clusters are. This aspect assesses the quality of clusters and helps ensure that the identified clusters are meaningful and well-separated. Furthermore, the research introduces fingerprint diversity scores for verification of the audio and image fingerprinting samples, which indicate the variability and distinctiveness of the fingerprints. A high diversity score (up to 94%) suggests that the sampled fingerprints are diverse and can capture a wide range of information.

The proposed variation of Canberra distance and KL divergence are reported to provide stable performance with a low standard deviation (SD=0.02). This stability ensures consistent results across different datasets and multimedia types with statistical significance. However, the limitation of this study is the choice of the various clustering algorithms may generate distinct results without standard selection criteria. According to the impossibility theorem, no single clustering algorithm can generate consistent and optimal results for a variety of problems. Hence, this aspect needs thorough investigation to determine the effect of clustering ensemble via detailed comparative analysis.

C. IMPLICATIONS

The proposed approach can also be adapted for different application scenarios. It can be utilized for enhanced content-based multimedia retrieval, exploration of big datasets, and summarization of multimedia result sets. The summarization creates a subset of information by reducing information computationally [17]. The subset signifies the most relevant and valuable information comprised of original content. In the image and video domain context, selecting the most representative images and frames can be depicted as the process of image summarization and video summarization, respectively [13]. The proposed approach can be practiced in image and video summarization as it provides a diverse and representative representation of multimedia objects. Dataset fingerprints can be classified as summaries of a collection.

The proposed approach can also be practiced in the context of Content-based retrieval. It is the process of retrieving contents via similar multimedia content, i.e., an acoustic query returns similar audio files [74]. The practice can be leveraged by matching the query fingerprint and retrieving cluster fingerprint results that demonstrate relatedness to the user query.

Information exploration is the process of searching for and discovering the required information. In the case of information exploration and discovery, users often need clarification and more skills in expressing their information needs via query [75]. To ease the user, relevant information and diverse content can be provided. According to our proposed approach, fingerprints can be provided to the user that contains a diverse collection of relevant information.

VIII. CONCLUSION AND FUTURE RESEARCH

The paper presented the framework for relevant sample identification as fingerprints. The proposed approach identified the m-most convergent items as multimedia fingerprints from n-most divergent clusters. The approach was instantiated across various multimedia datasets over widely recognized descriptors such as MFCC, SR, and SC for acoustic samples, and CEDD, HoG, and CC descriptors for visual samples. A detailed comparison was conducted for the proposed algorithm with existing state-of-the-art clustering techniques such as K-Means, DBSCAN, and Mean-Shift. On average, the proposed variation of Canberra distance and KL divergence achieved 80%, 77%, and 78% accuracy, precision, and recall, respectively, with the most stable clustering performance (SD=0.02) across all the descriptors. The fingerprints were further assessed in the context of diversity, obtaining surpassed scores of 94%. The proposed approach has implications in content-based multimedia retrieval, summarization, and multimedia exploration activities. While the choice of the various clustering algorithms may generate distinct results without standard selection criteria, according to the impossibility theorem, no single clustering algorithm can generate consistent and optimal results for a variety of problems. Hence, in the future, the effect of clustering ensemble can be identified via detailed comparative analysis. Furthermore, the fingerprinting approach can be adopted via deep learning-based embeddings to automate the multimodal feature description.

ACKNOWLEDGMENT

The authors would like to thank the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

REFERENCES

[1] S. Vrochidis, A. Moumtzidou, I. Gialampoukidis, D. Liparas, G. Casamayor, L. Wanner, N. Heise, T. Wagner, A. Bilous, E. Jamin, B. Simeonov, V. Alexiev, R. Busch, I. Arapakis, and I. Kompatsiaris, "A multimodal analytics platform for journalists analyzing large-scale, heterogeneous multilingual, and multimedia content," *Frontiers Robot. AI*, vol. 5, p. 123, Oct. 2018.

[2] S. Sedkaoui, R. Benaichouba, and K. M. Belkebir, "Web analytics and social media monitoring," in *Proc. Int. Conf. Manag. Bus. Web Anal.* Cham, Switzerland: Springer, 2022, pp. 179–192.

[3] D. Kumar, P. Kumar, and A. Ashok, "Introduction to multimedia big data computing for IoT," in *Multimedia Big Data Computing for IoT Applications: Concepts, Paradigms Solutions*. Singapore: Springer, 2020, pp. 3–36.

[4] U. Rashid, K. Saleem, and A. Ahmed, "MIRRE approach: Nonlinear and multimodal exploration of MIR aggregated search results," *Multimedia Tools Appl.*, vol. 80, no. 13, pp. 20217–20253, May 2021.

[5] A.-B. Djaker, B. Kechar, H. Afifi, and H. Moungla, "Maximum concurrent flow solutions for improved routing in IoT future networks," *Arabian J. Sci. Eng.*, vol. 48, no. 8, pp. 10079–10098, Aug. 2023.

[6] A. Al-Jawad, I.-S. Comsa, P. Shah, O. Gemikonakli, and R. Trestian, "An innovative reinforcement learning-based framework for quality of service provisioning over multimedia-based SDN environments," *IEEE Trans. Broadcast.*, vol. 67, no. 4, pp. 851–867, Dec. 2021.

[7] U. Rashid, "Multiple media information search framework," Ph.D. thesis, Quaid-I-Azam Univ., Islamabad, Pakistan, 2017.

[8] G. Chen, C. Wang, M. Zhang, Q. Wei, and B. Ma, "How 'small' reflects 'large'?—Representative information measurement and extraction," *Inf. Sci.*, vols. 460–461, pp. 519–540, Sep. 2018.

[9] D. Flores-Martin, J. Berrocal, J. García-Alonso, C. Canal, and J. M. Murillo, "Enabling the interconnection of smart devices through semantic web techniques," in *Proc. Int. Conf. Web Eng.* Cham, Switzerland: Springer, 2019, pp. 534–537.

[10] T. Storsul, "What, when and where is the Internet?" *Eur. J. Commun.*, vol. 34, no. 3, pp. 319–322, Jun. 2019.

[11] A. R. Khan, U. Rashid, and N. Ahmed, "An explanatory study on user behavior in discovering aggregated multimedia web content," *IEEE Access*, vol. 10, pp. 56316–56330, 2022.

[12] U. Rashid, M. Saddal, G. Farooq, M. A. Khan, and N. Ahmad, "An SUI-based approach to explore visual search results cluster-graphs," *PLoS ONE*, vol. 18, no. 1, Jan. 2023, Art. no. e0280400.

[13] K. Rematas, B. Fernando, F. Dellaert, and T. Tuytelaars, "Dataset fingerprints: Exploring image collections through data mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4867–4875.

[14] N. N. Vu, B. P. Hung, N. T. T. Van, and N. T. H. Lien, "Theoretical and instructional aspects of using multimedia resources in language education: A cognitive view," in *Multimedia Technologies in the Internet of Things Environment*, vol. 2. Singapore: Springer, 2022, pp. 165–194.

[15] U. Rashid, M. Viviani, and G. Pasi, "A graph-based approach for visualizing and exploring a multimedia search result space," *Inf. Sci.*, vols. 370–371, pp. 303–322, Nov. 2016.

[16] U. Rashid, M. Viviani, G. Pasi, and M. A. Bhatti, "The browsing issue in multimodal information retrieval: A navigation tool over a multiple media search result space," in *Proc. 11th Int. Conf. Flexible Query Answering Syst. (FQAS)*, Cracow, Poland. Cham, Switzerland: Springer, 2016, pp. 271–282.

[17] A. R. Khan, U. Rashid, K. Saleem, and A. Ahmed, "An architecture for non-linear discovery of aggregated multimedia document web search results," *PeerJ Comput. Sci.*, vol. 7, p. e449, Apr. 2021.

[18] J. Fan, Y. Gu, M. Hachimori, and Y. Miao, "Signature codes for weighted binary adder channel and multimedia fingerprinting," *IEEE Trans. Inf. Theory*, vol. 67, no. 1, pp. 200–216, Jan. 2021.

[19] K. J. R. Liu, *Multimedia Fingerprinting Forensics for Traitor Tracing*, vol. 4. London, U.K.: Hindawi, 2005.

[20] S. B. A. Khattak, Fawad, M. M. Nasralla, M. A. Esmail, H. Mostafa, and M. Jia, "WLAN RSS-based fingerprinting for indoor localization: A machine learning inspired bag-of-features approach," *Sensors*, vol. 22, no. 14, p. 5236, Jul. 2022.

[21] G. Kabatiansky and E. Egorova, "Adversarial multiple access channels and a new model of multimedia fingerprinting coding," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2020, pp. 1–5.

[22] L. Berriche, "Comparative study of fingerprint-based gender identification," *Secur. Commun. Netw.*, vol. 2022, pp. 1–9, Dec. 2022.

[23] E. E. Egorova, M. Fernandez, G. A. Kabatiansky, and Y. Miao, "Existence and construction of complete traceability multimedia fingerprinting codes resistant to averaging attack and adversarial noise," *Problems Inf. Transmiss.*, vol. 56, no. 4, pp. 388–398, Dec. 2020.

[24] K. Cengiz, R. Sharma, K. Kottursamy, K. K. Singh, T. Topac, and B. Ozyurt, "Recent emerging technologies for intelligent learning and analytics in big data," in *Multimedia Technologies in the Internet of Things Environment*. Singapore: Springer, 2021, pp. 69–81.

[25] F. Amato, A. Castiglione, V. Moscato, A. Picariello, and G. Sperlì, "Multimedia summarization using social media content," *Multimedia Tools Appl.*, vol. 77, no. 14, pp. 17803–17827, Jul. 2018.

[26] E. Chiodino, D. Di Luccio, A. Lieto, A. Messina, G. L. Pozzato, and D. Rubinetti, "A knowledge-based system for the dynamic generation and classification of novel contents in multimedia broadcasting," in *Proc. ECAI.* Amsterdam, The Netherlands: IOS Press, 2020, pp. 680–687.

[27] M. M. Nasralla, S. B. A. Khattak, I. Ur Rehman, and M. Iqbal, "Exploring the role of 6G technology in enhancing quality of experience for m-Health multimedia applications: A comprehensive survey," *Sensors*, vol. 23, no. 13, p. 5882, Jun. 2023.

[28] M. D. Abdulrahaman, N. Faruk, A. A. Oloyede, N. T. Surajudeen-Bakinde, L. A. Olawoyin, O. V. Mejabi, Y. O. Imam-Fulani, A. O. Fahm, and A. L. Azeez, "Multimedia tools in the teaching and learning processes: A systematic review," *Heliyon*, vol. 6, no. 11, Nov. 2020, Art. no. e05312.

[29] M. Gao, Y. Tang, H. Liu, and R. Ma, "Statistics of fingerprint minutiae frequency and distribution based on automatic minutiae detection method," *Forensic Sci. Int.*, vol. 344, Mar. 2023, Art. no. 111572.

[30] R. A. Priyadharshini, S. Arivazhagan, and M. Arun, "A deep learning approach for person identification using ear biometrics," *Int. J. Speech Technol.*, vol. 51, no. 4, pp. 2161–2172, Apr. 2021.

[31] J. Haitsma, T. Kalker, and J. Oostveen, "An efficient database search strategy for audio fingerprinting," in *Proc. IEEE Workshop Multimedia Signal Process.*, Dec. 2002, pp. 178–181.

[32] D. Cozzolino, G. Poggi, and L. Verdoliva, "Extracting camera-based fingerprints for video forensics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2019, pp. 130–137.

[33] A. Ross, S. Banerjee, and A. Chowdhury, "Security in smart cities: A brief review of digital forensic schemes for biometric data," *Pattern Recognit. Lett.*, vol. 138, pp. 346–354, Oct. 2020.

[34] A. Adibfar, A. Costin, and R. R. A. Issa, "Design copyright in architecture, engineering, and construction industry: Review of history, pitfalls, and lessons learned," *J. Legal Affairs Dispute Resolution Eng. Construct.*, vol. 12, no. 3, Aug. 2020, Art. no. 04520032.

[35] M. Mundher, D. Muhamad, A. Rehman, T. Saba, and F. Kausar, "Digital watermarking for images security using discrete slantlet transform," *Appl. Math. Inf. Sci.*, vol. 8, no. 6, pp. 2823–2830, Nov. 2014.

[36] T. Li, M. Jia, and X. Cao, "A hierarchical retrieval method based on hash table for audio fingerprinting," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2021, pp. 160–174.

[37] A. Wang, "The Shazam music recognition service," *Commun. ACM*, vol. 49, no. 8, pp. 44–48, Aug. 2006.

[38] A. Gupta, A. Rahman, and G. Yasmin, "Audio fingerprinting using high-level feature extraction," in *Computational Intelligence in Pattern Recognition*. Singapore: Springer, 2022, pp. 281–291.

[39] A. Messina, M. Montagnuolo, and M. L. Sapino, "Characterizing multimedia objects through multimodal content analysis and fuzzy fingerprints," in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.* Berlin, Germany: Springer, 2006, pp. 22–33.

[40] S. Khan, "Canopy approach of image clustering based on camera fingerprints," *Multimedia Tools Appl.*, vol. 81, no. 15, pp. 21591–21618, Jun. 2022.

[41] H. Tan and A. Kumar, "Towards more accurate contactless fingerprint minutiae extraction and pose-invariant matching," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3924–3937, 2020.

[42] H. Wang, Y. Kawahara, C. Weng, and J. Yuan, "Representative selection with structured sparsity," *Pattern Recognit.*, vol. 63, pp. 268–278, Mar. 2017.

[43] A. R. Khan, U. Rashid, and N. Ahmed, "AMED: Aggregated multimedia exploratory and discovery search software," *SoftwareX*, vol. 21, Feb. 2023, Art. no. 101312.

[44] V. Turner, J. F. Gantz, D. Reinsel, and S. Minton, "The digital universe of opportunities: Rich data and the increasing value of the Internet of Things," *IDC Analyze Future*, vol. 16, pp. 13–19, 2014.

[45] T. Beck, F. Böschen, and A. Scherp, "What to read next? Challenges and preliminary results in selecting representative documents," in *Proc. Int. Conf. Database Expert Syst. Appl.* Cham, Switzerland: Springer, 2018, pp. 230–242.

[46] J. Bien and R. Tibshirani, "Prototype selection for interpretable classification," *Ann. Appl. Statist.*, vol. 5, no. 4, pp. 2403–2424, Dec. 2011.

[47] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1600–1607.

[48] M. Page, J. Taylor, and M. Blenkin, "Uniqueness in the forensic identification sciences-fact or fiction?" *Forensic Sci. Int.*, vol. 206, nos. 1–3, pp. 12–18, 2011.

[49] J. P. Ogle and D. P. W. Ellis, "Fingerprinting to identify repeated sound events in long-duration personal audio recordings," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2007, pp. I–233.

[50] M. D. Kamaladas and M. M. Dialin, "Fingerprint extraction of audio signal using wavelet transform," in *Proc. Int. Conf. Signal Process., Image Process. Pattern Recognit.*, Feb. 2013, pp. 308–312.

[51] X. Anguera, A. Garzon, and T. Adamek, "MASK: Robust local features for audio fingerprinting," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 455–460.

[52] C. Yu, R. Wang, J. Xiao, and J. Sun, "High performance indexing for massive audio fingerprint data," *IEEE Trans. Consum. Electron.*, vol. 60, no. 4, pp. 690–695, Nov. 2014.

[53] C. Ouali, P. Dumouchel, and V. Gupta, "A robust audio fingerprinting method for content-based copy detection," in *Proc. 12th Int. Workshop Content-Based Multimedia Indexing (CBMI)*, Jun. 2014, pp. 1–6.

[54] M. Malekesmaeili and R. K. Ward, "A local fingerprinting approach for audio copy detection," *Signal Process.*, vol. 98, pp. 308–321, May 2014.

[55] C. Saravanos, D. Ampeliotis, and K. Berberidis, "Audio-fingerprinting via dictionary learning," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process. (MMSP)*, Sep. 2020, pp. 1–7.

[56] J. Meng, S. Wang, H. Wang, Y.-P. Tan, and J. Yuan, "Video summarization via multi-view representative selection," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 1189–1198.

[57] I. Kavati, A. M. Reddy, E. S. Babu, K. S. Reddy, and R. S. Cheruku, "Design of a fingerprint template protection scheme using elliptical structures," *ICT Exp.*, vol. 7, no. 4, pp. 497–500, Dec. 2021.

[58] B. Pandya, G. Cosma, A. A. Alani, A. Taherkhani, V. Bharadi, and T. M. McGinnity, "Fingerprint classification using a deep convolutional neural network," in *Proc. 4th Int. Conf. Inf. Manage. (ICIM)*, May 2018, pp. 86–91.

[59] X. Li, C. Guo, Y. Yang, and L. Xu, "Video fingerprinting based on quadruplet convolutional neural network," *Syst. Sci. Control Eng.*, vol. 9, no. sup1, pp. 131–141, Apr. 2021.

[60] B. Tseytlina and I. Makarova, "Content based video retrieval system for distorted video queries," in *Proc. MacsPro*, 2020, pp. 1–9.

[61] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, "Facing device attribution problem for stabilized video sequences," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 14–27, 2020.

[62] C. Ye, Z. Xiong, Y. Ding, G. Wang, J. Li, and K. Zhang, "Joint fingerprinting and encryption in hybrid domains for multimedia sharing in social networks," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 658–666, Dec. 2014.

[63] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.

[64] E. Egorova, M. Fernandez, G. Kabatiansky, and M. H. Lee, "Signature codes for the A-channel and collusion-secure multimedia fingerprinting codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 3043–3047.

[65] C. Ouali, P. Dumouchel, and V. Gupta, "Robust video fingerprints using positions of salient regions," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3041–3045.

[66] Q.-T. Phan, G. Boato, and F. G. B. De Natale, "Accurate and scalable image clustering based on sparse representation of camera fingerprint," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1902–1916, Jul. 2019.

[67] H. Chen, B. D. Rouhani, C. Fu, J. Zhao, and F. Koushanfar, "DeepMarks: A secure fingerprinting framework for digital rights management of deep learning models," in *Proc. Int. Conf. Multimedia Retr.*, Jun. 2019, pp. 105–113.

[68] F. Pandey, P. Dash, D. Samanta, and M. Sarma, "ASRA: Automatic singular value decomposition-based robust fingerprint image alignment," *Multimedia Tools Appl.*, vol. 80, no. 10, pp. 15647–15675, Apr. 2021.

[69] D. Sharma and A. Selwal, "An intelligent approach for fingerprint presentation attack detection using ensemble learning with improved local image features," *Multimedia Tools Appl.*, vol. 81, no. 16, pp. 22129–22161, Jul. 2022.

[70] C. Ye, S. Tan, Z. Wang, L. Shi, and J. Wang, "A secure social multimedia sharing scheme in the TSHWT_SVD domain based on neural network," *Multimedia Tools Appl.*, vol. 82, no. 10, pp. 15395–15414, Apr. 2023.

[71] G. Ning, Z. Zhang, X. Ren, H. Wang, and Z. He, "Joint audio-video fingerprint media retrieval using rate-coverage optimization," 2016, *arXiv:1609.01331*.

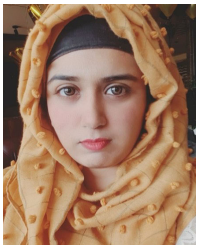
[72] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[73] S. Özarpacı, B. Kılıç, O. C. Bayrak, A. Özdemir, Y. Yılmaz, and M. Floyd, "Comparative analysis of the optimum cluster number determination algorithms in clustering GPS velocities," *Geophys. J. Int.*, vol. 232, no. 1, pp. 70–80, Sep. 2022.

[74] Y.-J. Zhang, "Content-based retrieval," in *Handbook of Image Engineering*. Singapore: Springer, 2021, pp. 1513–1548.

[75] T. Ruotsalo, J. Peltonen, M. J. A. Eugster, D. Głowacka, P. Floréen, P. Myllymäki, G. Jacucci, and S. Kaski, "Interactive intent modeling for exploratory search," *ACM Trans. Inf. Syst.*, vol. 36, no. 4, pp. 1–46, Oct. 2018.

UMER RASHID received the B.S. degree in computer sciences from the University of Lahore, Pakistan, in 2005, and the M.Phil. and Ph.D. degrees in computer sciences from Quaid-i-Azam University, Islamabad, Pakistan, in 2008 and 2017, respectively. He is currently an Assistant Professor in computer sciences with Quaid-i-Azam University. He has teaching and research experience in international organizations. His research interests include user-centered computing, multimedia information retrieval, and multimedia technology. His work is published in international journals and conferences. He is also the author of several book chapters. He also reviewed for several international journals and conferences.



SAMRA NASEER received the M.Phil. degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan, in 2022. Her current research interests include multimedia information retrieval systems, artificial intelligence, computer vision, human and computer interaction, and machine learning.

ABDUR REHMAN KHAN received the B.S. degree (Hons.) from the National College of Business Administration and Economics, Lahore, and the M.Phil. degree in computer science from Quaid-i-Azam University, Islamabad, Pakistan. His research interests include multimedia information retrieval, user-centered computing, AI, search engines, UI/UX, web, and data science. He possesses diverse working experience in private, semi-government, and government institutes as an Android Developer, a Graphic Designer, a Research Assistant, and an IT professional. He is currently a Gazetted Lecturer with the Faculty of Engineering and Computing, National University of Modern Languages, Pakistan.

MUAZZAM A. KHAN is a working Director Science and Technology and Director ICESCO Chair Big Data Analytics and Edge Computing, Quaid I Azam University, Islamabad, Pakistan. He received his Ph.D. degree in Computer Science in a sandwich program from IIUI, Pakistan, and UMKC, USA, in 2011. Later he also received his Post Doc from University of Missouri, KC, USA in 2016. He joined NUST University, Pakistan, as an Assistant Professor in 2013, and promoted to Tenured Associate Professor and Associate Dean Computing at NUST-SEECS, Pakistan, in 2017. He also worked as a Research Fellow at the Networking and Multimedia Laboratory, at UMKC, USA and University of Ulm Germany. His research interests include wireless sensor networks, body area networks, image compression, image encryption, and data network security. He has published more then 200 publications with 10 book chapters having total citations of 3500. He has served as the co-chair, publicity chair, TPC member and reviewer of the famous international conference on Smart cities and Information and communication technologies HONET-ICT 2017 till 2023.

GAUHAR ALI received the M.S. degree in computer science from the Institute of Management Sciences, Peshawar, Pakistan, in 2012, and the Ph.D. degree in computer science from the University of Peshawar, Peshawar. He is currently a Postdoctoral Researcher with the EIAS Data Science and Blockchain Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. His research interests include the Internet of Things, access control, blockchain, machine learning, wireless sensors networks, intelligent transportation systems, formal verification, and model checking.

NAVEED AHMAD received the B.S. degree in computer science from the University of Peshawar, Pakistan, in 2007, and the Ph.D. degree in computer science from the University of Surrey, U.K., in 2013. He is currently an Associate Professor with the College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. His research interests include security and privacy in emerging networks, such as VANETs, DTN, the Internet of Things (IoT), machine learning, and big data.

YASIR JAVED (Member, IEEE) received the Ph.D. degree. He is currently a highly qualified Data Scientist and a Senior Programmer/Developer with over 18 years' experience in research, security programming, software development, project management, and analytics. As a part of his research, he has interests in data analytics, forensics, smart cities, network security, and education sustainability, instructional development, learning and education sustainability, robotics, unmanned aerial vehicles, vehicular platoons, secure software development, signal processing, the IoT analytics, intelligent applications, and predictive computing inspired by artificial intelligence. He is an Active Member of the RIOTU Group, Prince Sultan University (PSU). In addition, he was awarded a Rector's Medal for his M.S. degree and a Distinguished Teaching Award from the President. He received the Outstanding Ph.D. Student Award from UNIMAS, Sarawak. Listed in the Top Researcher Award from PSU in recognition of his research contributions, he has published over 100 peer-reviewed papers in top-tier journals, conference proceedings, and book chapters. Additionally, he serves as a reviewer for several journals. With regard to his professional experience, he has undertaken a variety of national and international research funding projects and he has also served as an Analyst Programmer with the Prince Megren Data Center, the Center of Excellence, and the Research and Initiative Center, PSU. He serves as the Chair of the ACM Professional Chapter, Saudi Arabia.