Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2018 - Volume: 19 Number: 1
Page: 132 - 143
DOI: 10.18038/aubtda.330958
Received: 26 July 2017 Revised: 31 October 2017 Accepted: 25 December 2017

DISTRIBUTED VIDEO IDENTIFICATION WITH PERCEPTUAL TAGS IN PEER-TO-PEER NETWORKS

Alper KOZ *

Center for Image Analysis, Middle East Technical University, 06800, Ankara, Turkey

ABSTRACT

In this paper, a distributed solution is proposed for video identification and copy detection in P2P networks, which represents a video file in the network with a set of (64-256) bits, named as perceptual tags. As such information is derived from the perceptual content of the video rather than its bitstream representation as in the case of cryptographic hashes, it provides a robust identification after the alterations in the file names and formats provided that the visual quality of the video is at acceptable levels. The paper first briefly discusses the requirements for a distributed perceptual tagging system considering the low computational power and low bandwidth of internet users. Then, it presents the proposed perceptual tag extraction method using the temporal differences between the video frame averages and the proposed distributed searching scheme for a P2P implementation. The proposed extraction and searching methods provide robustness to the alterations in video formats and small additions and cuttings in the video content as the typical processing in P2P environment and also achieve uniform distribution and storage load between the peers.

Keywords: Perceptual tags, peer to peer networks, Content search, Distributed video identification, Distributed hash tables, Access right management

1. INTRODUCTION

File sharing in P2P networks has been reported as one of the major source of internet traffic in the last decade with its opportunities providing to its users in accessing free media files including movies, music and games [1-3]. However, the growth of these free mediums has also increased the necessity for access right management solutions for highly valuable copyrighted material and the means for proper robust identification of that content [4, 5].

Identification of video files in P2P networks has been performed in two ways until now. While the first method is to use the file tags to distinct two files, the other approach uses cryptographic hashes extracted from the whole bitstream representation of the files [1, 6, 7]. These methods however fail to identify a video when the file tag is changed or the video is encoded in a different format changing the bitstream representation, frame size, aspect ratio and other format parameters. In this paper, a complementary solution is proposed to identify a video in P2P networks with perceptual tags which are only a few hundreds of bits (64-256 bits) in size as file tags and crypto-hashes. As such information is derived from the video content rather than its bitstream representation, it enables a sustainable identification when the file names and formats are changed.

In our recent work [8], we present distributed solutions for video fingerprinting (perceptual hashing) in P2P networks for file identification. While this distributed fingerprint system [8] and the previous fingerprinting studies utilized in central servers [9-11] are designed to identify any possible segment of a video file with small granularities, the proposed approach in this paper aims to identify the entire file with only a few hundred bits of information and to find the modified versions of the entire file rather than its every possible segment.

*Corresponding Author: koz@metu.edu.tr

As in the case of distributed video fingerprinting system [8], a distributed perceptual tagging system should also provide robustness to the typical processing in P2P networks [12]. In addition to the changes in coding qualities (bitrates) and standard video formats, some small insertions and deletions at the beginning, end, or inside of a video for different purposes can also be commonly performed on the video content. These common insertions and deletions introduce a synchronization problem for perceptual tag extraction avoiding the trivial solutions such as using the frame numbers and other content independent solutions.

As a second requirement, a distributed perceptual tagging system should disseminate the load of the storage and searching of perceptual tags in the network among the peers, by the definition. Compared to the central server indexing system, distribution of the storage and searching provides advantages in balancing the network traffic, in sustainability of the resources, and in scalability and adaptation to the changes in the number of peers. However, besides its advantages, such a system brings also new challenges for the perceptual tagging methodology due to the low computational power, low bandwidth and low storage space of internet users.

First, the low computational power of P2P users with their ordinary personal computers avoids the usage of high level features or computationally loaded transforms [11, 13, 14] for the extraction of perceptual tags at the peers. Second, while the search of a perceptual tag is possible with its possible erroneous versions in a centralized database [15], a network implementation does not provide such an opportunity as the bandwidth of internet users are limited. Therefore, the perceptual tags must survive from typical processing in the network with only a few number of bit errors compared to the higher tolerable levels in a central server implementation. Finally, the storage load for the perceptual tag index should be uniformly disseminated to the peers. In order to construct a distributed index with the state of art distributed hash tables, such a uniform distribution requires the perceptual tags not to possess repetitive patterns.

The previous works specially designed for the detection of entire file as in the same framework of the proposed study are mainly tailored for central database implementations and therefore, are not suitable for distributed implementations. Among these works [11, 13, 14], Coskun et al. [11] use three dimensional cosine transform to extract video hashes, which would be computationally loaded for simple P2P users having personal computers. Hampapur and Bolle [13] use gradients of each video frame to construct hashes. A main challenge for their work is the adaptation of inverted video index to a distributed implementation. Finally, Cheung and Zakhor [14] propose different approaches using color histograms of the frame segments. However, the robustness, discrimination cabability, and indexing are all in questions for a distributed implementation of these early works.

In this paper, a distributed perceptual tagging system is proposed which addresses the mentioned requirements for a P2P implementation. The proposed perceptual tag extraction mainly uses the temporal difference of the video frame averages. In order to align the perceptual tag extraction from the different duplicates of a video, which can involve small insertions and deletions with respect to each other, a self-synchronization scheme is developed to determine reference positions on the sequence of video frame averages. Finally, the design parameters of the proposed method are selected to achieve a uniform distribution load among the peers and the performance of the whole distributed perceptual tag system is verified with the experiments.

The next section overviews the identification methods and search mechanisms using file tags and cryptographic hashes in P2P networks with a brief overview of a fundamental DHT system, Content-Addressable Networks (CAN) [16], which constitutes a baseline for the proposed distributed perceptual tagging system. The main steps of the proposed distributed tagging system are given in Section 3. The details of the proposed perceptual tag extraction with the proposed synchronization method are given in Section 4. Section 5 follows with the experimental tests for the robustness of the proposed method. In Section 6, the distribution of the perceptual tags is explained and the experimental verification of the complete distributed system is presented. Finally, the conclusions are given in Section 7.

2. IDENTIFICATION AND SEARCH MEACHANISMS IN P2P NETWORKS

The identification of a video file in P2P networks is achieved in two ways, whether by using its name or by using its derived cryptographic hash.

2.1. Identification with File Names and Search with Distributed Hash Tables

The first identification method for the files in P2P networks [8] is simply to use the file names to distinct two files from each other. A given query as a keyword is searched in the network by comparing the keyword with the file names. The search returns the peers possessing the files whose names match with the keyword. This search mechanism is achieved in the pioneer P2P networks by using a centralized index holding the information of which file is shared at which peer [17]. Although such an approach is easy to employ, the storage and traffic load at the server leads to distributed solutions.

The first solution was to flood the query in the network [18, 19]. A given query by a peer is recursively forwarded in the network for a given number of hops. Each visited peer compares the keyword with the name of the files it is sharing. If there is a match in the network, the information (IP address) of the peer possessing the matched file is returned. If not, a failure report is received. While flooding does not impose any hierarchy between the peers in the initial systems, the later approaches such as Kazaa adopt more structural searches, which uses the changes in bandwidth and computational power of the peers to define an upper network to increase the performance [18]. More recent approaches, such as Tribler and its derivatives, have further increase the search performance by using the similarities in the peer interests [19].

Flooding type of searches are robust against failures and node transience as well as avoid traffic congestion in the network. However, a main disadvantage is not to guarantee to find a desired file as the search is defined for a number of hops. The scalability of such systems is also linearly dependent to the number of peers, making the adaptation of the system to sudden changes in peer numbers more difficult [16, 20].

The stated lacks of the flooding approach in P2P search are addressed by developing distributed hash tables (DHTs) [16, 20]. DHT is a distributed implementation of hash tables which map the keys to the values by using a hash function. Keys and values correspond to the names of the files and the IP address of the peers storing the files, respectively. The usage of hash functions for such a mapping provides the uniform spreading of storage and routing burden in the network. In such a distributed implementation, the peers partition the hash table storing the (key, value) pairs and each peer is assigned as a responsible peer from a zone of the hash table. Each of the peer assigned to a zone keeps the information regarding the peers assigned to the neighbor zones. This neighbor information enables the routing of a request for a key to the neighbor peer, which is closest to the peer containing that key.

The proposed distributed system for the indexing of perceptual tags uses a basic DHT structure, namely Content-addressable Network (CAN), as a baseline. CAN first generates a Cartesian coordinate space to locate the peers as illustrated in Figure 1. The Cartesian space is partitioned and each peer is assigned to a zone of the space to store the (key,value) pairs and keeps the IP addresses and coordinates (in the Cartesian space) of the neighbor peers [16]. If a (key,value) pair, which is denoted as (K_1,V_1) in Figure 1, will be stored in the system, a uniform hash function is applied to key K_1 and the point C in the coordinate space is obtained. Then, the peer whose zone covers the point C

(Peer 8 for the given illustration in Figure 1) stores the (K_1,V_1) pair. If a query for K_1 is performed by a peer in the network, the peer first uses the uniform hash function and obtains the position C. The querying peer afterwards sends the request to its neighbor closest to the point C in the Cartesian space. For instance, the querying peer, Peer 1 in Figure 1, send the request for key K_1 to Peer 4. The neighbor forwards this request to its neighbor closest to the point C again and this process ends up with the peer storing the point C, which is Peer 8. Peer 8 returns the value, V_1 for key K_1, including the IP address of the peer holding the video file corresponding to K_1 to the querying peer, Peer 1.

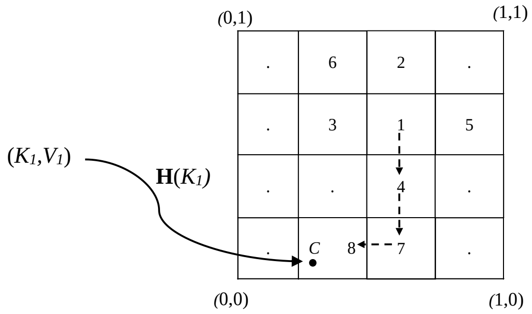


Figure 1. Partition of 2D coordinate space by the peers in CAN, mapping of a (key, value) pair and routing of a query (K_1) (from peer 1 to peer 8).

The proposed distributed indexing system for perceptual tags is designed as an analogue of the CAN design by replacing the keys representing the names of video files with the perceptual tags. In Section 3, the main stages and the requirements for such a perceptual tag based distributed identification system are briefly discussed.

2.2. Cryptographic Hash Based Identification in P2P Networks

Cryptographic hashes are used as a second method to identify the files in P2P networks. The hash value is extracted from the whole content by using a hash function. As the hash function is designed to avoid collision, which can be described as the situation of the same output for two different inputs, the cryptographic hashes provide a unique and robust identification for the changes in the file names. Given the hash value of a video file, the peers query the network with this hash value to find the peers having the desired video file [7, 21].

Magnet links [21] and torrents [1] are the main standards for the representation of video files with cryptographic hashes in P2P networks. While torrents are used mainly for the verification of data integrity after the file downloads, magnet links are utilized for also searching in addition to verification, due to their simple plain text structure and low storage size [7, 21] enabling a fast dissemination.

A magnet link is defined with a number of parameters which include name, file size, topics, and hash values for the entire file and for the file parts [7, 21]. The hash value for the entire file is employed for identification and the hash values for the file parts are used to check the integrity of each part after download. The utilized hashing algorithm, size of the file parts (chunks), and the size of the resulting hash set for a 350 MB video file, which is coarsely the size of a one-hour episode, for some of the P2P clients are given in Table 1 [6]. As an important conclusion from the table, the output size of the resulting hashes used for file identification is 128 to 160 bits. Based on these sizes of the cryptographic hashes, the size of the perceptual tags that will be used for the perceptual identification of the files is selected to be in the same levels in the proposed system. Such a perceptual tagging system compensates the disadvantages of the crypto hashes by enabling a further robust identification in case of the changes in format.

Table 1. Total crypto-hash sizes for a 350 MB video. A file is divided into chunks. For each chunk, a crypto hash is generated. MD4 and SHA1 refer to message digest algorithm 4 and secure hash algorithm 1.

P2P Network	Alg.	Output Size	Chunk Size	Total Hash Size (~ KB)
Emule [7] (main block)	MD4	128 bits	9.28 MB	0.59
Emule [7] (subblocks)	MD4	128 bits	180 KB	31.11
Bittorrent [1] (new)	SHA-1	160 bits	256 KB	27.34
Bittorrent [1] (old)	SHA-1	160 bits	1 MB	6.84
FastTrack [6]	MD4	128 bits	64 KB	87.49
Gnutella [6]	SHA-1	160 bits	500 KB (min)	14.00 (max)

3. PROPOSED DISTRIBUTED PERCEPTUAL TAGGING SYSTEM AND ITS REQUIREMENTS

The proposed distributed perceptual tagging system takes the CAN design as a baseline. Noting that the (key-value) pair, (K, V), corresponds to the video name and the internet protocol address of its owner in the network, the proposed system is constructed by replacing the file name with the perceptual tag (P), resulting (P, V) pairs (Figure 2). With such a notation, the distributed perceptual tagging system consists of these main phases:

- A Cartesian coordinate space to map and store the perceptual tags is formed. Similar to CAN, Cartesian space is divided and allocated by the peers in the network.
- The P2P client program at the peer sharing the video automatically computes the perceptual tag of the video file.
- The computed perceptual tag with the internet protocol (IP) address of the peer, (P, V), are mapped to the perceptual tag space. As in the CAN system, the peer responsible from the mapped position stores the (P, V) pair.
- The peers in the network collectively routes a query in the network like in CAN.

A distributed perceptual tagging system with these stages do not involve a new requirement as only the file names are replaced with the perceptual tags. The networking aspects such as joining the network, leaving the network, and bootstrapping can be performed by using similar strategies developed for CAN.

Such a system however imposes new requirements on the perceptual tagging methodology. First, the proposed perceptual tagging algorithm should be computationally not loaded for the simple internet users forming the P2P network and solve the synchronization problem that can occur due to additions and deletions of scenes in different versions of a video file. Second, the robustness of the perceptual tags should be high, resulting only a few bits of errors after the possible processing in a network. Finally, the extracted perceptual tags should yield a uniform distribution for a fair dissemination of storage and searching load.

4. PROPOSED PERCEPTUAL TAG EXTRACTION FROM VIDEO FILES

Based on the previous discussions, the proposed method uses the frame means, as a simple and robust feature of a video file against compression and format changes to address the computational complexity and robustness. The proposed approach determines robust reference positions on the one dimensional sequence formed by computing the pixel averages at each frame of the video to align the extraction of perceptual tags for different copies of the video. The main stages of the proposed perceptual tag extraction from a shared video at a peer are described as follows [12]:

- Compute the average of the pixel intensities of every frame for a video shared at a peer. The resulting one dimensional sequence is denoted as $I(n)$, where n is the frame number.

- Split the mean sequence, $I(r)$, into R segments, which have equal number of frames where $I_r(n)$ with $r = 1..R$ stands for the segments. For each segment, $I_r(n)$, take the difference of the consecutive samples.

- Take the magnitude of the difference signal and find the position where maximum is located for each segment. The position of the maximum, which is denoted as S_r, is regarded as the synchronization point to begin extraction.

- Define a vector, $W_r = [\ I_r(S_r\text{-}0.5xLxM)\ldots I_r(S_r+0.5xLxM)\ ]$, which is composed of LxM points in the neighborhood of S_r for the segment $I_r(n)$.

- Perform a subsampling on the vector, W_r, by M and compute the difference of the successive samples. Compute the sign of the vector elements to binarize the vector:
-

$$v_r = sign[W_r(1)\text{-}W_r(M+1)\ldots W_r((L\text{-}1)M+1)\text{-}W_r(LM+1)]$$

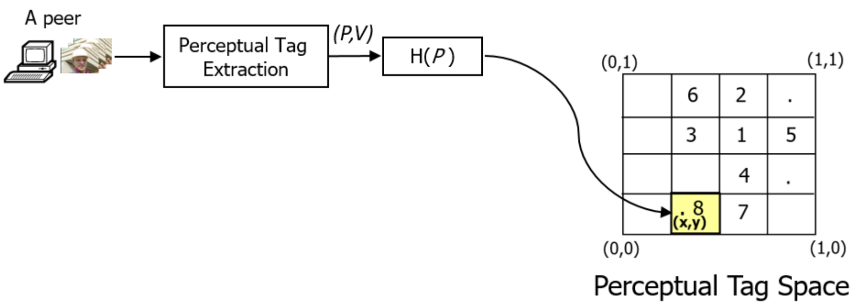


Figure 2. The proposed DHT based distributed perceptual tagging system as an analogue to the CAN design.

Each of the vector, v_r, is regarded as a descriptor of the video file used for identification. L corresponds to the length of the descriptor and M is the subsampling period during extraction. The total R descriptors form the perceptual tag of the video file with RxL bits. Each of the descriptor forming the perceptual tag is separately searched during a query, which results R searches for a file. The descriptor giving the minimum distance among all descriptors determines whether a match is found or not in the network.

5. EXPERIMENTAL RESULTS FOR THE ROBUSTNESS OF THE PROPOSED METHOD

The proposed perceptual tag extraction method is tested with respect to its robustness against typical processing in P2P environment. The experimental setup designed for this purpose is as follows:

• A collection of 337 films is gathered which has a coding bitrate of 800 kbps to 2 Mbps, a length of 52.682.104 frames (~585 hours), and a frame size changing in the range of (300-400) x (700-800) pixels. This set is regarded as the video set shared by the peers in P2P.

• This video set is processed to form a set of query video. The video files are encoded with a fixed bitrate of 128 kbps, the width and height of frames are decreased to half of the original sizes, and a different structure is chosen for group of size (GOP). These processing can be assumed worse than the typical changes for different editions of a video in P2P.

As the initial step of the verification of the proposed method, it is tested whether a found synchronization point on a video file changes its position or not after the video passes from the mentioned processing. For this purpose, video pairs of same duration are extracted from 585 hour of video. One of the pair is selected from the shared set and the other is selected from the query set. This is followed by computing the ratio of the video pairs (in percentage), which have the synchronization points at the same position.

The same process is repeated for the video clips with different durations. Table 2 shows the obtained percentages for the video clips with different durations. The synchronization points are obtained at the same position for the video files and their processed version in almost 100% of the all video pairs, showing the high robustness of selected points to typical video operations in P2P.

Detection and false positive rates are selected as the metrics to evaluate the performance of the proposed method. For this purpose, the histograms of the Hamming distance [22] between the perceptual tags of the video pairs are formed under binary hypothesis [23]. In the first hypothesis, denoted as H_0, the video pairs are copies of each other whereas the video pairs are not copy of each other in the second hypothesis (H_1). The histograms are formed for different number of descriptors (R), different subsampling periods (M) and different length of descriptor vectors (L). While the distance between the perceptual tags is selected as Hamming distance when $R=1$, it is selected as the minimum of Hamming distances calculated for all possible different selection of two descriptors from two video files when $R>1$. The decision is made with respect to the descriptor giving the minimum distance when multiple queries for a given tag are performed with multiple descriptors.

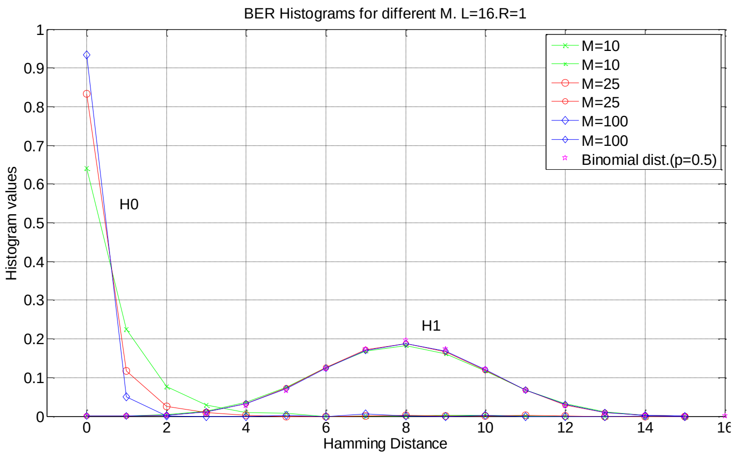
The effect of the subsampling period (M) and the length of the descriptor vector (L) to the detection rates is analyzed in the first experiment. The granularity of the proposed method, defined as the required duration of the video clip for identification, directly depends on the M and L parameters (MxL frames). For varying M values, the histograms of *Hamming* distance between the perceptual tags under H_0 and H_1 are illustrated in Figure 3 (a) for $L=16$ and in Figure 3 (b) for $L=64$. Only one segment (descriptor) for the entire file ($R=1$) is utilized for perceptual tag extraction. As the first observation, for no error case (Hamming Distance=0), the percentage of the video files under H_0 increases as the difference between consecutive mean values increases for higher subsampling period (M). On the other hand, higher M values also increase the minimum duration required for the identification (granularity).

The percentages of the video files having no error are about 83% for $L=16$ and about 48% for $L=64$ in Figure 3 (a) and (b) respectively. As the length of the perceptual tags (L) increases, the probability of having no error on a perceptual tag decreases as well. Therefore, using high number of bits is not preferable to correctly map and locate a perceptual tag in the virtual coordinate space of a DHT system [8]. However, the larger L values enable a better thresholding by increasing the gap between H_0 and H_1. As the distributions under H_1 also indicate binomial characteristics with a probability of 0.5 for different L values in the figures, the number of false positives can be made very small with a suitable selection of the thresholds for higher L values. Due to the mentioned tradeoff for the L values, the proposed solution is to use only a subset of the perceptual tag bits for distribution and mapping, while keeping the length of perceptual tags sufficiently high for an efficient detection. In order also to be compliant with 32-bit and 64-bit processor architectures, the length of the perceptual tags (L) are considered to be 32 or 64. Taking the higher false positive rates into account for L=32, the L value is selected as 64 in the experiments. Accordingly, subsampling period (M) is chosen as 25 to be able to detect video clips having duration more than one minute (64x25 frames).

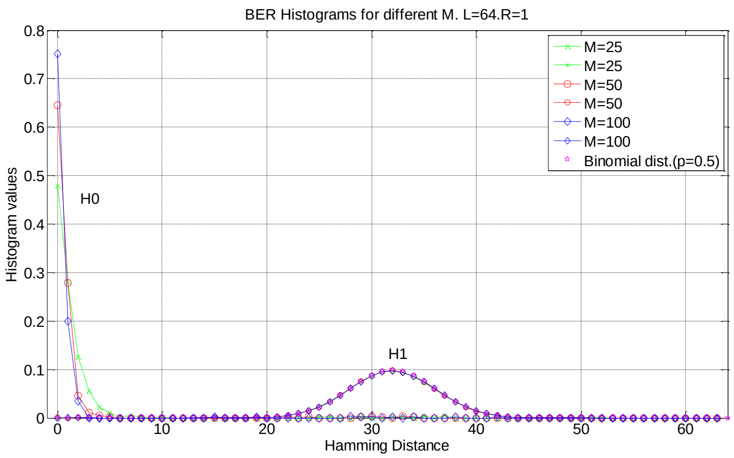
In the next experiment to analyze the interaction between the design parameters and detection performance, the number of segments (descriptors (R)) is changed and the subsampling period (M) and the length of the perceptual tag (L) are fixed to 25 and 64, respectively. Then, the detection rates are obtained by computing the percentage of the perceptual tag pairs in the histograms which is smaller than a given threshold. Please, note that the query perceptual tags are assumed correctly mapped to the peers in this section.

Table 2. Percentage of the video clips whose synchronization points are at the same position for different video durations.

# of video	2000	4000	8700	17500	35000
Dur(~min.)	17 min	8.5 min	4 min	2 min	1 min
# Frames	1024x25	512x25	256x25	128x25	64x25
Percent(%)	**99.9**	**99.8**	**99.5**	**99.08**	**97.20**



(a) L=16, R=1



(b) L=64, R=1

Figure 3. Histogram of Hamming distance for different *M* values

Figure 4 illustrates the detection rates with respect to the number of descriptors (*R*). Note that each video file is divided into R segments and totally R descriptors represent the whole video file extracted from each segment. The video file is queried with each of the descriptor in the network. As the queries increases with the number of descriptors forming a perceptual tag, the detection probability for the queried file also increases [12].

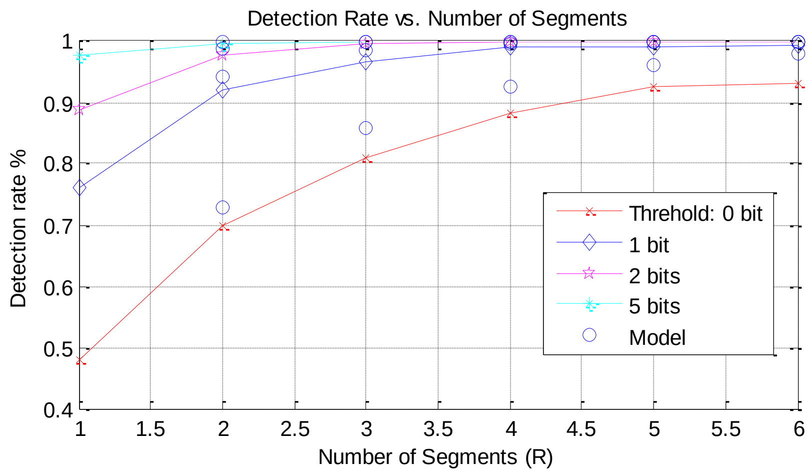


Figure 4. Experimental detection rates and the model for different *R*

Let us denote the probability that a video file is detected with only one descriptor as PD_1 for $R=1$. Then, the probability of detecting the video file with one of the R descriptors, PD_R, can be approximated as the complement of missing all descriptors for that perceptual tag:

$$PD_R = 1 - P(\text{missing all of the R descriptors}) \cong 1 - (1 - PD_1)^R \tag{1}$$

The approximation indicates a very similar characteristic to the experimental detection rates in the figure. When the number of segments is 2, the detection rate is about 99% with a threshold of 5 bits. The false positive rates (FP) due to higher R values can be neglected, as the FPs are in the range of 2^{-64} due to the binomial distribution under H_1 ($p=0.5$ and $L=64$).

In the final experiment, the performance of the proposed method is investigated with respect to the insertions and deletions in the video file. For this purpose, one of the video pairs is clipped with a varying ratio until 10%. Figure 5 illustrates the obtained detection rates as a function of temporal cropping rate for different R values. The detection rates linearly drops for increasing cropping rates for $R=1$, as the probability of missing the reference point in the cropped region also linearly drops with respect to size of the cropped region. The detection rates for the case of two descriptors (R=2) is also analytically calculated by using the overlapping regions between the segments before and after temporal cropping. The detection rates obtained with the experiments and the analytical model for the case of two descriptors give a close match in the figure. Using two segments (descriptors) in the proposed extraction method increases the detection performance to the rates higher than 96% up to the 10% cropping in the video files.

6. DISTRIBUTION OF PERCEPTUAL TAGS AND VERIFICATION OF THE DISTRIBUTED SYSTEM

The same strategy proposed for the distribution of fingerprints in [8] is also adopted here for the distribution of perceptual tags. First, the subsampling period, which achieves decorrelation of consecutive frames and gives a uniform distribution, is selected as 25 [8]. Second, only a subset of the bits of perceptual tags is utilized for mapping the perceptual tags in the coordinate space of DHT. Considering the peer numbers in ten thousand ranges in current P2P networks, 16 bits is selected from 64-bit perceptual tags to use in mapping. This 16-bits is selected as taking 1 bit from every 4 bits of the tags.

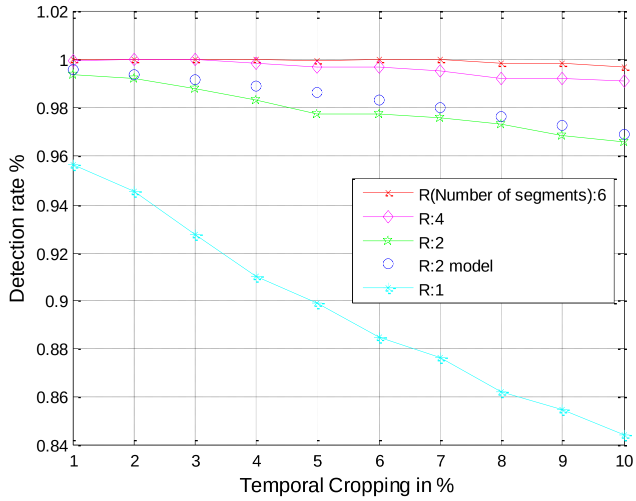


Figure 5. Detection Rate vs. Temporal Cropping. (Threshold: 5 bits)

The experimental verification begins with the extraction of 64-bits perceptual tags for different number of segments (R) from the shared video by a peer. The extracted tag (or its descriptors for $R>1$) from a video file and the internet protocol address of the peer sharing the video are mapped (linked) to the peer, which corresponds to its 16-bit mapping vector. The extraction is also repeated for the query set and the perceptual tags are assumed regularly routed from the querying peer to the targeted peer, corresponding to the 16-bit mapping vector of query perceptual tags. The query perceptual tag of L-bit is compared with the perceptual tags stored in the targeted peer and those having a Hamming distance smaller than a threshold are returned back to the querying peer. In the case of R>1, the same process is repeated for each descriptor and the decision is given with respect to the descriptor giving the minimum distance.

Table 3 gives the detection rates obtained with the experiments (PD_E) for different number of segments and cropping ratios in percentages. The detection rates are also analytically computed (PD_C) as a product of the detection of finding the correct peer (zone) during the mapping ($L=16$) and the detection of finding the correct perceptual tag of 64 bits which is stored at the targeted peer by using the histograms given for L=16 and L=64 in Figure 3 (a) and (b), when there is no cropping. The experimental results give a close match with the computed rates over histograms. The proposed distributed perceptual tagging system detects about 94% of the video files when there is no cropping, and about 88% of the video files up to the cropping ratios of 10%, by performing only two queries in the network.

The high detection rates obtained for the whole system first validate the accuracy of correct mapping to the peers storing the perceptual tags in the proposed distributed structure. Second, the experiments reveal the robustness and discriminability of the perceptual tags extracted from the temporal variation of the frame means to the typical processing in P2P environment. Finally, the proposed synchronization and segmentation based perceptual tag extraction successfully compensate the effect of shift and cropping that can occur in different versions of the same clips in the network.

7. CONCLUSIONS

A distributed system based on perceptual tags is proposed for the identification of video files after the modifications in the name and format of the files and small additions and cuttings in the video content. The proposed method forms an alternative robust identification in P2P networks compared to the conventional tag based and crypto-hash based methods, which fail in the case of mentioned modifications. The paper first reveals that the present methods for perceptual tagging are not suitable for distributed implementations due to their high complexity and BER levels. The proposed method, which extracts perceptual tags from the temporal variation of the video frame means, is shown to be robust to typical coding operations in P2P and is computationally low for the utilization of P2P clients. The proposed method is then adapted to the shifts and cropping by means of developing a novel synchronization method to align perceptual tag extraction from different versions of the same file in the network. In order to be convenient with 64-bit computer registers and to achieve a low false positive rate, the length of the perceptual tag is selected as 64. Choosing also 16-bits for mapping, the success of the proposed distributed video identification system is verified with the detection rates of 88 % to 94% for cropping ratios up to 10% by using only two 64-bit descriptors for video files. The proposed method is verified to be well suited for distributed P2P implementation and can also have a potential for the developing mobile and content aware networks with its light weight characteristics.

Table 3. Experimental and computational detection rates in percentage (%) for different number of segments (R) and cropping rates (%) in the proposed tagging system. (Threshold for L=16 is 0 bit (no query for the erroneous versions) and L=64 is 5 bits.)

Detection Rates	**Cropping Ratio**	**Number of Segments or Descriptors (R)**			
		1	**2**	**3**	**4**
PD_C	0%	81.0	96.4	99.3	99.9
PD_E	0%	80.4	94.2	98.4	99.1
PD_E	5%	73.9	89.9	95.7	97.2
PD_E	10%	69.9	87.5	92.6	96.1

REFERENCES

[1] Cohen B. The BitTorrent Protocol Specification. [Online]. Available: http://www.bittorrent.org/

[2] Pouwelse JA, Garbacki P, Epema DHJ, Sips HJ. The Bittorrent P2P File-sharing System: Measurements and Analysis. In: 4th International Workshop on Peer-to-Peer Systems; 24-25 February 2005; Ithaca, NY, USA. pp. 205-216.

[3] Cisco Visual Networking Index: Forecast and Methodology, 2013-2018. White paper. [Online]: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf

[4] Kalker T, Epema DHJ, Hartel PH, Lagendijk RL, Van Steen M. Music2Share—Copyright-Compliant Music Sharing in P2P Systems. Proc. IEEE 2004; 92 (6): 961-970.

[5] Kundur D, Liu Z, Merabti M, Yu H. Advances in Peer-to-Peer Content Search. In: IEEE International Conference on Multimedia and Expo; 2-5 July 2007; Beijing, China. pp. 404-407.

[6] Jantunen A, Peltotalo S, Peltotalo J. Peer-to-Peer Analysis: State-of-the-art. Techical Report; Feb. 2006; Tampere University of Technology [Online]. Available: http://delco.cs.tut.fi/doc/other/p2p_analysis_v01.pdf

[7] Kulbak Y, Bickson D. The eMule Protocol Specification. Technical Report; Jan. 2005; DANSS Lab, The Hebrew University of Jerusalem, Israel.

[8] Koz A, Lagendijk RL. Distributed Content Based Video Identification in P2P Networks: Requirements and Solutions. IEEE Trans. on Multimedia 2017; 19 (3): 475-491.

[9] Law-To J, Chen L, Joly A, Laptev I, Buisson O, Gouet-Brunet V, Boujemaa N, Stentiford F. Video Copy Detection: A Comparative Study. In: 6th ACM Conference on Image and Video Retrieval – CIVR; July 9-11 2007; Amsterdam, Netherlands. pp. 371-378.

[10] Zhao WL, Ngo CW. Flip-Invariant SIFT for Copy and Object Detection. IEEE Trans. on Image Processing 2013; 22 (3): 980-991.

[11] Coskun B, Sankur B, Memon N. Spatio-temporal transform-based video hashing. IEEE Trans. on Multimedia 2006; 8(6): pp. 1190-1208.

[12] Koz A, Lagendijk RL. Perceptual Tagging of Video Files in P2P Networks. In: IEEE International Conference on Image Processing; 26-29 September 2010; Hon Kong, China. pp. 193-196.

[13] Hampapur A, Balle R. VideoGREP: Video Copy Detection using Inverted File Indices. Techical Report, IBM Research, 2002.

[14] Cheung SC, Zakhor A. Estimation of web video multiplicity. In: Proc. SPIE 3964; 20 December 1999; 3964: 34-46.

[15] Oostveen J, Kalker T, Haitsma J. Feature Extraction and a Database Strategy for Video Fingerprinting. In: VISUAL 2002, LNCS 2314. pp. 117-128.

[16] Ratnasamy S, Francis P, Handley M, Karp RM, Shenker S. A scalable content-addressable network. In: Proc. of the 2001 conference on applications, technologies, architectures, and protocols for computer communications; San Diego, California, USA. pp. 161-172.

[17] Shirky C. Listening to Naspter. In: Oram A, editor. Peer-to-Peer: Harnessing the Power of Distruptive Technologies. Sebostopal, CA, USA: O'Reilly, 2001. pp. 21-38.

[18] Liang J, Kumar R, Ross KW. The Kazaa Overlay: A measurement Study. Computer Networks Journal Elsevier, Special Issue on Overlays, October 2005.

[19] Pouwelse J, Garbacki P, Wang J, Bakker A, Yang J, Losup A, Epema DHJ, Reinders M, Van Steen MR, Sips HJ. Tribler: A Social-Based Peer-to-Peer System. Journal Concurrency and Computation: Practice and Experience 2008; 20(2): 127-138.

[20] Stoica I, Morris R, Liben-Nowell D, Karger DR, Kaashoek MF, Dabek F, Balakrishnan H. Chord: A Scalable Peer-to-peer Lookup Protocol for Internet Applications. IEEE/ACM Trans. on Networking 2003; 11 (1): 17-32.

[21] Xinxing Z, Zhihong T, Luchen Z. A measurement Study on Mainline DHT and Magnet Link. In: IEEE International Conference on Data Science in Cyberspace; 13-16 June 2016; Changsa, China. pp. 11-19,

[22] Blahut RE. Theory and Practice of Error Control Codes. Addison Wesley, 1983.

[23] Van Trees HL. Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory. John Wiley and Sons, 1968.