Digital Fingerprinting for Media Content Monitoring and Copyright Control

By WebKontrol

Today, people are overwhelmed by vast volumes of media content, some of which is illegal, and a growing number of sites where anyone can upload and watch anything. New ways of recording, replicating and distributing content emerge every day. More photographs were taken in 2017 than over the entire history of photography. With so many titles, channels and individual uploaders, compliant content providers are not always able to track questionable copies to take them down.

However, this seemingly untamable situation can be controlled, organized and monetized. AI-based digital fingerprinting is a powerful, robust tool that detects various types of media content across the far-reaches of the web.

This document provides a high-level overview of how fingerprinting technologies allow producers and publishers to automatically monitor, identify and protect their media content (prefiltering by IDs). It also describes in technical detail how WebKontrol's proprietary neural networks make fingerprints and sweep the internet for millions of copies and duplicates on behalf of its clients.

Media fingerprinting – what it is

Digital fingerprints are data files assigned to individual pieces of media content, containing unique characteristics of that particular content. They can be assigned to everything from movie titles and TV programs to cartoons, and images of all kinds. These fingerprints are normally used to identify apiece of content that may have leaked online – a pre-release or an unauthorized copy of a popular film – even if the content has been distorted or altered in ways such as adding subtitles, cropping and dubbing.

Generated from a sequence of key frames and their dynamics, every video fingerprint is a specific pattern of a certain shape and length in an N-dimensional space containing inherent characteristics of a source media file – a cartoon, a short video clip, or a full movie. Along with the individual pattern, the fingerprint can contain metadata about the media, making the initial stage of automatic content search and identification fast and cost-effective.

Though video and other media content, including images, require different fingerprinting algorithms and approaches, the outcomes are similar: resolution and format-independent fingerprints of a relatively small size.

At WebKontrol, we believe digital fingerprinting is a powerful tool for online content security, monitoring and control. It comprises the following components:

- Reliable, highly accurate AI technologies generate fingerprints that can identify originals and copies of high to medium and even upper-low visual quality, without generating false matches.

- A database of millions of fingerprints against which to compare content detected on web pages, video portals, social networks, P2P sites.

- A comprehensive index of fingerprinted content from major video hosting sites and a rapid technology to accurately compare newly uploaded copies to fingerprints inside the database in automated mode.

These sophisticated techniques are powered by WebKontrol's proprietary neural networks and 3 complex modules – Feature calculator, Indexer and Finder – that are examined below in more detail.

What is the value of digital fingerprinting?

In these days of user generated content and digital distribution channels, our technologies allow content owners to:

- develop new revenue models to legalize what previously seemed to be "pirated content" for new audiences,
- screen the web in search of illegitimate retransmissions of live TV events,
- identify all unauthorized online copies of their original content in real time to enforce their copyright via take-down notices.

Web monitoring and content screening vendors value creating a digital fingerprint that facilitates a fast, accurate and cost-efficient content identification process.

For content owners, from Hollywood majors to streaming services, VOD providers and online cinema websites, it is important to facilitate critical content-related business functions: production, promotion and protection.

Production and promotion may be manageable locally, but protection presents a more complex problem. Digital fingerprinting makes the whole spectrum of content security services accessible, including:

- ✓ Content monitoring
- ✓ Content pre-filtering
- ✓ Copyright control
- ✓ Forensics (content manipulation)

As a part of video security techniques, video fingerprints help to identify complete videos, portions of videos, video mash ups and short video snippets, including animated GIFs (basically, short animations and low-resolution video clips).

Content distributors and stock footage providers use fingerprints to check whether their digital libraries contain unauthorized content, and to monitor how the content they license is spread or manipulated.

Video fingerprints can be used as content IDs to prevent unauthorized copying, replication and even uploading of the original content pieces. In the latter case, every piece is automatically compared against a fingerprint database before becoming visible, and when there is a match, a warning to a website host pops up, saying that they should not allow this video to be shown because it is legally unwanted.

At WebKontrol, we create concise fingerprints for a variety of media types upon request.

It takes a couple of seconds to generate a digital fingerprint of a complete video piece, and a couple of milliseconds to automatically compare it against the database of fingerprinted content that we use to protect our customers. Our internal database contains more than 35 million fingerprints and is steadily growing, with 15-50 thousand new fingerprints arriving daily.

We make nearly 3.8 million comparisons a day, so that each new fingerprint is cross-checked against all the digital assets we have in just 8 days.

TECHNOLOGIES

WebKontrol's AI-based fingerprinting at a glance

WebKontrol's proprietary fingerprinting technology permits the creation of comparable video fingerprints based on a title's individual characteristics, including key scenes and their dynamics. Our technology automatically detects and extracts key video frames from the video stream. For each frame it intelligently marks the item's distinctive features: the global ones, that individually characterize the frame at whole, and the local ones, which portrait only informative fragments of the frame. Every feature is then encoded into a binary representation, and these representations form a data array of a scene. Together, all arrays make up a fingerprint of a certain shape and length. This means a video and its copies have fingerprints of almost the same central contours, even if copies are somehow distorted, dubbed or cropped.

For other media content such as animations or images (TIFF, JPEG, PNG, BPM, etc.), we use the same fingerprinting technology as for videos, but there are numerous technical nuances, including the use of:

- other APIs for data entry and processing of results,
- alternative technological chains, and
- customized configurations and parameters of artificial neural networks.

Artificial neural networks powered by Graphical Processing Units streamline the complex process of fingerprint creation and comparison in an N-dimensional data space. For each type of media content we have a dedicated neural network that has been specifically trained on various relevant content pieces of this particular type – e.g. pictures, movies, cartoons, news.

When fingerprints of the original piece of content and its copies are compared, the technology is looking for maximum similarity instead of exact matching. For example, a specifically shaped fingerprint of an online copy may have rough edges or minor shape variations, but our technology will trace it down to the most similar fingerprint of an original piece with the same central contours of the pattern. If the discrepancies in patterns are beyond a threshold – say, the top has been severely cut off – the technology does not consider these fingerprints as similar.

During its lifetime, each data set is regularly cross-checked against eventual false responses and is updated to reflect emerging computer vision approaches and new knowledge about the domain area.

Such a fingerprint comparison approach proves effective in detecting and identifying pirated copies with various distorting attacks, including:

- playback speed changed
- aspect ratio changed
- black areas in the video
- frame jitter and displacement (mostly in cam rips)
- color and brightness modifications (cam rips again)
- changes in the playback angle (tilting and rotation)
- inserted logos, subtitles and other “improvements”.

In our experience, we have encountered copies so distorted that technology does not identified a match, and indeed, consumers would not want to watch such poor-quality content. But in the vast majority of cases our artificial intelligence can detect and identify a pirated or user generated copy, regardless of how far it is distorted:

- ✓ For geometric, perspective and technical distortions, the technology spots similar items as long as copied videos remain “watchable” for a human eye.

Geometric distortion example: a rectangular video frame becomes diamond-shaped or pillow-shaped. Perspective distortion example: a rectangular video frame becomes a trapezoid. Technical distortions examples: frame jitter, frame displacement, black bar on the screen.

- ✓ For position distortion of the frame, the angle of rotation is important. Small angles up to 30° and right angles (90°, 180° and 270°) are typical and not problematic for correct content identification. Other angles are questionable, but first from the human perspective – no one would be watching such videos because of their quality.

Position distortion, example: a rectangular video frame rotates through an angle of N.

- ✓ The same is true in the case of film transitions and information losses caused by frame cropping and/or frame compression, or any modifications of color, brightness, and frame proportions. While they maintain an acceptable video quality for watching, the technology will trace down all the copies of the original piece.

Film transitions, examples: wipes, fades, L cuts, and other effects.

- ✓ For third party content insertion we have a clear distinction. Logotypes and subtitles do not dramatically spoil the visual quality of videos, while the quality of film translations and dubbing is not crucial, since the technology does not rely on audiotracks when making fingerprints.

These and other technical aspects of the complex process of content fingerprinting, online monitoring and identification are all solved automatically by a combination of advanced WebKontrol's technologies and proprietary technology components we call modules. Let us look at them closely.

Going deeper in details: the technologies and the process

Once uploaded to major hosting and sharing sites, illegal content is often promoted via thousands of landing pages, in many languages. With a manual or half-automated search, it's almost impossible to issue a takedown notice on all copies, let alone renamed, cropped and distorted versions.

WebKontrol's proprietary AI technologies hound the root file down to its hosting server – no matter its quality and nicknames – and when it is removed after a takedown notice, all related copies embedded to landing pages vanish too.

This is how it works in practice:

1. The technology creates fingerprints of our clients' content using API, which means we never have access to an original video file and never store digital originals inside our servers.

2. Another set of technologies then compares these fingerprints to our fingerprint database and our index of fingerprinted content from major hosting, streaming and file sharing sites. This ensures rapid response rates once matches are detected.

3. This index is constantly updated, and content identification takes place 24/7, making detection of unauthorized copies effective and cost-efficient.

4. On a parallel track, our web crawler automatically screens all video-related services and sites on the global web, including social media platforms and torrents.

5. Once there is a match, a take-down notice is automatically sent out to a website owner, whose physical address is known. The evidence and its details (including the time of issue) is recorded, both the event and response are monitored, and all relevant data is stored in a digital archive for any legal proceedings, should the client decide to take further action.

WebKontrol's web crawler is based on advanced computer vision algorithms. We call it unstoppable, because it can pass over complicated hosting algorithms of the top video storage sites to screen vast volumes of data from the depths of multiple servers.

A cost-effective multi-layered search technology allows different types of search to focus on exact content locations – first by different types of metadata, then by fingerprints.

It also includes ***Indexer*** and ***Finder*** components for higher precision, just as the AI-based fingerprinting technology contains a ***Feature calculator*** module to rapidly create high-quality compact digital signatures of content.

The modules: their functions and technical description

The Feature Calculator extracts key features of a piece of content and encodes them into the N-dimensional fingerprint representing that content the way that ensures similar pieces of content to have similar fingerprints. The calculator has a set of phases, which can be described more technically in the following way:

1.0. The video is decoded and divided into frames. Not all frames are chosen, only those that differ in histogram types (YUV or RGB). While the frames are extracted, a range of filters suppresses the noise and rejects the connected areas of pixels on the edges of each frame, where the overall brightness is below a threshold. For reference, consider the following papers [5], [6].

1.1. A global feature is calculated with the help of a convoluted network. This network is an "alexnet" modification (see [4]). Let us denote it by V.

1.2. Local features are detected (see [1]). Let us denote these spots by p_i.

1.3. A set of local features is detected by using these spots: namely the color feature (see [2]) and the feature in charge of the gradient behavior (see [1], [3]). Let us denote these vectors by u_i and w_i. They correspond with p_i.

Thus, each frame denoted by I has a corresponding set of features generated at the above phases. In other words, a frame's representation F can be described in the following way: I → {V, p_i, u_i, w_i}. These are the so-called primary features, which are supplemented by additional features upon the search request.

References:

[1] Pablo F. Alcantarilla, Adrien Bartoli and Andrew J. Davison: KAZE Features

[2] Xiang-Yang Wang, Jun-Feng Wu1 and Hong-Ying Yang: Robust image retrieval based on color histogram of local feature regions

[3] Herbert Bay, Tinne Tuytelaars and Luc Van Gool: Speeded Up Robust Features

[4] Alex Krizhevsky, Ilya Sutskever: ImageNet Classification with Deep Convolutional Neural Networks

[5] Shapiro, Linda G. and Stockman, George C. «Computer Vision»

[6] Novak, C.L.; Shafer, S.A. «Anatomy of a color histogram»

Indexer creates a specific set of data structures which provide quick search and access to videos which are similar in features. Feature indexation has its own phases:

1.0. A hybrid structure is generated on the basis of three components: a locality-sensitive hash coding, a K-dimensional tree, and a clustering over the range of vectors (created at phase 1.1. of the feature calculation). See in detail in reference papers [7], [8], [9].

1.1. Local features (calculated at phase 1.3 of the feature calculation) are clustered (quantized) using the methods described in [10] and [11].

1.2. Based on the results of the previous phase, the corpus of documents (a frame is considered a document) is marked and the so-called index is created (a cluster's center acts as a word) [12].

1.3. Using the clusters' centers, a local descriptor is projected into a binary signature, according to papers [13], [14].

Feature indexation is over when the supplemental structures are ready and the list of correspondence between the index documents and the videos is created.

References:

[7] Marius Muja and David G. Lowe, "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration", in International Conference on Computer Vision Theory and Applications (VISAPP'09), 2009

[8] Marius Muja and David G. Lowe: "Fast Matching of Binary Features". Conference on Computer and Robot Vision (CRV) 2012

[9] Marius Muja and David G. Lowe: "Scalable Nearest Neighbor Algorithms for High Dimensional Data". Pattern Analysis and Machine Intelligence (PAMI), Vol. 36, 2014

[10] McCallum, A.; Nigam, K.; and Ungar L.H. (2000) Efficient Clustering of High-Dimensional Data Sets with Application to Reference Matching

[11] D. Arthur, S. Vassilvitskii - k-means++: The Advantages of Careful Seeding

[12] Herve Jegou, Matthijs Douze: Packing bag-of-features

[13] Herve Jegou, Matthijs Douze: Improving bag-of-features for large scale image search

[14] Matthijs Douze, Adrien Gaidon, Herve Jegou, Marcin Marszalek: INRIA-LEARs video copy detection system

Finder is a specific search module which compares a video request against a corresponding index document. The whole search procedure is performed as follows:

1.0. Upon the video request, direct features are created by the Feature calculator.

1.1. Indirect features – those dependent on the “request-index” match, see [15] – are built upon the direct ones. In this case the process includes the bag-of-words [15] and the calculation of signatures, which is performed for each key frame of the video request.

1.2. For the requested video frame, possible similarity candidates are chosen by nearest neighbor searching.

1.3. The requested frame is measured (ranked) against each similarity candidate using the following methods:

a) Ranking by the Okapi BM25 formula [16]

b) Signature-based ranking [15], [17]

c) Ranking based on the geometric transformation evaluation [18], [19]

d) Color-information based ranking [20]

1.4. Then the “unfit” candidates are removed, and the remaining candidates are considered adequate.

1.5 Times shifts are calculated for the requested frame and all the frames ranked as adequate from the index. The resulted shifts are quantized, and the candidate frames belonging to the same source (videos from the index) is connected to a separate histogram cell.

1.6. Phases 1.0 - 1.5 are repeated for each video frame. The results are: a) each histogram cell relates to a set of frames of the same video from the index and b) we have the total vote called an assumption [15].

1.7. Assumptions are ranged by decreasing of votes and provided as search results.

References:

[15] Matthijs Douze, Adrien Gaidon, Herve Jegou, Marcin Marszalek: INRIA-LEARs video copy detection system

[16] Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. Okapi at TREC-3. In Proceedings of the Third Text REtrieval Conference (TREC 1994). Gaithersburg, USA, November 1999

[17] Herve Jegou, Matthijs Douze: Hamming embedding and weak geometric consistency for large scale image search

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60:91–110, 2004.

[19] E. Dubrofsky and R. J. Woodham. Combining line and point correspondences for homography estimation. In ISVC ’08: International Symposium on Visual Computing, 2008.

The afterword

WebKontrol’s technologies and the process of content fingerprinting, monitoring, and identification have been successfully tested on movies, cartoons and series as the most complex and large type of media content. However, we can fingerprint and detect all other types of online content, including images, TV programs, texts, and music.

ABOUT WEBKONTROL

WebKontrol is a worldwide technology provider and anti-piracy safeguard, with 8 years' experience protecting content across the global web for numerous content developers, including the Hollywood majors.

We know content is precious. Safeguarding content matters. It allows creators to regain control of their work, their reputations and the revenues that fuel future innovation. Consumers can be confident they are accessing safe, legal and high-quality content they can trust and enjoy.

WebKontrol pairs cutting-edge AI technology with deep expertise in content protection. We scan vast volumes of online content and verify its legal use across the globe. We pair our technology with rigorous enforcement actions, giving global content owners the power to protect the distribution of their video, music and text products in the digital universe.

Today, WebKontrol uses a database of 35 million digital fingerprints to cancel the unauthorized uploading of titles to sites (prefiltering by content ID) or to accurately identify unwanted copies online. Hundreds of websites worldwide are screened 24/7, and tens of thousands of illegal items are detected every month, including those on P2P sites and social media platforms.

WebKontrol unites high-level specialists – a team of talented lawyers and IT professionals – to evolve proprietary AI technologies and develop services that bring the digital content distribution market into line with international copyright standards and laws.

Contact us:

https://www.webkontrol.com
phone: +7 (495) 663 9361