Video Copy Detection providing Localized Matches

Damian Borth, Adrian Ulges, Christian Schulze, Thomas M. Breuel
d_borth@cs.uni-kl.de
German Research Center for Artificial Intelligence (DFKI)
and University of Kaiserslautern

Abstract:
With the availability of large scale online video platforms like YouTube, copyright infringement becomes a severe problem, such that the demand for robust copy detection systems is growing. Such system must find multiple occurrence of copyright protected material within video clips that are created, modified, remixed and uploaded by the user. A particular challenge is to find the exact position of a copy in a – potentially huge – reference database. For this purpose, this paper presents a Content Based Copy Detection system that both detects copies in query videos against a reference database and gives an exact alignment between them. For finding and aligning a matching shot, a fast search for candidates is conducted, and as a second step an exact alignment is found using a dynamic programming minimization of the well-known edit distance from text retrieval. The introduced approach was evaluated on the public available MUSCLE-VCD-2007 [LTJB07] data corpus and showed competitive alignment results compared to the ACM CVPR 2007 evaluation.

GI-Topic: KI-BV (artificial intelligence - image understanding)

1 Introduction

Large scale online video platforms like YouTube rises new critical issues for content owner and platform provider when it come to the illegal distribution of copyright protected video content by its users. Furthermore the nature of such platforms encourage the reediting, modification and mixing of self-created video content with already available video content not created by the user. Often the reused video snipplets consist of copyright protected video material that is used without the knowledge of the content owner and therefore should be identified and deleted by the platform provider. Content Based Copy Detection (CBCD), as an alternative to a watermarking technique, solves this issue.

As stated by [LTCJ+07], it is important to not confuse similarity driven search in Content Based Video Retrieval (CBVR) with the task of finding a copy in CBCD. A copy is not only an identical or near duplicate video sequence, it is rather a transformed video sequence resulting in a distortion of its original appearance. In Fig. 1 possible video transformations are shown.

2 Our Approach

A common setup of a CBCD systems holds a database of known copyright protected videos (reference database) and checks questioned video (query videos) against it [DLÁ+07]. In case a query video contains a reference video the CBCD system should match both

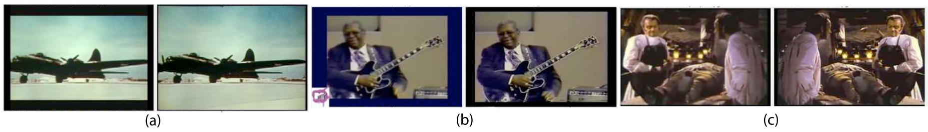

Figure 1: Common video transformations: a) cropping, b) logo insertion and blurring, c) flipping

video sequence regardless of the applied transformations. Our method is utilizing an edit distance [Lev66] based similarity measurement on extracted video signatures to detect and localize copies within query videos. As CBCD systems consist of common components [GB08] [DGJ+08], only the differences to those are explained as following:

Preprocessing & Feature Extraction The preprocessing step performs an adaptive keyframe extraction [BUSB08], which additionally segments the processed video into shots. Having these structural information of the video we extract color histograms from the keyframes and MPEG-7 Color Layout Descriptors (CLD) [MOVY01] from the shots. This process is performed on the entire reference database and for each query video.

Candidate Computation & Similarity Check Due to efficiency reasons we first compute k shot candidates using Nearest Neighbor Matching on keyframe level with color histograms features. Then we perform the computational more expensive edit distance calculation on the shot candidates using CLD features, mainly due to their robustness and efficiency in CBVR [KY01]. In Fig. 2 the resulting trellis images of similarity check between a query and a reference shot is shown. The left image visualize the direction selection during the edit distance calculation. A distinct localization of a copy would appear as a diagonal line indicating an identical sequence given by the lowest cost for its frame-by-frame comparison. It can be seen that the left image is too noisy to make a distinct localization possible. An additional Least Square Fitting of the blue trellis path in the right image improves the localization precision dramatically. In combination with the edit distance value and the trellis information we are able to make a global decision if a positive match was detected and a local decision where the match occurred.

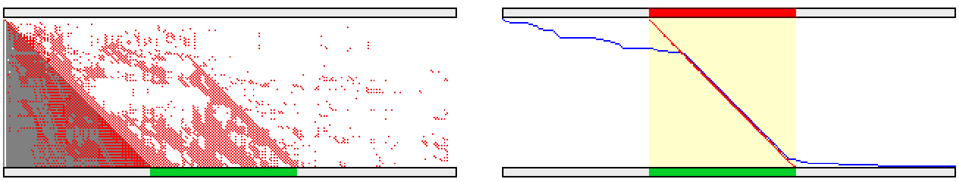

Figure 2: The left image visualize the direction selection during the edit distance calculation. The right image displays the result of the Least Square Fitting of the blue trellis path. The green bar below the images indicates the true copy position, the red bar at the top displays the localization result.

Match Marking & Post Processing In this processing step the decision must be made if a query video contains sequences of a reference video. For the global decision we threshold the calculated edit distance of the similarity check against a predefined value where low costs indicate high similarity. In case of a positive match, adjacent query shots with their edit distance result are checked against the threshold and potentially merged together. This leads to a localization based on the entire query video and on reference video level.

3 Experiments & Results

For experimental evaluation of the copy detection system we used the MUSCLE-VCD-2007 data corpus [LTJB07]. The dataset consist of a reference database with 100 hours of video materials and a set of 3 queries with total length of 45 min. This query set belongs to the **ST2** task of the **CVPR 2007 Copy Detection Live Benchmark**, which deals with localized CBCD. The structure of the provided queries can be seen in Fig. 3, where each query consist of randomly selected video sequences form the reference database mixed with non-reference material and modified by one or several of the following transformations: cropping, fade cuts, insertion of logos, moving texts, change of contrast, gamma. Sequences belonging to the reference database must be detected and localized in the query and the reference video with their start and end time code.

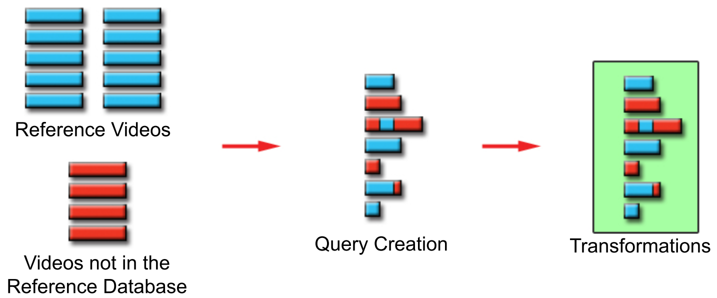

Figure 3: Query structure, mixing random video sequences from a reference and non-reference database together and applying random transformation to each of the query

The used performance criteria for the CBCD evaluation were defines as the following two separate measurements:

$$QualitySegment = \frac{Ncorrect - FalseAlarm}{Nsegments}$$

$QualitySegment$ is computed from the percentage of mismatched segments in all queries, where $Ncorrect$ is the number of correct matches, $FalseAlarm$ is the number of false positive matches and $Nsegments$ represents the number of all detected segments.

$$QualityFrame = 1 - (\frac{Nmis}{Nframes})$$

$QualityFrame$ represents the percentage of mismatches frames in all queries including non-detected segments, imprecision within correctly detected segments and false positive matches. Here $Nmis$ stands for the amount of mismatches frames and $Nframes$ for the total number of query frames.

Our system obtained a value of **0.79** for $QualitySegment$, compared with 0.86 as best and 0.33 as worst result of the 2007 evaluation and **0.35** for $QualityFrame$, compared with 0.76 as best and 0.17 as worst result of the 2007 evaluation. An analysis of the $QualityFrame$ results identified false negative, like depicted in Fig. 5, as major source of frames mismatches (700 sec.). Frame mismatches for correctly detected copy sequences were in the range of 120 sec. out of a total query length of 45 min. containing 21 min. of video from the reference database.

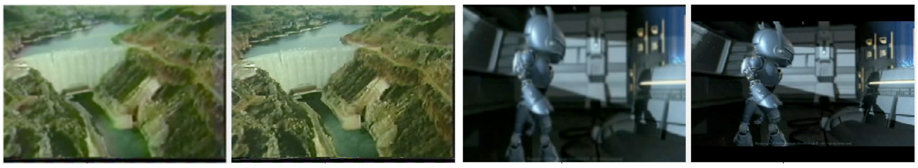

Figure 4: Two examples of correctly detected query sequence which were modified by blur, change of color (left) and crop, blur, zoom (right) transformations

Figure 5: An example of a not detected query sequence modified by gamma transformation, a vertical shift and insertion of subtitles. The lack of color information additionally challenged the detection system due to its focus on color features.

References

[BUSB08] Damian Borth, Adrian Ulges, Christian Schulze, and Thomas M. Breuel. Keyframe Extraction for Video Taggging and Summarization. In *Informatiktage 2008*, pages 45–48, 2008.

[DGJ+08] M. Douze, A. Gaidon, H. Jegou, M. Marszałek, C. Schmid, and F. Grenoble. INRIA-LEAR's Video Copy Detection System. In *TRECVID-Workshop*, 2008.

[DLÁ+07] K. Dadason, H. Lejsek, F. Ásmundsson, B. Jónsson, and L. Amsaleg. Videntifier: identifying pirated videos in real-time. In *Proceedings of the 15th international conference on Multimedia*, pages 471–472. ACM Press New York, NY, USA, 2007.

[GB08] N. Gengembre and S.A. Berrani. The Orange Labs Real Time Video Copy Detection System-TrecVid 2008 Results. In *TRECVID-Workshop*, 2008.

[KY01] E. Kasutani and A. Yamada. The MPEG-7 color layout descriptor: a compact image featuredescription for high-speed image/video segment retrieval. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 1, 2001.

[Lev66] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10:707â710, 1966.

[LTCJ+07] J. Law-To, L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet-Brunet, N. Boujemaa, and F. Stentiford. Video copy detection: a comparative study. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 371–378. ACM Press New York, NY, USA, 2007.

[LTJB07] J. Law-To, A. Joly, and N. Boujemaa. Muscle-VCD-2007: a live benchmark for video copy detection, 2007. http://www-rocq.inria.fr/imedia/civr-bench/.

[MOVY01] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada. Color and Texture Descriptors. *IEEE Trans. on Circuits Syst. for Video Techn.*, 11(6), 2001.