Video Phylogeny Tree Reconstruction Using Aging Measures

Simone Milani
Department of Information Engineering
University of Padova, Padova, Italy
Email: simone.milani@dei.unipd.it

Paolo Bestagini, Stefano Tubaro
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano, Milan, Italy
Email: paolo.bestagini@polimi.it, stefano.tubaro@polimi.it

***Abstract*—The increasing diffusion of user-friendly editing software and online media sharing platforms has brought forth a growing on-line availability of near-duplicate (ND) videos. The need of authenticating these contents and tracing back their history has led to the investigation of forensic algorithms for the reconstruction of the video phylogeny tree (VPT), i.e., an acyclic directed graph summarizing video genealogical relationships. Unfortunately, state-of-the-art solutions for VPT reconstruction suffer from strong computational requirements.**

In this paper, we propose a processing age measure based on video DCT coefficients and motion vectors statistics, which enables to provide preliminary information about possible video parent-child relationship. The use of processing age allows a forensic analyst to blindly select a smaller amount of significant video pairs to be compared for VPT reconstruction. This solution grants computational complexity reduction to the overall VPT reconstruction pipeline.

I. Introduction

The recent disposal of versatile acquisition, editing, and sharing tools has led to the spreading of multiple versions of the same multimedia objects, which are called near-duplicates (NDs). This has brought several new issues and problems concerning the discrimination of the originating file, the identification of the owner, or the reconstruction of the processing history of each copy [1], [2], [3]. In these tasks, multimedia forensics research has usually focused on the detection of footprints left on images [4] or video sequences [5] by each editing step. This analysis is significantly affected by the modelling accuracy and the amount of noise affecting the data under analysis (which could erase or alter these traces).

As a matter of fact, recent researches have been focusing the analysis on the relations between different versions of the same content [6], [7], [8]. The underlying idea is that multimedia contents evolve like DNA sequences of organisms mutate in biology. This process can be well-described by means of a structure called phylogeny tree (PT), and phylogenetic analysis permits reconstructing it by analyzing similarities between the nucleotides sequences of different organisms. Similarly, multimedia phylogeny solutions aim at building a complete relational graph, in which edge labels model the similarity/dissimilarity between every pair of ND images or videos [9], [6]. Then, the underlying PT is estimated by means of graph optimization strategies that identify which dependency relations among the different contents are the most plausible.

Unfortunately, the accuracy of the PT reconstruction is degraded by several factors such as the noise affecting the similarity/dissimilarity measures, or the missing of some objects/nodes in the analysis pool. Moreover, in order to build the relational graph, current solutions prove to be computationally expensive due to the need of comparing every pair of ND objects in the analysis set. This is a problem especially when video sequences are taken into account, rather than still images.

The current paper aims at reducing the computational burden of the typical state-of-the-art video phylogeny tree (VPT) reconstruction pipeline. To this purpose, we present a set of processing age metrics for video sequences that are based on statistics of DCT coefficients and motion vector differences. By including the proposed metrics in the video phylogeny tree reconstruction process, it is possible to check the feasibility of graph edges before running the optimization routine on them. Experimental results performed on 2.800 ND video sequences, show that the proposed solution permits improving the accuracy of the identification of the root sequence (i.e., the original one used to generate all the other ND in a set), and it reduces the computational complexity of the overall VPT reconstruction scheme.

In the following, Section II presents the problem of VPT reconstruction and overviews some of the works published on the subject. Section III describes how the proposed processing age metric is computed, and reports how to include it in the VPT reconstruction strategy. Finally, Section IV verifies the performance of our algorithm by means of thorough empirical testing, whereas Section V draws the final conclusions.

II. Problem statement and related works

Two video sequences S_i and S_j are considered NDs if they can be generated applying some content preserving editing

This work has been partially supported by the project Phylo4and6 BIRD165882/16 funded by the University of Padova, Padova, Italy.

This material is based on research sponsored by DARPA and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and Air Force Research Laboratory (AFRL) or the U.S. Government.

ISBN 978-0-9928626-7-1 © EURASIP 2017

operations (e.g., blurring, coding, brightness adjustment, etc.) to the same originating video sequence S_0. Solving the problem of VPT reconstruction means finding the genealogical relationships between ND videos in a pool in order to infer which sequence generated another one. In order to solve this problem, state-of-the-art solutions are inspired by works originally proposed for still images [9], [10], and basically follow a common pipeline [6], [11].

First, video phylogenetic strategies start building a complete relational graph where each node corresponds to a different video sequence, and edge labels denote dissimilarity (or alternatively similarity) relations between nodes [6], [11]. Specifically, given a pair of ND video sequences S_i and S_j, dissimilarity is defined as

$$\mathcal{D}(S_i, S_j) = \arg\min_{\mathcal{T}} \mathcal{L}\left(S_i,\, \mathcal{T}(S_j)\right), \tag{1}$$

where $\mathcal{L}$ computes the mean squared error, and $\mathcal{T}$ is the combination of editing operations (such as cropping, resizing, logo addition, rotation, color enhancement, etc.) that best maps S_j into S_i. The rationale behind dissimilarity is that, if a transformation $\mathcal{T}$ mapping S_j into S_i exists (i.e., low dissimilarity value), then S_j may have been used to generate S_i. Conversely, if this transformation does not exists (i.e., high dissimilarity value), the two sequences are surely not in parent-child relationship. The most time consuming operation in VPT reconstruction is the estimation of $\mathcal{T}$.

Then, the underlying VPT is estimated by means of optimization strategies that identify the maximum/minimum spanning tree, like Oriented Kruskal (OK) [12] or Optimum Branching (OB) [13].

Unfortunately, the accuracy of the reconstruction can be significantly impaired by several factors. Often the adopted similarity/dissimilarity metric is highly noisy, leading to several reconstruction errors. This fact is more evident whenever video sequences have been significantly edited at every ND generation, and therefore, several equalization and synchronization steps need to be applied in order to have a meaningful measurement [11]. One of the most frequent errors is parent-child inversion, which takes place whenever the editing operations that generate the child do not significantly change the visual information of the father (e.g., in the case of minor cropping). Moreover, many reconstruction errors arise whenever some nodes of the VPT trees are missing, which causes the estimation algorithm to approximate ancestry relations via the similarity of non-directly related nodes. Additionally, computational complexity is another crucial issue since the dissimilarity needs to be computed for every pair of videos; thus, the overall amount of calculation scales quadratically with the number of videos in the analysis pool.

Problems related to noisy dissimilarity values can be effectively mitigated by including additional redundancy in the reconstruction process [14]. Conversely, problems related to high computational burden can be mitigated by techniques enabling to pre-emptively select subsets of video pairs to analyze. To this purpose, the approach in [15] introduces a no-reference quality metric that models the *processing age* (PA) of images, i.e., the amount of editing that has been applied on every ND image in the dataset. By comparing the PA of the images, it is possible to exclude *a-priori* some parent-child relations that appear to be unfeasible (i.e., a parent with a PA lower than his child). This operation permits reducing the computational complexity of the overall PT reconstruction, and improves accuracy.

In this paper, leveraging findings of [15], we propose a processing age measure for video sequences, which enables to reconstruct the VPT with decreased computation complexity.

III. Video Phylogeny Tree Reconstruction Using processing age measure

Because of the massive amount of data that need to be stored, video sequences are usually available in compressed format. As a matter of fact, every editing performed on a video sequence needs to be followed by a compression operation. Estimating the number of coding steps permits placing the analyzed video sequence at the correct depth of the reconstructed VPT (i.e., a video compressed many times cannot be parent of a video compressed less times). This permits detecting wrong dependencies and removing unfeasible links (thus reducing the computational complexity since their similarities/dissimilarities do not need to be computed anymore).

Video processing age. The forensic community has faced the problem of double or multiple video compression detection before [16], [17], [18]. Anyway, most of the approaches rely on training a machine-learning classifier on a set of video sequences which were edited and coded according to a finite set of possible parameters. Since in a real scenario the range of possible editing and coding choices is quite wide, we investigate a more general no-reference metric that permits comparing and ordering different ND sequences according to their creation time rather then identifying the exact number of compressions operated on each of them.

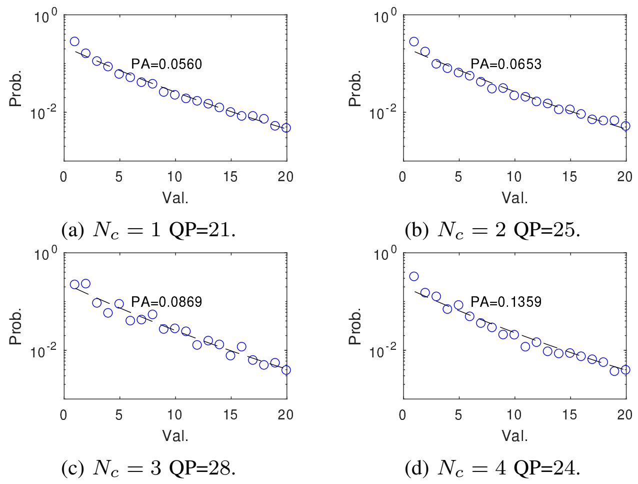
We call this metric *processing age* (PA), and compute it analyzing the statistics of the DCT coefficients of prediction residuals of video frames, as well as motion vector statistics, leveraging the findings in [15].

Aging metric based on DCT coefficient statistics

Given a video sequence S_i, the n-th frame $S_i(n)$ is predicted from the frame $S_i(n-1)$ using a motion estimation routine. The generated prediction $P_i(n)$ is then subtracted to it generating the prediction residual $R_i(n) = S_i(n) - P_i(n)$. Then, $R_i(n)$ is partitioned into 4×4 blocks x, and each one of them is transformed using the 4×4 DCT-like transform of H.264/AVC and quantized into the integer output coefficients X_q. In the analysis, we adopted 4×4 blocks since it is the smallest transform size adopted by the existing video coding standards (and therefore, it grants the finer granularity on frame analysis).

For every spatial frequency (u, v), it is possible to compute a histogram of the absolute coefficient values $c = |X_q(u, v)|$. This empirical statistics, which will be referenced

ISBN 978-0-9928626-7-1 © EURASIP 2017



(a) $N_c = 1$ QP=21.

(b) $N_c = 2$ QP=25.

(c) $N_c = 3$ QP=28.

(d) $N_c = 4$ QP=24.

Fig. 1. Probabilities $P_{0,2}(c)$ (blue values) and the fitted model $\tilde{P}_{0,2}(c)$ (dashed line) for frame 1 of sequence `soccer` compressed N_c times. The adopted video coded is H.264/AVC with quantization parameter QP.

with the symbol $P_{u,v}(c)$, can be well-approximated by an exponentially-decreasing model. As a matter of fact, several DCT coefficients fitting models have been proposed in literature, such as Laplacian [19], generalized Gaussian [20], laplacian+impulsive [21], and Cauchy [22]. In this work, we simplified all these models with the function

$$\tilde{P}_{u,v}(c) = \Gamma e^{-\pi(c)}, \tag{2}$$

where $\pi(\cdot)$ is a polynomial of third degree and Γ is a normalizing constant. In this way, it is possible to include both a Laplacian and a Gaussian model for the absolute value of quantized coefficients avoiding the fitting problems related to the generalized Gaussian.

The statistics $\tilde{P}_{u,v}(c)$ is obtained by fitting $c > 0$ values to the given model. Null coefficients are omitted since many video coders adopt dead-zone quantizers and non-linear coefficient cancellation strategies driven by rate-distortion optimization routine. This alters coefficient statistics making the fitting more complex.

Fig. 1 reports the statistics of quantized DCT coefficients (on semi-logarithmic axes) for the sequence `soccer` coded N_c times with varying quality parameters. It is possible to notice that as N_c increases, the empirical $P_{u,v}(c)$ deviates from the fitted $\tilde{P}_{u,v}(c)$ model. Therefore, it is possible to associate the processing age metric with a divergence metric. To this purpose, we considered the Jensen-Shannon divergence

$$a_{u,v} = \frac{1}{2}\sum_c P_{u,v}(c)\log_2\frac{P_{u,v}(c)}{\tilde{P}_{u,v}(c)} + \frac{1}{2}\sum_c \tilde{P}_{u,v}(c)\log_2\frac{\tilde{P}_{u,v}(c)}{P_{u,v}(c)} \tag{3}$$

The graphs in Fig. 1 reports the PA values $a_{u,v}$ for different N_c. It is possible to notice that the values $a_{u,v}$ increase as the number of compression increases.

Experimental results showed that this property is verified for low-frequencies coefficients; as a matter of fact, processing age computation was limited to a subset $\mathcal{U}$ of $N_{\mathcal{U}}$ spatial frequencies corresponding to the first 9 AC coefficients (following a zig-zag scan).

Aging metric based on motion vector statistics

A similar analysis can be performed on the statistics of motion vectors (MVs). At first, each frame is partitioned into 4×4 blocks and a displacement vector $\mathbf{v} = [v_x, v_y]$ is assigned to each block from motion vector values coded in the coded stream. Whenever motion vectors are referred to larger blocks, displacement vectors are obtained by replicating the corresponding MV. Then, for every MV of the frame, motion vector difference $\mathbf{d}_{MV}$ is computed as follows:

$$\mathbf{d}_{MV} = [|d_x|, |d_y|] = \left[\left|v_t - \frac{v_t^A + v_t^B}{2}\right|\right]_{t=x,y} \tag{4}$$

where $\mathbf{v}^A$, $\mathbf{v}^B$ are the displacement vectors related to the left and upper 4×4 blocks.

The statistics of $|d_x|$, $|d_y|$ can be well-characterized by a second-order description defined by the averages $m_x = E[|d_x|]$, $m_y = E[|d_y|]$ and the corresponding variances $\sigma_x = E[|d_x| - m_x]$, $\sigma_y = E[|d_y| - m_y]$. As a matter of fact, it is possible to define two motion vector based aging metrics as

$$a_{avg} = \frac{m_x + m_y}{2}, \qquad a_{var} = \frac{\sigma_x + \sigma_y}{2}. \tag{5}$$

PA-based video phylogeny tree estimation. Given the previously-described metrics, it is possible to generate for the i-th video sequence an age vector

$$\mathbf{a}_i = [a_{i,k}] = \left[[a_{u,v}]_{(u,v)\in\mathcal{U}}, \quad a_{avg}, \quad a_{var}\right] \tag{6}$$

that groups the different metrics.

For every pair of nodes/videos S_i, S_j in the dissimilarity graph, it is possible to check the hypotheses $H_1 = \{S_i \text{ is younger than video } S_j\}$ and $H_2 = \neg H_1$ for every component $a_{i,k}$. More precisely, if $a_{i,k} - a_{j,k} < \gamma$, the hypothesis H_1 is verified by the k-th aging metric; if $a_{i,k} - a_{j,k} > \gamma$, H_2 is considered valid; otherwise the situation is doubtful and nothing is done. The threshold γ can be chosen upon training depending on how much we trust PA for the considered video tree. Composing the outcomes for all the ages via a majority voting strategy, it is possible to determine which hypothesis between H_1 and H_2 is more likely. In case H_1 obtains the majority of votes, the link $S_j \to S_i$ is removed from the graph; in case H_2 wins, link $S_i \to S_j$ is erased; otherwise, nothing is removed.

Then, the underlying minimum spanning tree (MST) can be estimated from the resulting dissimilarity graph using a standard optimum branching strategy (like in [13]).

Note that, in case a link is removed, dissimilarity computation for that link is skipped reducing the overall computational complexity. Moreover, whenever the noise level affecting the dissimilarity is high, final accuracy can improve as well. These advantages will be highlighted in the following section.

IV. Experimental results

In this section we describe the performed experimental campaign and the achieved results in order to validate the proposed algorithm.

TABLE I
EMPLOYED OPERATIONS, CODECS, QP VALUES AND GOP SIZES.

Operation	Parameters
Resize	New size in $[90\%, 110\%]$
Crop	New size in $[90\%, 98\%]$
Brightness	Luminance increased or decreased up to 10%
Contrast	Luminance mapped in ranges $\{10\%, 80\%\}$ or $\{20\%, 90\%\}$
MPEG2	QP $\in [2, 10]$ and GOP $\in [15, 30]$
MPEG4	QP $\in [2, 10]$ and GOP $\in [15, 30]$
H264	QP $\in [5, 25]$ and GOP $\in [15, 30]$

TABLE II
BREAKDOWN OF THE VIDEO PHYLOGENY TREE DATASETS (10 NODES).

Dataset	N. Trees	Topology	Operations	Time Clip
$\mathcal{D}_{\text{MPEG2}}$	30	Chain	MPEG2	No
$\mathcal{D}_{\text{MPEG4}}$	30	Chain	MPEG4	No
$\mathcal{D}_{\text{H264}}$	30	Chain	H264	No
$\mathcal{D}_{\text{cod}}$	30	Chain	Random Codec	No
$\mathcal{D}_{\text{geom}}$	30	Chain	Crop or Resize + Random Codec	No
$\mathcal{D}_{\text{luma}}$	30	Chain	Brightness or Contrast + Random Codec	No
$\mathcal{D}_{\text{tree}}$	100	Random	Random Operation + Random Coding	Yes

Datasets. To test the proposed method under diverse working conditions, we built different datasets for a total number of 2.800 near-duplicate videos. As original videos, we selected 15 well-known sequences at CIF resolution (352x288) of approximately 300 frames each, namely: *city*, *crew*, *news*, *foreman*, *hall*, *akiyo*, *coastguard*, *container*, *flower*, *mobile*, *mother*, *paris*, *salesman*, *soccer*, and *table*[1]. Near-duplicates have been generated applying to a video one editing operation and a coding step. Optionally, temporal clipping was also applied by removing 10% of frames from the head or the tail of a video. The breakdown of editing operations and parameters is reported in Table I.

Considering these transformations, we generated different realizations of video phylogeny trees of 10 nodes each, randomly selecting a root among the 15 original videos and mixing different operations with different tree topologies. Table II reports the list of all generated datasets, reporting the number of realizations, the used operations, and considered topology (i.e., random or chain). Datasets $\mathcal{D}_{\text{MPEG2}}$, $\mathcal{D}_{\text{MPEG4}}$, $\mathcal{D}_{\text{H264}}$ and $\mathcal{D}_{\text{cod}}$ are composed by chains (i.e., each video only generates one near-duplicate sequence) of only coded videos, using MPEG2, MPEG4, H264 and mixing them, respectively. Conversely, $\mathcal{D}_{\text{geom}}$ and $\mathcal{D}_{\text{luma}}$ are composed by chains of videos to which only one operation (geometrical or luminance-based) and one coding step were applied. Finally, $\mathcal{D}_{\text{tree}}$ contains randomly shaped trees of videos to which any operation and coding scheme has been randomly applied, in addition to possible temporal clipping.

Methodology. We measure the performance of the proposed algorithm according to different indicators.

In order to evaluate how many video pair comparisons we can successfully avoid thanks to the use of aging metrics, we compute the `Parent-Child Link Loss`. This is the percentage of parent-child relationships that are mistakenly removed and not considered for dissimilarity computation. Clearly, discarding parent-child edges from the dissimilarity matrix hinder the reconstruction of the VPT. Conversely, as long as we avoid comparing videos not in parent-child relationship, VPT reconstruction is not negatively affected. Therefore, the `Parent-Child Link Loss` measures how reliable is the proposed method.

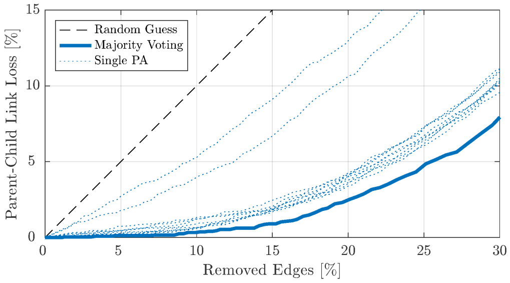


Fig. 2. Parent-child link loss using different coefficients separately or their average. Results are averaged on different datasets. Majority voting (thick blue) provides better performance than any single PA (dotted blue). All solutions strongly deviates from random guess (dashed black).

To evaluate the final effect on VPT reconstruction, we make use of the standard graph-matching metrics used in other phylogeny works [6], [11], namely: `Root`, which is the percentage of correctly reconstructed tree roots (i.e., the originating node); `Edges`, which is the percentage of correctly reconstructed directional edges; `Leaves`, which is the percentage of correctly reconstructed leaves, i.e., the furthest nodes in a tree; `Ancestry`, which measures the percentage of correctly identified ancestral relationships among videos. All these metrics assume values in $[0, 1]$, where 0 is the worst result and 1 means perfect reconstruction.

Results. Our first experiments aims at validating the increased robustness given by voting on processing ages computed from different DCT coefficients and motion vectors, rather than simply using single PAs. To this purpose, we computed processing ages for all videos in $\mathcal{D}_{\text{MPEG2}}$, $\mathcal{D}_{\text{MPEG4}}$, $\mathcal{D}_{\text{H264}}$, $\mathcal{D}_{\text{cod}}$, $\mathcal{D}_{\text{luma}}$, $\mathcal{D}_{\text{geom}}$, using either one of the first nine DCT coefficients read in zig-zag mode separately, motion vectors statistics, or voting among all of them. Fig. 2 shows the average `Parent-Child Link Loss` obtained for each coefficient separately (dotted blue lines) and using voting (thick blue line) while increasing the percentage of removed edges (i.e., by increasing and decreasing the processing age confidence threshold γ). It is possible to notice that each curve based on processing age (blue lines) strongly deviates from the one obtained by removing edges randomly (dashed black line). Moreover, voting always grants a smaller `Parent-Child Link Loss` for each given percentage of removed edges, thus making it the best choice for our algorithm. As a matter of fact, the plot shows that it is possible to avoid almost 15% video pair dissimilarity computations (i.e., save 15% of computational time), with almost lossless results. By reducing the computational complexity by 25%, the link loss is about 5% only, still making the algorithm pretty accurate.

In order to provide a better insight on processing age reliability when different transformations are involved, Fig. 3 reports `Parent-Child Link Loss` results for each tested

[1] Available at https://media.xiph.org/video/derf/

ISBN 978-0-9928626-7-1 © EURASIP 2017

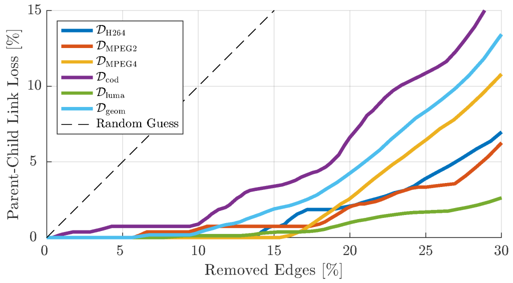


Fig. 3. Parent-child link loss for different percentages of removed edges. Each curve is obtained for a different dataset.

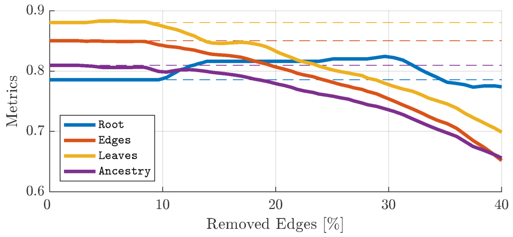


Fig. 4. VPT metrics for different percentages of removed edges considering dataset $\mathcal{D}_{\mathrm{tree}}$. Baseline is dashed.

dataset separately. This test highlights that luminance based transformations are strongly characterized by processing age. Conversely, randomly switching codecs or applying geometric transformations make processing age less reliable, due to changes of coding block sizes.

Finally, we tested the effect of processing age on VPT reconstruction. To this purpose we applied the proposed VPT reconstruction algorithm obtained by integrating processing age results within the pipeline proposed in [11] (not considering the temporal constraint of [11]). Results in terms of `Root`, `Edges`, `Leaves` and `Ancestry` while increasing the percentage of removed edges (i.e., changing γ) are reported in Fig. 4. In this case, avoiding edge removal (i.e., the leftmost point of each curves) coincides with using the baseline [11]. It is therefore interesting to notice that, using processing ages, it is possible to decrease the computational complexity of the baseline solution of about 10% with no significant VPT reconstruction accuracy loss. Even more interesting, the use of processing age seems to "denoise" dissimilarity matrix making the root more easily identifiable during VPT reconstruction. For this reason, `Root` metric assumes values even higher than those obtained by the baseline [11], even considering that computational complexity is decreased by more than 30%.

V. Conclusions

In this paper we proposed a processing age metric that enables reducing VPT reconstruction computational complexity, still granting accurate results. The idea is that it is possible to approximately correlate deviations in DCT and motion vector statistics and the amount of processing operations applied to videos. Given two near-duplicate videos, this enables to understand which one may have generated the other one, thus providing a rough directionality indication that is useful to avoid meaningless dissimilarity computations.

Even though a thorough theoretical validation of the proposed idea is left for future work, results obtained on a dataset of 2.800 video sequences preliminary validate the methodology, especially when some processing operations are considered. Therefore we consider the proposed algorithm as a possible solution towards VPT reconstruction computational complexity reduction.

References

[1] L. Kennedy and S.-F. Chang, "Internet image archaeology," in *Proc. of ACM MM 2008)*, 2008, pp. 349–358.

[2] A. De Rosa, F. Uccheddu, A. Costanzo, A. Piva, and M. Barni, "Exploring image dependencies: A new challenge in image forensics," in *SPIE Conference on Media Forensics and Security*, 2010.

[3] J. R. Kender, M. L. Hill, A. Natsev, J. R. Smith, and L. Xie, "Video genetics: a case study from YouTube," in *Proc. of ACM MM 2010*, Oct. 2010, pp. 1253–1258.

[4] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, pp. 22, 2013.

[5] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics," *APSIPA Transactions on Signal and Information Processing*, vol. 1, 2012.

[6] Z. Dias, A. Rocha, and S. Goldenstein, "Video Phylogeny: Recovering near-duplicate video relationships," in *IEEE WIFS*, Nov. 2011, pp. 1–6.

[7] S. Lameri, P. Bestagini, A. Melloni, S. Milani, A. Rocha, M. Tagliasacchi, and S. Tubaro, "Who is my parent? Reconstructing video sequences from partially matching shots," in *Proc. of IEEE ICIP 2014*, Oct 2014, pp. 5342–5346.

[8] S. Lameri, P. Bestagini, and S. Tubaro, "Video Alignment for Phylogenetic Analysis," in *Proc. of EUSIPCO 2016*, Aug 2016, pp. 2255–2259.

[9] Z. Dias, A. Rocha, and S. Goldenstein, "First steps toward image phylogeny," in *Proc. of IEEE WIFS 2010*, Dec. 2010, pp. 1–6.

[10] Zanoni Dias, Siome Goldenstein, and Anderson Rocha, "Large-Scale Image Phylogeny: Tracing Image Ancestral Relationships," *IEEE MultiMedia*, vol. 20, no. 3, pp. 58–70, jul 2013.

[11] F. O. Costa, S. Lameri, P. Bestagini, Z. Dias, A. Rocha, M. Tagliasacchi, and S. Tubaro, "Phylogeny reconstruction for misaligned and compressed video sequences," in *IEEE ICIP*, Sept 2015, pp. 301–305.

[12] Z. Dias, A. Rocha, and S. Goldenstein, "Image Phylogeny by Minimal Spanning Trees," *IEEE Trans. Inf. Forensics Security*, vol. 7, pp. 774–788, 2012.

[13] Z. Dias, S. Goldenstein, and A. Rocha, "Exploring heuristic and optimum branching algorithms for image phylogeny," *J. of Visual Comm. and Image Representation*, vol. 24, pp. 1124–1134, 2013.

[14] A. Melloni, P. Bestagini, S. Milani, M. Tagliasacchi, A. Rocha, and S. Tubaro, "Image phylogeny through dissimilarity metrics fusion," in *Proc. of EUVIP 2014*, Dec 2014, pp. 1–6.

[15] S. Milani, M. Fontana, P. Bestagini, and S. Tubaro, "Phylogenetic analysis of near-duplicate images using processing age metrics," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 2054–2058.

[16] W. Chen and W. Q. Shi, *Detection of Double MPEG Compression Based on First Digit Statistics*, pp. 16–30, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[17] D. Vazquez-Padin, M. Fontani, T. Bianchi, P. Comesana, A. Piva, and M. Barni, "Detection of video double encoding with GOP size estimation," in *Proc. of IEEE WIFS 2012*, Dec 2012, pp. 151–156.

[18] S. Milani, P. Bestagini, M. Tagliasacchi, and S. Tubaro, "Multiple compression detection for video sequences," in *Proc. of MMSP 2012*, Sept. 2012, pp. 112–117.

[19] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, pp. 1661–1666, 2000.

[20] G. Calvagno, C. Ghirardi, G.A. Mian, and R. Rinaldo, "Modeling of subband data for buffer control," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 7, pp. 402–408, 1997.

[21] S. Milani, L. Celetto, and G.A. Mian, "An Accurate Low-Complexity Rate Control Algorithm Based on (ρ, E_q)-Domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 2, pp. 257–262, Feb. 2008.

[22] Y. Altunbasak and N. Kamaci, "An analysis of the DCT coefficient distribution with the H.264 video coder," in *Proc. of ICASSP 2004*, May 2004, vol. 3, pp. iii–177–80 vol.3.