Image and Video Phylogeny Reconstruction

Filipe de Oliveira Costa, Zanoni Dias and Anderson Rocha
Institute of Computing, University of Campinas, Brazil
Email: {fcosta, zanoni, anderson.rocha}@ic.unicamp.br

***Abstract*—In this thesis, we designed and developed approaches for solving Image and Video Phylogeny problem, in which we aim at finding the ancestral relationship between the near duplicates and the original source of images and videos. For images, we proposed approaches to deal with the phylogeny problem considering two main points: (i) the forest reconstruction, an important task when we consider scenarios in which there is a set of semantically similar images, but generated by different sources or at different times; and (ii) new measures for dissimilarity calculation between near-duplicates, given that it directly impacts the quality of the phylogeny reconstruction. For video phylogeny, we developed a new approach to temporally align two video sequences before calculating the dissimilarity between them as, in real-world conditions, a pair of videos can be temporally misaligned, one video can have some frames added or removed and compression can also take place, for instance.**

I. Introduction

Multimedia documents (e.g., images and videos) are powerful communication tools living up to the classical adage comparing them to a thousand words when conveying any information. This communication power was multiplied significantly with the advent of social networks. Within this new reality, multimedia documents are published, shared, modified, and often republished effortlessly and, depending on their contents, they can easily *go viral*, being republished by many other users in different channels trough the Web. This scenario easily leads to copyright infringement, sharing of illegal or abusive contents (e.g., child pornography) and, in some cases, negatively affect or impersonate the public image of people or corporations.

When small changes are applied during the redistribution, usually without interfering with their semantic meaning, they are called *near-duplicate* objects [1]. A far more challenging task, however, has been overlooked thus far in which we also want to find the ancestral relationship between the near duplicates and the original source (*root* or *patient zero*), estimating the transformations (e.g., geometric transformations, cropping, color changing, compression, etc.) that originally created the near duplicates in a set and reconstruct the order of them. This new research field is called *Multimedia Phylogeny* [2] and has several applications. For instance, the relationship structure of a set of documents provides information of suspects' behavior, and points out the directions of content distribution; traitor tracing can be performed without the requirement of source control techniques such as watermarking or fingerprinting; we could use phylogeny to point out group's reuse of illegal material online.

The multimedia phylogeny problem can be separated into two basic steps: the *dissimilarity calculation* between the duplicates, in which a dissimilarity function is used to compare each pair of images, returning small values for similar images and large values for more distinct images, and the *phylogeny reconstruction*, considering one dissimilarity matrix that represents the dissimilarity between each pair of documents [2]. The definition of reliable dissimilarity measure is paramount for document phylogeny research, given that the dissimilarity calculation directly affects the result of the final phylogeny reconstruction.

In this thesis, we designed and developed some solutions that aim at solving the multimedia phylogeny problem for images and videos. More specifically, for images, we present solutions that aim at reconstructing phylogeny trees and forests for representing the relationship between the duplicates. Moreover, we also develop new dissimilarity measures based on gradient and mutual information that significantly improve the quality of the phylogeny reconstruction process. Finally, for video phylogeny, we introduce new methods considering the temporal misalignment of the videos and different parameters of coding used for creating the near duplicates.

This paper is structured as follow. Section II presents a background for Multimedia Phylogeny. Section III describes the proposed solutions for image phylogeny forest reconstruction. In Section IV, we present new dissimilarity measures for image phylogeny. Proposed solutions for video phylogeny are described on Section V. Finally, we present the conclusions and the obtained publications on Section VI.

II. Background

The main hypothesis of Multimedia Phylogeny assumes that the transformations (e.g., color changing, compression, geometric transformation, cropping, etc.) used for creating near duplicates of a document (image or video) often leave irreversible artifacts in the data that allow us to point out the direction of the transformations that the documents have undergone and, ultimately, create a phylogeny map coding the evolutionary structure of such a set of documents.

Dias et al. [2], [3] formally defined the problem of Image Phylogeny following two steps: the calculation of the dissimilarity between each pair of near-duplicate images and the reconstruction of the phylogeny tree. Considering $\mathcal{T}$ a family of image transformations, T a transformation such that

Ph.D. Thesis. The authors thank the financial support of Unicamp, Politecnico di Milano (Italy), São Paulo Research Foundation – FAPESP, (Grant #2013/05815-2), CAPES PDSE program (Grant #99999.003836/2014-02) and CAPES DeepEyes Project and Microsoft Research.

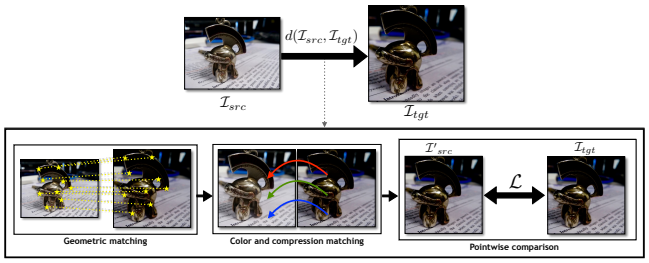


Fig. 1. Dissimilarity calculation process. The mapping of image $\mathcal{I}_{src}$ onto $\mathcal{I}_{tgt}$'s domain involves a three-step process: geometric, color and compression matching. Afterwards, it is possible to directly compare the images using any point-wise comparison algorithm.

$T_{\overrightarrow{\beta}} \in \mathcal{T}$ parameterized by $\overrightarrow{\beta}$, and two near-duplicate images $\mathcal{I}_{src}$ (source) and $\mathcal{I}_{tgt}$ (target), the dissimilarity function $d(.,.)$ between them is defined as the lowest value of $d(\mathcal{I}_{src}, \mathcal{I}_{tgt})$, such that

$$d(\mathcal{I}_{src}, \mathcal{I}_{tgt}) = \min_{T_{\overrightarrow{\beta}} \in \mathcal{T}} |\mathcal{I}_{tgt} - T_{\overrightarrow{\beta}}(\mathcal{I}_{src})|_{\text{ point-wise comparison } \mathcal{L}}. \quad (1)$$

Equation 1 calculates the dissimilarity between the best transformation mapping $\mathcal{I}_{src}$ onto $\mathcal{I}_{tgt}$ parameterized by $\overrightarrow{\beta}$, according to the family of transformations $\mathcal{T}$. The comparison between the images can be performed by any point-wise comparison method $\mathcal{L}$.

Given a set of near duplicates, the estimation of the transformation T, parameterized by $\overrightarrow{\beta}$ used to map an image $\mathcal{I}_{src}$ onto an image $\mathcal{I}_{tgt}$'s domain follows a three-step method generating $\mathcal{I}'_{src} = T_{\overrightarrow{\beta}}(\mathcal{I}_{src})$:

1) **Geometric matching:** also known as Image Registration. The image registration is calculated by finding interest points in each pair of images, using SURF (Speeded-Up Robust Features) [4], which will be used to estimate warping and cropping parameters robustly using RANSAC [5];
2) **Color matching:** it is performed for adjusting the color of the source image $\mathcal{I}_{src}$ to the target image $\mathcal{I}_{tgt}$ by normalizing each channel of $\mathcal{I}_{src}$ by the mean and standard deviation of the respective channel in $\mathcal{I}_{tgt}$ [6];
3) **Compression matching:** the image $\mathcal{I}_{src}$ is compressed with $\mathcal{I}_{tgt}$'s JPEG compression parameters.

Then, a comparison between the estimated $\mathcal{I}'_{src} = T_{\overrightarrow{\beta}}(\mathcal{I}_{src})$ and $\mathcal{I}_{tgt}$ is performed point-wise. The authors estimated it using the Mean Squared Error (MSE) technique. Figure 1 depicts this dissimilarity calculation process.

Note that the dissimilarity is not a metric. Once the transformation T used for mapping $\mathcal{I}_{src}$ to $\mathcal{I}_{tgt}$ can insert irreversible artifacts in $\mathcal{I}_{src}$ for mapping it to $\mathcal{I}_{tgt}$ (e.g., remove pixels after spatial cropping, JPEG compression, etc.) the inverse transformation T^{-1} will not recover the lost information when performing the mapping in the opposite way. Thus, $d(\mathcal{I}_{src}, \mathcal{I}_{tgt}) \neq d(\mathcal{I}_{tgt}, \mathcal{I}_{src})$. Low values of $d(\mathcal{I}_{src}, \mathcal{I}_{tgt})$ denote a good transformation and the resultant image $\mathcal{I}'_{src}$ is a close approximation of $\mathcal{I}_{tgt}$, which is a strong evidence that $\mathcal{I}_{src}$ is the father of $\mathcal{I}_{tgt}$ in the phylogeny tree.

After calculating the dissimilarity for each pair of images, we have a dissimilarity matrix $M_{n \times n}$, where n is the number of near duplicates and each position of the matrix represents the dissimilarity between one pair of images.

For reconstructing the *Image Phylogeny Tree* (IPT) associated with the dissimilarity matrix, the authors first proposed the Oriented Kruskal algorithm (OK), an extension of the classic Kruskal Minimal Spanning Tree algorithm [7], adapted for oriented graphs.

Since the first work in image phylogeny [3], several branches to this research field have been developed for image phylogeny [8]–[10], videos (Video Phylogeny) [11] and audio phylogeny [12]. In addition to our pioneer work in this field, there are other important works in the literature following a similar trend in the literature, aiming at finding the structure of the evolution of images on the Internet [13]–[15].

III. Image Phylogeny Forest Reconstruction

In some cases in multimedia phylogeny, instead of one tree, we may find m trees representing the ancestry relationship in a set of near-duplicate images. This happens when we have multiple images with the same semantic content, but that are not directly related to each other (e.g., images from the same scene taken with different cameras or in slightly different positions). In this case, each tree in the forest represents the structure of transformations and the evolution of one subset of near-duplicate images, while the forest comprises distinct subsets of near-duplicate images which are semantically similar.

To enable automatic IPF reconstructions, Dias et al [16] presented a modified version of the Oriented Kruskal algorithm originally used to reconstruct IPTs. To create the new approach, named *Automatic Oriented Kruskal* (AOK), three parameters are required as input: the number of semantically similar images n, an $n \times n$ dissimilarity matrix M built upon these images and a parameter γ_{AOK}, calculated beforehand and defined as the number of standard deviations used to limit the number of edges to be included in the forest.

The AOK algorithm keeps track of the variance of processed edges and only adds a new one to the forest if the weight of the current edge is lower than γ_{AOK} times the standard deviation of the processed edges up to that point. This parameter γ_{AOK} is related to a threshold point τ_{AOK} that selects only edges that belong to valid trees. To define its value, a study about the behavior of the dissimilarity values of valid trees and forests was performed. It was observed that a Log-Normal distribution can reasonably describe the data regardless of the number of trees in the forest and the type of image capture (single/multiple cameras). The threshold $\tau_{\text{AOK}} = \mu_{\text{AOK}} + \gamma_{\text{AOK}} \times \sigma_{\text{AOK}}$ was used, where μ_{AOK} represents the average and σ_{AOK}, the standard deviation of the weight of the edges already selected. After testing for different values, it was defined that $\gamma_{\text{AOK}} = 2$.

Following this idea, we apply a similar process to the Optimum Branching (OB) algorithm [17], proposing the Automatic Optimum Branching algorithm (AOB), with the necessary modifications to deal with its particularities. In a few words, we create an optimum tree considering the OB algorithm and select the edges of the forest as is done for the AOK algorithm, but considering the edges of the optimum tree instead all the edges of the graph.

After some experiments, we noticed that the IPF reconstruction could be further improved by also executing the OB algorithm on each tree belonging to this forest recursively. The AOB algorithm considers all edges to construct the minimum branching. Once we remove some edges to build the forest, we create several partitions that are independent of each other. If these partitions are analyzed separately, the OB algorithm will choose edge connections that are optimal considering only the edges that belong to the current partition. This re-execution characterizes our Extended Automatic Optimum Branching (E-AOB) algorithm.

Considering a dissimilarity matrix M, AOK algorithm follows a greedy heuristic, while AOB and E-AOB searches for the best global solution for the phylogeny reconstruction. However, all these algorithms assume a perfect dissimilarity calculation, which is not always true, since the dissimilarity calculation involves the estimation of the transformations that map a source image onto a target image, which is not exact. Thus, in some cases, a greedy heuristic may present better results than a global solution. Aiming at exploring these different properties and their complementarity, we propose a combination among the results given by each approach, in such a way that errors introduced by one method can be fixed by other method(s).

First, we apply different amounts of perturbations through noise addition to the dissimilarity matrix M relating a set of images, generating 100 different variations of M and using them to reconstruct the IPFs. Once the number of roots and edges for all forests have been found, we calculate the final number of roots r by choosing the median of the number of votes received by all methods in each of the 100 executions.[1] Then, to decide which nodes are the roots, we select the r nodes having the highest number of votes. In case there is a tie among the votes, we randomly choose one or some of them, depending on how many roots are yet to be decided. For the edges selection, we sum up the number of times each edge connecting two nodes in each forest appears, constructing a matrix of votes for edges. Once the roots are chosen, we fix these roots and give the matrix of votes for edges as input to a maximum branching algorithm, resulting in the sought combined forest.

A. Experiments and results

For evaluating the proposed methods, we proposed a dataset that comprises images randomly selected from a set of 20 different scenes, 10 different cameras, 10 images per camera, 10 different tree topologies, 10 random variations of parameters for creating the near duplicate images. We consider images taken with a single camera (OC) and with multiple cameras (MC) having similar scene semantics (the main content of the image is the same, but with small variations in the camera parameters, such as viewpoint, zoom, etc.). For generating the near duplicates, we considered the following transformations: resampling, rotation, scaling, off-diagonal correction, cropping, brightness, contrast and gamma correction and re-compression.

For evaluating the reconstructed IPFs, we consider the scenarios where the ground truth is available. We use the following evaluation metrics: ***roots***, which measures if the reconstructed forest contains exactly the same roots as the ground-truth forest; ***edges***, which measure how well the algorithm finds the kinship relationships between two near duplicates; ***leaves***, which compares the leaves (most modified images in a given branch of the tree) found by an algorithm with the original ones in the ground-truth; and ***ancestry***, which measure how well the algorithm finds the kinship relationships along the whole tree.

Table I shows the results of the phylogeny reconstruction, considering 10 trees for each forest (other results can be found in the thesis). Using the AOK method as baseline, these results show that AOB is only able to improve the results of AOK regarding the metrics edges and leaves. On the other hand, the results confirm the E-AOB method has better performance than the AOK method. Furthermore, Table I shows that we can improve the results when combining different approaches for the phylogeny reconstruction.

IV. New dissimilarity measures for image phylogeny reconstruction

Given that the dissimilarity calculation directly impacts the phylogeny reconstruction, we propose new approaches to the standard formulation of the dissimilarity measure employed in image phylogeny, aiming at improving the reconstruction of the tree structure that represents the generational relationships between semantically similar images.

a) Gradient Comparison: As contrast enhancement and color transformations are often used when creating near duplicates, directly affecting the gradients of the image, this becomes an important information to add to the dissimilarity calculation. Here we filter an image by using a convolution with a 3×3 *Sobel* kernel [18] for gradient estimation, while the image comparison metric $\mathcal{L}$ stays the same (i.e., Minimum Square Error) [2]. We considered each color channel separately and the final phylogeny is the average of MSE for each channel.

b) Mutual Information Comparison: in Information Theory, mutual information (MI) is a measure of statistical dependency of two random variables, which represents the

[1]There are several alternative strategies to calculate r, such as the average and the most frequent values. In our experiments, the median presented better results during training, but there is still room for further exploration.

[2]We performed exploratory experiments with different sizes o Kernel and also with Histogram of Oriented Gradient [19], but these approaches reported lower effectiveness comparing to Sobel's filtering.

TABLE I
COMPARISON AMONG AOK [16], THE VARIATIONS OF AOB ALGORITHM AND THE FUSION APPROACH

	AOK				AOB				E-AOB				Fusion approach			
	Roots	**Edges**	**Leaves**	**Ancestry**	**Roots**	**Edges**	**Leaves**	**Ancestry**	**Roots**	**Edges**	**Leaves**	**Ancestry**	**Roots**	**Edges**	**Leaves**	**Ancestry**
OC	0.755	0.883	0.854	0.784	0.682	0.898	0.877	0.793	0.802	0.908	0.890	0.823	0.796	0.909	0.890	0.820
MC	0.885	0.888	0.862	0.859	0.768	0.896	0.875	0.852	0.909	0.908	0.889	0.886	0.917	0.910	0.891	0.890

amount of information that one random variable contains about the other [20]. Calculating the mutual information of two duplicates instead MSE let us avoiding effects caused by slight misalignment during the mapping.

c) Gradient Estimation and Mutual Information Combined: first, we calculate the gradient of the images $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$ as we described before. Afterwards, we compare the gradient of both images with mutual information, instead of using the image comparison metric $\mathcal{L}$ based on the standard Minimum Square Error. With this approach, we aim at better capturing the information about variation in certain directions of the image (gradient information), as well as at seeking to avoid effects caused by slight misalignments during the mapping (mutual information estimation). This method also takes into consideration the amount of texture information preserved between two near duplicates for calculating the dissimilarity. Unfortunately, the combined method slightly increases the computational cost of the dissimilarity calculation, given that we need to estimate the mutual information after the gradient calculation for each color channel. However, this method yields better reconstruction results.

A. Experiments and results

Table II presents the results for the different approaches considered herein for calculating the dissimilarities for OC and MC scenarios. In all cases, the geometrical mapping of one source image onto a target image is performed following the procedure discussed in the beginning of Section II. The phylogeny reconstruction part uses the E-AOB algorithm for all methods. For this experiment, we considered the dataset described on Section III-A. Here, we considered phylogeny forests with 10 trees (additional results are described in the thesis).

The baseline dissimilarity calculation considered is the MSE, the state of the art, which compares two images point-wise using the pixel intensities. The proposed modifications are:

1) Gradient estimation (GRAD), which still compares the images point-wise but using image gradients instead of pixel intensities;
2) Mutual information (MINF), which replaces the point-wise comparison using pixel intensities with the mutual information calculation of pixel intensities;
3) Gradient estimation plus comparison with mutual information (GRMI), incorporating the calculus of dissimilarities using mutual information of image gradients.

TABLE II
COMPARISON BETWEEN DIFFERENT DISSIMILARITY MEASURES, CONSIDERING FORESTS WITH 10 TREES AND E-AOB PHYLOGENY RECONSTRUCTION ALGORITHM.

	One Camera				Multiple Cameras			
	Roots	**Edges**	**Leaves**	**Ancestry**	**Roots**	**Edges**	**Leaves**	**Ancestry**
MSE	0.802	0.908	0.890	0.823	0.909	0.908	0.889	0.886
GRAD	0.630	0.805	0.803	0.638	0.645	0.808	0.806	0.653
MINF	0.664	0.940	0.927	0.761	0.890	0.949	0.937	0.889
GRMI	0.906	0.962	0.954	0.922	0.958	0.965	0.956	0.949

The GRAD approach only captures directional variations and small misalignment when comparing two gradient images affect the results more than when comparing the images through pixel intensities. With MINF, small misalignment are not as important as for the GRAD case. The results improve when combining the gradient calculation with mutual information (GRMI). The first reason is that, by not comparing the pixel intensities directly, the color information artifacts are not as strong. Second, the comparison is not done in a point-wise fashion but rather, in a probability distribution-like form, better capturing the different variations of the gradient images as well as accounting for possible small misalignment.

V. VIDEO PHYLOGENY

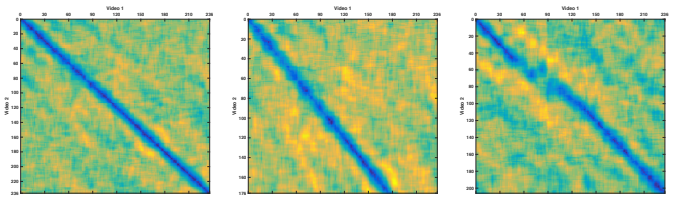
Considering videos, Dias et al. [11] proposed an initial approach to deal with the video phylogeny tree reconstruction problem. However, only temporally coherent videos (i.e., temporally aligned videos with the same number of frames) were considered thus far. Furthermore, the authors considered only videos compressed with the same standard and parameters without explicitly taking into account any other compression scheme in their reconstruction pipeline.

To deal with these more challenging setups, we propose a modification to the pipeline used in [11] for solving the problem of VPT reconstruction. To compute dissimilarity, it is necessary to achieve temporal synchronization and estimate $T_{\overrightarrow{\beta}}$.

We proposed the using of two different methods for temporal alignment for videos. For the first method, we resort to a 1-dimensional description of a video over time, obtained through computing the difference between the average of luminance values of adjacent frames of a video and estimating the phase-correlation between them. The index of the highest value of the phase-correlation indicates the number of misaligned frames. With this, temporally align the videos becomes straightforward. Unfortunately, this method works

only in cases we have temporal clipping only in the beginning or the end of the video sequence.

The second alignment method addresses this problem as follows: first, we extract 64 Discrete Cosine Transform (DCT) coefficients and compute a binary hash for each group of frames of the video sequence, binarizing the DCT coefficients according to their median value. By calculating the Hamming distance for every pair of clusters of two sequences, we obtain a distance matrix in which it is stored the Hamming distances of all pairs of clusters, respecting their temporal order. Figure 2 shows examples of such distance matrices. The axes represent two video sequences, and darker colors mean lower values for Hamming distance between them. Note that low values in the matrices are found to be aligned along a segment. The higher the number of matching frames, the longer the "dark blue" segment.



(a) (b) (c)

Fig. 2. Example of distance matrices, in which (a) both video sequences do not have temporal clipping and (b) Sequence 2 has temporal clipping in the end; and (c) Sequence 2 has temporal clipping in the middle. The axes represents the index of the groups of 64 frames for each video.

Finally, given a distance matrix, we extracted the matching blocks of near-duplicates and, for each matching block and we execute, for each correspondent block of frames, the temporal alignment based on phase correlation.

After synchronize the videos, we estimate $T_{\overrightarrow{\beta}}$ as is done for images, but considering only the selected frames after alignment. The geometric and color matching steps are performed for each correspondent frame separatelly. We proposed a coding/compression matching for videos, in which a video $\mathcal{V}_{\text{src}}$ is encoded using the same coding scheme and parameters used to encode the video $\mathcal{V}_{\text{tgt}}$. This transformation is denoted as T_{cod}. We consider that the bitstream of $\mathcal{V}_{\text{tgt}}$ is available, thus the used codec, quantization parameter (QP), group of pictures (GOP) size (i.e., the distance between consecutive intra-coded frames) can be extracted from header information. Nonetheless, if this is not possible (e.g., the bitstream is not available, the video is distributed in the decoded domain, etc.) these parameters can be estimated according, for instance, to [21], [22].

After the aforementioned transformations are estimated, they are stacked up to obtain $T_{\overrightarrow{\beta}}$. Dissimilarity is finally computed according to Eq. (1), by averaging the frame-wise MSE computed only on ND frame pairs selected during the video temporal alignment step [3]. Then, a phylogeny tree is reconstructed using the optimum branching algorithm [17].

[3]Due to the computational cost of GRMI dissimilarity for images, we decide to use MSE dissimilarity measure for videos.

A. Experiments and results

The validation of the proposed video phylogeny tree reconstruction algorithm was carried out on a set with 300 phylogeny trees comprising a total number of 3,000 near-duplicate videos. More specifically, we started from eight well known uncompressed sequences at CIF resolution (i.e., 352×288 pixels) of 300 frames each[4]. Then, for creating the duplicates, we considered the following possible transformations: contrast enhancement, brightness adjustment, spatial cropping and spatial resizing in any combination. As video codecs, we selected MPEG-2, MPEG-4 Part 4, and H.264/AVC.

We separated the into three subsets: near duplicates without any temporal clipping ($\mathcal{D}^{\text{no clip}}$), near duplicates with possible temporal clipping but only at the beginning or at the end of the stream ($\mathcal{D}^{\text{no clip}}$) and near duplicates with temporal clipping in the middle of the stream ($\mathcal{D}^{\text{clip middle}}$). We consider 100 trees for each one of five different scenarios and probability of temporal clipping and frame rate changing of 50%.

To estimate the effectiveness of the estimated phylogeny trees, we used the following metrics: *root*, *edges*,*leaves*, *ancestry*. We also consider a new metric *depth*, which indicates the depth of the ground truth root in the reconstructed tree.

Aiming at evaluating the importance of temporal alignment and compression matching, Table III clearly shows the need of explicitly considering coding, temporal clipping and temporal alignment in the reconstruction process. As a matter of fact, when these operations are not taken into account, the reconstruction accuracy drops significantly.

TABLE III
RESULTS OBTAINED WITH AND WITHOUT TEMPORAL ALIGNMENT AND CODING MATCHING.

Alignment	Coding matching	Dataset	*Root*	*Depth*	*Edges*	*Leaves*	*Ancestry*
None	No	$\mathcal{D}^{\text{no clip}}$	0.63	0.50	0.77	0.83	0.67
Phase-correlation	No	$\mathcal{D}^{\text{no clip}}$	0.63	0.50	0.77	0.83	0.67
Phase-correlation	Yes	$\mathcal{D}^{\text{no clip}}$	0.87	0.14	0.84	0.86	0.81
Hash-based	No	$\mathcal{D}^{\text{no clip}}$	0.65	0.44	0.78	0.83	0.70
Hash-based	Yes	$\mathcal{D}^{\text{no clip}}$	0.69	0.40	0.78	0.83	0.71
None	No	$\mathcal{D}^{\text{clip border}}$	0.60	0.72	0.66	0.74	0.56
Phase-correlation	No	$\mathcal{D}^{\text{clip border}}$	0.71	0.35	0.79	0.83	0.72
Phase-correlation	Yes	$\mathcal{D}^{\text{clip border}}$	0.84	0.17	0.86	0.90	0.83
Hash-based	No	$\mathcal{D}^{\text{clip border}}$	0.71	0.33	0.81	0.85	0.74
Hash-based	Yes	$\mathcal{D}^{\text{clip border}}$	0.66	0.40	0.78	0.82	0.72
Phase-correlation	No	$\mathcal{D}^{\text{clip middle}}$	0.62	0.56	0.72	0.76	0.63
Phase-correlation	Yes	$\mathcal{D}^{\text{clip middle}}$	0.72	0.39	0.80	0.83	0.70
Hash-based	No	$\mathcal{D}^{\text{clip middle}}$	0.67	0.45	0.76	0.79	0.69
Hash-based	Yes	$\mathcal{D}^{\text{clip middle}}$	0.66	0.47	0.77	0.81	0.70

Hash-based alignment has lower performance when considering the coding matching, compared to the alignment based on difference of luminance. One justification for these results is the nature of video compression. Considering that

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

some frames can be discarded, the *inter-frame* compression may not be assigned correctly. Updated results shows that using only *intra-frame* protocol for compression matching increases the results.

VI. Conclusions and Future Research Directions

In this thesis, we introduced four techniques for solving the image and video phylogeny reconstruction problems. For image phylogeny, our contributions focus on solving two main problems: the phylogeny forest reconstruction and dissimilarity calculation for each pair of compared images. In video phylogeny, we introduced two approaches to video phylogeny tree reconstruction starting from the analysis of a pool of near-duplicate video sequences, accommodating the cases of time clipped, misaligned and compressed video sequences.

Each one of the proposed methods has its scientific contributions, as well as its own limitations. Therefore, it is important to develop and combine complementary solutions in order to better reconstruct image and video phylogenies. For future research, we suggest the analysis of the behavior of the phylogeny algorithms using other statistical measures, to consider other types of image and video transformations, consider spatio-temporal features for dissimilarity calculation and investigate different phylogeny reconstruction strategies.

Scientific production

Finally, this research has results in the following publications:

- Costa et al. (2014). *Image Phylogeny Forest Reconstruction*. IEEE Transactions on Information Forensics and Security (T-IFS), vol. 9, n.10, pp.1533–1546. (Impact Factor: 4.332)
- Costa et al. (2015). *Phylogeny reconstruction for misaligned and compressed video sequences*. IEEE International Conference on Image Processing (ICIP). pp. 301 – 305.
- Costa et al. (2016). *Hash-Based Frame Selection for Video Phylogeny*. IEEE Workshop on Information Forensics and Security (WIFS), pp. 1–5.
- Costa et al. (2017). *New dissimilarity measures for image phylogeny reconstruction*. Accepted on Springer Pattern Analysis and Applications (PAA). DOI: 10.1007/s10044-017-0616-9 (Impact Factor: 1.352).

References

[1] B. S. Alsulami, M. F. Abulkhair, and F. E. Eassa, "Near duplicate document detection survey," *International Journal of Computer Science and Communications Networks*, vol. 2, no. 2, pp. 147–151, 2012.

[2] Z. Dias, A. Rocha, and S. Goldenstein, "Image phylogeny by minimal spanning trees," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 7, no. 2, pp. 774–788, 2012.

[3] ——, "First steps toward image phylogeny," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2010, pp. 1–6.

[4] H. Bay, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Elsevier Computer Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[5] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *ACM Communications*, vol. 24, no. 6, pp. 381–395, 1981.

[6] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer Graphics Applications*, vol. 21, pp. 34–41, 2001.

[7] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proceedings of the American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.

[8] M. Oikawa, Z. Dias, A. Rocha, and S. Goldenstein, "Manifold learning and spectral clustering for image phylogeny forests," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 5–18, 2016.

[9] A. Oliveira, P. Ferrara, A. D. Rosa, A. Piva, M. Barni, S. Goldenstein, Z. Dias, and A. Rocha, "Multiple parenting phylogeny relationships in digital images," *IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 11, no. 2, pp. 328–343, 2016.

[10] A. Melloni, P. Bestagini, S. Milani, M. Tagliasacchi, A. Rocha, and S. Tubaro, "Image phylogeny through dissimilarity metrics fusion," in *IEEE European Workshop on Visual Information Processing (EUVIP)*, 2014, pp. 1–6.

[11] Z. Dias, A. Rocha, and S. Goldenstein, "Video phylogeny: Recovering near-duplicate video relationships," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2011, pp. 1–6.

[12] M. Nucci, M. Tagliasacchi, and S. Tubaro, "A phylogenetic analysis of near-duplicate audio tracks," in *IEEE International Workshop on Multimedia Signal Processing*, 2013, pp. 99–104.

[13] L. Kennedy and S.-F. Chang, "Internet image archaeology: Automatically tracing the manipulation history of photographs on the web," in *ACM International Conference of Multimedia*, 2008, pp. 349–358.

[14] A. D. Rosa, F. Uccheddu, A. Costanzo, A. Piva, and M. Barni, "Exploring image dependencies: a new challenge in image forensics," *SPIE Media Forensics and Security*, vol. 7541, no. 2, pp. 1 – 12, 2010.

[15] J. R. Kender, M. L. Hill, A. P. Natsev, J. R. Smith, and L. Xie, "Video genetics: a case study from youtube," in *International Conference on Multimedia*, 2010, pp. 1253–1258.

[16] Z. Dias, S. Goldenstein, and A. Rocha, "Toward image phylogeny forests: Automatically recovering semantically similar image relationships," *Elsevier Forensic Science International (FSI)*, vol. 231, pp. 178–189, 2013.

[17] J. Edmonds, "Optimum branchings," *Journal of Research of National Institute of Standards and Technology*, vol. 71B, pp. 48–50, 1967.

[18] I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," *a talk at the Stanford Artificial Project in*, pp. 271–272, 1968.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893.

[20] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[21] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Estimating QP and motion vectors in H. 264/AVC video from decoded pixels," in *ACM Workshop on Multimedia in Forensics, Security and Intelligence*, 2010, pp. 89–92.

[22] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Codec and GOP identification in double compressed videos," *IEEE Transactions on Image Processing*, vol. 25, pp. 2298–2310, 2016.