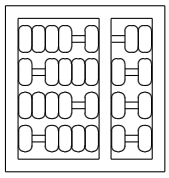

INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

New Dissimilarity Measures for Image Phylogeny Reconstruction

Filipe de O. Costa *Alberto Oliveira*
Pasquale Ferrara *Siome Goldenstein*
Zanoni Dias *Anderson Rocha*

Technical Report - IC-01-45 - Relatório Técnico

December - 2001 - Dezembro

New Dissimilarity Measures for Image Phylogeny Reconstruction

Filipe de O. Costa* Alberto Oliveira* Pasquale Ferrara†
Siome Goldenstein* Zanoni Dias* Anderson Rocha*

Abstract

Image Phylogeny is the problem of reconstructing the structure that represents the history of generation of semantically similar images (e.g., near-duplicate images). Typical Image Phylogeny approaches break the problem into two steps: the estimation of the dissimilarity between each pair of images, and the actual reconstruction of the phylogeny structure. Given that the dissimilarity calculation directly impacts the phylogeny reconstruction, in this paper, we propose new approaches to the standard dissimilarity calculation formulation in image phylogeny, aiming at improving the accuracy when estimating the historical relationships between near-duplicates. The new formulations explore a different family of color adjustment, local gradient, and mutual information processes. The obtained results with the new formulations remarkably outperform the existing counterparts in the literature allowing a much better analysis of the parenthood relationships in a set of images better enabling the deployment of phylogeny solutions to tackle traitor tracing, copyright enforcement, and digital forensics problems.

1 Introduction

Undoubtedly, images are powerful communications 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, images are published, shared, modified, and often republished effortlessly. Frequently, reposting and sharing will happen after myriad small modifications (near duplicates), such as cropping, resampling, affine-warping, and color adjustments. Sometimes content sharing might be illegal, however, such as in cases of copyright infringement and public defamation. Occasionally, simply possessing the content (for example, images depicting child pornography) already constitutes a crime. Considering the aforementioned scenarios, it is often important to develop means to track and monitor how images are shared and evolve on the internet over time.

*Institute of Computing, University of Campinas, Brazil. Thanks to FAPESP for the financial support (Grant #2013/05815-2 and #2013/21251-1)

†University of Firenze, Italy

In this vein, *Image Phylogeny* has been developed recently [1, 2, 3, 4] in an attempt to find the relationship structure among near-duplicate images. According to [5], an image is a near duplicate of another if it shares similar content differing up to some editing transformations. In other words, the two images contain a kinship relationship.

For the case of image phylogeny, we model the kinship relationships as a tree, whereby the root is the patient zero (the original image), the edges represent "father-child" relationships, and where the leaves of the tree represent "terminal" images that have more modifications than their ancestors. In some cases, the near-duplicate set does not come from a single original document, but rather from images with the same semantic content generated either from different sources (cameras) or from the same source in distinct moments in time. (Two pictures virtually at the same viewpoint taken by different cameras or by the same camera in different moments of time are considered semantically similar images rather than near duplicates). In these cases, the set of near duplicates can be represented by a forest correlating semantically similar images [6, 7]. Figure 1 depicts an example of the image phylogeny problem.

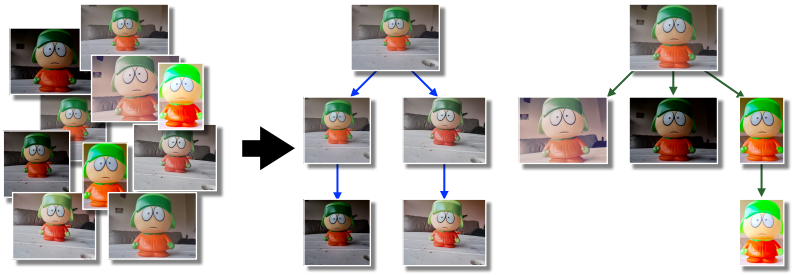

Figure 1: Image phylogeny problem. Given a set of semantically-similar images, our objective is to reconstruct a structure that represents the historical relationships among the images. In this example, we have a forest with two trees, which means that the group of semantically similar images has two original images (with similar content) and each one spurs its own near duplicates (descendants).

Once we trace the past history of the near duplicates, image phylogeny can be useful for aiding (allied with additional side information) to discover, for instance, who was the first user that did publish an image containing illegal or abusive content (e.g., fake and defamatory image of celebrities or politicians, child pornography, etc.), which were redistributed after being modified by different users.

In the literature, there are other approaches following a similar trend, aiming at finding the structure of the evolution of images on the Internet [8, 9, 10, 11, 12], extensions to the original image phylogeny algorithm for reconstructing the tree of evolution of a set of near-duplicate videos (Video Phylogeny) [13, 14, 15], as well for dealing with more than one phylogeny tree [2, 6, 7]. In addition, the phylogeny of audio clips were also investigated by [16].

Dias et al. [1, 2] 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 $T_{\overrightarrow{\beta}} \in \mathcal{T}$ parameterized by β, 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 $\mathcal{L}$ (e.g., minimum squared error).

Since the first image phylogeny work [1], several research branches have been developed such as video [13], audio phylogeny [16], as well as the study of multiple parent-child relationships (images obtained through the modifications of more than one image) [17]. In addition, improvements on the image phylogeny original framework have been proposed for dealing with large-scale scenarios [18] and semantically similar images [6]. Recently, new improvements on the construction of the parent-child relationships have also been studied and proposed such as optimum-branching solutions [3].

As discussed in those works, phylogeny solutions have several important applications in security (for finding the modification's graph of a set of documents hinting at information about suspects' behavior and the directions of content distribution), forensics (enabling the forensic analyst to focus on original versions of documents instead of their descendants), copyright enforcement (powering new passive traitor tracing techniques), and news tracking services (feeding news tracking services with key elements for mining opinion forming processes along time), among others.

Although the field has been developing significantly over the past years, thus far researchers mainly focused on proposing different phylogeny reconstruction approaches [1, 13, 2, 3, 6, 7, 18] often using a standard methodology for dissimilarity calculation as originally proposed by [2]. This dissimilarity calculation involves the estimation of the transformations that map a source image onto a target image, followed by their comparison in a point-wise fashion. As the transformations estimation is not exact, the point-wise comparison method $\mathcal{L}$ is strongly affected by artifacts generated in these processes. Given that the dissimilarity calculation directly affects the result of the final phylogeny reconstruction [2], the definition of reliable dissimilarity measure is paramount for the image phylogeny research field.

Aiming at solving those problems and increasing the quality of the phylogeny reconstruction, in this paper, we introduce new methods to perform the dissimilarity calculation between images for the phylogeny reconstruction process. Firstly, we employ an histogram-based method to match color histograms between two near-duplicate images better capturing possible color differences between them. Secondly, we develop a new comparison metric working on images gradients, rather than directly on the pixels domain, and finally, we use the mutual information to compare them. The new comparison metrics

aim at better tackling possible image misalignment during the mapping process of one image onto another's domain.

Finally, we organized this paper into four more sections. Section 2 presents details about the novel methods proposed herein for dissimilarity calculation. Section 3 presents the methodology that we use for carrying out the experiments and the used datasets. Section 4 presents the performed experiments and obtained results. Finally, Section 5 concludes the article and shows some possible future work worth pursuing.

2 New Dissimilarity Calculation Techniques

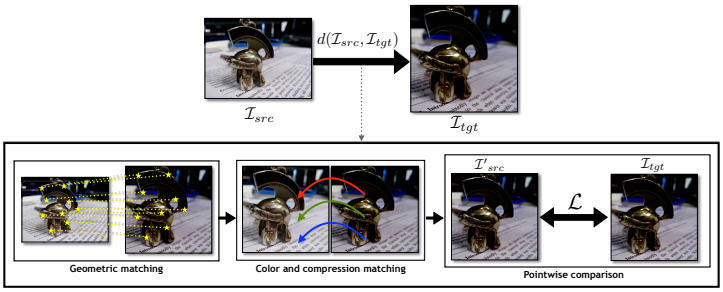

Figure 2: Dissimilarity calculation process. The mapping of image $\mathcal{I}_{src}$ onto $\mathcal{I}_{tgt}$'s domain involves a three-step process: geometric, color and compression matching. Afterwards, it is possible to directly compare the images using any pointwise comparison algorithm.

As proposed in [2], 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 $\mathcal{I}'_s = T_{\overrightarrow{\beta}}(\mathcal{I}_{src})$:

1. **Geometric matching:** also known as Image Registration. Among several different approaches known in the literature [19], a point-based registration process has been employed by using Speeded-Up Robust Features (SURF) [20] with a robust estimation [21] of warping and cropping parameters;

2. **Color matching:** it is performed for adjusting the color of the source image $\mathcal{I}_{src}$ to the target image $\mathcal{I}_{tgt}$;

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. There are also different approaches for calculating the point-wise dissimilarity between two images [22] and the authors opted to estimate it using Mean Squared Error (MSE). Figure 2 depicts the dissimilarity calculation process.

Differently from this standard dissimilarity pipeline, we propose here improvements for the dissimilarity calculation process. For that, we propose the replacement of the color

matching step, which was not very accurate, and also the metric used for performing the comparison between two images. We now turn our attention to these new approaches for improving the dissimilarity calculation.

2.1 Histogram Color Matching

The second step of the transformation estimation T (after geometric matching) consists of mapping the color space of the source image $\mathcal{I}_{src}$ onto the target's image $\mathcal{I}_{tgt}$ color space. Previous work on image phylogeny [1, 2, 3, 7, 4] performed the color matching between two images by normalizing each channel of $\mathcal{I}_{src}$ by the mean and standard deviation of the $\mathcal{I}_{tgt}$'s corresponding channel [23]. This method, although simplistic, works reasonably well when the color changes are minor. However, it leads to some problems when the transformations applied to the image when generating a child are stronger, specially in the case of contrast changes (e.g., gamma correction) or non-linear color mappings, which affect the distribution of pixel intensities throughout the image.

For a better color matching step, we propose to use the histogram matching technique [24]. This technique transforms the source image colors in such a way that their distribution acquires a form closer to the color distribution of the target image, by using the target image's color distribution information. Figure 3 shows two examples of color matching algorithms.

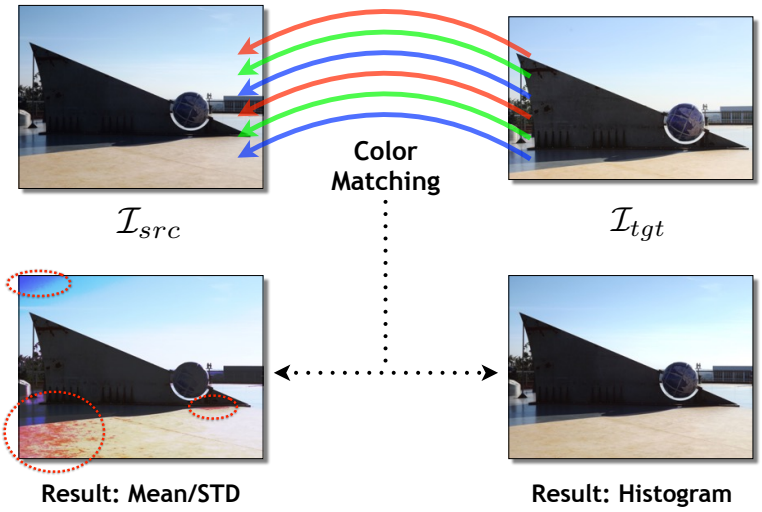

Figure 3: Matching the colors of the source image according to the color of the target image. The result of the color matching algorithm based on mean and standard deviation normalization [23] presents undesired artifacts that cannot be simply neglected, as can be noted in the marked regions of the picture. This problem does not happen when we perform a better color matching through histogram analysis.

To match the histograms of two images $\mathcal{I}_{src}$ and $\mathcal{I}_{tgt}$, we compute their histograms, H_{src} and H_{tgt} and compute their *Cumulative Distribution Function* (CDF) [25]. For a gray-scale image $\mathcal{I}$, with L gray levels, the gray level i has the probability of

$$p^{\mathcal{I}}(i) = \frac{n_i}{n}, \quad 0 \leq i < L \tag{2}$$

where n is the number of pixels in the image and n_i is the number of pixels of gray value i in the histogram of the image. The CDF of an image $\mathcal{I}$ is

$$\mathcal{C}^{\mathcal{I}}(i) = \sum_{k=0}^{i} p^{\mathcal{I}}(k). \tag{3}$$

With $\mathcal{C}^{\mathcal{I}_{src}}$ and $\mathcal{C}^{\mathcal{I}_{tgt}}$, the CDFs for $\mathcal{I}_{src}$ and $\mathcal{I}_{tgt}$, respectively, we find a transformation $\mathcal{M}$ that maps $\mathcal{C}^{\mathcal{I}_{src}}$ onto $\mathcal{C}^{\mathcal{I}_{tgt}}$. For each gray level i of $\mathcal{I}_{src}$, we find the gray level j of $\mathcal{I}_{tgt}$ whose $\mathcal{C}^{\mathcal{I}_{tgt}}(j)$ is the closest in $\mathcal{C}^{\mathcal{I}_{tgt}}$ to $\mathcal{C}^{\mathcal{I}_{src}}(i)$. Once the mapping is found, each pixel with gray level i in $\mathcal{I}_{src}$ has its value replaced by j.

We treat each color channel of these images independently, matching their histograms individually.

2.2 Gradient Comparison

Image gradients describe the value and direction of pixel intensity variation. They can be used to extract different information about the image, such as texture and location of edges. Here we filter an image by using the convolution with a *Sobel* [26] kernel for gradient estimation [24]. The convolution of an image $\mathcal{I}(x, y)$ with an $m \times n$ kernel $K(x, y)$ is given by:

$$K(x,y) * \mathcal{I}(x,y) = \sum_{i=\lfloor -m/2 \rfloor}^{\lfloor m/2 \rfloor} \sum_{j=\lfloor -n/2 \rfloor}^{\lfloor n/2 \rfloor} K(i,j)\mathcal{I}(x-i, y-j) \tag{4}$$

where '$*$' denotes the convolution operator. This equation is evaluated for all values of displacement variables x and y [24].

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. By comparing the gradients of a transformed image $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$, it is possible to compare both the intensity values (encoded in the gradient), as well as their variation throughout the image.

While the image comparison metric $\mathcal{L}$ stays the same (i.e., Minimum Square Error), we first compute the gradients in the horizontal and vertical directions, by convolving the images to be compared with the 3×3 Sobel kernels S_h (horizontal direction) and S_v (vertical direction)[1]. The R, G and B channels of $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$ are treated separately resulting in a total of six gradient images (two directions per color channel). The image comparison metric $\mathcal{L}$ is applied to each respective pair of gradient images of $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$, and the mean of the six values obtained in each position is taken as the final dissimilarity value.

[1] In our experiments, we have used the 3×3 Sobel kernel. We performed some exploratory tests with other kernel sizes (e.g., 3×3, 5×5 and 7×7) but their performance was similar for the problem herein.

2.3 Mutual Information Comparison

In Information Theory, mutual information (MI) is a measure of statistical dependency of two random variables, which represents the amount of information that one random variable contains about the other [27]. The mutual information between two random variables X and Y is given by:

$$MI(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y), \tag{5}$$

where $H(X) = -E_x[\log(P(X))]$ is the entropy (i.e., the expected value of the information associated to a random variable) of X and $P(X)$ is the probability distribution of X. In the case of discrete random variables, MI is defined as:

$$MI(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)(p(y)} \right), \tag{6}$$

where $p(x,y)$ is the joint Probability Distribution Function (PDF) [25] of X and Y, and both $p(x)$ and $p(y)$ are the marginal PDFs of X and Y, defined, respectively, as:

$$p(x) = \sum_y p(x,y), \tag{7}$$

$$p(y) = \sum_x p(x,y). \tag{8}$$

MI has been widely employed in several image applications such as gender identification [28], multi-modal data fusion [29], feature selection [30], and in image registration problems [31, 32] as a similarity measure (or cost function) to maximize when aligning two images (or volumes).

Applying MI to images means that the two random variables are the image $X = \mathcal{I}'_{src}$ and the image $Y = \mathcal{I}_{tgt}$ and x and y are the values of two pixels belonging to $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$, respectively. Thus, $p(x,y)$ is the joint PDF of the images $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$, evaluated for the values (x,y), where $x, y \in [0 \dots 255]$.

Clearly, the previous definitions involve the knowledge of the PDFs of pixels and, in particular, the joint PDF $p(x,y)$, from which it is easy to obtain $p(x)$ and $p(y)$ by marginalization (Equations 7 and 8). In general, such joint PDF is not known *a priori*, and needs to be estimated. Several methods [33] have been conceived to estimate the PDF of one or more random variables from a finite set of observations, such as the approximation of the joint PDF by the joint histogram

$$\hat{p}(x,y) = \frac{h(x,y)}{\sum_{x,y} h(x,y)}, \tag{9}$$

where $h(x,y)$ is the joint histogram of the images X and Y, namely the number of occurrences for each couple of gray level values (x,y), evaluated on the same (i,j) position on both images. MI has the following property: given two images $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$, $MI(\mathcal{I}'_{src}, \mathcal{I}_{tgt})$ is bounded as

$$0 \leq MI(\mathcal{I}'_{src}, \mathcal{I}_{tgt}) \leq \min(H(\mathcal{I}'_{src}), H(\mathcal{I}_{tgt})). \tag{10}$$

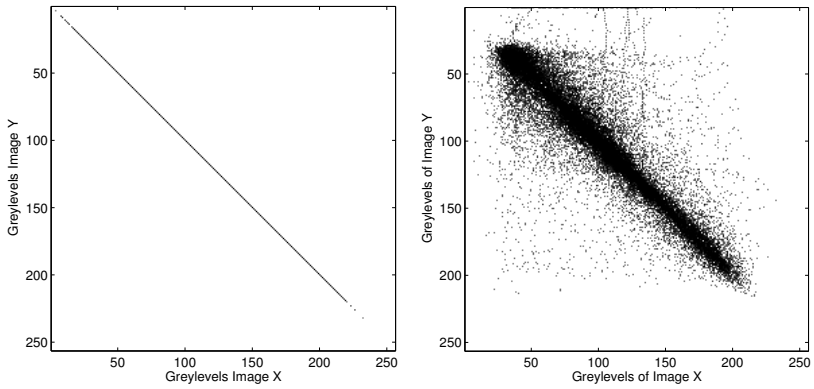

Figure 4: Bi-dimensional representation of two joint histograms. White pixels mean zero values while the other pixels represent values greater than zero (the images were inverted for viewing purposes). (a) Joint histogram of two (gray-scale) images perfectly aligned. (b) Joint histogram of two slightly misaligned images.

It can be demonstrated that MI is maximum when the two images are completely aligned (in terms of geometrical, color and compression transformation). Figure 4(a) shows a perfectly aligned case. If we assume a good transformation $T_{\overrightarrow{\beta}}$ that maps an image $\mathcal{I}_{src}$ onto an image $\mathcal{I}_{tgt}$'s domain, the mutual information $MI(T_{\overrightarrow{\beta}}(\mathcal{I}_{src}), \mathcal{I}_{tgt})$ is maximum. However, since each transformation is not completely reversible, if we apply the inverse transformation $T^{-1}_{\overrightarrow{\beta}}$ to $\mathcal{I}_{tgt}$ to obtain $\mathcal{I}_{src}$, their joint histogram is similar to Figure 4(b).

2.4 Gradient Estimation and Mutual Information Combined

The Gradient and Mutual Information comparison, presented in Sections 2.2 and 2.3, respectively, can be further combined into a single way to compute the dissimilarity value between two images. First, we calculate the gradient of the images $\mathcal{I}'_{src}$ and $\mathcal{I}_{tgt}$ as we described in Section 2.2. Afterwards, we compare each correspondent gradient of both images with mutual information, instead of using the image comparison metric $\mathcal{L}$ based on the standard Minimum Square Error. The final dissimilarity is the average of mutual information values for each gradient image.

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 misalignment 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, since we need to estimate the mutual information six times after the gradient calculation. However, this method provides better reconstruction results as we

shall discuss in Section 4. Finally, these two methods can also be combined with a better color matching approach (c.f., Section 2.1) further improving the dissimilarity calculation between pairs of images.

3 Experimental Setup

In this section, we discuss the evaluation setup and the used dataset and metrics for validating the methods discussed in this work.

3.1 Phylogeny reconstruction

As the actual phylogeny reconstruction is not a focus of this paper, after estimating the dissimilarity matrix, we apply an algorithm for reconstructing the phylogeny forest. For that, we use the Extended Automatic Optimum Branching (E-AOB) algorithm proposed by Costa et al. [7] currently the state-of-the-art for phylogeny reconstruction. This method is based on an optimum branching algorithm [34]. In short, the E-AOB algorithm works as follows. Consider a dissimilarity matrix $M_{n\times n}$. After calculating an optimum branching and sorting its $n-1$ edges into non-decreasing order according to their weight w, the algorithm selects the edges for the final forest one by one, from the lowest to the highest cost. After selecting $i-1$ edges, for $i = 1 \ldots n-1$, if $w(e_i) - w(e_{i-1})$, i.e., the difference of costs between the next edge to be selected and the last selected edge is higher than $\gamma \times \sigma$ (where σ is the standard deviation of all selected edges up to that point), the algorithm stops and returns the branching with $i-1$ edges. Afterwards, we find the optimum local branching in each group of nodes. We here used the best parameter reported by the authors ($\gamma = 2.0$) [7].

3.2 Dataset

For validation, we employed the freely available dataset introduced by Costa et al. [7]. This set comprises semantically similar images randomly selected from a set of 20 different scenes generated by 10 different acquisition cameras, 10 images per camera, 10 different tree topologies (i.e., the form of the trees in a forest) and 10 random variations of parameters for creating the near-duplicate images. We considered 2,000 forests of images generated by a single camera (Scenario *One Camera* – OC) and 2,000 forests generated by multiple cameras (Scenario *Multi Camera* – MC). The forests vary in the number of trees (size) $|\mathcal{F}| = \{1 \ldots 10\}$. The dataset has $2 \times 2,000 \times 10 = 40,000$ test cases in total.

The image transformations used to create the near duplicates are the same used in [7, 2]: re-sampling, cropping, affine warping, brightness and contrast adjustment, and lossy compression using the standard JPEG algorithm.

3.3 Evaluation metrics

For a better assessment of the proposed methods, we consider scenarios in which the *ground truth* is available. We used the metrics introduced by Dias et al. [6] to evaluate the proposed

approach: *Roots, Edges, Leaves* and *Ancestry* given by:

$$\mathrm{EM}(\mathrm{IPF}_R, \mathrm{IPF}_G) = \frac{S_R \cap S_G}{S_R \cup S_G} \tag{11}$$

where EM is the evaluation metric, IPF_R is the reconstructed forest with elements represented by S_R, and IPF_G is the forest ground truth with elements S_G. For instance, when considering the Edges metric, we calculate the intersection of the set of reconstructed edges with the set of edges in the ground truth normalized by all edges present in the union of the groups. The Roots metric measures if the reconstructed forest contains exactly the same roots as the ground-truth forest, i.e., the algorithm was able to find the very original images used to start the near-duplicate generation processes. Edges and Ancestry measure how well the algorithm finds the kinship relationships along time. While edges assess this information only locally and independently, Ancestry assesses the entire evolutionary process of a given image (a full branch in the tree). Finally, the leaves metric 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 forest. Figure 5 illustrates the calculation of these evaluation metrics.

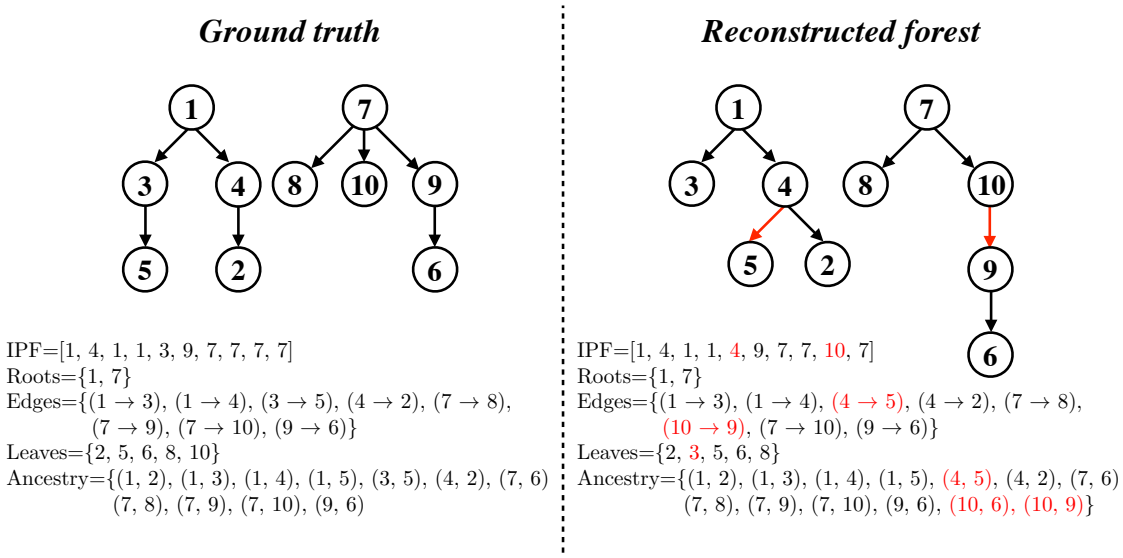

Figure 5: Evaluation metrics: roots, edges, leaves and ancestry. We represent the IPF as a vector, where $\mathrm{IPF}[v] = u$ means that exists the edge $(u \rightarrow v)$ in the forest, and one node v is a root only if $IPF[v] = v$. The differences between the reconstructed forest and the ground truth forest are highlighted in red.

3.4 Real cases

We also performed experiments and qualitative analysis considering two real datasets available in the literature.

- *The Situation Room* [6]: It comprises an image taken on May 1st, 2011, by the White House photographer Peter Souza and its variants, collected from the Internet. We performed the dissimilarity matrix calculation and the phylogeny reconstruction considering 98 near-duplicate images collected through Google Images and manually classified them in different groups considering (a) cases of inserting the Italian soccer player Mario Balotelli, (b) text overlay, (c) watermarking , (d) face swapping, (e) insertion of a joystick, (g) hats, and (n) changes in the image size without splicing operations.

- *The Ellen DeGeneres' selfie* [35]: this dataset comprises near-duplicates images related to the *selfie* taken by the host Ellen DeGeneres and some famous actors on March 2nd, 2014, during the 86th Academy Awards. The original image became viral after it was published on her Twitter account. Since then, it has been copied, modified and republished several times, with cases of text overlay, insertion of other people and animals in the picture and face swap. The dataset has 44 pictures from the internet and it is divided in five groups:

  (a) Edited versions of the original image posted at DeGeneres' Twitter account (@TheEllenShow[2]);

  (b) The moment that the picture has been taken but from a different point of view (another camera);

  (c) Group similar to group (b), but with slight differences on the posture of the people in the picture;

  (d) Similar to group (b) and (c), but with slight differences on the facial expression and posture of the people;

  (e) The moment before the acquisition of the *selfie* when the artists were gathering for taking the picture.

4 Results and discussion

In this section, we show the performed experiments to compare the proposed methods with the state-of-the-art MSE method, which has been the "de facto" dissimilarity calculation method thus far for image phylogeny [1, 13, 2, 3, 6, 7]. We analyze the impacts of calculating the dissimilarities using image gradients instead of image intensities, the replacement of the standard point-wise comparison metric minimum squared error with a mutual information dissimilarity calculation, and the incorporation of color matching for better representing the mapping of a source image onto a target image before actually calculating the dissimilarity.

4.1 Quantitative Experiments

Figures 6 and 7 show the results for the different approaches considered herein for calculating the dissimilarities for OC and MC scenarios, respectively. In all cases, the geometrical

[2]https://twitter.com/TheEllenShow/status/440322224407314432/photo/1

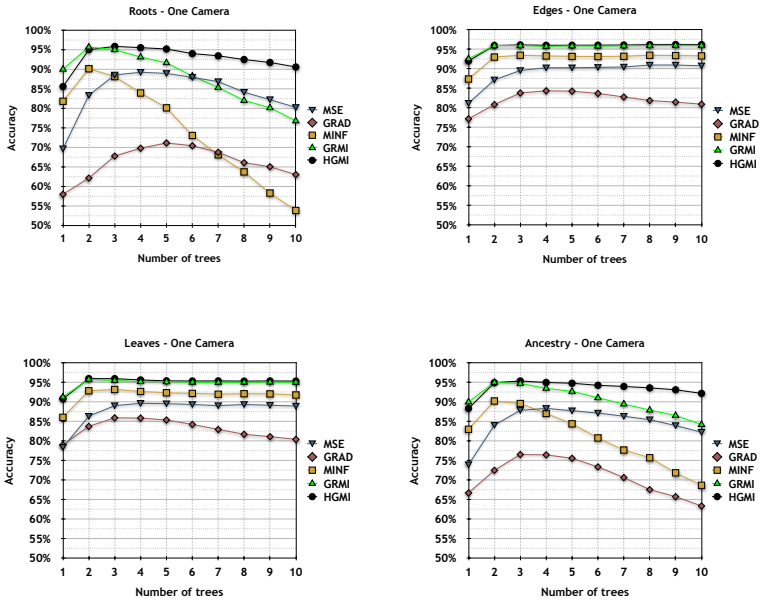

Figure 6: Results of forest reconstruction in the one camera (OC) scenario, considering the metrics Roots, Edges, Leaves and Ancestry.

mapping of one source image onto a target image is performed following the procedure discussed in the beginning of Section 2. The phylogeny reconstruction part uses the E-AOB algorithm for all methods, regarded as the state of the art in the literature for the reconstruction part [7, 3].

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; and, finally,

4. histogram color matching plus gradient estimation with mutual information (HGMI), extending upon GRMI to incorporate a better color matching before comparison.

First of all, the dissimilarity calculation does not benefit directly from the replacement of point-wise pixel intensity comparison by a point-wise comparison of image gradients as the results show MSE outperforming GRAD for OC and MC scenarios. The gradient itself only captures directional variations and small misalignments when comparing two gradient images affect the results more than when comparing the images through pixel intensities.

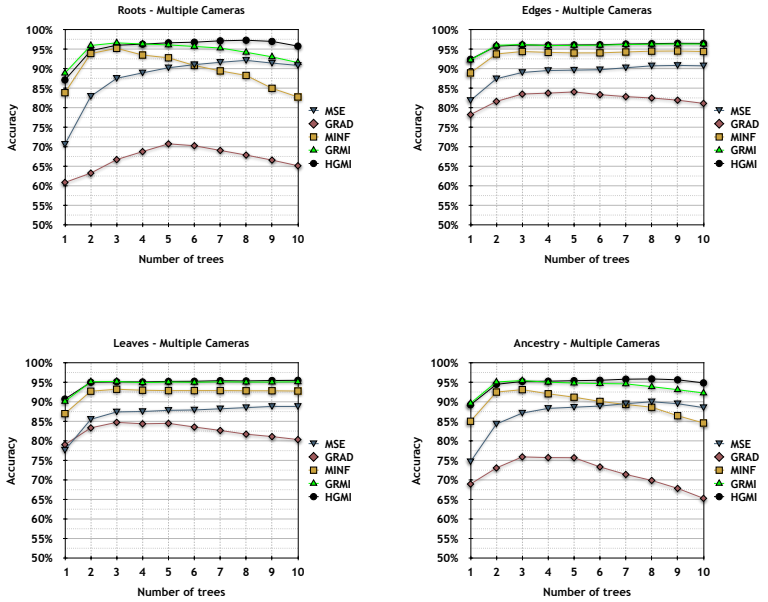

Figure 7: Results of forest reconstruction in the multiple cameras (MC) scenario, considering the metrics Roots, Edges, Leaves and Ancestry.

If we change the point-wise comparison method to mutual information but still use the pixel intensities, we have MINF outperforming MSE for the MC case. With MINF, small misalignments are not as important as for the GRAD case. One interesting behavior, however, is the improved performance for the OC case (Root and Ancestry metrics). In the OC case, as all of the images come from the same camera, the color matching for such images should be more refined than just the mapping using the mean and standard deviation to differentiate an image and its descendant. A point-wise comparison, in this case, is more effective for small differences (MSE method).

The results improve when combining the gradient calculation with mutual information (GRMI). The first reason is that, by not comparing the intensities, 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 in the gradient images as well as accounting for possible small misalignment. Finally, combining histogram color matching, gradient estimation and mutual information yields the final method HGMI, which solves the former color matching problem when using MINF. As we can see, HGMI outperforms the MSE baseline for all cases. With HGMI, we can reduce the dissimilarity errors by better matching the color transformations involved in the process of near-duplicate generation, by comparing the images using gradients instead of pixel intensities and in a distribution-like form instead of a point-wise one.

4.2 Error Reduction

To directly compare the approaches, we also calculate the error variation $\Delta error$ with respect to each metric (roots, edges, leaves and ancestry), using the same equation introduced in [3]:

$$\Delta error_{metric}(\text{M1,M2}) = \left(\frac{1 - \text{M1}_{metric}}{1 - \text{M2}_{metric}} \right) - 1 \tag{12}$$

where $M1$ represents the method being evaluated in comparison to method $M2$. Figure 8 depicts the average error reduction for HGMI when compared to the baseline MSE. In this case, there is an error reduction in about 45% in the OC scenario and more than 50% in the MC scenario for all evaluation metrics, clearly showing that the proposed HGMI dissimilarity measure is remarkably superior to the standard MSE procedure.

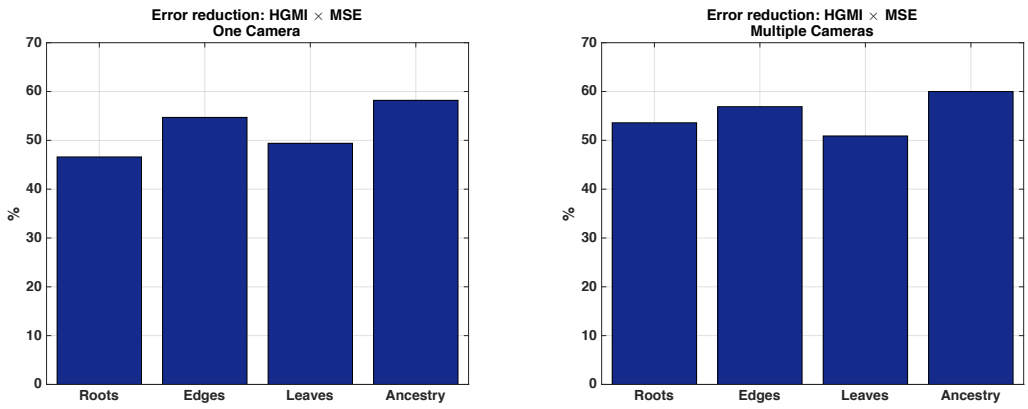

Figure 8: Error reduction: HGMI × MSE.

A Wilcoxon signed-rank test [36] shows that the best proposed approach, HGMI, is statistically better than the state-of-the-art MSE method for all cases and metrics, with 95% of confidence and a p-value of 0.002. Other possible combinations of the methods discussed herein are presented in the supplemental material along with this paper but none of them is more effective than the ones presented and discussed here.

4.3 Efficiency

To compare a pair of typical images (each with about one megapixel), including the time to register both images, MSE takes about 0.6s, GRAD takes 0.8s, and MINF takes 0.7s. The best performing methods GRMI and HGMI take both about 1.5s. However, all methods can be optimized to compensate their additional computational requirement using GPUs and parallel computing. The experiments were performed in a machine with an Intel Xeon E5645 processor, 2.40GHz, 16GB of memory, and Ubuntu 12.04.5 LTS.

4.4 Effects of Dissimilarity Errors on the Reconstruction

The dissimilarity errors directly affect the selection of the edges by the E-AOB reconstruction algorithm, as this process is done by comparing the difference of edge weights and the standard deviation of edges already selected. Considering that the forest needs to have 90 edges[3]. However, this event does not happen (on average) for GRAD-MC, GRAD-OC and MINF-OC, showing that a wrong number of trees is calculated for these cases, as presented in Figure 9. Note that, for GRMI and HGMI cases, in most of the cases, a correct number of trees is selected. Specifically for the HGMI case, the correct size of the forests outperform the baseline (MSE) in approximately 10 percentage points in MC scenario and 20 percentage points in the OC scenario.

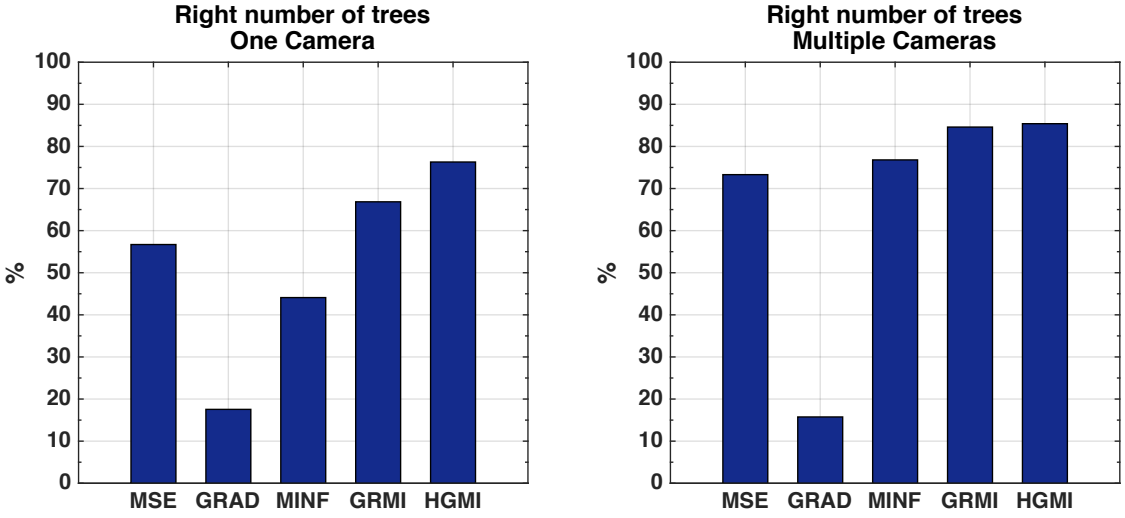

Figure 9: Average result (%) of correct number of trees calculated by E-AOB algorithm, for 2,000 test cases, considering forests with 10 trees.

4.5 Qualitative Experiments with Real Cases

We now turn our attention to assessing the behavior of the best performing method (HGMI) considering two real cases from the internet: *The Situation Room* [6] and *The Ellen DeGeneres' selfie* [35] (c.f., Section 3.4.)

For real cases, the feedback of a forensic expert for evaluating the quality of an algorithm is essential as there is no ground-truth. In this case, we empirically define γ parameter of the E-AOB algorithm for each case ($\gamma = 2.0$ for the case in *The Situation Room* and $\gamma = 0.5$ for the case of *The Ellen DeGeneres' selfie*). Figures 10 and 11 show the reconstructed forests for these cases.

[3]For cases with $n = 100$ images, the initial branching has $n - 1 = 99$ edges. For creating a forest $\mathcal{F}$ where $|\mathcal{F}| = 10$ tress, the number of total edges is $n - |\mathcal{F}| = 100 - 10 = 90$.

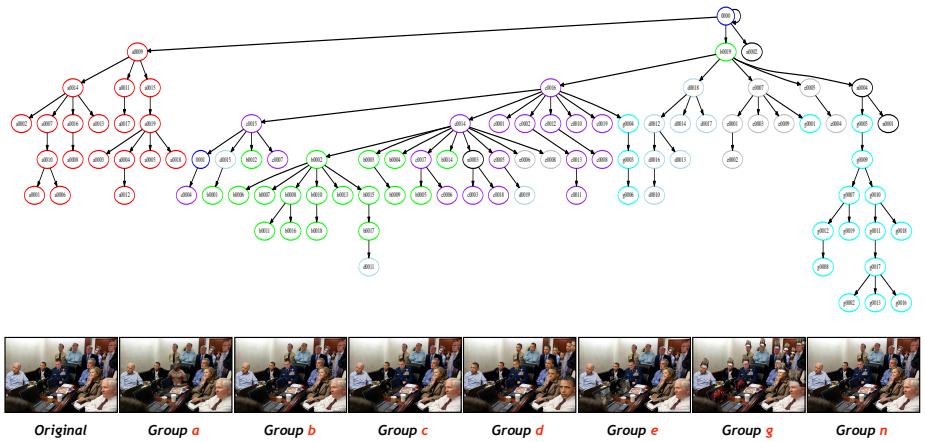

Figure 10: Reconstructed phylogeny for *The Situation Room* scenario.

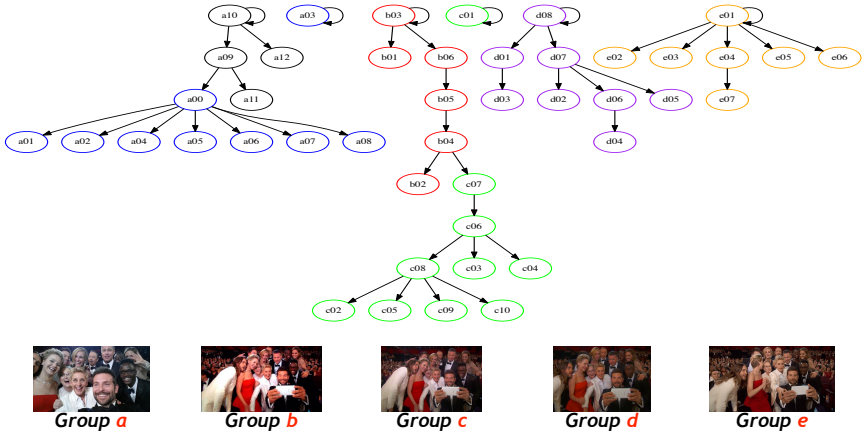

Figure 11: Reconstructed phylogeny for *The Ellen DeGeneres' selfie* scenario.

For *The Situation Room* scenario, the algorithm correctly identified image with ID *0000* (the White House version) as the root of the tree. Furthermore, as we expected, the result was that all images were grouped under the same tree (with image *0000* as the root). Although there are some images in wrong groups (sub-trees) in the reconstructed phylogeny, it is important to note that this dataset is mostly composed by images generated by splicing operations, which is in fact a special case of IPFs (multiple parenting phylogeny [17]). However, the E-AOB could separate these groups in different sub-trees with good effectiveness.

Considering the *Ellen DeGeneres' selfie* scenario, we have a forest with five trees. The near-duplicates are correctly organized according to their groups. The node *a00* is the picture originally posted at DeGeneres' Twitter account, and it was not selected here as the root of the group. However, the node is only two-edges of distance to the root. The tree with images *a09, a10, a11* and *a12* should also be placed as a child of node *a00*, but it has

a splicing of a cat in the picture, and the algorithm ended up classifying *a09* and *a10* as ancestors of *a00* and the nodes *a11* and *a12* as nodes not related to *a00*.

The nodes *a09*, *a10*, *a11* and *a12* are correctly grouped, since image *a09* is actually a montage also extracted from a Twitter's official account (@RealGrumpyCat[4]). The images *a10*, *a11*, and *a12* are all variants of this image. The image *a03* also should be classified as a child of *a00*, but it was separated in a single tree. This image was generated by splicing, in which all the faces in the picture were replaced by DeGeneres' face. Groups *b*, *c* and *d* are the hardest to analyze, since there is a subtle difference among them. As we can see, group *d* was correctly separated in a different tree. Although the groups *b* and *c* are placed on the same tree, it is possible to note that most of the images that belongs to the same group are together (with the exception of image *c01*, which is in a single tree). This structure certainly would help the work of a forensics expert. The group *e* was also correctly classified in a different tree.

5 Conclusion

In this paper, we presented novel approaches for computing the dissimilarity between two images, applied to the problem of image phylogeny forest reconstruction. Our approaches rely on the incorporation of a different color matching approach for better estimating the involved changes during the generation of near duplicates and the comparison between two images using gradient calculation and mutual information estimation.

This paper shows that comparing distributions is better than direct point-wise comparisons (with mutual information outperforming MSE as the comparison approach), gradient distributions are more appropriate than direct color distributions (with GRAD outperforming pixel-based comparisons when combined with mutual information), and it also shows that a more powerful family of color transformations enables a better tree reconstruction at the end of the dissimilarity calculation pipeline (with the incorporation of the histogram matching approach).

As discussed earlier, in the supplemental material, we provide direct comparisons, using the Wilcoxon signed-rank test, between the GRMI/HGMI and all combinations of these methods. These improvements are not marginal and certainly will significantly boost the current existing image phylogeny solutions as the dissimilarity calculation step, although overlooked thus far, is as important to the whole process as is the actual tree reconstruction step, if not more important. The HGMI method also presented good results in real-case setups, with good separation of different groups of near-duplicate images showing good potential for real-world deployment when analyzing the relationship among images.

For future work, we intend to investigate the use of mutual information for estimating the step of image registration [32] and also evaluate the impacts of new dissimilarity calculations to phylogeny estimation for different multimedia content such as videos and texts.

[4]https://twitter.com/RealGrumpyCat/status/440335332265848835/photo/1

Acknowledgment

We would like to thank the CAPES PDSE program (#99999.003836/2014-02), CAPES DeepEyes project, FAPESP (#2013/05815-2), Microsoft Research and the European Union through the REWIND (REVerse engineering of audio-VIsual coNtent Data) project for the financial support.

References

[1] Z. Dias, A. Rocha, S. Goldenstein, First steps towards image phylogeny, in: IEEE International Workshop on Information Forensics and Security (WIFS), 2010, pp. 1–6.

[2] Z. Dias, A. Rocha, S. Goldenstein, Image phylogeny by minimal spanning trees, IEEE Transactions on Information Forensics and Security (TIFS) 7 (2) (2012) 774–788.

[3] Z. Dias, S. Goldenstein, A. Rocha, Exploring heuristic and optimum branching algorithms for image phylogeny, Elsevier Journal of Visual Coimmunication and Image Representation 24 (2013) 1124–1134.

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

[5] A. Joly, O. Buisson, C. Frélicot, Content-based copy retrieval using distortion-based probabilistic similarity search, IEEE Trans. Multimedia 9 (2) (2007) 293–306.

[6] Z. Dias, S. Goldenstein, A. Rocha, Toward image phylogeny forests: Automatically recovering semantically similar image relationships, Elsevier Forensic Science International 231 (2013) 178–189.

[7] F. O. Costa, M. Oikawa, Z. Dias, S. Goldenstein, A. Rocha, Image phylogeny forest reconstruction, IEEE Transactions on Information Forensics and Security (TIFS) 9 (10) (2014) 1533–1546.

[8] L. Kennedy, S.-F. Chiang, Internet image archaeology: Automatically tracing the manipulation history of photographs on the web, in: Proceedings of the 16th ACM International Conference of Multimedia, 2008, pp. 349–358.

[9] A. D. Rosa, F. Uccheddu, A. Costanzo, A. Piva, M. Barni, Exploring image dependencies: a new challenge in image forensics, Proceedings of SPIE – Media Forensics and Security 7541 (2) (2010) 1 – 12.

[10] Z. Fan, R. L. Queiroz, Identification of bitmap compression history: Jpeg detection and quantizer estimation, IEEE Transactions on Image Processing 12 (2) (2003) 230–235.

[11] J. Mao, O. Bulan, G. Sharma, S. Datta, Device temporal forensics: an information theoretic approach, in: IEEE International Conference on Image Processing, 2009, pp. 1485–1488.

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

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

[14] S. Lameri, P. Bestagini, A. Melloni, S. Milani, A. Rocha, M. Tagliasacchi, S. Tubaro, Who is my parent? reconstructing video sequences from partially matching shots, in: IEEE Intl. Conference on Image Processing (ICIP), 2014, pp. 5342–5346.

[15] F. Costa, S. Lameri, P. Bestagini, Z. Dias, A. Rocha, M. Tagliasacchi, S. Tubaro, Phylogeny reconstruction for misaligned and compressed video sequences, in: IEEE Intl. Conference on Image Processing (ICIP), 2015, p. To appear.

[16] M. Nucci, M. Tagliasacchi, S. Tubaro, A phylogenetic analysis of near-duplicate audio tracks, in: IEEE Intl. Workshop on Multimedia Signal Processing, 2013, pp. 99–104.

[17] A. Oliveira, P. Ferrara, A. De Rosa, A. Piva, M. Barni, S. Goldenstein, Z. Dias, A. Rocha, Multiple parenting identification in image phylogeny, in: Image Processing (ICIP), 2014 IEEE International Conference on, 2014, pp. 5347–5351.

[18] Z. Dias, S. Goldenstein, A. Rocha, Large-scale image phylogeny: Tracing back image ancestry relationships, IEEE Multimedia 20 (2013) 58–70.

[19] B. Zitová, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (2003) 977–1000.

[20] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-up robust features (SURF), Elsevier Computer Vision Image Understanding 110 (3) (2008) 346–359.

[21] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.

[22] A. A. Goshtasby, Image Registration: principles, tools and methods, 1st Edition, Advances in Computer Vision and Pattern Recognition - Springer, 2012.

[23] E. Reinhard, M. Ashikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Computer Graphics Applications 21 (2001) 34–41.

[24] R. Gonzalez, R. Woods, Digital Image Processing, 3rd Edition, Prentice-Hall, 2007.

[25] J. G. MacKinnon, Numerical distribution functions for unit root and cointegration tests, Institute for Economic Research, Queen's University, 1995.

[26] I. Sobel, G. Feldman, A 3x3 isotropic gradient operator for image processing, a talk at the Stanford Artificial Project in (1968) 271–272.

[27] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 379–423, 623–656.

[28] J. Tapia, C. Perez, Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape, IEEE Transactions on Information Forensics and Security (TIFS) 8 (3) (2013) 488–499.

[29] R. Bramon, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, M. Sbert., Multimodal data fusion based on mutual information, IEEE Transactions on Visualization and Computer Graphics (TVCG) 18 (9) (2012) 1574–1587.

[30] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Transactions on Neural Networks (TNN) 5 (4) (1994) 537–550.

[31] P. Viola, W. M. Wells, Alignment by maximization of mutual information, International Journal of Computer Vision 24 (1997) 137–154.

[32] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, P. Suetens, Multimodality image registration by maximization of mutual information, Medical Imaging, IEEE Transactions on 16 (2) (1997) 187–198.

[33] K. A. Brownlee, Statistical theory and methodology in science and engineering, Wiley series in probability and mathematical statistics: Applied probability and statistics, Wiley, 1965.

[34] J. Edmonds, Optimum branchings, Journal of Research of National Institute of Standards and Technology 71B (1967) 48–50.

[35] M. Oikawa, Z. Dias, A. Rocha, S. Goldenstein, Manifold learning and spectral clustering for image phylogeny forests, Accepted on IEEE Transaction on Information Forensic and Security.

[36] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics bulletin (1945) 80–83.