



UNIVERSIDADE ESTADUAL DE CAMPINAS SISTEMA DE BIBLIOTECAS DA UNICAMP REPOSITÓRIO DA PRODUÇÃO CIENTIFICA E INTELECTUAL DA UNICAMP

Versão do arquivo anexado / Version of attached file:

Versão do Editor / Published Version

Mais informações no site da editora / Further information on publisher's website: https://link.springer.com/article/10.1007/s10044-017-0616-9

DOI: 10.1007/s10044-017-0616-9

Direitos autorais / Publisher's copyright statement:

©2017 by Springer. All rights reserved.

DIRETORIA DE TRATAMENTO DA INFORMAÇÃO

Cidade Universitária Zeferino Vaz Barão Geraldo CEP 13083-970 – Campinas SP Fone: (19) 3521-6493 http://www.repositorio.unicamp.br INDUSTRIAL AND COMMERCIAL APPLICATION



New dissimilarity measures for image phylogeny reconstruction

Filipe Costa¹ · Alberto Oliveira¹ · Pasquale Ferrara² · Zanoni Dias¹ · Siome Goldenstein¹ · Anderson Rocha¹

Received: 15 July 2016/Accepted: 21 March 2017/Published online: 30 March 2017 © Springer-Verlag London 2017

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: (1) estimating the dissimilarity between each pair of images and (2) reconstructing the phylogeny structure. Given that the dissimilarity calculation directly impacts the phylogeny reconstruction, in this paper, 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. These new formulations exploit a different method of color adjustment, local gradients to estimate pixel differences and mutual information as a similarity measure. The results obtained with the proposed formulation remarkably outperform the existing counterparts in the literature, allowing a much better analysis of the kinship relationships in a set of images, allowing for more accurate deployment of phylogeny solutions to tackle traitor tracing, copyright enforcement and digital forensics problems.

Electronic supplementary material The online version of this article (doi:10.1007/s10044-017-0616-9) contains supplementary material, which is available to authorized users.

Anderson Rocha anderson.rocha@ic.unicamp.br

- ¹ Institute of Computing, University of Campinas, Campinas, Brazil
- ² Università degli Studi di Firenze, Florence, Italy

Keywords Digital forensics \cdot Image phylogeny reconstruction \cdot Mutual information \cdot Dissimilarity calculation

1 Introduction

Undoubtedly, images are powerful communication tools living up to the classical adage comparing them to a thousand words when conveying any information. Their communication power has raised significantly with the advent of social media. Within this new reality, images are published, shared, modified and often republished effortlessly. Frequently, reposting and sharing will happen after myriad small modifications, such as cropping, resampling, affine warping and color adjustments, resulting in what is called a near duplicate of the original image. Sometimes content sharing might be illegal, however, such as in cases of copyright infringement or public defamation. On other occasions, simply possessing the content (e.g., images depicting child pornography) already constitutes a crime. Considering the aforementioned scenarios, it is often important to develop appropriate solutions to track and monitor how images are shared and evolve on the Internet over time.

In this vein, *Image Phylogeny* [8, 14, 16] has been developed recently in an attempt to find the relationship structure among near-duplicate images. According to [23], 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 the leaves of the tree represent

Filipe Costa fcosta@ic.unicamp.br

"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 but in distinct moments in time. Both cases represent a generalization of the near-duplicate concept, referred to by Dias et al. [13] as **Semantically Similar Images**. In this case, the set of semantically similar images can be represented by a forest comprising different trees, each one correlating the near duplicates which originated from the same source image [8, 13]. Figure 1 depicts an example of the image phylogeny problem.

Once we trace back the past history of the near duplicates, image phylogeny can be useful for aiding (allied with additional side information) in the discovery, for instance, of who was the first user that published an image containing illegal or abusive content (e.g., fake and defamatory image of celebrities or politicians, and child pornography), which in turn was redistributed after being modified by different users. Image phylogeny solutions can also be useful for detecting image restaging and repurposing as well as propaganda effects in the Internet.

Dias et al. [14, 16] formally defined the problem of Image Phylogeny in two steps: (1) the calculation of the dissimilarity between each pair of near-duplicate images and (2) the reconstruction of the phylogeny tree. Let \mathcal{T} be a family of image transformations, T a transformation such that $T_{\overrightarrow{\beta}} \in \mathcal{T}$ is parameterized by β . Considering two nearduplicate 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

$$d(\mathcal{I}_{\text{src}}, \mathcal{I}_{\text{tgt}}) = \min_{\substack{T \xrightarrow{\beta} \in \mathcal{T}}} \left| \mathcal{I}_{\text{tgt}} - T_{\overrightarrow{\beta}}(\mathcal{I}_{\text{src}}) \right| point - wisecomparison \mathcal{L}.$$
(1)

Equation 1 calculates the dissimilarity between the best transformation mapping \mathcal{I}_{src} onto \mathcal{I}_{tgt} parameterized by $\vec{\beta}$, according to the family of transformations \mathcal{T} . After the proper mapping, the comparison between the images can be performed by any pointwise comparison method \mathcal{L} (e.g., minimum squared error).

Since the first work in image phylogeny [14], several branches to this research field have been developed. Extensions to the original image phylogeny algorithm were also proposed for reconstructing the tree of evolution of a set of near-duplicate videos (video phylogeny) [6, 15, 26, 31]. In addition, the phylogeny of audio clips were also investigated by [33]. The study of multiple parent–child relationships (images obtained through the modifications of more than one image) was also explored [35, 36]. Moreover, improvements on the image phylogeny original framework have been proposed for dealing with



Fig. 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 (\mathbf{b} , \mathbf{c}) with similar content, each of which spurring its own near duplicates (descendants). After a transformation applied on image \mathbf{b} , we generate the near duplicate

d. The near duplicate **e** is the result of a different transformation over **b**. The near duplicates **h**, **i** are created considering different transformations applied over the near duplicate **d**). The near duplicates **f**, **g** and **j** are the result of transformations applied over the original image **c**. Finally, the near duplicate **a** is the result of a transformation over **f**

large-scale setups [12] and semantically similar images [13]. Recently, new improvements on the construction of the parent-child relationships have also been studied and proposed such as the using of multiple dissimilarity matrices for the phylogeny reconstruction [30] and optimum branching solutions [11]. 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 [18, 24, 25, 29, 38].

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 (strengthening new passive traitor tracing techniques) and news tracking services (feeding news tracking services with key elements for mining opinion forming processes along time).

Although the field has been developing significantly over the past few years, thus far researchers mainly focused on proposing different phylogeny reconstruction approaches [8, 11, 13, 14] often using a standard methodology for dissimilarity calculation as originally proposed by [16]. This dissimilarity calculation involves the estimation of the transformations that map a source image onto a target image, followed by their comparison in a pointwise fashion. As the transformation estimation is not exact, the pointwise comparison method \mathcal{L} is strongly affected by artifacts generated in such processes. Given that the dissimilarity calculation directly affects the result of the final phylogeny reconstruction [16], the definition of a 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, with the intent of improving the phylogeny reconstruction as a whole. First, we employ a histogram-based method to match color histograms between two near-duplicate images better capturing possible color differences between them. Then, we develop a new comparison metric working on images gradients, rather than directly on the pixels domain. Finally, we use the mutual information technique to compare them. The new comparison metrics aim at better tackling possible image misalignments during the mapping process of one image onto another's domain.

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, Sect. 5 concludes the paper and shows some possible future work worth pursuing.

2 New dissimilarity calculation techniques

As proposed in [16], 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, which results in the generation of $\mathcal{I}'_s = T_{\overrightarrow{\beta}}(\mathcal{I}_{src})$:

- Geometric matching also known as Image Registration. Among several different approaches known in the literature [47], the image registration is computed by finding keypoints in each pair of images using SURF (Speeded-Up Robust Features) [2], which are in turn used to estimate warping and cropping parameters robustly using RANSAC [19];
- 2. Color matching performed to adjust the color of the source image \mathcal{I}_{src} according to the color of the target image \mathcal{I}_{tgt} . It is done through normalization of each channel of \mathcal{I}_{src} by the mean and standard deviation of the respective channel in $\mathcal{I}_{tgt}[37]$;
- 3. Compression matching the image \mathcal{I}_{src} is compressed with \mathcal{I}_{tgt} 's JPEG compression parameters. Considering that near duplicates may be recompressed, this process might result in the generation of artifacts over the target image. This step is important for simulating this aspect, aiming at inserting the same compression artifacts present in \mathcal{I}_{tgt} into \mathcal{I}_{src} , which may improve the quality of the estimation of $T_{\overrightarrow{R}}(\mathcal{I}_{src})$.

Then, a comparison between the estimated $\mathcal{I'}_{src} =$

 $T_{\overrightarrow{\beta}}(\mathcal{I}_{\mathrm{src}})$ and $\mathcal{I}_{\mathrm{tgt}}$ is performed using a pointwise image

comparison measure. There are many different approaches for calculating the pointwise dissimilarity between two images [21], though the authors opted to estimate it using the mean squared error (MSE). Figure 2 depicts the dissimilarity calculation process.

In this work, we propose several improvements over the individual steps of the aforementioned general pipeline for dissimilarity calculation such as the replacement of the color matching step, which was not very accurate, and also the method used for performing the pointwise comparison between two images. We now turn our attention to these new approaches for improving the dissimilarity calculation. The main goal of this paper is to show that it is indeed possible to consider alternative and promising options for obtaining better results in terms of phylogeny graph



Fig. 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. Afterward, it is possible to directly compare the images using any pointwise comparison algorithm

reconstruction and in finding the actual kinship relationships among semantically similar images. MSE comparison is considered the state of the art as can be seen in several recent publications such as [3, 6, 7, 15, 30, 32, 38].

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 work phylogeny space. Previous on image [8, 11, 13, 14, 30] performed the color matching between two images by normalizing each color channel of \mathcal{I}_{src} by the mean and standard deviation of the \mathcal{I}_{tgt} 's corresponding channel [37]. 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 descendent are stronger, especially in the case of contrast changes, gamma correction or nonlinear color mappings, which affect the distribution of pixel intensities throughout the image.

For a better color matching step, we propose to use a histogram matching technique [20]. 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.

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) [27]. For a grayscale image \mathcal{I} , with *L* gray levels, the gray level *i* has the probability of

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



Fig. 3 Matching the colors of the source image according to the color distribution of the target image. The result of the color matching algorithm based on mean and standard deviation normalization [37] presents undesired artifacts that cannot be simply neglected, as can be noted in the marked regions of the picture. This problem is lessened when we perform a better color matching through histogram analysis

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

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

With $C^{\mathcal{I}_{src}}$ and $C^{\mathcal{I}_{tgt}}$, the CDFs for \mathcal{I}_{src} and \mathcal{I}_{tgt} , respectively, we find a transformation \mathcal{M} that maps $C^{\mathcal{I}_{src}}$ onto $C^{\mathcal{I}_{tgt}}$. For each gray level *i* of \mathcal{I}_{src} , we find the gray level *j* of \mathcal{I}_{tgt} whose $C^{\mathcal{I}_{tgt}}(j)$ is the closest in $C^{\mathcal{I}_{tgt}}$ to $C^{\mathcal{I}_{src}}(i)$. 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 a convolution with a *Sobel* [41] kernel for gradient estimation [20]. It is worth mentioning that other filters could be used as well but we opted to work with Sobel because it is separable in the vertical and horizontal directions and it is reasonably efficient. The convolution of an image $\mathcal{I}(x, y)$ with an $m \times n$ kernel K(x, y) is given by:

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

where '*' denotes the convolution operator. This equation is evaluated for all values of displacement variables x and y [20].

As contrast enhancement and color transformations are often used when creating near duplicates, directly affecting the gradients of an 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 the intensity values (encoded in the gradient) and 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.

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 [40]. The mutual information between two random variables X and Y is given by:

$$MI(X, Y) = H(Y) - H(Y|X) = H(X) - H(X|Y),$$
 (5)

where $H(X) = -E_x[\log(P(X))]$ is the entropy (i.e., the expected value of the information associated with 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:

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

where p(x, y) is the joint probability distribution function (PDF) [27] of X and Y, and both p(x) and p(y) are the marginal PDFs of X and Y, defined, respectively, as:

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

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

MI has been widely employed in several image applications such as gender identification [42], multi-modal data fusion [4], feature selection [1], and in image registration problems [28, 43] 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 $x, y \in [0...255]$.

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 (Eqs. 7 and 8). In general, such joint PDF is not known *a priori* and needs to be estimated. Several methods [5] 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

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

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

It can be demonstrated that MI is maximum when the two images are completely aligned (in terms of geometrical, color and compression transformation). Figure 4a shows a perfectly aligned case. If we assume a perfect transformation $T_{\overrightarrow{\beta}}$ that maps an image \mathcal{I}_{src} onto an image

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.

 \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 $T_{\overrightarrow{\beta}}^{-1}$ to \mathcal{I}_{tgt} to obtain \mathcal{I}_{src} , their joint histogram is similar to Fig. 4b.

2.4 Gradient estimation and mutual information combined

The Gradient and Mutual Information comparison, presented in Sects. 2.2 and 2.3, respectively, can be further combined into a single form of computing 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 Sect. 2.2. Afterward, 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 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 six times after the gradient calculation. However, this method leads to better reconstruction results as we discuss in Sect. 4. Moreover, the additional computational time can be easily compensated by parallel implementation as the different calculations are independent and can take advantage of modern GPUs and multi-core technologies. Finally, these two methods can also be combined with a better color matching approach (c.f., Sect. 2.1) further improving the dissimilarity calculation between pairs of images.

3 Experimental setup

In this section, we discuss the evaluation setup, including the used datasets and validation metrics for the methods discussed in this work.

3.1 Dataset

For validation, we the two datasets introduced by Costa et al. [8], which are freely available:

- *Training Dataset* it represents a small exploratory set containing images in two setups: *One Camera* and *Multiple Cameras*. The images were taken from three different cameras, three different scenes, three images per camera, four forest sizes $|F| = \{2..5\}$, one topology² and 10 random variations of parameters for creating the near-duplicate images, totaling $2 \times 3^3 \times$ $4 \times 1 \times 10 = 2$, 160 forests.
- *Test Dataset* it also comprises cases for two different setups, *One Camera* and *Multiple Cameras*, considering forests with size $|\mathcal{F}| = 1...10$. More specifically, 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 2000 forests of images generated by a single camera (Scenario *One Camera*—OC) and 2000 forests generated by multiple cameras (Scenario *Multi Camera*—MC). The forests vary in the number of trees (size) $|\mathcal{F}| = \{1...10\}$. Therefore the dataset has $2 \times 2000 \times 10 = 40,000$ test cases in total. As we evaluate each dissimilarity measure and each color matching approach, in this dataset, the final number of test cases is 320,000.

The image transformations used to create the near duplicates present in the used datasets and described in [16] are: geometric transformations, brightness and contrast adjustment, and lossy compression using the standard lossy JPEG algorithm. Table 1 details the transformations and their operational ranges for creating the controlled dataset. The near-duplicate generation process uses the algorithms implemented in the ImageMagick Library³. Figure 5 depicts some examples of scenes we considered in this work.

3.2 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. [13] to evaluate the proposed approach: *Roots, Edges, Leaves* and *Ancestry*. 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

 $^{^2}$ A topology refers to the form of the trees in a forest. For instance, Fig. 1 depicts two different topologies for the set of images present on its left side.

³ http://www.imagemagick.org/script/index.php.



Table 1 Transformations and their operational ranges for creating controlled dataset

Resampling (up/down)	[90%, 110%]
Rotation	$[-5^\circ,5^\circ]$
Scaling by axis	[90%, 110%]
Off-diagonal correction	[0.95, 1.05]
Cropping	[0%, 5%]
Brightness adjustment	[-10%, 10%]
Contrast adjustment	[-10%, 10%]
Gamma correction	[0.9, 1.1]
Recompression	[50%, 100%]

metric measures whether or not 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. The metrics Edges and Ancestry, in turn, measure how well the algorithm finds the kinship relationships along time. While the Edges metric assesses this information only locally and independently, the Ancestry one 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.



Fig. 5 Examples of pictures present in the datasets described in this work

Figure 6 illustrates the calculation of these evaluation metrics.

The evaluation metrics are formally defined as:

Root	$R(\mathrm{IPF}_1, \mathrm{IPF}_2) = \frac{R_1 \cap R_2}{R_1 \cup R_2};$
Edges	$E(\operatorname{IPF}_1, \operatorname{IPF}_2) = \frac{ E_1 \cap E_2 }{ E_2 };$
Leaves	$L(\mathrm{IPF}_1, \mathrm{IPF}_2) = \frac{ L_1 \cap L_2 }{ L_1 \cup L_2 };$
Ancestry	$A(\mathrm{IPF}_1, \mathrm{IPF}_2) = \frac{ A_1 \cap A_2 }{ A_1 \cup A_2 }.$

In all cases, N is the number of nodes in a tree, IPF_1 is the reconstructed image phylogeny forest (IPF) with elements represented by R_1 (roots), E_1 (edges), L_1 (leaves) and A_1 (ancestry) and IPF₂ is the forest ground truth with elements R_2 , E_2 , L_2 and A_2 . The roots, leaves and ancestry metrics calculate the intersection of the results returned by IPF₁ with respect to the reference forest IPF2 and normalizes it by the union of both sets, while the edges metric calculates the score of correct edges by normalizing the intersection of the result returned by IPF₁ with respect to the reference forest IPF₂ by the ground-truth set. For instance, in the example of Fig. 6, the Root metric vields $R(IPF_1, IPF_2) = 2/2 = 100\%$, the Edges metric yields $E(\text{IPF}_1, \text{IPF}_2) = 6/8 = 75\%$, the Leaves metric yields $L(IPF_1, IPF_2) = 4/6 = 66.6\%$ and the Ancestry metric yields $A(IPF_1, IPF_2) = 10/14 = 71.4\%$.

3.3 Phylogeny reconstruction

As the actual phylogeny reconstruction process is not a focus of this paper, after estimating the dissimilarity matrix, we apply an already proposed algorithm to reconstruct the phylogeny forest. Our choice was the Extended Automatic Optimum Branching (E-AOB) algorithm proposed by Costa et al. [8], which is currently the state-of-the-art for phylogeny reconstruction. This method is based on an optimum branching algorithm [17]. In short, the E-AOB algorithm works as follows. Consider a dissimilarity matrix $M_{n \times n}$ representing the pairwise relationships of *n* images. After calculating an optimum branching and sorting its n-1edges 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 - 1edges, for $i = 1 \dots n - 1$, 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. Afterward, we find the optimum local branching in each group of nodes (further refining each tree).

The parameter γ_{E-AOB} was found considering a training dataset and each one of the proposed dissimilarity measures. From these experiments, we also defined

 $\tau_{AOB} = \mu_{AOB} + (2.0 \times \sigma_{AOB})$, and as a consequence, $\gamma_{E-AOB} = 2.0$, following the best parameter reported by the authors in [8].

3.4 Real cases

In addition to the experiments previously outlined, we also performed experiments and qualitative analysis considering two real datasets available in the literature.

- The Situation Room [13]: It comprises an image taken on May 1st, 2011, by the White House photographer Pete 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 [34] this dataset comprises near-duplicate images related to the selfie taken by the host Ellen DeGeneres and some famous actors on March 2, 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 into five groups:
 - (a) Edited versions of the original image posted at DeGeneres Twitter account (@TheEllenShow⁴;
 - (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 groups (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 for analyzing image dissimilarities in a phylogeny setup with the state-of-the-art

⁴ http://migre.me/vTYN7 (secure shortened link).



MSE method, which has been the "de facto" dissimilarity calculation method thus far for image phylogeny [8, 11, 13–16]. We analyze the impacts of calculating the dissimilarities using image gradients instead of image intensities, the replacement of the standard pointwise 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 7 and 8 show the results for the different approaches considered herein for calculating the dissimilarities for OC (images taken with a single camera) and MC (images taken with multiple cameras) scenarios, respectively. In all cases, the geometrical mapping of one source image onto a target image is performed following the procedure discussed in the beginning of Sect. 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 [8, 11].

The baseline dissimilarity calculation considered is the MSE, the state of the art, which compares two images pointwise using the pixel intensities after the proper mapping (transformation) of one image onto the other's target domain. The proposed modifications are:

- 1. gradient estimation (GRAD), which still compares the images pointwise but using image gradients instead of pixel intensities;
- 2. mutual information (MINF), which replaces the pointwise 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 pointwise pixel intensity comparison by a pointwise comparison of image gradients as the results show MSE outperforming GRAD for OC and MC scenarios. The gradient itself only captures directional variations; small misalignments when comparing two gradient images affect the results more than when comparing the images through pixel intensities.

If we change the pointwise 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 descendants. A pointwise 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 directly, the color information artifacts are not as strong. Second, the comparison in this case is no longer done in a pointwise 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 misalignments after the image mapping (registration). Finally, combining histogram color matching, gradient estimation and mutual information leads to the final method HGMI, which solves the former color matching problem when using MINF. As



Fig. 7 Forest reconstruction results for the one camera (OC) scenario, considering the metrics Roots, Edges, Leaves and Ancestry

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 pointwise one. Although GRMI outperforms HGMI when we have few trees, HGMI excels at this task when the size of the forest increases. Furthermore, considering the cases that GRMI is better, the difference is not significant, according to the Wilcoxon signed-rank test. For more results, please refer to the supplemental material.

4.2 Error reduction

To directly compare the approaches, we also calculate the error variation Δ error with respect to each metric (roots, edges, leaves and ancestry), using the same equation introduced in [11]:

$$\Delta \operatorname{error}_{\operatorname{metric}}(M1, M2) = \left(\frac{1 - M1_{\operatorname{metric}}}{1 - M2_{\operatorname{metric}}}\right) - 1 \tag{11}$$

where M1 represents the method being evaluated in comparison to method M2. Table 2 shows the average error reduction for HGMI when compared to the baseline MSE. In this case, there is an error reduction of about 45% in the OC scenario and more than 50% for the MC scenario, for all evaluation metrics, clearly showing that the proposed HGMI dissimilarity measure is remarkably superior to the standard MSE procedure.

A Wilcoxon signed-rank test [44] shows that the bestproposed 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.



Fig. 8 Forest reconstruction results for the multiple cameras (MC) scenario, considering the metrics Roots, Edges, Leaves and Ancestry

Table 2 Error reduction (%): HGMI versus MSE

	Roots	Edges	Leaves	Ancestry
One camera	46.6	54.7	49.4	58.2
Multiple cameras	53.6	56.9	50.9	60.0

4.3 Run-time efficiency

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

4.3.1 Registration efficiency

Although the efficiency of the dissimilarity calculation is not the primary focus of this work, someone could argue what would happen if we also optimize the dissimilarity calculation process by selecting, for instance, a faster keypoint detector and descriptor for the registration step. Taking this into account, we performed a performance test comparing two descriptor extractors: SURF (that was used in this work and has been the standard in image phylogeny solutions thus far) and ORB (*Oriented Fast and Rotated Binary Robust Independent Elementary Features*), a binary descriptor extractor based on the Harris corner detector [39].

For the performance test, we considered 50 examples, comprising trees with 10 nodes each. We evaluate, for these examples, the time (in seconds) of each step of the dissimilarity calculation process. Table 3 shows the time spent by each step of the dissimilarity calculation, comparing the descriptor extraction using SURF and the

Table 3 Time analysis (inseconds) of each step of HGMIdissimilarity calculation,considering SURF and ORB forthe descriptor matching in theregistration step

	SURF	ORB
Keypoints and descriptors extraction (for each image)	0.831	0.030
Descriptors cross-check matching (for each pair of images)	0.077	0.101
Image registration $(\mathcal{I}_{src} \rightarrow \mathcal{I}_{tgt})$	0.138	0.166
Color and compression matching $(\mathcal{I}_{src} \rightarrow \mathcal{I}_{tgt})$	0.100	0.102
Dissimilarity calculation $(\mathcal{I}_{src} \rightarrow \mathcal{I}_{tgt})$	0.911	0.965
Total execution time (full 10×10 matrix)	112.790	105.410

The reported execution times refer to the full-time required to compare the 10 input images pairwise in both directions

descriptor extraction made using ORB. For this test, we considered the HGMI dissimilarity calculation, which was the best approach presented in Sect. 4.1.

Table 3 shows that the ORB descriptor extractor is more efficient than SURF for finding the keypoints and for describing them. However, its efficiency does not influence the other steps.

Moreover, to analyze the effectiveness of the phylogeny reconstruction, we used 1000 samples of test cases (500 for the OC scenario and 500 for the MC scenario), considering the HGMI method for dissimilarity calculation and forests with 10 trees. Figure 9 depicts the difference in the quality of reconstruction for roots and ancestry, considering different γ_{E-AOB} parameters for the phylogeny forest reconstruction. The results for edges and leaves are similar.

Figure 9 shows that the registration step using SURF as the descriptor extractor is more appropriate than ORB. While SURF is invariant to rotation, scale and color changing, ORB is only invariant to rotation and Gaussian noise. Considering the family of transformations presented in the datasets, it is only natural to expect SURF to outperform ORB in the registration step and, consequently, in the phylogeny forest reconstruction process as a whole.

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 edges⁵. However, this event does not happen (on average) for GRAD-MC, GRAD-OC and MINF-OC, showing that the wrong number of trees is found in these cases, as Fig. 10 shows. Note that, for GRMI and HGMI cases, in most of the cases, the correct number of trees is selected.

Specifically for the HGMI case, the correct size of the forests outperforms the baseline (MSE) in approximately 10 percentage points in MC scenario and 20 percentage points in the OC scenario.

4.5 Exploring other gradients

As previously mentioned, in this paper, we considered the 3×3 Sobel filter for extracting the gradient of the near duplicates when exploring gradient-based dissimilarity calculations. However, as one would expect, other methods can be considered as well. For the sake of comparison, we performed an exploratory experiment considering the gradient estimation using Histogram of Oriented Gradients (HoG) [9] instead of Sobel filters. For these experiments, we considered 500 dissimilarity matrices for the MC setup and 500 cases for OC setup, each of which comprising 10 trees for each forest. Table 4 presents the results for the HoG estimation for dissimilarity calculation, compared to the GRAD method.

Table 4 shows that HoG is not as effective as GRAD at the task of finding the original images (roots) of the forests, which also affects negatively the ancestry measure. In this case, we believe the main problem is due to the nature of the E-AOB reconstruction algorithm which reconstructs only one tree instead a forest when using the HoG-based dissimilarity. Nevertheless, HoG-based dissimilarity leads to good trees in general, correctly finding the relationship between direct ancestors (edges).

Considering this positive aspect of HoG, we went on and performed another exploratory experiment to check what would happen when using a Sobel-based gradient method for finding the roots and a HoG-based method for calculating the remaining relationships in the tree (edges). In this case, for each dissimilarity tree, we reconstruct a forest with one of the proposed methods and, then, we reconstruct another forest considering the HoG-based dissimilarity matrix, but removing any edge that points to the node which was already selected as root in the first reconstruction. With this experiment, we seek to maintain the best of both worlds, the performance for roots and ancestry of

⁵ For cases with n = 100 images, the initial branching has n - 1 = 99 edges. For creating a forest \mathcal{F} where $|\mathcal{F}| = 10$ trees, the number of total edges is $n - |\mathcal{F}| = 100 - 10 = 90$.



Fig. 9 Phylogeny reconstruction test considering ORB and SURF for the registration step



Right number of trees One Camera 100 90 80 70 60 50 % 40 30 20 10 0 MSE GRAD MINF **GRMI HGMI**



 Table 4 Comparison of GRAD versus HoG for gradient-based dissimilarity calculation

	Method	Roots	Edges	Leaves	Ancestry
OC	GRAD	0.693	0.835	0.836	0.708
	HoG	0.130	0.974	0.970	0.559
MC	GRAD	0.666	0.819	0.816	0.672
	HoG	0.139	0.974	0.964	0.579

Best results are highlighted in bold

GRAD and, at the same time, to improve the results for edges and leaves when using HoG in the dissimilarity. Table 5 presents the results of this fusion approach, considering the combined GRAD & HoG methods for gradient estimation, the HGMI + HoG combination and the bestproposed combination HGMI, from previous sections.

The results of Tables 4 and 5 show that using HoG is not strictly better than using Sobel for gradient estimation.

However, this experiment shows that exploring other gradient estimation methods for dissimilarity measures is an endeavor worth pursuing and it holds potential to push the results even further. At this point, the choice of Sobel for extracting the gradient of the images is mainly motivated by its efficiency when compared to other filters or gradientbased methods such as HoG, especially when we consider that Sobel can be implemented with two separable filters.

4.6 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* [13] and *The Ellen DeGeneres' selfie* [34] (c.f., Sect. 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 available. In this case, we empirically define the γ parameter of the E-AOB algorithm for each

Table 5 Results for thereconstruction consideringdifferent gradient methods andtheir combination with the othermethods proposed in this paper

	Method	Roots	Edges	Leaves	Ancestry
One camera	HGMI	0.953	0.970	0.963	0.949
	GRAD + HoG	0.675	0.956	0.948	0.828
	HGMI + HoG	0.953	0.974	0.965	0.955
Multiple cameras	HGMI	0.905	0.970	0.964	0.929
	GRAD + HoG	0.713	0.956	0.956	0.849
	HGMI + HoG	0.905	0.974	0.969	0.933



Fig. 11 Reconstructed phylogeny for The Situation Room scenario

case ($\gamma = 2.0$ for the case in *The Situation Room* and $\gamma = 0.5$ for the case of *The Ellen DeGeneres' selfie*). Figures 11 and 12 show the reconstructed forests for these cases.

For *The Situation Room* case, 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 [10, 35]). However, the E-AOB could separate these groups in different sub-trees with good effectiveness.

Considering the *Ellen DeGeneres selfie* case, 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 a11and 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 (@RealGrumpyCat⁶). 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 belong to the same group are together (with the exception of image c01, which is in a single tree). This structure

⁶ http://migre.me/vTYLt (secure shortened link).



Fig. 12 Reconstructed phylogeny for The Ellen DeGeneres selfie scenario

certainly would help the work of a forensics expert. The group e was also correctly classified in a different tree.

5 Conclusion and future work

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

This paper shows that comparing distributions is more appropriate to this problem than direct pointwise comparisons (with mutual information outperforming MSE as the comparison approach), gradient distributions are more also more adequate 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 signedrank 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. The HGMI method also presented good results in real-case setups, with good separation of different groups of nearduplicate images showing good potential for real-world deployment when analyzing the relationship among images. Furthermore, a series of experiments shows that the choice of Sobel for the gradient calculation is just one option out of many other alternatives. For instance, HoG has shown to be equivalent, although more computationally expensive. The positive result of HoG shows that exploiting other alternatives might be a worthwhile effort.

For future work, we intend to investigate the use of mutual information for estimating the step of image registration [28] and also evaluate the impacts of new dissimilarity calculations to phylogeny estimation for different multimedia content such as videos and texts. We also want to investigate other measures for dissimilarity calculation and forest reconstruction, as multi-view features [22, 45] and deep multi-modal features [46]. Furthermore, we intent to investigate new temporal features for video phylogeny reconstruction.

Acknowledgements We would like to thank the Brazilian Coordination for Higher Education and Personnel (CAPES) through the CAPES DeepEyes Project, the São Paulo Research Foundation (Grants #2013/05815-2 and the DéjàVu Project #2015/19222-9), Microsoft Research and the European Union through the REWIND (REVerse engineering of audio-VIsual coNtent Data) project for the financial support. Finally, it is important to mention that this material is also 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.

References

- Battiti R (1994) Using mutual information for selecting features in supervised neural net learning. IEEE Trans Neural Netw (TNN) 5(4):537–550
- Bay H, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). Elsevier Comput Vis Image Underst 110(3):346–359
- Bestagini P, Tagliasacchi M, Tubaro S (2016) Image phylogeny tree reconstruction based on region selection. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2059–2063
- Bramon R, Boada I, Bardera A, Rodriguez J, Feixas M, Puig J, Sbert M (2012) Multimodal data fusion based on mutual information. IEEE Trans Vis Comput Graph (TVCG) 18(9):1574–1587
- Brownlee KA (1965) Statistical theory and methodology in science and engineering. Wiley series in probability and mathematical statistics: applied probability and statistics. Wiley, New York
- Costa F, Lameri S, Bestagini P, Dias Z, Rocha A, Tagliasacchi M, Tubaro S (2015) Phylogeny reconstruction for misaligned and compressed video sequences. In: IEEE international conference on image processing (ICIP), pp 301–305
- Costa F, Lameri S, Bestagini P, Dias Z, Tubaro S, Rocha A (2016) Hash-based frame selection for video phylogeny. In: IEEE international workshop on information forensics and security (WIFS)
- Costa F, Oikawa M, Dias Z, Goldenstein S, Rocha A (2014) Image phylogeny forests reconstruction. IEEE Trans Inf Forensics Secur (TIFS) 9(10):1533–1546
- Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE computer society conference on computer vision and pattern recognition (CVPR), pp 886–893
- de Oliveira A, Ferrara P, De Rosa A, Piva A, Barni Mauro, Goldenstein S, Dias Z, Rocha A (2016) Multiple parenting phylogeny relationships in digital images. IEEE Trans Inf Forensics Secur(TIFS) 11(2):328–343
- Dias Z, Goldenstein S, Rocha A (2013) Exploring heuristic and optimum branching algorithms for image phylogeny. Elsevier J Vis Commun Image Represent 24:1124–1134
- Dias Z, Goldenstein S, Rocha A (2013) Large-scale image phylogeny: tracing back image ancestry relationships. IEEE Multimed 20:58–70
- Dias Z, Goldenstein S, Rocha A (2013) Toward image phylogeny forests: automatically recovering semantically similar image relationships. Elsevier Forensic Sci Int (FSI) 231:178–189
- Dias Z, Rocha A, Goldenstein S (2010) First steps toward image phylogeny. In: IEEE international workshop on information forensics and security (WIFS), pp 1–6
- Dias Z, Rocha A, Goldenstein S (2011) Video phylogeny: recovering near-duplicate video relationships. In: IEEE international workshop on information forensics and security (WIFS), pp 1–6

- Dias Z, Rocha A, Goldenstein S (2012) Image phylogeny by minimal spanning trees. IEEE Trans Inf Forensics Secur (TIFS) 7(2):774–788
- Edmonds J (1967) Optimum branchings. J Res Natl Inst Stand Technol 71B:48–50
- Fan Z, De Queiroz RL (2003) Identification of bitmap compression history: Jpeg detection and quantizer estimation. IEEE Trans Image Process 12(2):230–235
- Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. ACM Commun 24(6):381–395
- 20. Gonzalez R, Woods R (2007) Digital image processing, 3rd edn. Prentice-Hall, New Jersey
- 21. Goshtasby AA (2012) Image registration: principles, tools and methods. advances in computer vision and pattern recognition, 1st edn. Springer, New York
- Hong C, Yu J, Tao D, Wang M (2015) Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval. IEEE Trans Ind Electron 62(6):3742–3751
- Joly A, Buisson O, Frélicot C (2007) Content-based copy retrieval using distortion-based probabilistic similarity search. IEEE Trans Multimed 9(2):293–306
- Kender JR, Hill ML, Natsev AP, Smith JR, Xie L (2010) Video genetics: a case study from youtube. In: International conference on multimedia, pp 1253–1258
- Kennedy L, Chang S-F (2008) Internet image archaeology: automatically tracing the manipulation history of photographs on the web. In: ACM international conference of multimedia, pp 349–358
- Lameri S, Bestagini P, Melloni A, Milani S, Rocha A, Tagliasacchi M, Tubaro S (2014) Who is my parent? Reconstructing video sequences from partially matching shots. In: IEEE international conference on image processing (ICIP), pp 5342–5346
- 27. MacKinnon JG (1996) Numerical distribution functions for unit root and cointegration tests. J Appl Econom 11(6):601–618
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997) Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 16(2):187–198
- Mao J, Bulan O, Sharma G, Datta S (2009) Device temporal forensics: an information theoretic approach. In: IEEE international conference on image processing, pp 1485–1488
- Melloni A, Bestagini P, Milani S, Tagliasacchi M, Rocha A, Tubaro S (2014) Image phylogeny through dissimilarity metrics fusion. In: IEEE European workshop on visual information processing (EUVIP), pp 1–6
- Melloni A, Lameri S, Bestagini P, Tagliasacchi M, Tubaro S (2015) Near-duplicate detection and alignment for multi-view videos. In: IEEE international conference on image processing (ICIP), pp 1–4
- 32. Milani S, Fontana M, Bestagini P, Tubaro S (2016) Phylogenetic analysis of near-duplicate images using processing age metrics. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 2054–2058
- Nucci M, Tagliasacchi M, Tubaro S (2013) A phylogenetic analysis of near-duplicate audio tracks. In: IEEE international workshop on multimedia, signal processing, pp 99–104
- 34. Oikawa M, Dias Z, Rocha A, Goldenstein S (2016) Manifold learning and spectral clustering for image phylogeny forests. IEEE Trans Inf Forensics Secur 11(1):5–18
- Oliveira A, Ferrara P, De Rosa A, Piva A, Barni M, Goldenstein S, Dias Z, Rocha A (2014) Multiple parenting identification in image phylogeny. In: IEEE international conference on image processing (ICIP), pp 5347–5351
- Oliveira A, Ferrara P, De Rosa A, Piva A, Barni M, Goldenstein S, Dias Z, Rocha A (2016) Multiple parenting phylogeny

relationships in digital images. IEEE Trans Inf Forensics and Secur (TIFS) 11(2):328-343

- Reinhard E, Ashikhmin M, Gooch B, Shirley P (2001) Color transfer between images. IEEE Comput Graph Appl 21:34–41
- De Rosa A, Uccheddu F, Costanzo A, Piva A, Barni M (2010) Exploring image dependencies: a new challenge in image forensics. SPIE Med Forensics Secur 7541(2):1–12
- Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In: IEEE international conference on computer vision (ICCV), pp 2564–2571
- Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(379–423):623–656
- Sobel I, Feldman G (1968) A 3 × 3 isotropic gradient operator for image processing. In: Artificial Project in a talk at the Stanford, pp 271–272
- 42. Tapia JE, Perez CA (2013) Gender classification based on fusion of different spatial scale features selected by mutual information

- Viola P, Wells WM (1997) Alignment by maximization of mutual information. Int J Comput Vis 24:137–154
- Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1(6):80–83
- 45. Yu J, Rui Y, Chen B (2014) Exploiting click constraints and multi-view features for image re-ranking. IEEE Trans Multimed 16(1):159–168
- 46. Yu J, Yang X, Gao F, Tao D (2016) Deep multimodal distance metric learning using click constraints for image ranking. IEEE Trans Cybern. doi:10.1109/TCYB.2016.2591583
- Zitová B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21:977–1000