Image Provenance Analysis

Jakob Veselsky & Jesus Cantu

What is Provenance Analysis?

- “Forensics analysts are interested not only in determining if a digital object is fake or real but also in pinpointing who created it, what happened, when and how an asset was created.
- Understanding these journeys, a process known as “provenance analysis,” provides rich insights into the use, motivation, and authenticity underlying any given work. Provenance analysis provides a snapshot of the chronology and validity of content as it is uploaded, re-uploaded, and modified over time.

The Provenance Framework

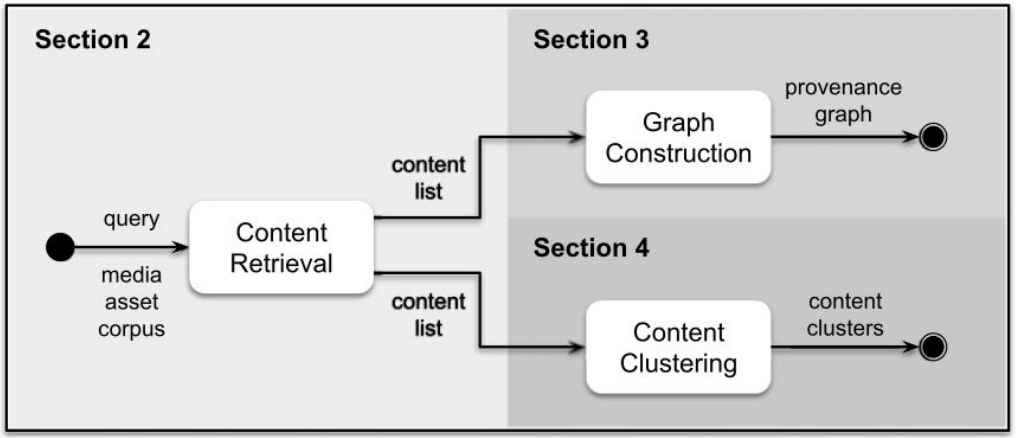


Fig. 15.5 The provenance framework. Provenance analysis is usually executed in two stages. Starting from a given query and a corpus of media assets, the first stage is always related to the content retrieval activity, which is herein explained in Sect. 15.2. The first stage's output is a list of assets of interest (content list), which is fed to the second stage. The second stage, in turn, may either comprise the activity of graph construction (discussed within Sect. 15.3) or the activity of content clustering (discussed within Sect. 15.4). While the former activity aims at organizing the retrieved assets in a provenance graph, the latter focuses on establishing meaningful asset clusters

Provenance Filtering For Multimedia Phylogeny

Pinto et al. 2017

Multimedia Phylogeny?

A relatively new discipline that studies the evolutionary process that influence multimedia objects and collections, as well as the relationship among transformed versions of an object, looking for casual and ancestry relationships, the types of transformations, and the order in which they were applied to objects.

In essence, we want to know if this composition

- Fake or Real?
- Who created it & how?
- How did it change?

Problem

Before analyzing a pool of objects for possible kinship relationships, we need to be able to comb through large quantities of data looking for the pieces potentially associated with a given query.

This task, provenance filtering, needs to be performed prior to subsequent multimedia phylogeny steps– pairwise image dissimilarity calculations and the phylogenetic graph analysis/construction.

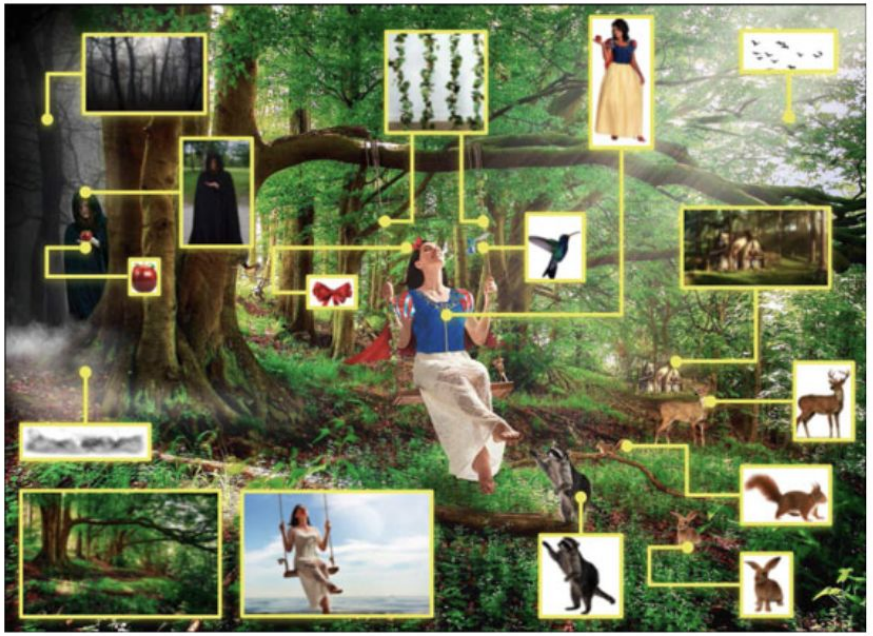
Most of the work thus far in multimedia phylogeny has overlooked the provenance filtering task, considering it to be a well solved problem.

Problem

Near-duplicate detection (NDD) methods work properly for the task of finding semantically-similar images but fail in the presence of multiple donors.

Likewise, the relations of the composite with the images donating small objects challenge the retrieval capabilities of a typical CBIR (content-based image retrieval) method.

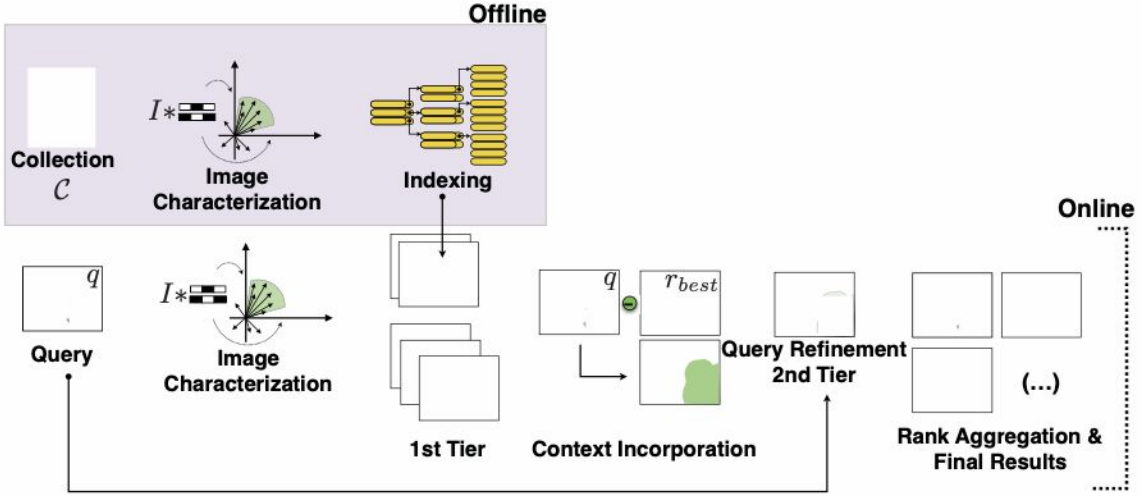
Important advances have been made on finding ancestral relationships between pairs of images; nevertheless, the performance of such algorithms is significantly degraded if a good set of potentially related images is not found beforehand.



Proposed Method

The authors extend upon image representation and indexing techniques (common in NDD and CBIR areas) to deal with provenance filtering for multiple donor and composite images.

Two Stage Technique



The authors propose a two-tiered provenance filtering approach to find all the potential images that might have contributed to the creation process of a given query q. In their solution, the first (coarse) tier aims to find the most likely “host” images — the major donor or background — contributing to a composite/doctored image. The search is then refined in the second tier, in which we search for more specific (potentially small) parts of the query that might have been extracted from other images and spliced into the query image

Dataset & Training

Datasets: Nimble Challenge 2016 (NC2016) & 2017 (NC2017) datasets, provided by the National Institute of Standards and Technology (NIST).

These datasets comprise a query set containing different kinds of manipulated images (e.g., copy-move and compositions), and a gallery set containing the source images used to produce the queries. The datasets also comprise distractor images. The probe sets of NC2016 and NC2017 datasets contain 288 and 16 composite images, respectively. The gallery sets contain 874 and 10446 images, respectively. The authors also embed the datasets within one million images (distractors) provided by RankOne Inc., as recommended by NIST for evaluating scalability (hereafter referred to as World1M dataset) .

The authors report the quality of the results in terms of Recall@k that measures the fraction of correct images at the top-k retrieved results.

Experiments and Results

Indexing Method

Table 1. Runtime (in seconds) and memory usage (GB), per query, in the first tier, for different indexing techniques in the NC2017 and NC2017 + World1M datasets. KD-Forest comprises two trees. * denotes the method did not scale.

Method	KD-Tree	KD-Forest	PQ	HCAL
Runtime	0.69 s	0.72 s	13.96 s	0.85 s
Memory	1.48 GB	10.69 GB	0.02 GB	5.38 GB
Runtime (World1M)	8.8 s	7.61 s	*	*
Memory (World1M)	34.99 GB	66.42 GB	*	*

Although PQ is more efficient in terms of storage for a small scale, it does not scale for World1M. The clustering in HCAL prevented it from scaling for 1M images. More work involving approximate clustering and sampling would be necessary in this case. KD-Tree shows a good storage and efficiency tradeoff.

Context Incorporation and Ranking Aggregation

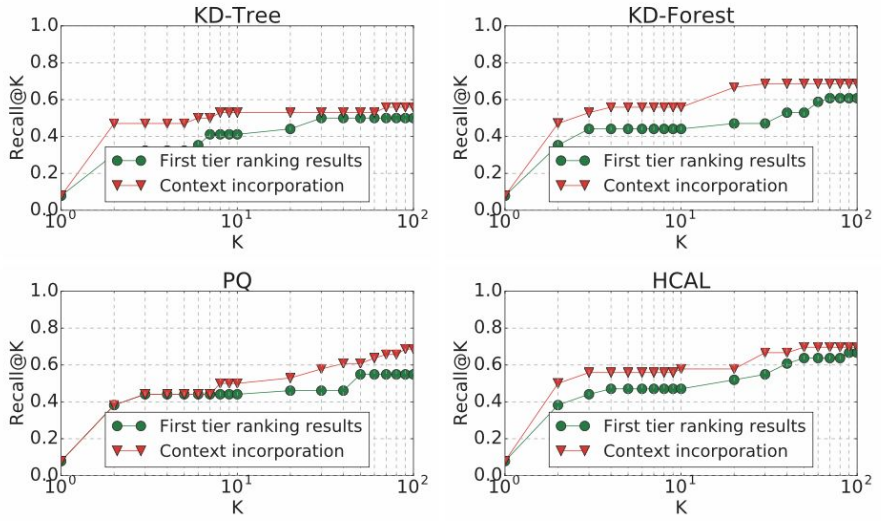


Fig. 4. First- and second-tier results for the NC2017 dataset in terms of Recall@k. The context incorporation is important regardless of the used indexing technique.

In this section, the authors evaluate the proposed approach to improve ranking results for donor images. Fig. 4 shows the performance results in terms of recall at the top-k retrieved images, considering the retrieval of donor images in the first and second tiers of the proposed method.

Although not shown here, the performance for retrieving the host image is always above 95% as it shares much content with q. The challenge in provenance filtering is in retrieving the donors.

Large Scale Image Retrieval

Table 2. Performance results for NC2016 and NC2017 datasets embedded in one million images and KD-Forest (2 trees). Bold highlights improvements in the second tier.

Dataset	Type	Tier	Recall@10
NC2016 + World1M	Host	1st	99.65%
		2nd	**100.00%**
NC2016 + World1M	Donor	1st	63.00%
		2nd	**67.71%**
NC2017 + World1M	Host	1st	88.24%
		2nd	88.24%
NC2017 + World1M	Donor	1st	25.49%
		2nd	25.49%

Here, the authors evaluate the proposed approach, considering a more challenging scenario, in which they embed the NC2016 and NC2017 datasets into one million images. The World1M dataset contains several images that are semantically similar to the images that compose both datasets.

Table 2 shows the obtained results in this experiment. There is a gain of about 7% when retrieving donors for NC2016 when they compare the obtained results in the first and second tiers. The results for NC2017 are slightly lower given that the composite images in this dataset are more difficult, more photorealistic and smaller with respect to the whole image.

Key Takeaways

By incorporating the context of the top results with respect to the query itself, the authors improve the retrieval results and better find possible donors of a given composite (forged) query q. Experiments with different indexing techniques have also shown that KD-forests seem to be the most effective but not the most efficient. KD-trees, on the other hand, are more efficient but less effective. In their experiments, PQ did not perform well for large galleries.

The contributions of this work are:

- The exploration of different querying and indexing techniques for the new problem of provenance filtering
- The incorporation of provenance context to single out possible candidate regions related to donors in the creation process of a query
- The study of the efficiency and effectiveness tradeoffs involved in the provenance filtering task while dealing with very large collections of images

Image Provenance Analysis at Scale

Moreira et al. 2018

Goal

- “recover the graph of relationships between plausibly connected images”
  - undirected edges
    - neighboring transformations are identified
  - directed edges
    - the order of neighboring transformations is expressed

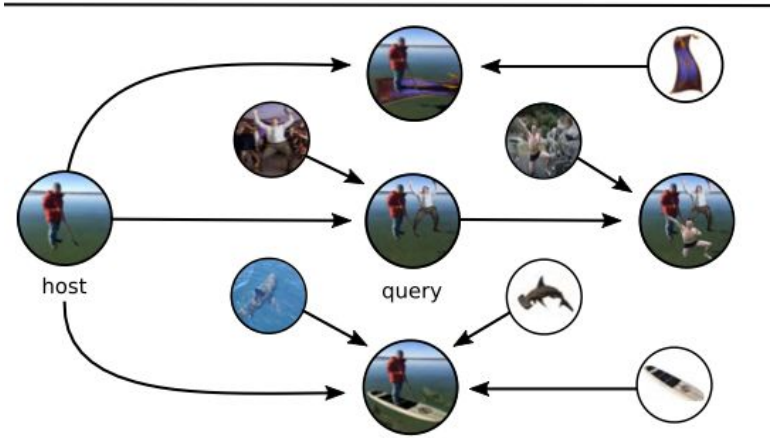


Image Provenance Analysis at Scale

- r/photoshopbattles/
- We have
  - Image to be queried
  - Database of images

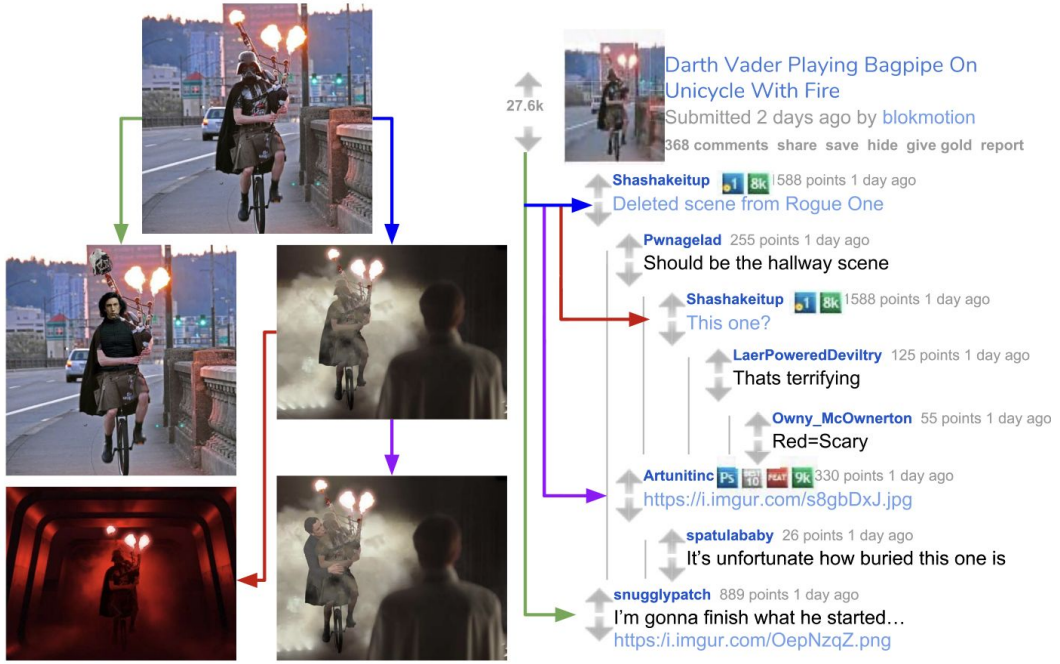
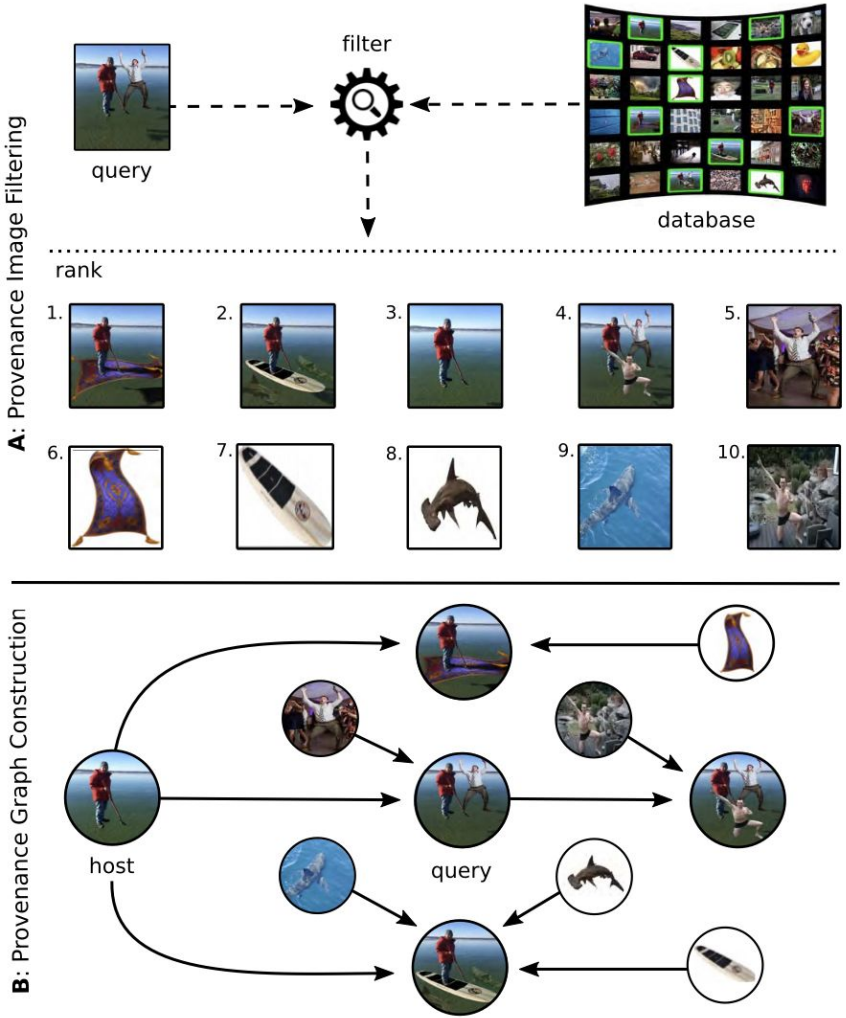
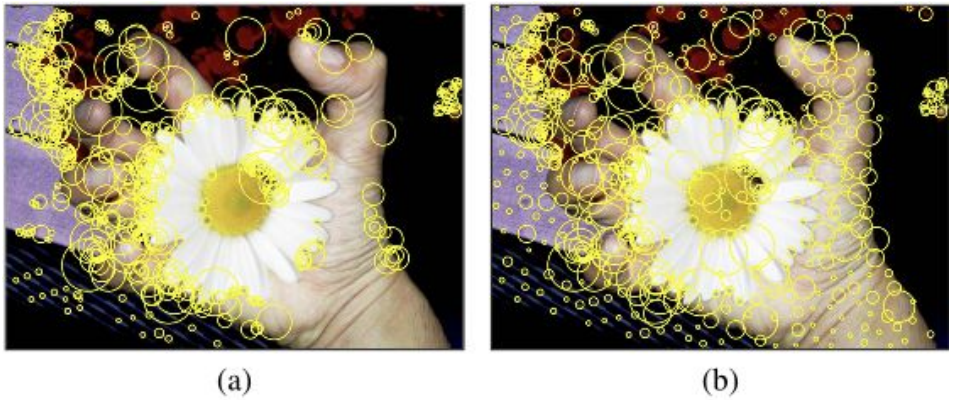


Image Provenance Analysis at Scale

- First
  - Provenance Image Filtering
    - Semantically similar
    - Near duplicates
    - Image compositions
- Second
  - Provenance Graph Construction
    - Host Image
      - Background
    - Donor Image
      - Some elements



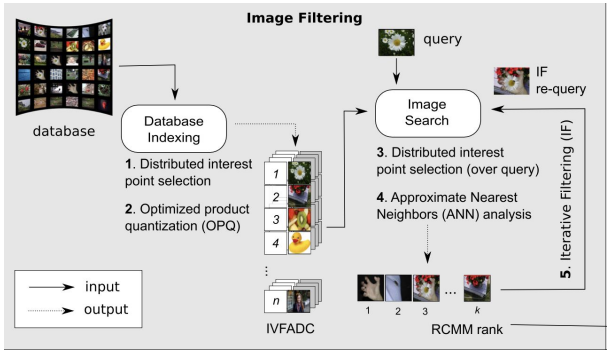
Provenance Image Filtering



- We do not want a list with:
  - Many near duplicates
  - Many duplicate donors
- We will look for key points
- SURF is ok
  - Pays more attention to edges then surfaces
- We want to look at all of a surface so we want something else to extract keypoints
- Distributed Interest Point Selection

Provenance Image Filtering

- So we have these feature vectors
- Transform these vectors to a new space using Optimized Product Quantization
  - method for dramatically compressing high-dimensional vectors
- “We refer to this new rotated feature set as Fr . From a random sample of Fr , a coarse set of representative centroids O is generated using PQ. A subsequent Inverted File System with Asymmetric Distance Computation (IVFADC) [55]) is generated from O, allowing for fast and efficient search.”
- Approximate Nearest Neighbor Search



Provenance Graph Construction

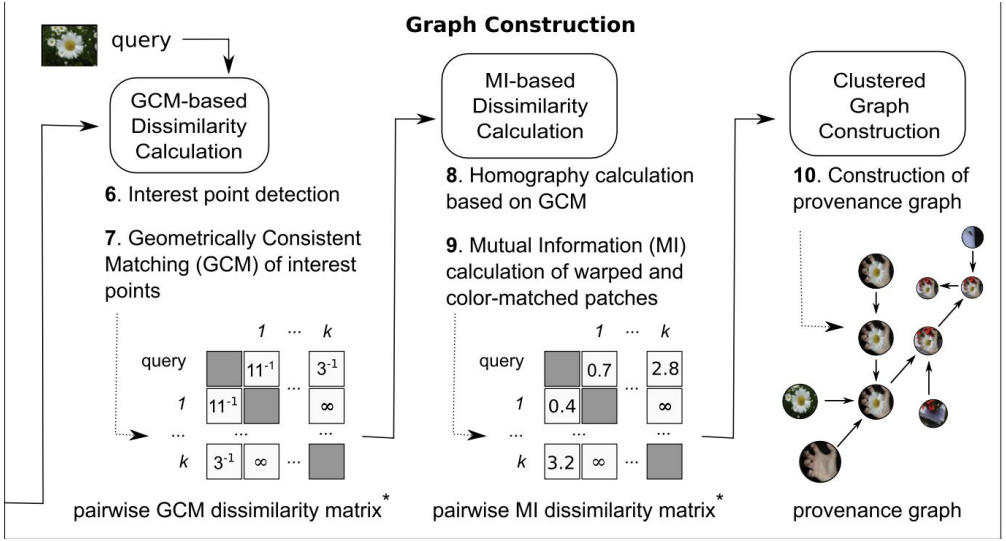
- Dissimilarity Matrices
  - Measure shared visual content
- Geometrically-Consistent Model
  - Measures shared geometric prop.
- Mutual-information Dissimilarity

variables.

$$MI(R_1, R_2) = \sum_{x \in R_1} \sum_{y \in R_2} p(x,y) \log\left(\frac{p(x,y)}{\sum_x p(x,y) \sum_y p(x,y)}\right), \quad (5)$$

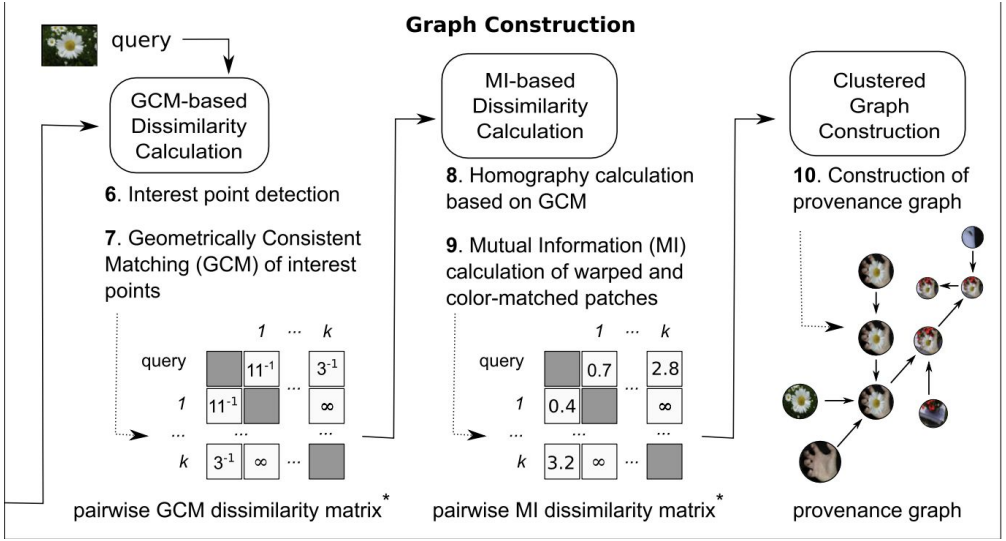
where $x \in [0, .., 255]$ refers to the pixel values of R_1, and $y \in [0, .., 255]$ refers to the pixel values of R_2. The $p(x, y)$ value regards the joint probability distribution function of R_1 and R_2. As explained in [36], it can be approximated by:

$$p(x,y) = \frac{h(x,y)}{\sum_{x,y} h(x,y)}, \quad (6)$$

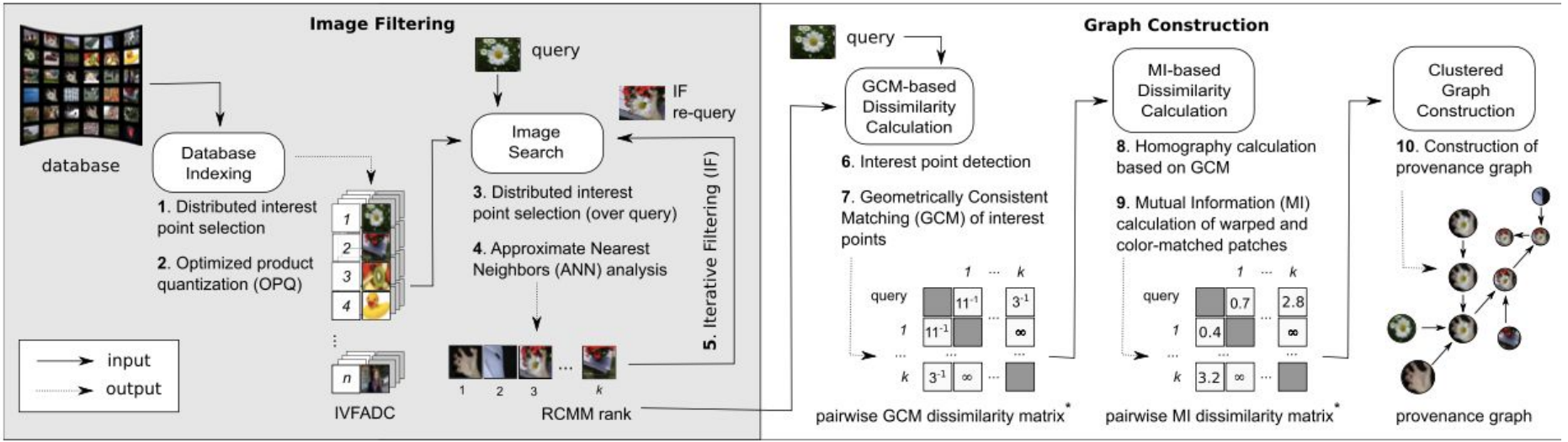


Provenance Graph Construction

- Using GCM and MI Construct graph
- If they share enough context add them to list



Provenance Analysis Methodology



Results

TABLE II

RESULTS OF PROVENANCE GRAPH CONSTRUCTION OVER THE NIST DATASET. WE REPORT THE AVERAGE VALUES ON THE PROVIDED 65 QUERIES

	Solution	VO	EO	VEO
End-to-end analysis	Kruskal-SURF [37]	0.638	0.429†	0.537†
	Kruskal-MSER	0.257	0.140†	0.199†
	Cluster-SURF	**0.853**	**0.353**	**0.613**
	Cluster-MSER	0.835	0.312	0.585
Oracle-filter analysis	**Kruskal-SURF** [37]	**0.933**	**0.256**†	**0.609**†
	Kruskal-MSER	0.902	0.239†	0.585†
	Cluster-SURF	0.931	0.124	0.546
	Cluster-MSER	0.892	0.123	0.525

†: Values for undirected edges. In bold, the solutions with the best VEO.

TABLE I

RESULTS OF PROVENANCE IMAGE FILTERING OVER THE NIST DATASET. WE REPORT THE AVERAGE VALUES ON THE PROVIDED 65 QUERIES

Solution	R@50	R@100	R@200	Query time (min)
KDF-SURF2k [15]	0.609	0.633	0.649	0.15
IVFADC-SURF2k	0.713	0.722	0.738	0.17
IVFADC-SURF5k	0.876	0.881	0.883	0.55
IVFADC-DSURF	0.882	0.895	0.899	0.54
IVFADC-SURF5k-IF	0.895	0.901	0.919	2.53
IVFADC-DSURF-IF	**0.907**	**0.912**	**0.923**	**2.20**

In bold, the solution with highest recall values.

Beyond Pixels: Image Provenance Analysis Leveraging Metadata

Bharati et al. 2019

Problem

We have reached a point where digital forgeries can be produced with fine-grained detail, down to photographic style and sensor noise.

These advancements in anti-forensics undermine the content's credibility, ownership, and authenticity.

The current scale at which images and videos are shared requires an automated way of answering such questions. Image processing and computer vision techniques can be employed to detect correspondences between images or other digital art forms (e.g., object matching in images and comparing the style/semantics).

Problem

Provenance analysis can be thought of as ordering pair similarities between multiple image pair sets, and is therefore a natural extension to pairwise image comparison.

However, due to the vast range of possible versions of a single original image, the metrics for quantifying the similarity between pairs of images can be noisy.

Matching difficulty can also arise within sets of near-duplicate images, which are generated from a single origin having undergone a series of transformations (e.g., crop → saturate → desaturate).

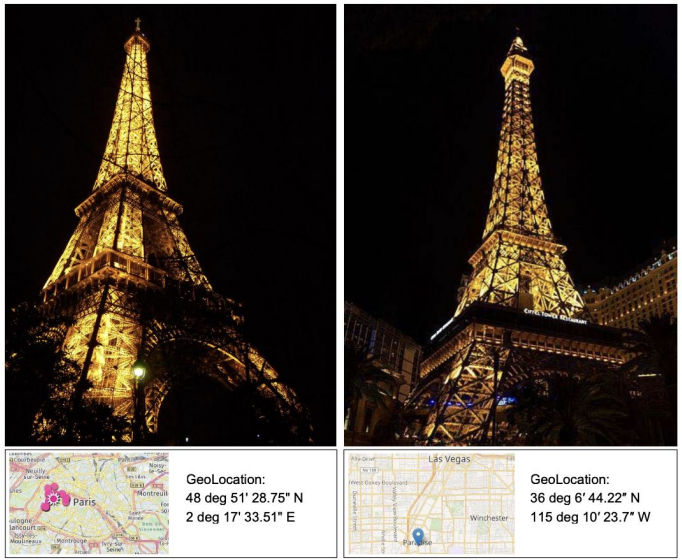


Figure 2. *Left*: Photo of the Eiffel Tower taken at night in Paris. *Right*: Photo of the replica of the monument in Las Vegas taken at night.

Proposed Method

To handle scenarios where image content fails to explain image evolution, file metadata can be used to fill in the gaps.

Image provenance analysis algorithms aim at constructing a provenance graph with related images, given a query image.

The provenance graph is a Directed Acyclic Graph (DAG) where each node corresponds to an image in the set of related images and the edges stand for the relationship of sharing duplicate content.

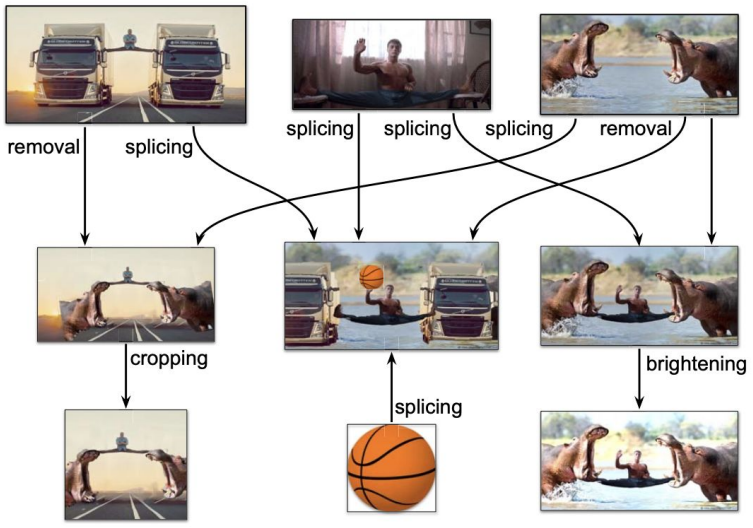


Figure 1. Example of an Image Provenance Graph (IPG) showing some common operations performed on images and how they are manifested in the case of provenance.

Proposed Method

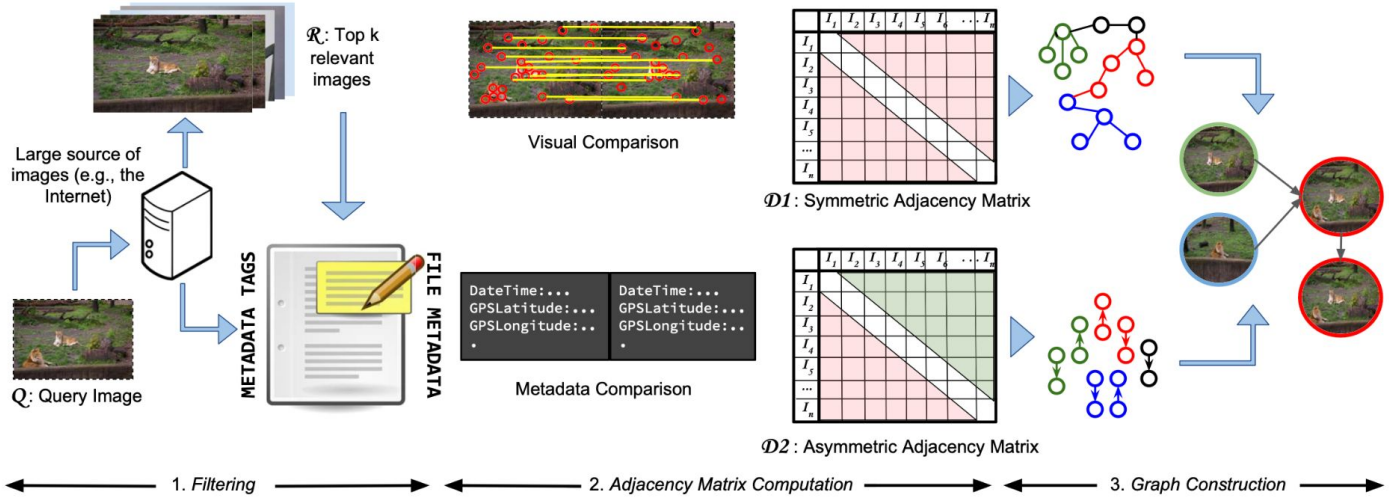


Figure 3. Stages of image provenance analysis. The proposed method starts with filtering images related to the provided query image Q. The 'k' most relevant images are selected for pairwise image comparison. This step is not present in an oracle scenario where we assume to have been provided with the perfectly correct set of 'k' related images. The images are compared in terms of visual content and metadata, yielding two types of adjacency matrices. The obtained matrices are then combined in the graph construction step to form an IPG.

Proposed Method

In this work, in order to incorporate metadata information, the authors introduce a heuristic-based normalized voting attribute weights to each pairwise image relationship.

The voting method is chosen as a complement to the similarity comparison in the visual domain. The heuristics used to obtain the scores for each pair are straight forward metadata-related assumption in the context of image provenance and rely upon the content of the tages. They include: Date, Location, Camera (Make, model, and software tags), Editing (Processing Software, Artist, Host Computer) and Thumbnail.

Proposed Method

These heuristics are used to generate a metadata-based image pairwise adjacency matrix M.

Once the vision-based and metadata-based adjacency matrices are available, one can either individually use them to directly generate a provenance graph, through, for example, the application of Kruskal's Maximum Spanning Tree (MST) algorithm, or as the authors propose, use a specialized algorithm for constructing a directed provenance graph, such as clustered provenance expansion.

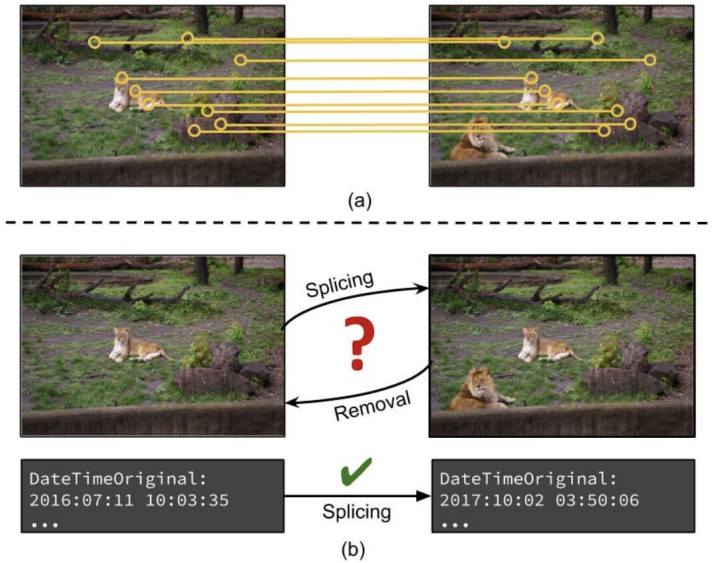


Figure 4. Usage of metadata information for determining direction in image pairwise provenance relationships. In (a), the output of interest-point-based analysis between two images is shown. The operation can be either a splicing or removal of the male lion. In (b), according to the date-based metadata, the operation is revealed to be a splice, since the image on the left is older.

Datasets & Training

NC2017-Dev1-Beta4 (from NIST): contains 65 queries and the ground-truth is released in the form of journals depicting provenance graphs. The graphs include links corresponding to simple image transformations (e.g., cropping, blurring, rotation, etc.) to complex ones (e.g., splicing from multiple sources and object removal).

Following the protocol proposed by NIST, the authors perform both *end-to-end* and *oracle-filter* provenance analysis over this dataset. End-to-end analysis requires performing provenance filtering prior to graph construction. Orthogonal to the *end-to-end* versus *oracle* comparison, the authors also compare results for both *metadata only* and *visual + metadata* solutions.

The provenance graphs generated using the proposed approach for both oracle and end-to-end scenarios are evaluated using the metrics proposed by NIST for the provenance task. The metrics focus on comparing the nodes and edges from both ground-truth and candidate graphs. The corresponding measures of Vertex Overlap (VO) and Edge Overlap (EO) are the harmonic mean of precision and recall (F1 score) for the nodes and edges retrieved by our method. In addition to these, a unified metric representing one score for the graph overlap namely the Vertex Edge Overlap (VEO) is also reported. The VEO is the combined F1 score for nodes and edges. All the metrics are computed through the NIST *MediScore* tool. The values of these metrics lie in the range [0, 1] where higher values are better.

Datasets & Training

Reddit Dataset: This dataset contains provenance cases created from images extracted from the *photoshopbattles* community on the Reddit. For the purpose of provenance, Moreira et al. 2018 utilize this comment structure to obtain 184 provenance graphs with an average graph order of 56. For the sake of fair comparison, the authors evaluate the variants of the proposed approach on the exact same set. The full set of images from Reddit do not contain distractors. This restricts the experiments for provenance analysis in this setting to *oracle-filter* analysis only, in contrast to the NC2017-Dev1- Beta4 dataset.

To evaluate the experimental results on the Reddit dataset, the authors employ the same metrics and scorer used in the case of the NC2017-Dev1-Beta4 dataset.

Experimental Results

The experiments performed on both datasets show that utilizing knowledge from metadata helps in the process of edge inference for provenance. As it can be observed from the values reported in Table 1, the proposed method significantly improves total edge overlap, and thereby total graph overlap, since it uses image-content-based information to initially establish connections between images, then relies on metadata to refine edge direction.

Table 1. Results of provenance graph construction over the NIST NC2017-Dev1-Beta4 dataset. We report the mean and the standard deviation for the metrics on the provided 65 queries. Visual results are from Moreira et al. [49]. Best results are in bold.

Data Modality	Solution	Oracle Filtering			End-to-End Analysis		
		VO	EO	VEO	VO	EO	VEO
Visual [49]	Cluster-SURF	0.931±0.075	0.124±0.166	0.546±0.096	0.853±0.157	0.353±0.236	0.613±0.163
	Cluster-MSER	0.892±0.154	0.123±0.161	0.525±0.129	0.835±0.180	0.312±0.252	0.585±0.177
Metadata	Kruskal	0.999±0.003	0.117±0.099	0.577±0.053	0.249±0.115	0.009±0.016	0.130±0.057
Visual + Metadata	Cluster-SURF	**0.931±0.075**	**0.445±0.266**	**0.699±0.148**	**0.853±0.157**	**0.384±0.248**	**0.628±0.169**
	Cluster-MSER	0.891±0.154	0.389±0.254	0.651±0.176	0.838±0.182	0.345±0.232	0.603±0.174

Experimental Results

Table 2. Ablation results for oracle and end-to-end provenance. We repeat the experiments seven times for the best solution presented in Table 1 (Visual + metadata, Cluster-SURF) in both scenarios, keeping only a subset of heuristics activated at a time. Best results in bold.

Heuristic	Oracle Filtering			End-to-End Analysis		
	VO	EO	VEO	VO	EO	VEO
Date only	**0.931±0.075**	**0.446±0.265**	**0.700±0.147**	0.853±0.157	0.389±0.244	0.630±0.169
Location only	0.931±0.075	0.394±0.282	0.674±0.154	0.853±0.157	0.348±0.241	0.611±0.164
Camera only	0.931±0.075	0.388±0.269	0.672±0.147	0.853±0.157	0.350±0.234	0.612±0.164
Editing only	0.931±0.075	0.396±0.281	0.675±0.153	0.853±0.157	0.353±0.237	0.613±0.163
Thumbnail only	0.931±0.075	0.411±0.285	0.683±0.155	0.853±0.157	0.363±0.238	0.618±0.167
All but Date	0.931±0.075	0.394±0.280	0.675±0.152	0.853±0.157	0.345±0.247	0.610±0.168
Date + Thumbnail	0.931±0.075	0.444±0.268	0.699±0.148	**0.853±0.157**	**0.391±0.245**	**0.632±0.169**

In the oracle scenario, while all five tags individually benefit graph EO, the date-based one performs best, followed by thumbnail usage.

Experimental Results

Table 3. Results of provenance graph construction over the Reddit dataset. We report the average values of the metrics over the 184 cases, as well as the standard deviations. This dataset only allows us to report oracle-filtering results. Visual results are from Moreira et al. [49]. Best results are in bold.

Solution	VO	EO	VEO
Visual [49]:			
Cluster-SURF	0.757±0.341	0.037±0.034	0.401±0.181
Cluster-MSER	0.509±0.388	0.027±0.034	0.271±0.207
Metadata:			
Kruskal	**0.969±0.073**	0.034±0.086	**0.506±0.056**
Visual + Metadata:			
Cluster-SURF	0.757±0.341	**0.085±0.065**	0.424±0.193
Cluster-MSER	0.509±0.388	0.061±0.063	0.288±0.220

Provenance analysis becomes significantly more difficult when dealing with real-world scenarios, such as those presented in the Reddit dataset. Although metadata doubles the number of correctly retrieved edges, as seen in Table 3, the edge overlap is still much lower than for the NC2017- Dev1-Beta4 dataset.

Key Takeaways

- This work only presents a preliminary exploration of utilizing metadata in provenance analysis. While the results show improvement, metadata-based approaches have higher chances of being rendered unreliable due to their absence or manipulation.
- Further advancements in solving the problem must focus on the examination of content-derived metadata as well. Future work could include estimating missing metadata information from the content and available tags.
- For now, the findings suggest that image-content-based methods should be the fallback option, as metadata alone is more useful for determining edge directions instead of edge selection.

Transformation-Aware Embeddings for Image Provenance

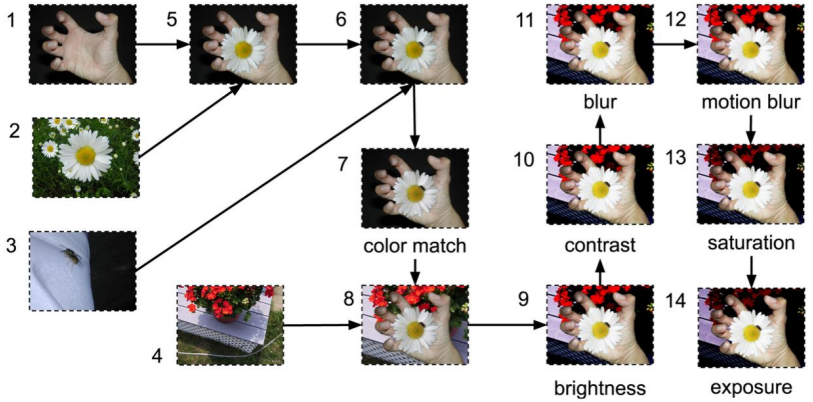
Bharati et al. 2021

Overview

“a step-by-step analysis of how the current version of a manipulated image or video was generated helps us in answering more holistic and contextual questions than just whether it is real or fake.”

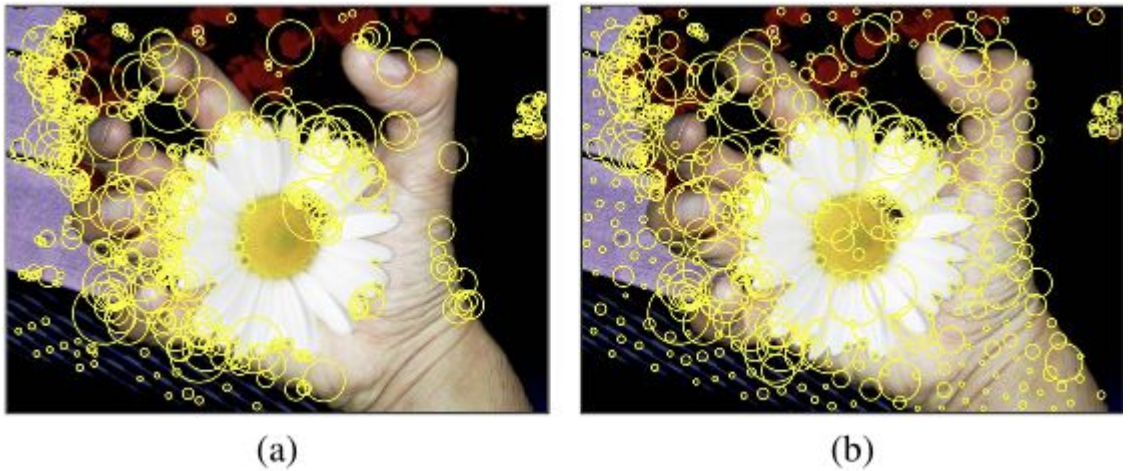
Types of Provenance Graph

- Content-based
  - “creation of composite images using image splicing or removal of content”
  - Well studied
- Transformation-based
  - “created from one another through operations such as simple geometric or color-based transforms”
  - Not well studied



Dissimilarity Matrix

- In the past handcrafted SURF and MSER vectors have been used
- “a novel data-driven framework to learn embeddings useful for provenance analysis”

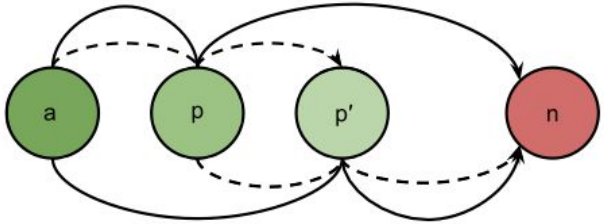


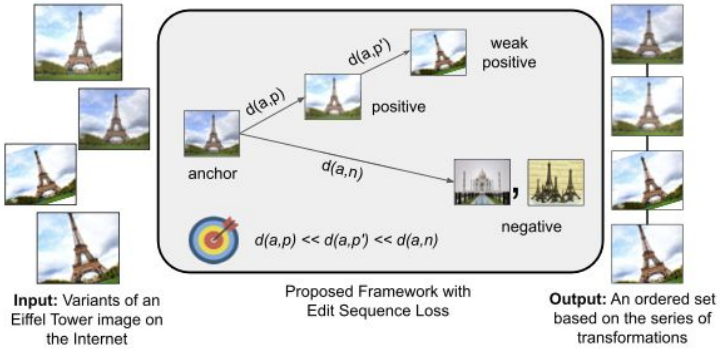
Goal

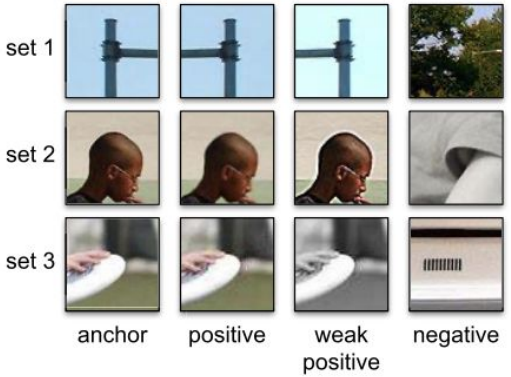
“improving the fidelity of reconstructing the chains of globally-related images in the provenance graphs by encoding this awareness in the first stages of graph construction”

Approach

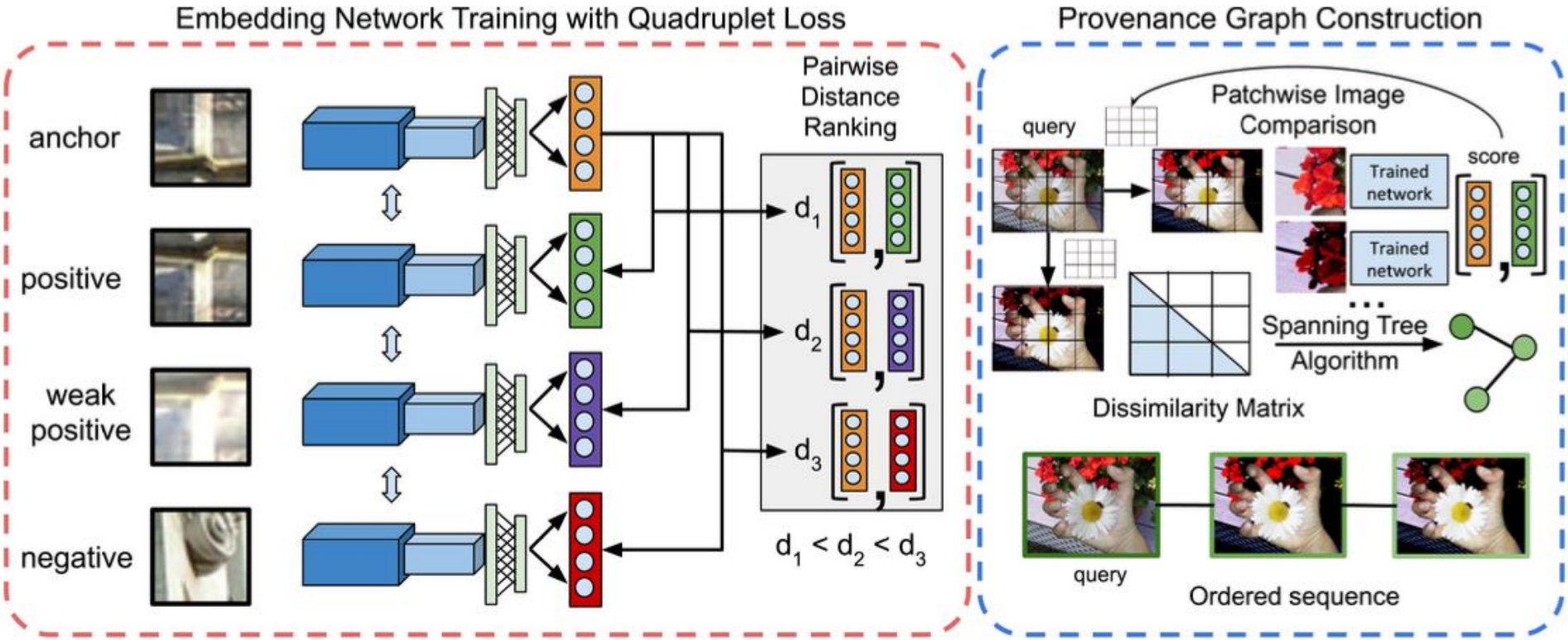
- “deep distance learning”
  - Learning how dissimilar an image
  - Resistant to transformation based changes







Approach


Embedding Network Training with Quadruplet Loss
anchor
positive
weak positive
negative
Pairwise Distance Ranking
d_1
d_2
d_3
$d_1 < d_2 < d_3$
Provenance Graph Construction
query
Patchwise Image Comparison
Trained network
Trained network
score
Spanning Tree Algorithm
Dissimilarity Matrix
query
Ordered sequence

Results

TABLE II

PROVENANCE GRAPH CONSTRUCTION OVER THE NC2017-DEV1-BETA4 DATASET (ORACLE MODE). WE REPORT THE MEAN AND THE STANDARD DEVIATION OF 65 CASES FOR THE METRICS PRESENTED IN SEC. IV. IN BOLD: BEST RESULTS. TAE STANDS FOR TRANSFORMATION AWARE EMBEDDINGS LEARNED USING THE PROPOSED APPROACH. N/A STANDS FOR "NOT APPLICABLE"

Description Method	Description Type	Local Feature Type	Feature Vector Size (#)	Disk Size (MB)	VO	EO	VEO
SURF [3], [13]	handcrafted	keypoints	64	851	0.90 (±0.08)	0.65 (±0.16)	0.78 (±0.11)
LIFT [23]	learned	keypoints	128	89	0.79 (±0.18)	0.39 (±0.23)	0.60 (±0.19)
DELF [58]	learned	keypoints	40	205	0.86 (±0.18)	0.59 (±0.23)	0.73 (±0.19)
DeepMatching [26]	handcrafted	keypoints	N/A	230	0.59 (±0.37)	0.28 (±0.25)	0.44 (±0.30)
AlexNet [59]	learned	image patches	4096	19000	1.00 (±0.00)	0.64 (±0.15)	0.83 (±0.08)
DeepRanking [14]	learned	image patches	4096	19000	1.00 (±0.00)	0.62 (±0.17)	0.82 (±0.08)
ResNet-18 [60]	learned	image patches	512	2400	1.00 (±0.00)	0.65 (±0.17)	0.83 (±0.08)
ForSim [48]	learned	image patches	N/A	less than 1	1.00 (±0.00)	0.30 (±0.15)	0.66 (±0.08)
TAE (ours)	**learned**	**image patches**	**256**	**1200**	**1.00 (±0.00)**	**0.68 (±0.15)**	**0.85 (±0.07)**

TABLE V

PROVENANCE GRAPH CONSTRUCTION OVER THE MFC19-EVAL-PART1 DATASET. IN THE FIRST THREE ROWS, THE PERFORMANCE OF THE OFFICIAL PARTICIPANTS OF THE MEDIA FORENSICS PROGRAM IN 2019, IN AN END-TO-END (HENCE MORE CHALLENGING) SCENARIO. IN THE LAST ROW, THE RESULTS OF THE PROPOSED SOLUTION (ORACLE SCENARIO). MEAN AND STANDARD DEVIATION OF 1025 CASES ARE REPORTED FOR THE METRICS PRESENTED IN SEC. IV, EXCEPT FOR THE PARTICIPANTS' SUBMISSIONS, WHOSE STANDARD DEVIATIONS WERE NOT AVAILABLE

Description	VO	EO	VEO
Submission #2038	0.83	0.58	0.72
Submission #2039	0.70	0.48	0.60
Submission #2044	0.72	0.13	0.48
TAE (ours)	0.97 (±0.08)	0.69 (±0.20)	0.84 (±0.12)

Open Questions

- What other datasets exist?
- Why was SURF used over DELF?
- Is the activation of region based on the single or composite composition of objects?

The end