U-Phylogeny: Undirected Provenance Graph Construction

A. Bharati[1], D. Moreira[1], A. Pinto[2], J. Brogan[1], K. Bowyer[1], P. Flynn[1], W. Scheirer[1] and A. Rocha[1,2]

[1]University of Notre Dame, USA   [2]University of Campinus, Brazil

Problem

Provenance: tracing the place of origin or earliest known history of something

Why solve Image Provenance?
- Existence of multiple copies of the same media
- Prevalence of photo manipulation in popular culture
- Which is the original image and where did it come from?

Limitations of Existing Methods
- No multi-composite scenarios
- Only works in constrained environments
- Not much exploration of dissimilarity computation
- Not invariant to many transformations
- Unavailability of sufficient information to perform mappings (compression artifacts for non-JPEG images)

What is U-Phylogeny*?
- Undirected provenance graph construction method that can be extended to obtain directed phylogenetic tree
- Generalized and more efficient method for > 2 donors without image format restrictions
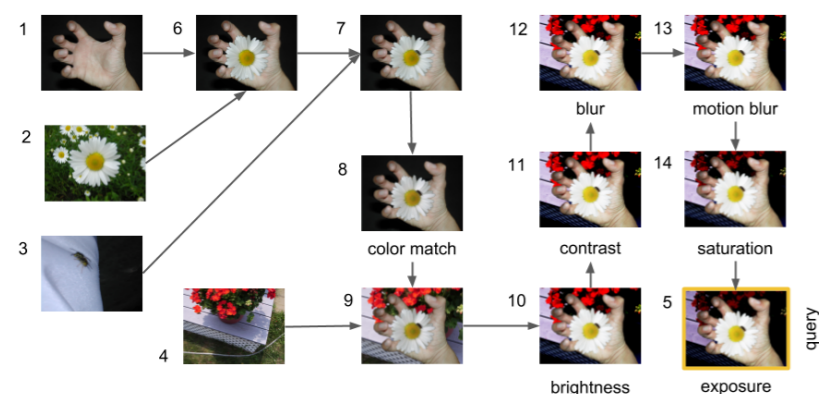
* Bharati, A., Moreira, D., Pinto, A., Brogan, J., Bowyer, K., Flynn, P., Scheirer, W. and Rocha, A., 2017. U-Phylogeny: Undirected Provenance Graph Construction in the Wild. IEEE International Conference on Image Processing, 2017. Code: https://gitlab.com/notredame-provenance/u-phylogeny

Methodology

Pipeline:
1. Detection and description of points of interest (SURF)
2. Keypoint matching using L2 distance
3. Match filtering
    - Geometrically Consistent Matching (GCM)
4. Dissimilarity computation
    - Number of GCM's, Distance of GCM's
    - Difference in pixel distribution(MSE, Mutual Information) (extended U-phylogeny)
5. Kruskal's min/max spanning tree algorithm
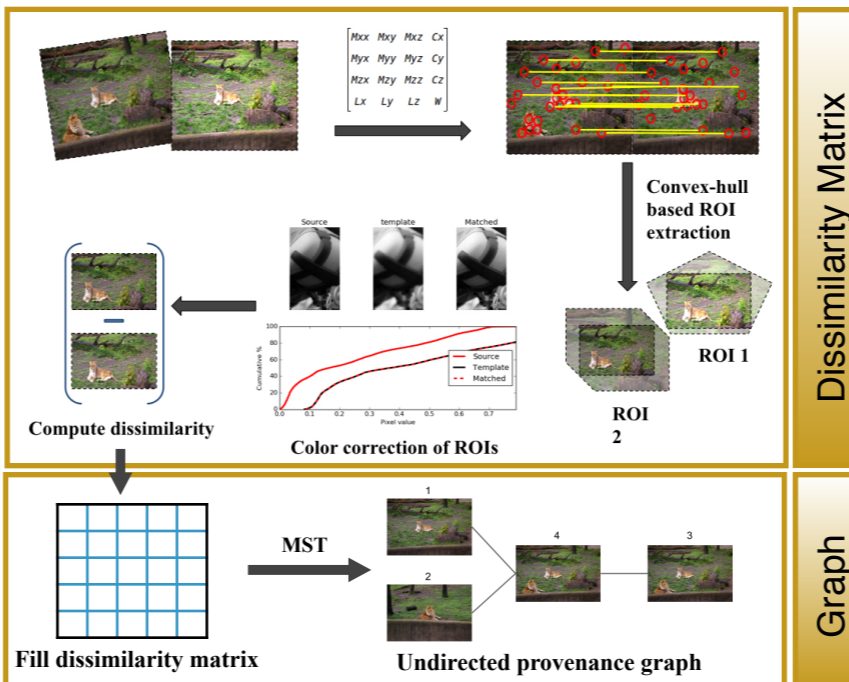
Dataset Used: 59 provenance cases from NIST's Nimble 2017 challenge dataset
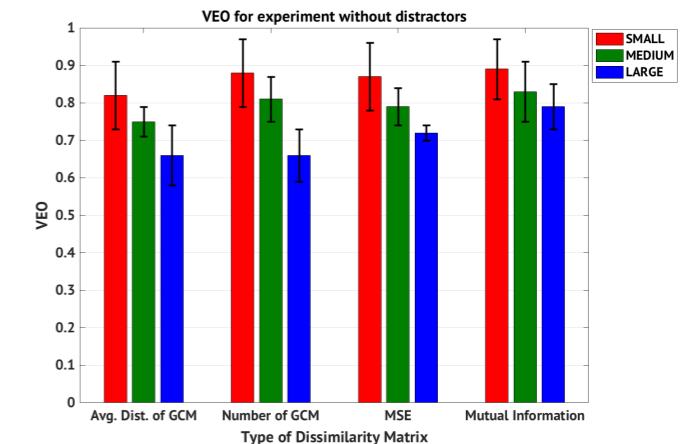- 59 queries and 10446 gallery images (750 related)
- Average graph order = 12.7, Range - [3, 82]
- Small (<=12 nodes) Medium (13-20 nodes) Large (>20 nodes)

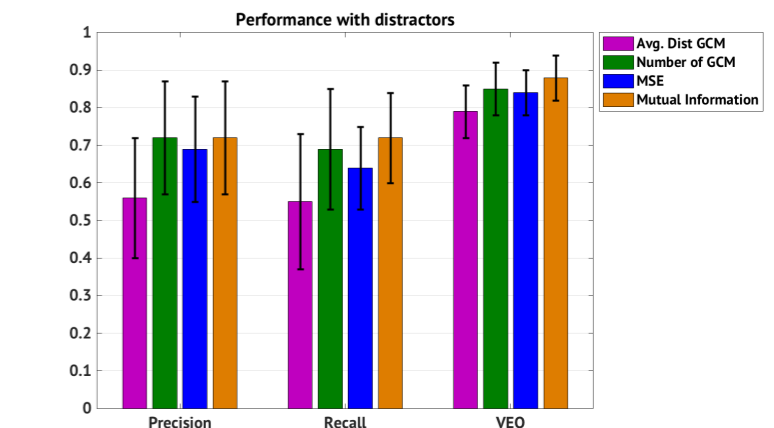### Evaluation Metrics: web graph similarity metrics

$$VEO(G,G') = \frac{|nodes(G') \cap nodes(G)| + |edges(G') \cap edges(G)|}{|nodes(G')| + |nodes(G)| + |edges(G')| + |edges(G)|}$$

Results

Experiment 1: Without distractors

Experiment 2: With distractors (25 images/case)

Conclusion
- Effectiveness decreases with the size of graphs irrespective of dissimilarity metrics
- Extended U-Phylogeny performs better on average but reduced efficiency
- U-Phylogeny is robust to the presence of distractors (VEO = 88%)
- Mutual Information yields better performance than MSE