Image Phylogeny Tree Reconstruction Based on Region Selection

P. Bestagini, M. Tagliasacchi, S. Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy

1-Introduction

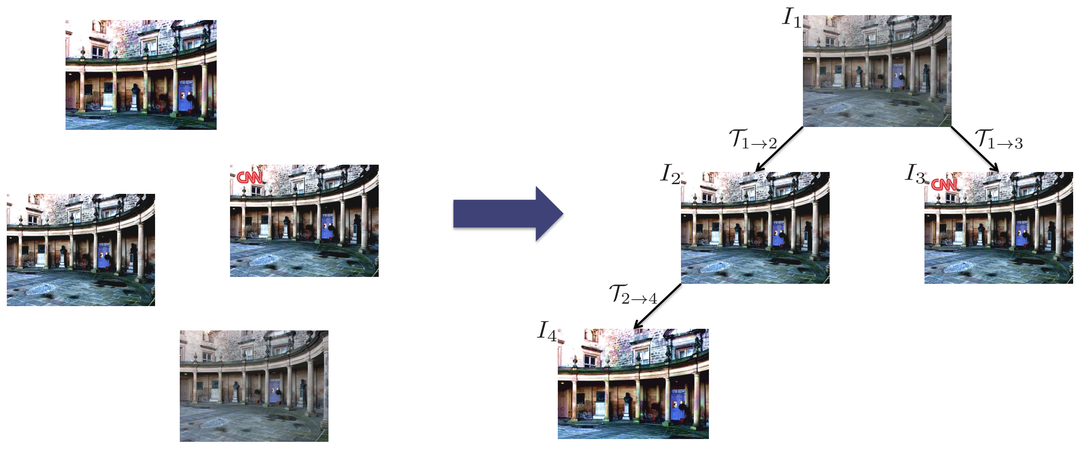
Nowadays, everyone can download, edit and republish any picture on the web, thus contributing to the diffusion of near-duplicate (ND) images. Studying the relationship between ND images enables to gain an interesting insight on the way NDs are distributed online.
In this work, we propose an algorithm to reconstruct ND images relationships even when local editing operations are applied.

2-Problem formulation

Problem
Given a set of ND images, reconstruct the image phylogeny tree (IPT).

Image Phylogeny Tree (IPT)
The IPT is an acyclic directed graph representing parental relationships.

- Each node represents a near-duplicate image.
- Each branch represents a transformation, or a combination of them.



3-IPT reconstruction

Rationale
If I_n has been generated from I_m it is possible to find a transformation $\mathcal{T}_{m\to n}$ to map I_m to I_n.

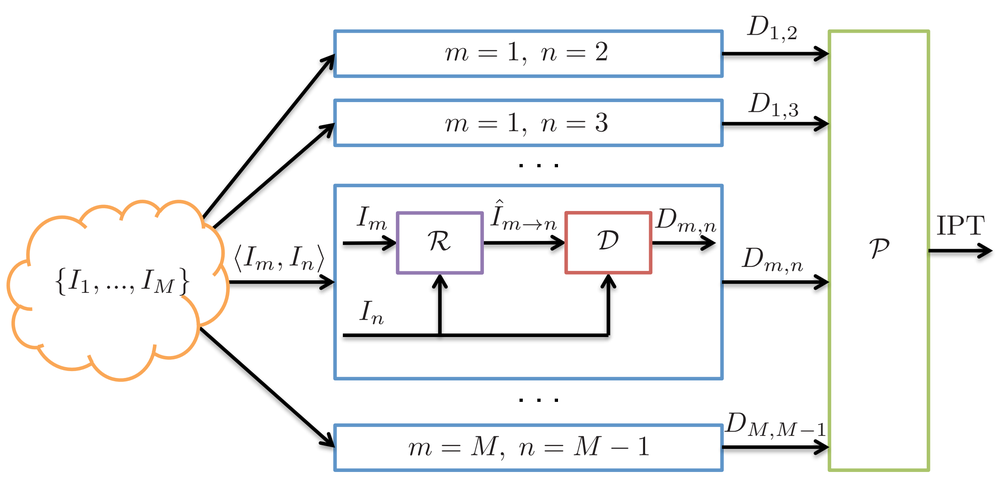
Pipeline

- Image registration ($\mathcal{R}$): search for the transformation $\hat{\mathcal{T}}_{m\to n}$ that possibly maps I_m to I_n.
- Dissimilarity computation ($\mathcal{D}$): evaluate the goodness of the estimated $\hat{\mathcal{T}}_{m\to n}$ computing image dissimilarity

$$D_{m,n} = \frac{1}{XY} \sum_{(x,y)} |R_{m,n}(x,y)|^2,$$

where $R_{m,n} = \hat{\mathcal{T}}_{m\to n}(I_m) - I_n$.

- IPT reconstruction ($\mathcal{P}$): dissimilarity matrix D can be interpreted as a fully connected graph. Several tree reconstruction algorithms [1] can be used to generate an IPT from D.



4-Region selection

Motivation

- $\hat{\mathcal{T}}_{m\to n}$ is typically estimated considering only global transformations (i.e., compression, colour and geometric).
- When local transformations are used (e.g., logo addition) dissimilarity is negatively affected.

Proposed solution

- Compute **local affected** $R_{m,n} = \hat{\mathcal{T}}_{m\to n}(I_m) - I_n$.
- Compute **non local affected** $\hat{R}_{m,n} = \hat{\mathcal{T}}_{m\to n}(\hat{\mathcal{T}}_{n\to m}(I_n)) - I_n$.
- Compare $R_{m,n}$ and $\hat{R}_{m,n}$ to estimate the non-locally manipulated region $\mathcal{M}_n$.
- Compute dissimilarity only on the non-locally manipulated region

$$D^{\text{loc}}_{m,n} = \frac{1}{|\mathcal{M}_n|} \sum_{(x,y)\in\mathcal{M}_n} |R_{m,n}(x,y)|^2.$$

5-Experiments

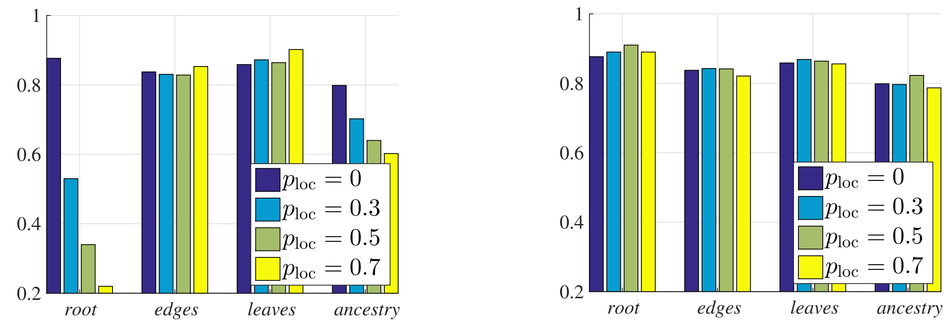
Dataset

- Different IPTs changing number of nodes (i.e., 10 or 20), probability p_{loc} of adding logos (i.e., 0, 0.3, 0.5, 0.7) and topology for a total number of 12000 images.

Region estimation

p_{FA}	Morph.	TPR	TNR	FPR	FNR
0.02	☐	0.957	0.961	0.039	0.043
	☑	0.978	0.978	**0.022**	0.022
0.04	☐	0.966	0.932	0.068	0.034
	☑	0.984	0.962	**0.038**	0.016
0.06	☐	0.970	0.908	0.092	0.030
	☑	0.987	0.947	**0.053**	0.013

IPT reconstruction



[baseline] [proposed]

References

[1] D. Zanoni, G. Siome, R. Anderson, "Exploring heuristic and optimum branching algorithms for image phylogeny", Journal of Visual Communication and Image Representation (JVCIR), 2013.