

La phylogénie des images dans les réseaux sociaux

Noé LE PHILIPPE

Équipe ICAR - William Puech

17 mars 2016

Sommaire

- 1 Le sujet de stage
- 2 État de l'art
- 3 Analyse des recompressions
- 4 Notre approche

Le sujet de stage

Le sujet

La phylogénie des images dans les réseaux sociaux

Définition

“La phylogenèse ou phylogénie est l'étude des relations de parenté entre êtres vivants.”

— Wikipedia

Les applications

Réduire le nombre de versions de la même image pour optimiser l'espace de stockage

Les applications

Réduire le nombre de versions de la même image pour optimiser l'espace de stockage

Suivre l'évolution et la diffusion d'images sur les réseaux sociaux

Les applications

Réduire le nombre de versions de la même image pour optimiser l'espace de stockage

Suivre l'évolution et la diffusion d'images sur les réseaux sociaux

Détecter l'altération d'images

Définitions

Near-Duplicate Image (NDI)

Une image I_1 est le near-duplicate^[1] d'une image I si :

$$I_1 = T(I), T \in \mathcal{T}$$

où \mathcal{T} est un ensemble de transformations autorisées

Dans le cas général,

$$\mathcal{T} = \{ \text{resampling, cropping, affine warping,} \\ \text{color changing, lossy compression} \}$$

mais dans le cadre du stage, $\mathcal{T} = \{ \text{lossy compression} \}$

1. Alexis Joly, Olivier Buisson et Carl Frélicot. "Content-based copy retrieval using distortion-based probabilistic similarity search". In : *Multimedia, IEEE Transactions on* 9.2 (2007), p. 293–306.

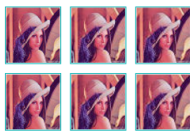
Définitions

Image Phylogeny Tree (IPT)

C'est l'arbre retraçant la parenté des images



Image phylogeny tree



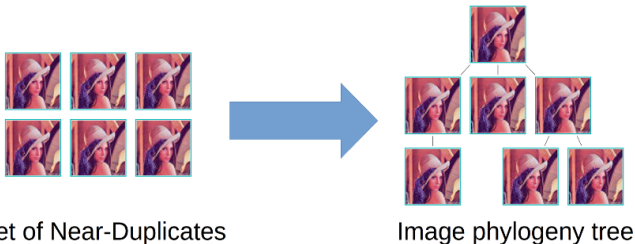
Set of Near-Duplicates



Image phylogeny tree

Deux parties importante lors de la reconstruction de l'arbre phylogénétique :

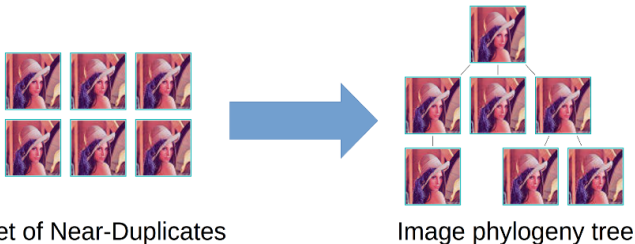
Image phylogeny tree



Deux parties importante lors de la reconstruction de l'arbre phylogénétique :

- Correctement identifier la racine

Image phylogeny tree



Deux parties importante lors de la reconstruction de l'arbre phylogénétique :

- Correctement identifier la racine

- Estimer au mieux l'arborescence

Sommaire

- 1 Le sujet de stage
- 2 État de l'art**
- 3 Analyse des recompressions
- 4 Notre approche

Du plus ancien

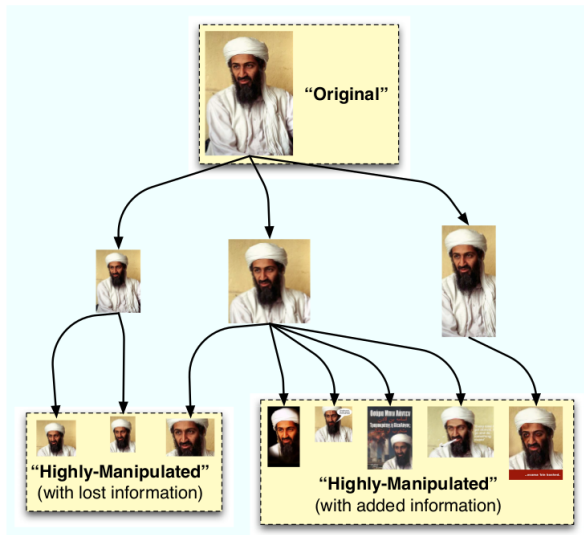
Visual Migration Map

- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

[2]

2. [Lyndon Kennedy et Shih-Fu Chang](#). "Internet image archaeology : automatically tracing the manipulation history of photographs on the web". In : *Proceedings of the 16th ACM international conference on Multimedia*. ACM. 2008, p. 349–358.

Du plus ancien



Du plus ancien

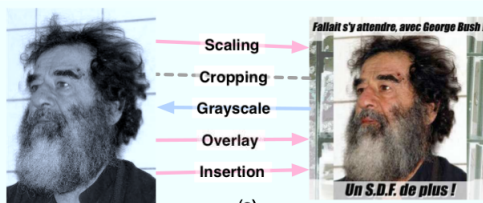
Visual Migration Map

- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

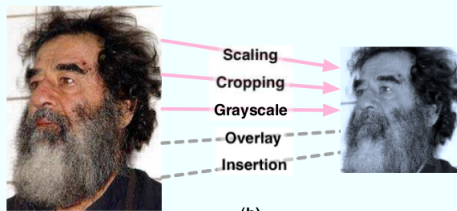
[2]

2. [Lyndon Kennedy et Shih-Fu Chang](#). "Internet image archaeology : automatically tracing the manipulation history of photographs on the web". In : *Proceedings of the 16th ACM international conference on Multimedia*. ACM. 2008, p. 349–358.

Du plus ancien



Inconsistent directions from individual manipulations.
(Neither image is parent)



Consistent directions from individual manipulations.
(Left image is parent of right)

Du plus ancien

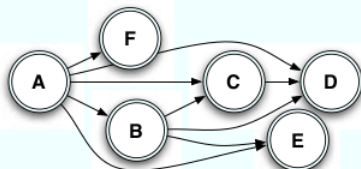
Visual Migration Map

- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

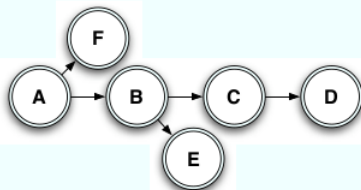
[2]

2. Lyndon Kennedy et Shih-Fu Chang. "Internet image archaeology : automatically tracing the manipulation history of photographs on the web". In : *Proceedings of the 16th ACM international conference on Multimedia*. ACM. 2008, p. 349–358.

Du plus ancien



(a) All Plausible Edits



(b) Simplified Structure

Au plus récent

Image phylogeny tree

- Calcul d'une *dissimilarity matrix*
- Calcul d'un arbre couvrant de poids min (Kruskal ou autre)

[2] [3]

2. Zanoni Dias, Anderson Rocha et Siome Goldenstein. "First steps toward image phylogeny". In : *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*. IEEE. 2010, p. 1–6.

3. Zanoni Dias, Anderson Rocha et Siome Goldenstein. "Image phylogeny by minimal spanning trees". In : *Information Forensics and Security, IEEE Transactions on* 7.2 (2012), p. 774–788.

Sommaire

- 1 Le sujet de stage
- 2 État de l'art
- 3 Analyse des recompressions**
- 4 Notre approche

Convergence des blocs lors de compressions successives

But

Compter le nombre de compressions

3 types de blocs

- Les blocs plats
- Les blocs stables
- Les blocs cycliques

Comment ?

plus petit commun multiple de la longueur des cycles

[4]

4. Matthias Carnein, Pascal Schöttle et Rainer Böhme. "Telltale Watermarks for Counting JPEG Compressions". In : *Proceedings of the Electronic Imaging 2016*. Publication status : Published. San Francisco, USA, 2016.

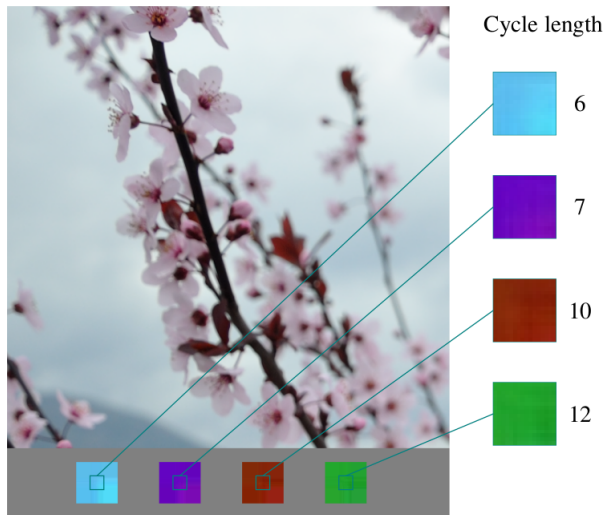
Convergence des blocs lors de compressions successives

Utilisation des blocs

- Les blocs de l'image
- Insérer des blocs

5. Matthias Carnein, Pascal Schöttle et Rainer Böhme. "Telltale Watermarks for Counting JPEG Compressions". In : *Proceedings of the Electronic Imaging 2016*. Publication status : Published. San Francisco, USA, 2016.

Convergence des blocs lors de compressions successives



[5]

5. Matthias Carnein, Pascal Schöttle et Rainer Böhme. "Telltale Watermarks

Convergence des blocs lors de compressions successives

Utilisation des blocs

- Les blocs de l'image
- Insérer des blocs

Les inconvénients

- Nécessite du padding
- Limité à la même table de quantification
- Résultats moyens pour $Q < 100$

[5]

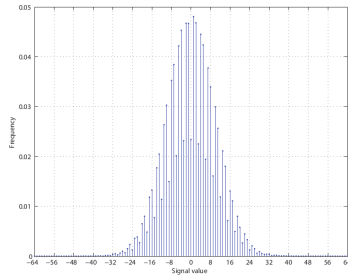
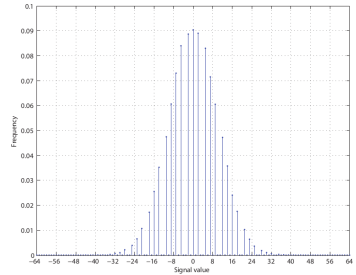
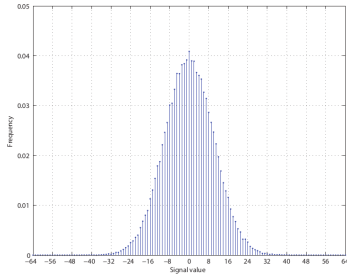
5. Matthias Carnein, Pascal Schöttle et Rainer Böhme. "Telltale Watermarks for Counting JPEG Compressions". In : *Proceedings of the Electronic Imaging 2016*. Publication status : Published. San Francisco, USA, 2016.

Estimation de la matrice de compression primaire

Analyse des valeurs manquantes de l'histogramme

Artefacts distincts pour $Q^1 > Q^2$ et $Q^1 < Q^2$

Estimation de la matrice de compression primaire



Estimation de la matrice de compression primaire

Analyse des valeurs manquantes de l'histogramme

Artefacts distincts pour $Q^1 > Q^2$ et $Q^1 < Q^2$

Limites

- $Q^1 = Q^2$
- Q^1 est facteur de Q^2

Estimation de la matrice de compression primaire

Principe de leur méthode

Comparer l'histogramme de l'image originale et l'histogramme des images compressées avec des tables de quantification modèles puis compressées avec Q^2 et enfin garder la table pour laquelle la différence entre histogramme est la plus faible

Sommaire

- 1 Le sujet de stage
- 2 État de l'art
- 3 Analyse des recompressions
- 4 Notre approche**

Notre approche

Matrice de parenté

- Tentative de preuve qu'une image n'est pas le parent d'une autre
- Si c'est impossible, l'image doit alors être le parent
- Extraction d'une *matrice de parenté*
- Calcul de l'arbre

Calcul de l'IPT

Data: M a $n \times n$ parentage matrix

Result: the root of the tree

```
1 nextRoot  $\leftarrow$  row with min sum of elements;
2 treeRoot  $\leftarrow$  nextRoot;
3 forall the rows row of  $M$  do
4   root  $\leftarrow$  nextRoot;
5   mark root as done;
6   for  $i \leftarrow 0$  to  $n$  do
7     row[ $i$ ]  $\leftarrow 0$ ;
8     if sum of elements of row == 0 then
9       | add  $i$  as child of root;
10    end
11    if row has the smallest sum of elements and is not marked
       as done then
12      | nextRoot  $\leftarrow i$ ;
13    end
14  end
15 end
16 return treeRoot
```

La suite

La problématique

Identifier un ensemble de marqueurs qui permettraient de réfuter qu'une image est le parent d'une autre

Les pistes

- Distance entre les histogrammes des coefficients DCT
- Valeurs manquantes à cause des compressions successives

Point clé

Réduction d'un problème de reconstruction d'un arbre de phylogénie à un problème de négation de parenté

La suite

