

# La phylogénie des images dans les réseaux sociaux

Noé LE PHILIPPE

Équipe ICAR - William Puech

15 juin 2016

# Sommaire

- 1 Présentation
- 2 Le sujet de stage
- 3 État de l'art
- 4 Notre approche

# Qui suis-je ?

Mon nom

Noé Le Philippe - 22 ans

# Qui suis-je ?

## Mon nom

Noé Le Philippe - 22 ans

## Formations

- Master IMAGINA à l'Université de Montpellier ( ?)
- Licence Informatique à l'Université de Montpellier
- DUT Informatique à l'Université de Montpellier
- Baccalauréat en 2011

# Qui suis-je ?

## Mes projets

- 2016 - shifumi / compresseur / moteur de jeu
- 2015 - CDD à We Are Learning
- 2015 - (TER) Développement d'un jeu sur mobile (disponible sur le playstore) avec Unity
- 2014 - (TER) Développement d'un éditeur XML avec Qt
- 2013 - Stage puis CDD au LIRMM dans l'équipe SMILE pour la création d'un jeu sérieux de rééducation de patients post-AVC

# Sommaire

- 1 Présentation
- 2 Le sujet de stage**
- 3 État de l'art
- 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<sup>[1]</sup> 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,

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

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

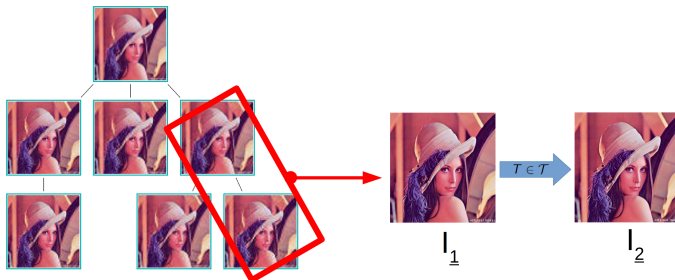
---

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.

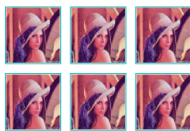
# Définitions

## Image Phylogeny Tree (IPT)

C'est l'arbre retraçant la parenté des images



# Image phylogeny tree



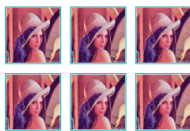
Set of Near-Duplicates



Image phylogeny tree

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

# Image phylogeny tree



Set of Near-Duplicates

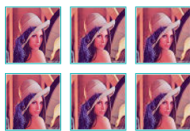


Image phylogeny tree

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

- Correctement identifier la racine

# Image phylogeny tree



Set of Near-Duplicates



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 Présentation
- 2 Le sujet de stage
- 3 État de l'art**
- 4 Notre approche



# Du plus ancien

## Visual Migration Map

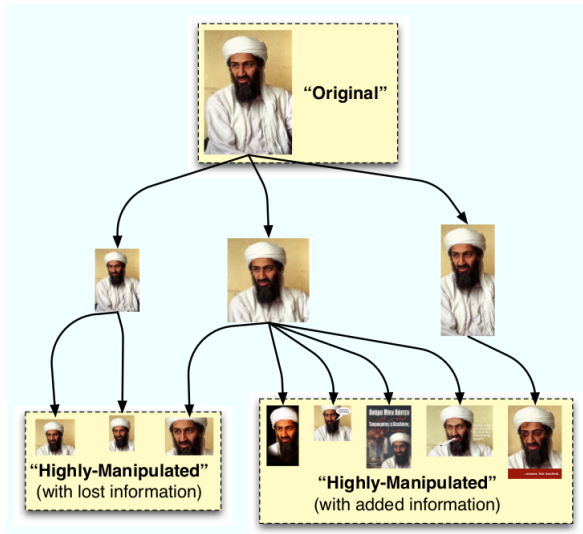
- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

[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

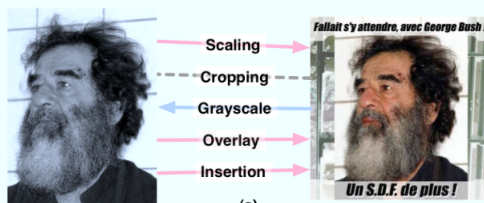
- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

[2]

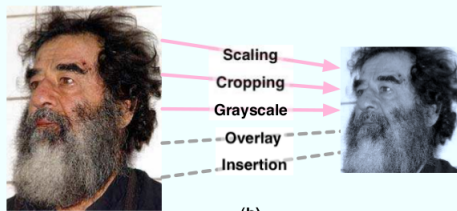
---

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

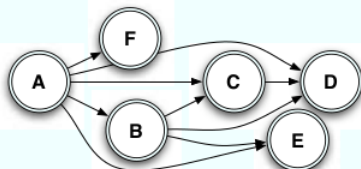
- Les transformations sont directionnelles
- Relation parent-enfant si tous les détecteurs s'accordent sur la direction
- Simplification du graphe par sélection des plus longs chemins

[2]

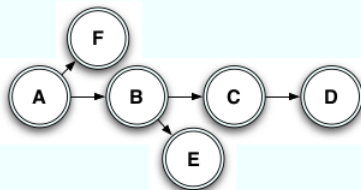
---

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 Présentation
- 2 Le sujet de stage
- 3 État de l'art
- 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 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

