

# La phylogénie des images dans les réseaux sociaux

Noé LE PHILIPPE

Équipe ICAR - William Puech

13 juin 2016

# Sommaire

- 1 Introduction
- 2 État de l'art
- 3 Notre approche
- 4 Résultats
- 5 Conclusion

# 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

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

Détecter l'altération d'images

# Définitions

## Near-Duplicate Image (NDI)

Une image  $I_n$  est le near-duplicate<sup>[1]</sup> d'une image  $I_m$  si :

$$I_n = T(I_m), T \in \mathcal{T}$$

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

Dans le cas général,

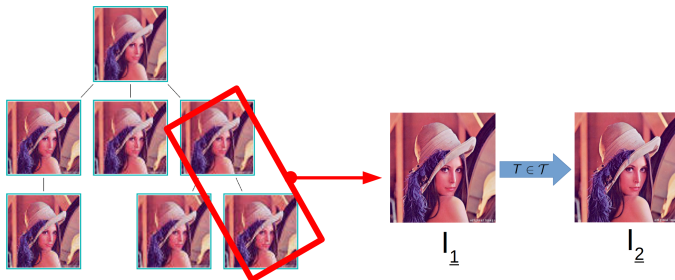
$$\mathcal{T} = \{\textit{resampling}, \textit{cropping}, \textit{affine warping}, \\ \textit{color changing}, \textit{lossy compression}\}$$

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

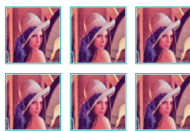
# Définitions

## Image Phylogeny Tree (IPT)

C'est l'arbre retraçant la parenté des images



# Image phylogeny tree



Set of Near-Duplicates

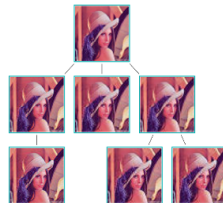


Image phylogeny tree

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

- Correctement identifier la racine

- Estimer au mieux l'arborescence

# Sommaire

- 1 Introduction
- 2 État de l'art**
- 3 Notre approche
- 4 Résultats
- 5 Conclusion

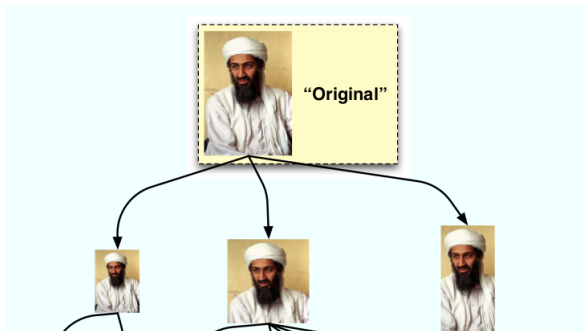


# Estimation de l'arbre de phylogénie

## 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]



# Estimation de l'arbre de phylogénie

## Image phylogeny tree

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

[3] [4]

- 
3. dias2010first.
  4. dias2012image.

# 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

[5]

# 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  $Q < 100$



Cycle length



6



7

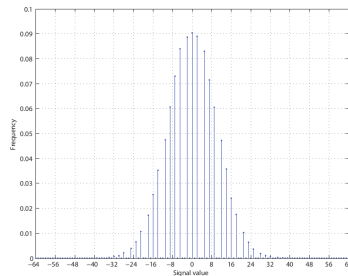
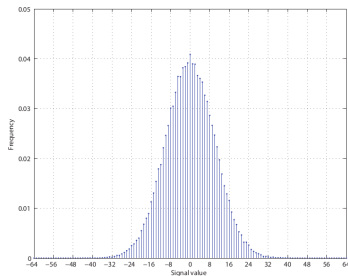
# Estimation de la matrice de compression primaire

## 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

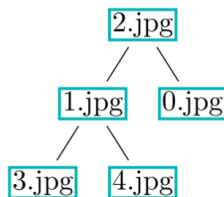
# Sommaire

- 1 Introduction
- 2 État de l'art
- 3 Notre approche**
- 4 Résultats
- 5 Conclusion

# Matrice de parenté

Matrice binaire de taille  $n \times n$

Construction de l'arbre à partir de la matrice



(a) Arbre de phylogénie

-	$I_0$	$I_1$	$I_2$	$I_3$	$I_4$
$I_0$	-	0	0	0	0
$I_1$	0	-	0	1	1
$I_2$	1	1	-	1	1
$I_3$	0	0	0	-	0
$I_4$	0	0	0	0	-

(b) Matrice de parenté



# Notre approche

## Marqueur

Caractéristique de l'image qui indique qu'une certaine opération a été effectuée et qui va se transmettre aux enfants

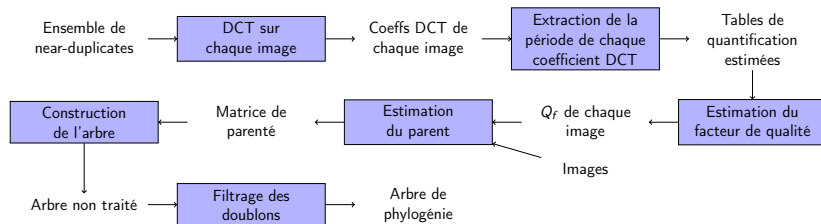
## Fonction de négation

$f(I_m, I_n)$  est une fonction qui pour tout couple d'images  $(I_m, I_n)$  détecte à chaque fois qu'il est présent un marqueur visible dans une image et pas dans l'autre, et donc prouve qu'il n'y a pas de relation de parenté entre  $I_m$  et  $I_n$ .

## Théorème

Pour tout couple d'images  $(I_m, I_n)$  d'un ensemble de near-duplicates, s'il n'existe pas de marqueur prouvant que  $I_m$  n'est pas le parent de  $I_n$ , alors il y a une relation parent-enfant entre  $I_m$  et  $I_n$ ,  $I_m \rightarrow I_n$ .

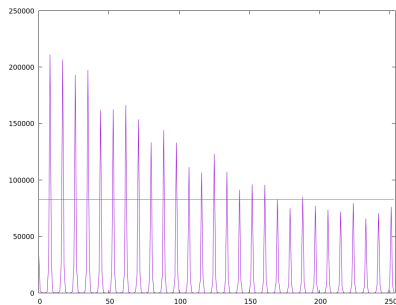
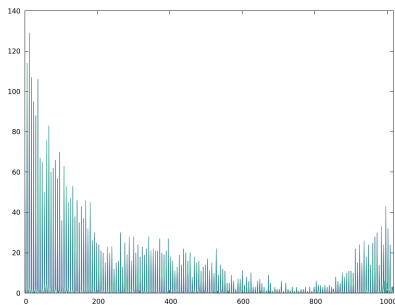
# Schéma de notre approche



# Extraction de la période

Qu'est ce que la période

Delta entre chaque pic de l'autocorrélation



# Extraction de la période

9	6	5	9	13	22	22	31
6	6	8	10	14	16	28	0
8	7	9	13	20	-1	0	0
8	9	12	19	-1	-1	0	0
10	12	-1	31	30	0	0	0
12	13	28	30	0	0	0	0

Figure – Exemple de table de quantification retournée par l'estimation de la période,  $\hat{q}(u, v)$ .

# Estimation du facteur de qualité : estimation primaire

## Estimation primaire

Calcul de distance entre  $\hat{q}(u, v)$  et  $\text{table}(i)$

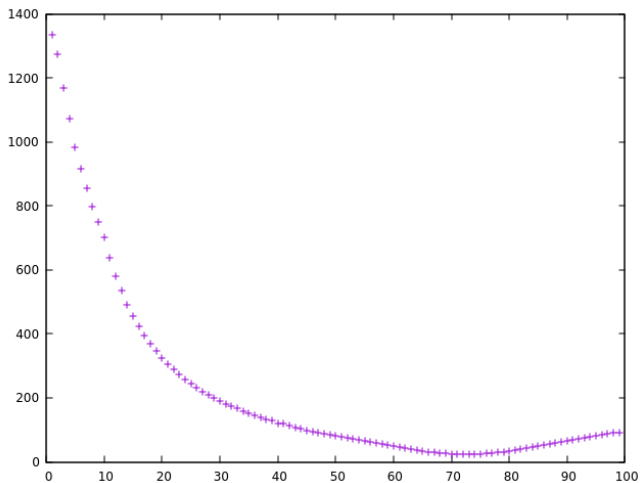
## Avantage

Donne directement  $Q_f$

## Inconvénients

- Lent
- Imprécise

# Estimation du facteur de qualité : estimation primaire



# Estimation du facteur de qualité : estimation secondaire

## Estimation secondaire

### Utilisation des formules

### Avantages

- Rapide
- Précise

### Inconvénients

$Q_f$  est nécessaire pour calculer  $Q_s$

$$\text{Si } Q_f < 50 \quad Q_s = 5000/Q_f \quad \text{sinon } Q_s = 200 - (Q_f \times 2) \quad (1)$$

$$q(u, v) = \frac{(\text{base}(u, v) \times Q_s) - 50}{100} \quad \text{avec } 1 \leq q(u, v) \leq 255 \quad (2)$$

# Estimation des ancêtres

Comparaison des images deux à deux

Filtrage des images ne pouvant pas être un ancêtre

Décision binaire

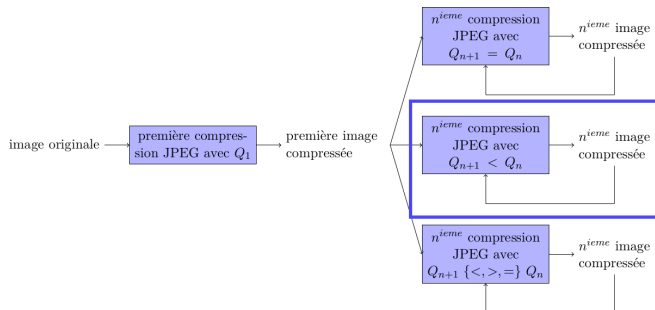


# Points clés de notre approche

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

Facilement extensible

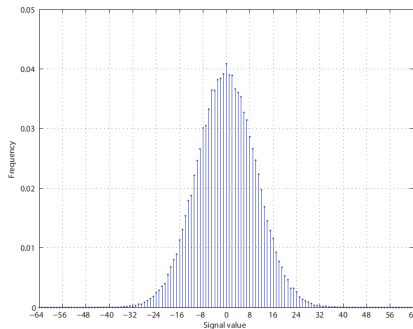
# Les marqueurs : facteur de qualité



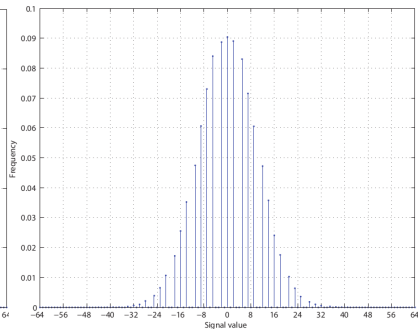
Estimation du facteur de qualité

# Les marqueurs : valeurs manquantes dans l'histogramme des coefficients DCT

Les coefficients sont des multiples des valeurs de la table de quantification



*Image simplement compressée*



*Image doublement compressée,  $Q_1 > Q_2$*

# Sommaire

- 1 Introduction
- 2 État de l'art
- 3 Notre approche
- 4 Résultats**
- 5 Conclusion

# Arbres complets

Dataset Métrique	15 images	25 images	50 images
<b>Erreur moyenne d'estimation de <math>Q_f</math></b>	0.42	0.64	0.83
<b>Sur-estimation moyenne de <math>Q_f</math></b>	1.61	1.86	1.94
<b>Sous-estimation moyenne de <math>Q_f</math></b>	1.07	2.12	6.68
<b>roots</b>	95.83	88.88	84.72
<b>edges</b>	99.70	99.24	98.97
<b>leaves</b>	99.59	99.15	98.74
<b>ancestry</b>	99.44	96.88	96.96

# Arbres avec une image manquante

<div>Dataset</div> <div>Métrique</div>	15 images	25 images	50 images
<b>roots</b>	69.60	33.33	49.01
<b>edges</b>	88.86	92.40	95.07
<b>leaves</b>	91.30	93.00	94.72
<b>ancestry</b>	78.06	82.22	88.21

# Arbre avec des images en couleurs

Dataset Métrique	15 images	25 images	50 images
<b>Erreur moyenne d'estimation de <math>Q_f</math></b>	1.15	1.29	1.42
<b>Sur-estimation moyenne de <math>Q_f</math></b>	1.85	2.70	2.64
<b>Sous-estimation moyenne de <math>Q_f</math></b>	2.97	3.79	4.94
<b>roots</b>	93.94	81.82	87.88
<b>edges</b>	99.35	98.61	99.38
<b>leaves</b>	99.62	98.66	99.78
<b>ancestry</b>	98.79	94.13	98.82

# Sommaire

- 1 Introduction
- 2 État de l'art
- 3 Notre approche
- 4 Résultats
- 5 Conclusion**



# Conclusion - perspectives

Une méthode prometteuse

Trouver d'autres marqueurs

Traiter tous les cas de la compression JPEG

Ne pas se limiter à la compression

## Conclusion - perspectives

Des questions ?