

La phylogénie des images dans les réseaux sociaux

Noé LE PHILIPPE

Équipe ICAR - William Puech

20 mai 2016

Sommaire

- 1 Introduction
- 2 Notre approche
- 3 Résultats
- 4 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^[1] 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,

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

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

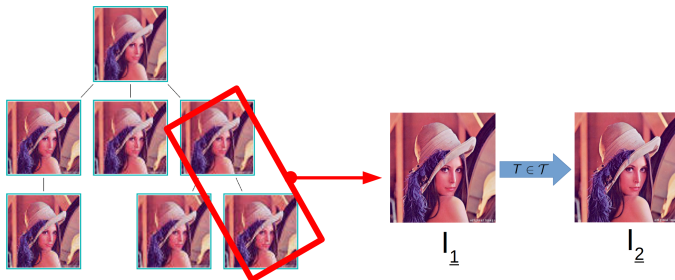
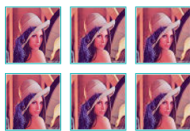


Image phylogeny tree



Set of Near-Duplicates



Image phylogeny tree

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

- Correctement identifier la racine

- Estimer au mieux l'arborescence

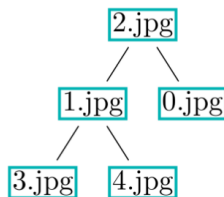
Sommaire

- 1 Introduction
- 2 Notre approche**
- 3 Résultats
- 4 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 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 indiquant qu'il n'y a pas de relation de parenté entre I_m et I_n .

Théorème

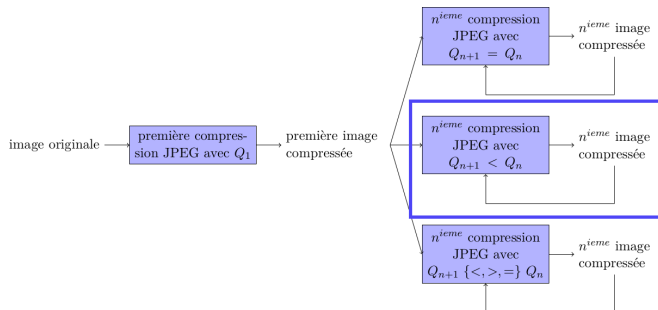
Pour tout couple d'images (I_m, I_n) d'un ensemble de near-duplicates, s'il n'existe pas de marqueur prouvant que I_m n'est pas le parent de I_n , alors il y a une relation parent-enfant entre I_m et I_n , $I_m \rightarrow I_n$.

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

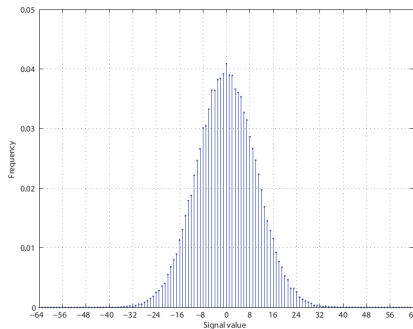


Image simplement compressée

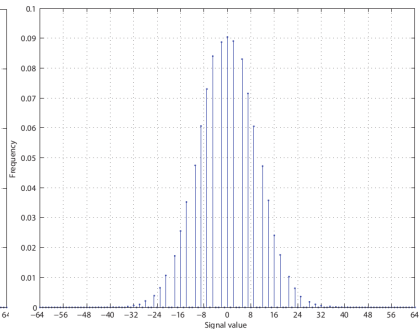


Image doublement compressée, $Q_1 > Q_2$

Sommaire

- 1 Introduction
- 2 Notre approche
- 3 Résultats**
- 4 Conclusion

Des résultats mitigés

Détection du parent direct très fiable

Mauvaise détection des parents éloignés

Arbre incomplet dans le cas d'images manquantes

Sommaire

- 1 Introduction
- 2 Notre approche
- 3 Résultats
- 4 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 ?