

17. Données visuelles massives

Nombre de participants : 12

Données visuelles massives

GIF-4105/7105 Photographie Algorithmique
Jean-François Lalonde

Image : *Little Infinity* (Jason Salavon)
Merci à Derek Hoiem, James Hays, Alexei Efros.

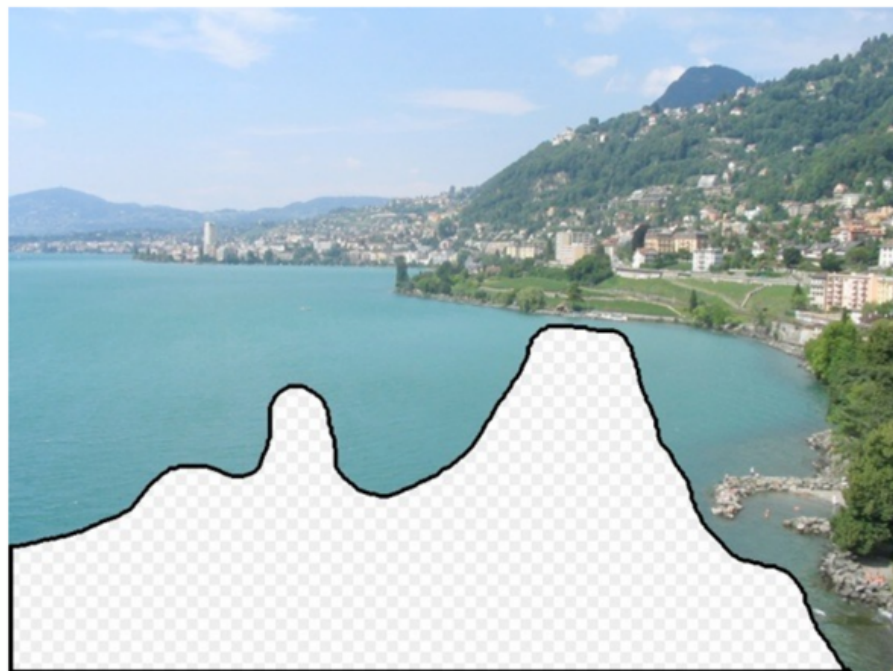
Exemple



3

J. Hayr

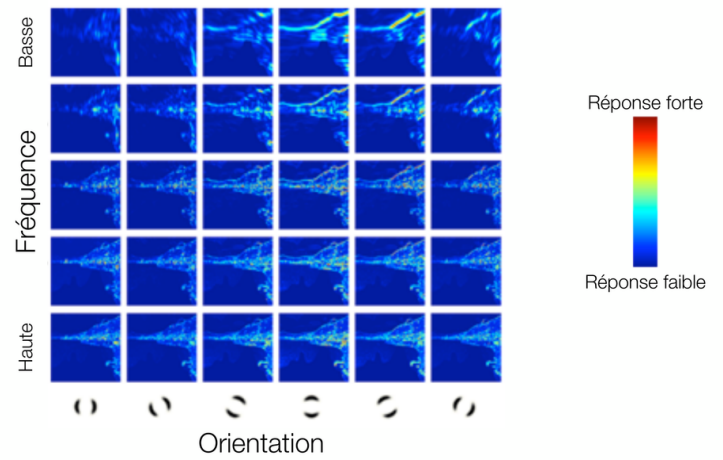
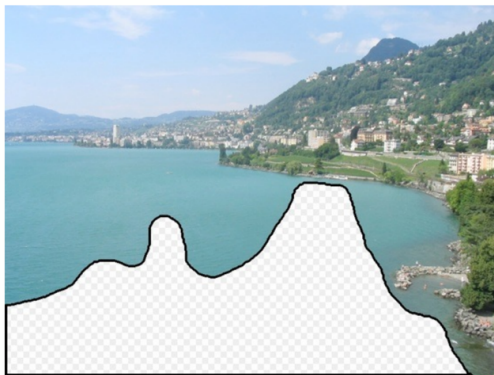
Exemple



4

J. Hayr

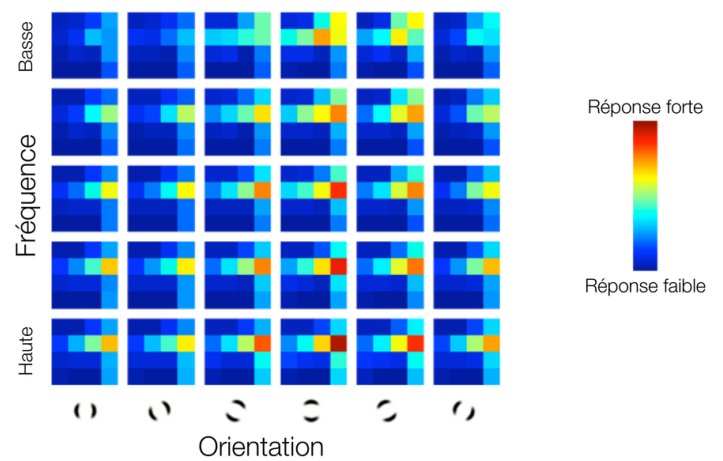
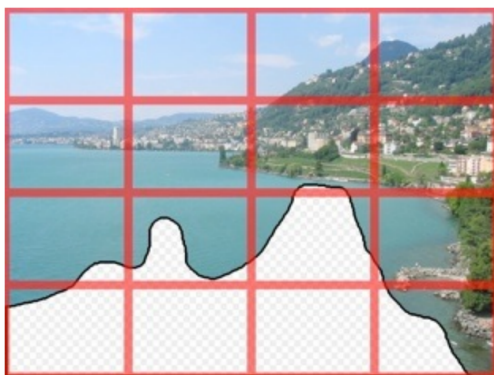
Descripteur de scène



5

J. Hayr

Descripteur de scène

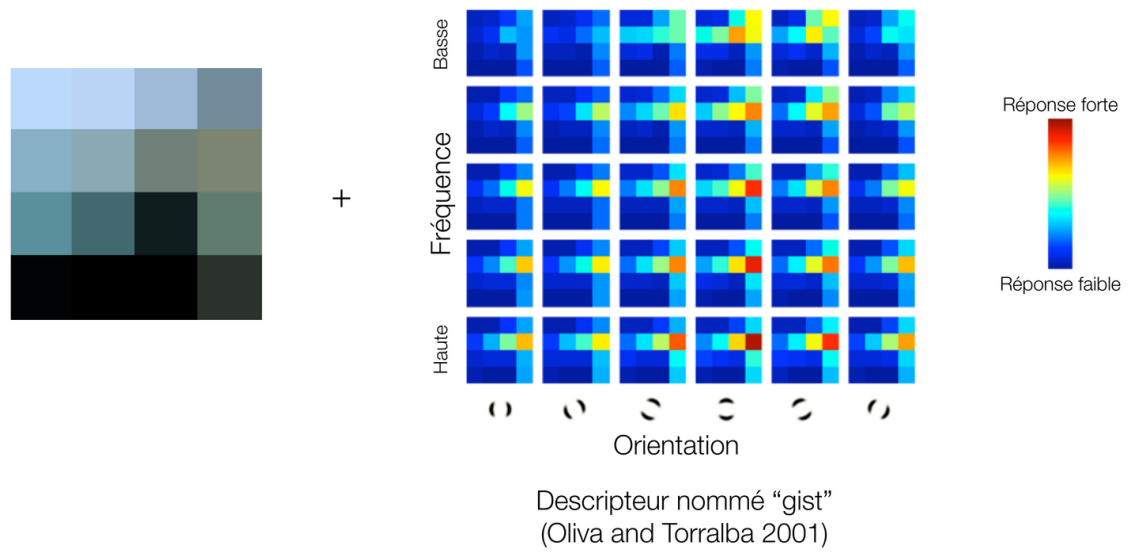


Descripteur nommé "gist"
(Oliva and Torralba 2001)

6

J. Hayr

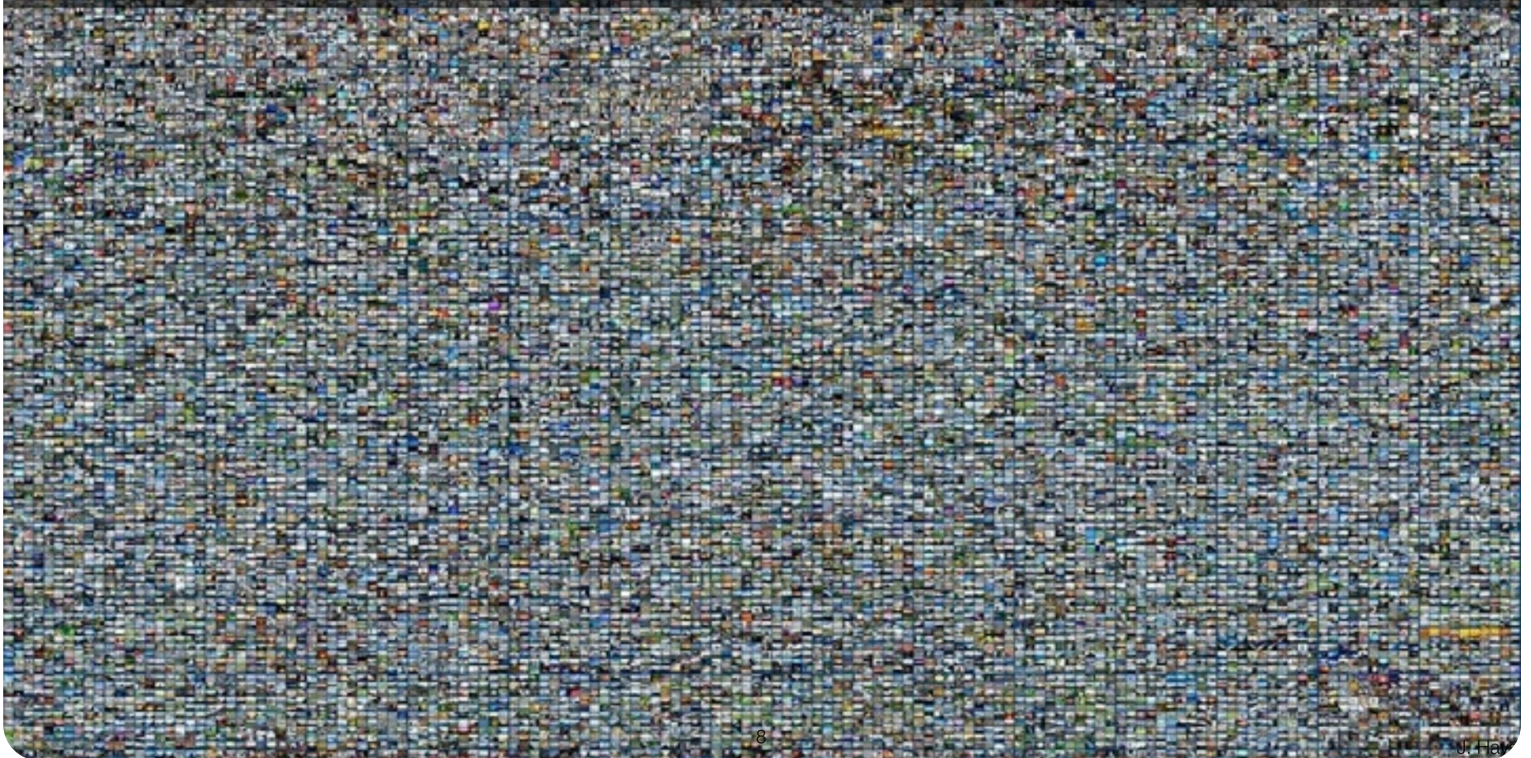
Descripteur de scène



7

J. Hayr

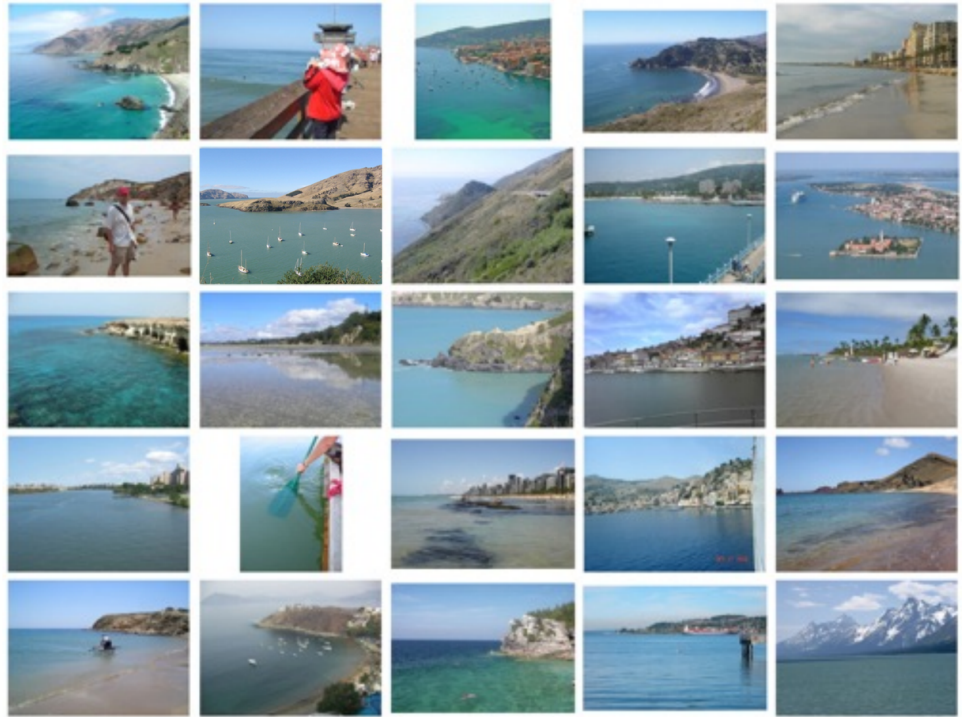
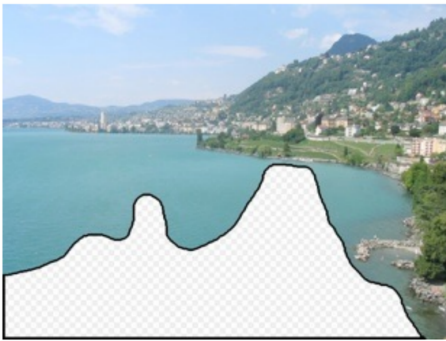
2 millions d'images de Flickr



8

J. Hayr

200 plus proches voisins

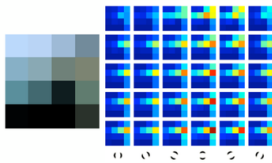


9

J. Hayr

Ordonner les résultats

Score final est la somme de:



L'appariement de scènes



L'appariement local (couleur + texture)



Le coût de la coupure de graphe

12

J. Hayr



13

J. Hayr

Pourquoi ça marche?



14

J. Hayr



1. Pourquoi les images sont similaires même si l'algorithme d'appariement est très simple (plus proche voisin) ?

5 répondants

Pourquoi ça marche?



14

J. Hays

quantité de données

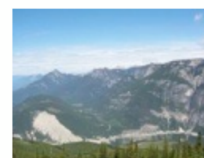
Son jeu de données

Car il y a un nombre assez grand d'images pour avoir un voisin suffisamment proche

Quantités de comparables

volume de données important

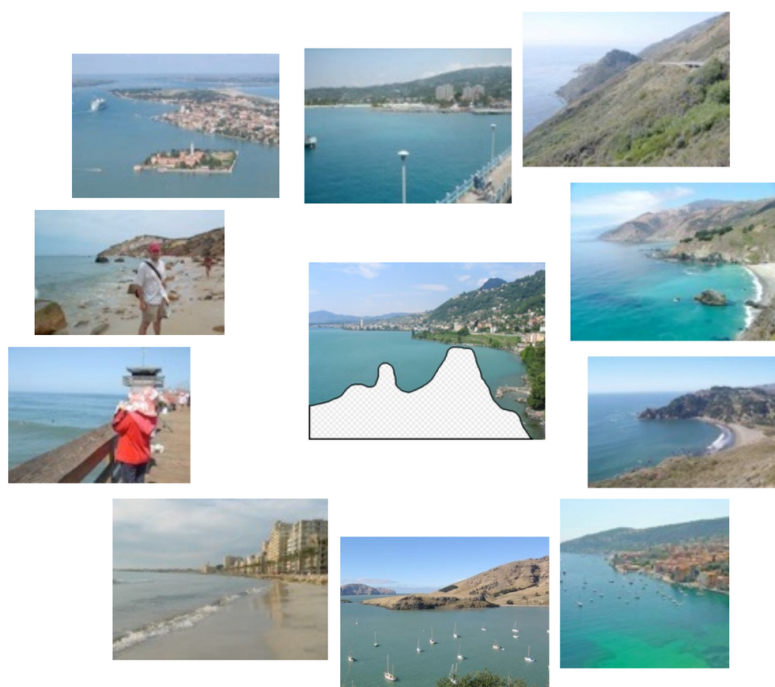
2 000 images



15

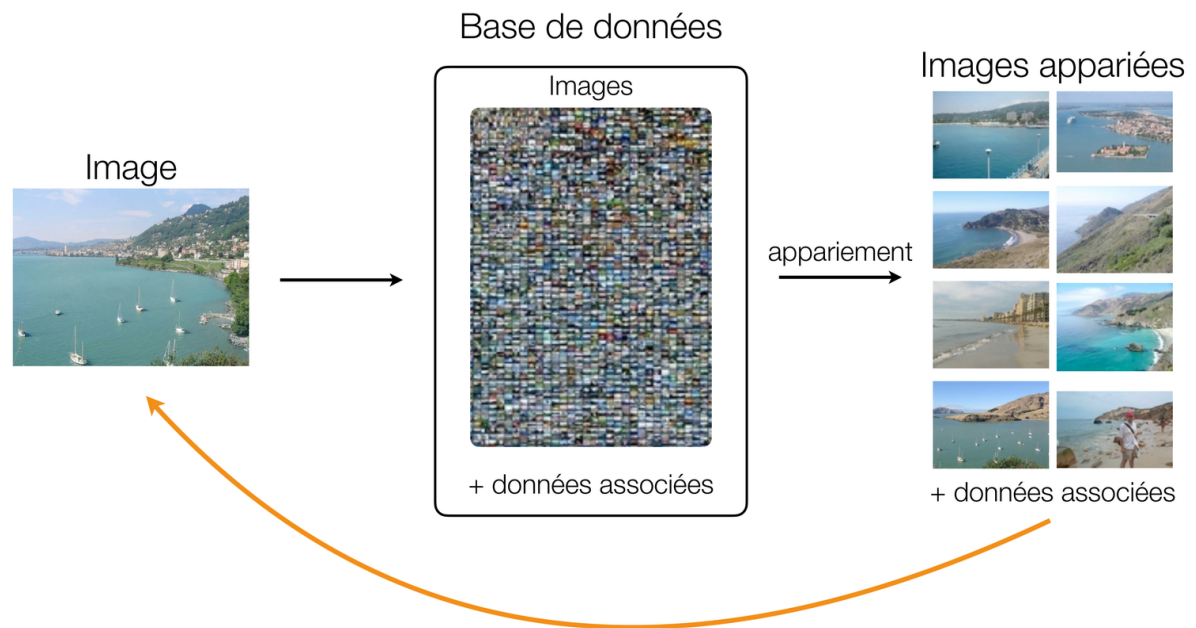
J. Hay^c

2 000 000 images



16

J. Hay^c



Si base de données est assez grande,
elle devrait contenir des images similaires!

17

Données visuelles massives

Géolocalisation d'images

GIF-4105/7105 Photographie Algorithmique
Jean-François Lalonde

Image : *Little Infinity* (Jason Salavon)
Merci à Derek Hoiem, James Hays, Alexei Efros

18

im2gps [Hays & Efros, CVPR 2008]

6 millions d'images avec coordonnées GPS



19

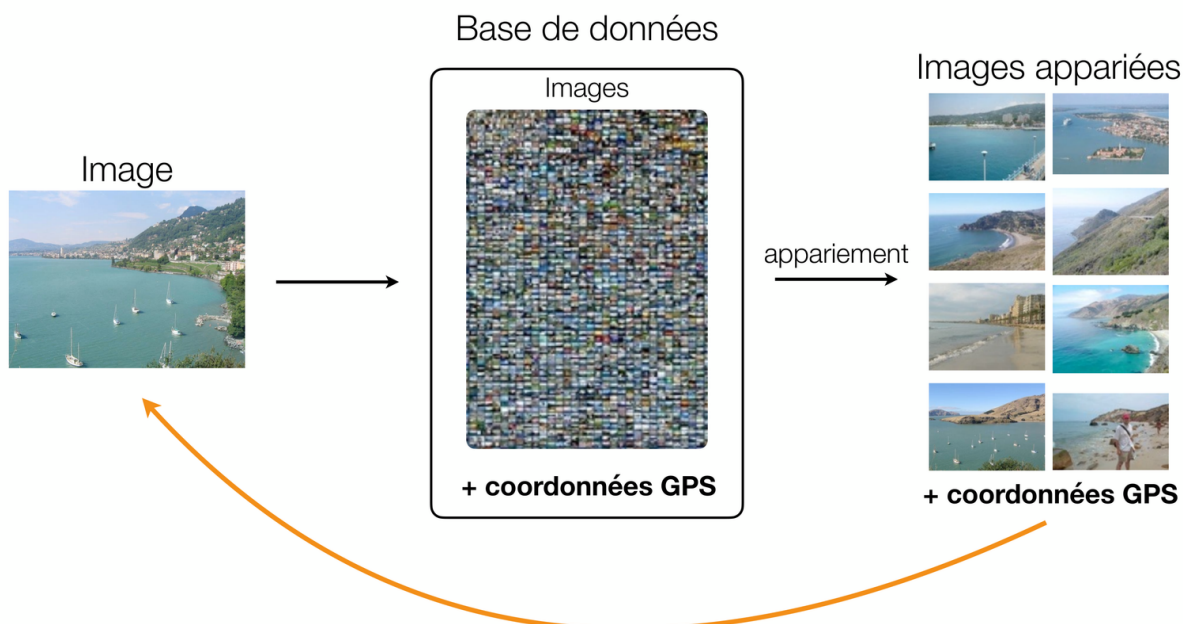
J. Hayr



3. Comment peut-on adapter l'approche précédente à la localisation d'images ?

1 répondant

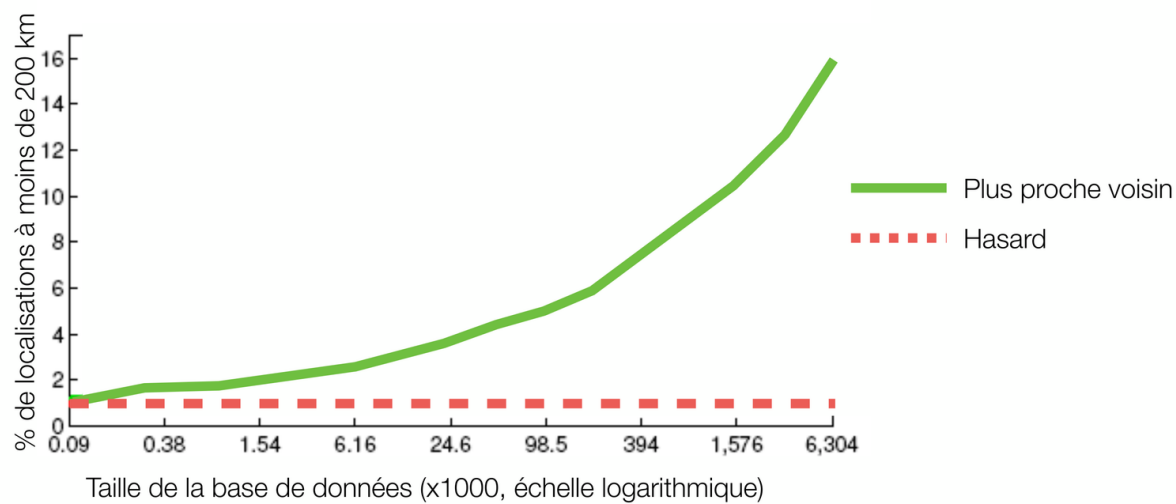
Ajouter des étiquettes aux données



Si base de données est assez grande,
elle devrait contenir des images similaires!

20

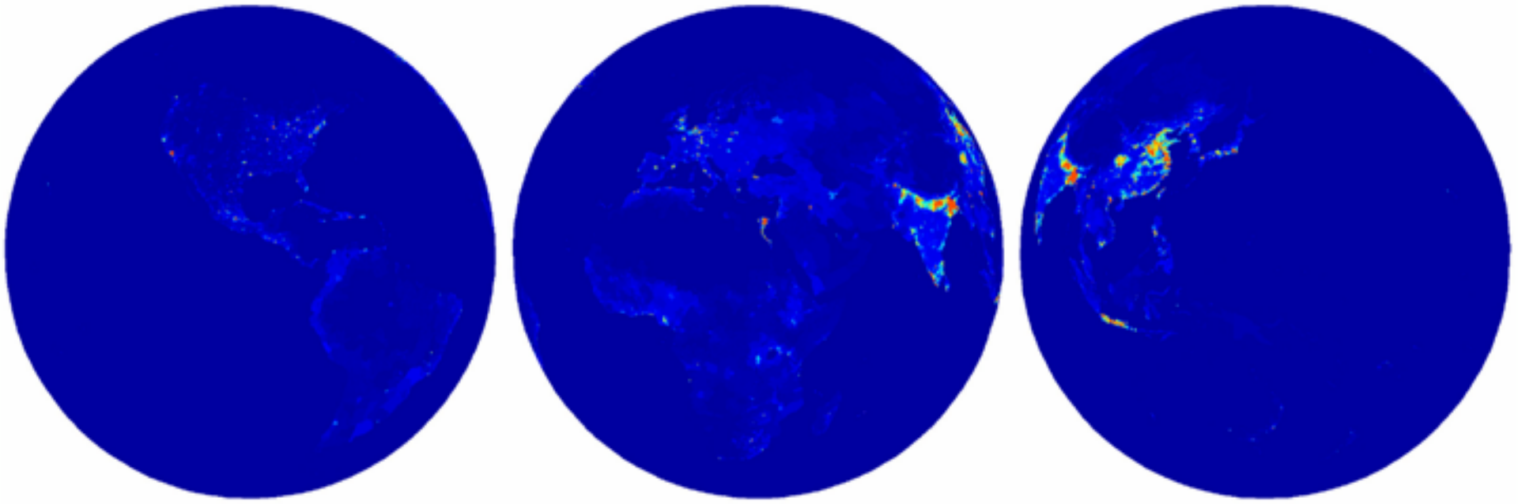
L'importance des données



21

J. Hayr

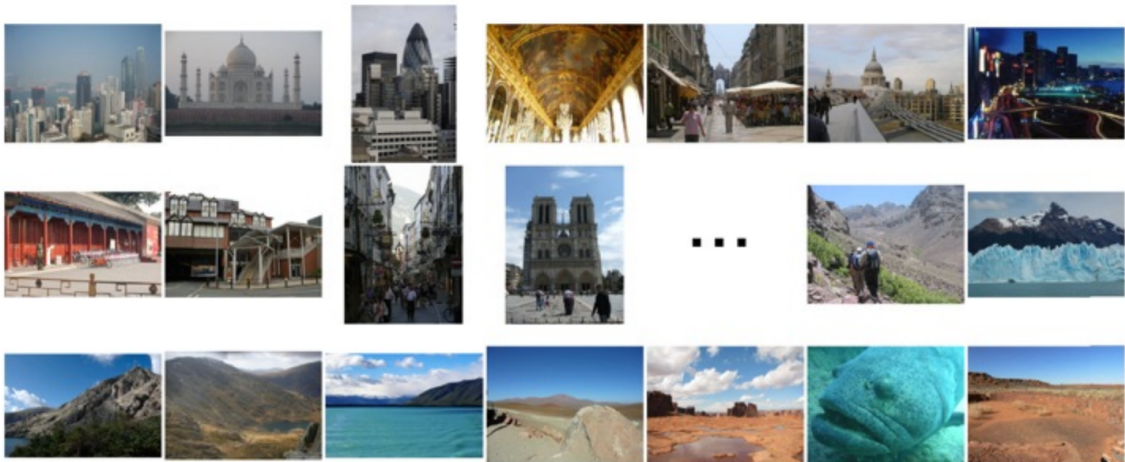
Densité de population



22

J. Hayr

Ordonner par densité de population



23

J. Hayr

Données visuelles massives

80 millions de petites images

GIF-4105/7105 Photographie Algorithmique
Jean-François Lalonde

Image : *Little Infinity* (Jason Salavon)

Tiré de : "80 million tiny images: a large dataset for non-parametric object and scene recognition" par Torralba, Fergus et Freeman



2. Combien existe-t-il d'images RGB 8-bits à résolution de 32x32 ?

7 bonnes réponses
sur 7 répondants

32*32*3

0%

0 votes



$256^{(32*32*3)}$

100%

7 votes

256*32*32*3

0%

0 votes

256*32*32

0%

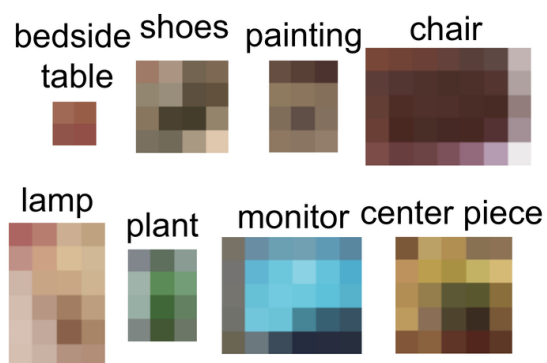
0 votes

De plus en plus d'images



Torralba, Fergus et Freemar

Images de résolution 32x32



Torralba, Fergus et Freemar

Colorisation automatique

Image



Plus proches voisins



Plus proches voisins (couleur)



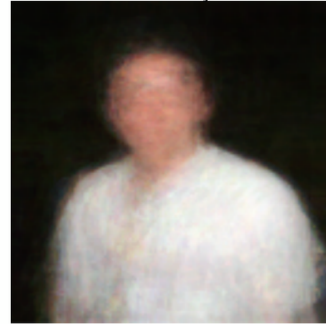
Autres options



Coloré avec couleur moyenne



Couleur moyenne



Torralba, Fergus et Freemar

Données visuelles massives

Les dangers des données

GIF-4105/7105 Photographie Algorithmique
Jean-François Lalonde

Image : *Little Infinity* (Jason Salavon)
Merci à Derek Hoiem, James Hays, Alexei Efros



Pas un échantillon aléatoire!



4. Nommez un ou plusieurs biais qui peuvent être présents dans un jeu de données d'images.

6 répondants

Biais de sous-représentation, biais de sélection, biais d'échantillonnage

Biais du photographe

Sur représentativité, absence d'exemples de certains type de données

Biais social

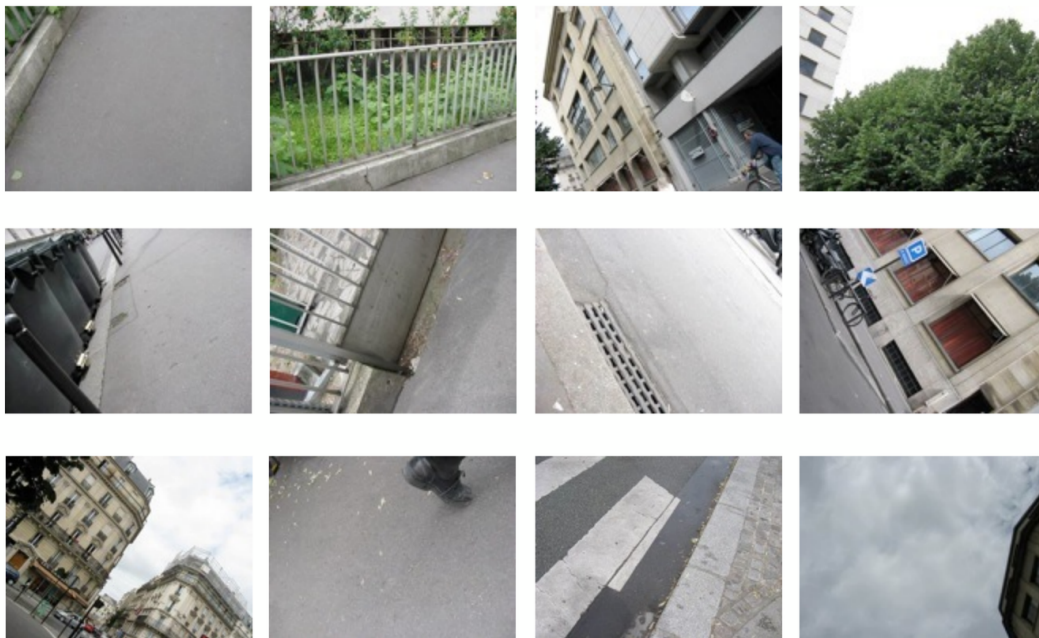
redondance

biais social

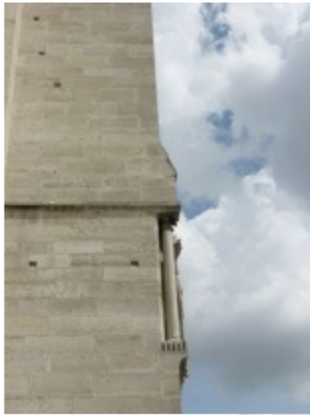
Paris selon Flickr



Le « vrai » Paris



La « vraie » Notre Dame



36

Photos : A. Efror



5. Biais de position géographique : à quels endroits y a-t-il plus de photos prises ?

1 répondant

Biais d'échantillonnage



37

J. Heys

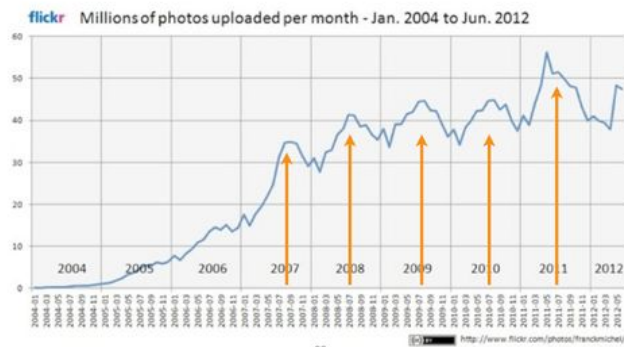
Lieu touristique



6. Biais temporel : quand avons-nous tendance à prendre plus de photos ? Pourquoi ?

4 répondants

Biais d'échantillonnage



A. Etros

Golden hour

Été, de jour

été

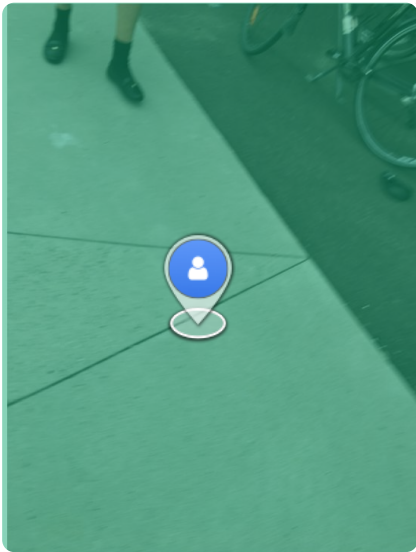
Voyages/Juillet



7. Laquelle de ces deux photos préférez-vous ?

6 répondants

Biais du photographe



vs.





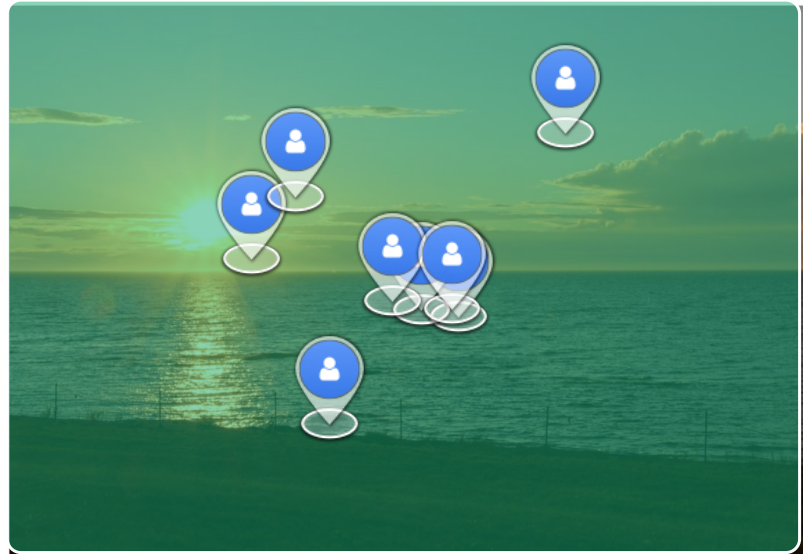
8. Laquelle de ces deux photos préférez-vous ?

8 répondants

Biais du photographe



vs.

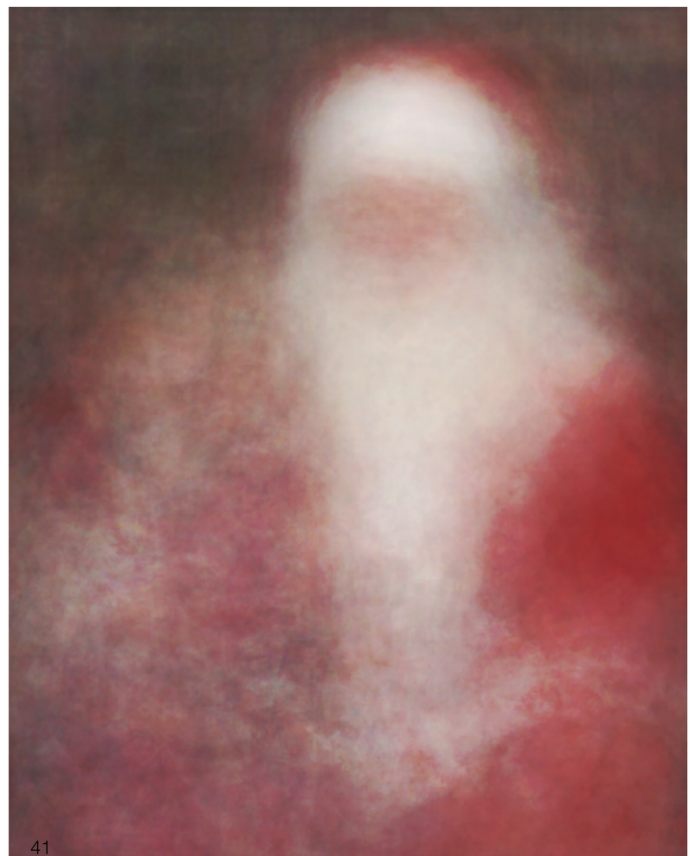


40

Biais social

Conventions sociales

« Kid with Santa »



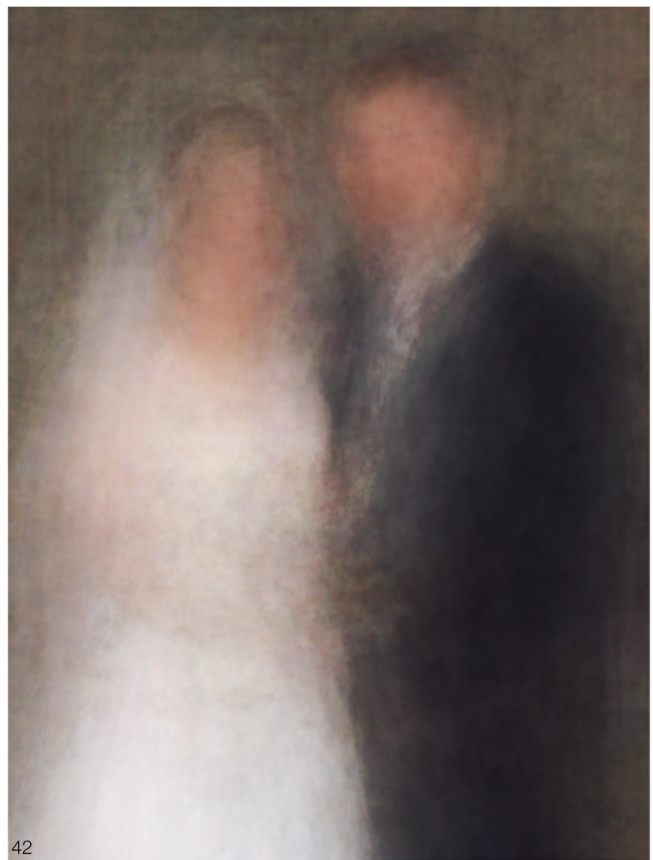
41

Biais social

Conventions sociales

« Newlyweds »

· 100 special moments » de Jason Salavon



42

Biais social

Conventions sociales

« The graduate »

· 100 special moments » de Jason Salavon



43

Biais social

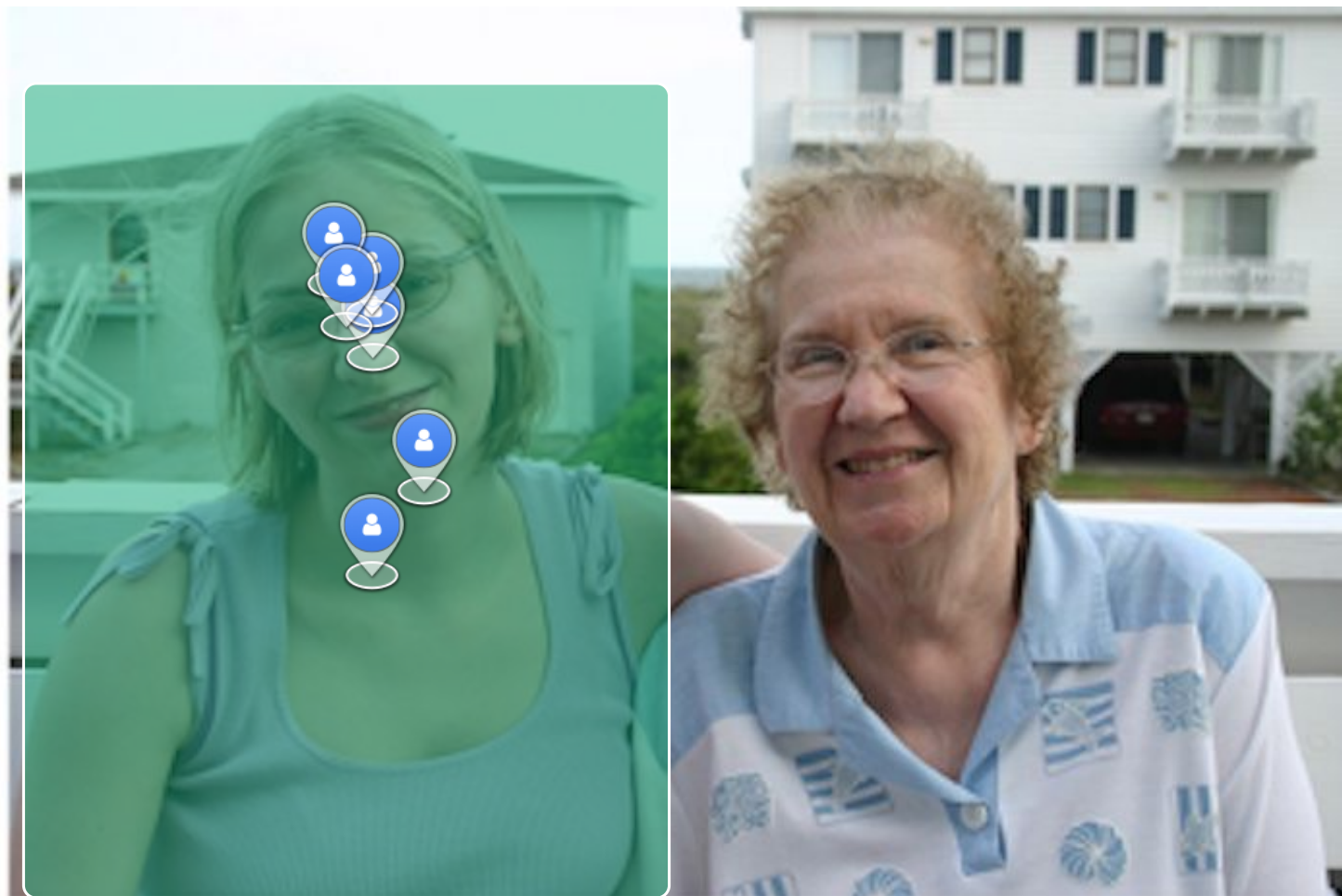
Conventions sociales





9. Je vous présente Mildred et Lisa. Identifiez Lisa.

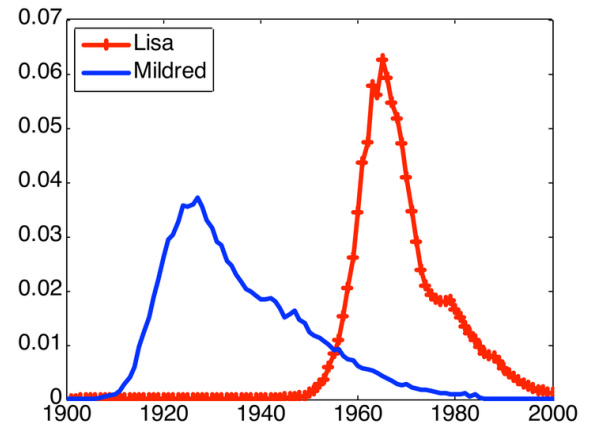
6 répondants



Biais social

Conventions sociales

Je vous présente Mildred et Lisa.
Qui est qui?



Source: U.S. Social Security
Administration

45

Gallagher et al CVPR 2008

Limiter le biais ?

- Capture autonome réduit le biais
- On en a toujours un peu...

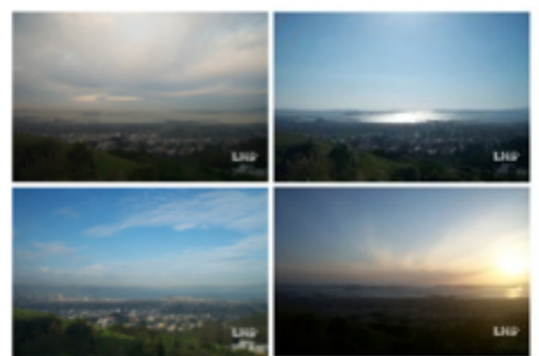
StreetView



Satellite



Webcams



46

Plusieurs autres dangers...

Pyrrhic win (victoire à la Pyrrhus)
une victoire obtenue à un prix tel qu'elle
n'en vaut pas la peine

Large image datasets: A pyrrhic win for computer vision?

Abeba Birhane*
School of Computer Science
Lero & University College Dublin, Ireland
abeba.birhane@ucdconnect.ie

Vinay Uday Prabhu*
UnifyID AI Labs
Redwood City, USA
vinay@unify.id

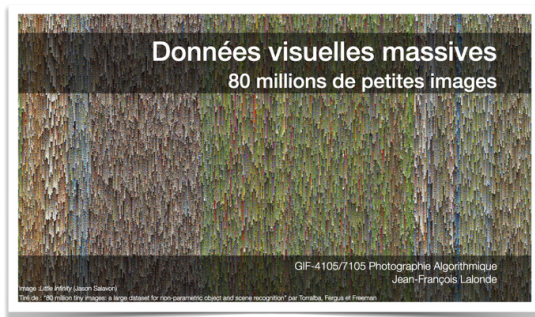
Nudité et voyeurisme

Présence de termes discriminatoires

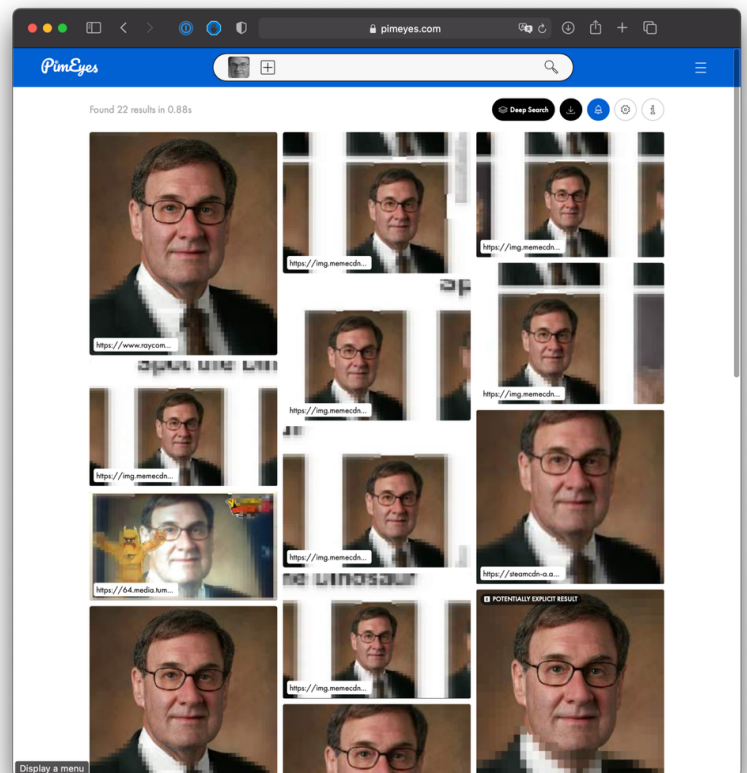
Biais en défaveur des minorités

47

Vous vous rappelez?



Source : pimeyes.com



48

Vie privée

- Les bases de données d'images contiennent beaucoup de personnes
- Ces personnes n'ont pas donné leur consentement!
- Mais on peut savoir qui c'est... voir pimeyes.com

