

# Utilisation des Descripteurs SIFT pour la Reconnaissance Faciale

M'Hand Kedjar

Montréal, le 03 août 2016

# Plan

- Introduction
- Contexte et objectifs
- Méthodes
  - Détection
  - Descripteurs SIFT
  - Classification
- Résultats
- Discussion
- Conclusion

# Introduction

- But de la reconnaissance faciale (RF): identifier ou vérifier un ou plusieurs individus à partir d'images fixes ou des séquences animées
- Autres techniques biométriques:
  - Iris: très précise, mais dispendieuse sur une grande échelle. Non acceptée généralement
  - Empreintes digitales: fiable et non intrusive. Ne convient pas pour les personnes non collaboratives.
- RF est un bon compromis
- Multitude d'applications: contrôle d'accès, systèmes d'identification, surveillance vidéo, ....



Source: youtube



Source: wikipedia.org

# Reconnaissance Faciale - Difficultés

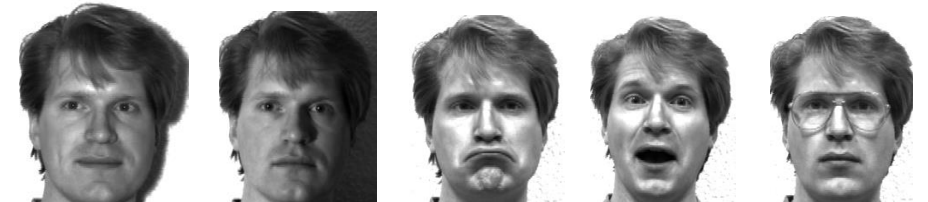
- Identifier des visages similaires (similarité entre les classes)
- Gérer la variabilité intra-classe:
  - Position de la tête
  - Conditions d'illumination
  - Expression
  - Port d'accessoires: chapeau, lunettes, ...
  - Variations avec l'âge



Source: pottermore.com



Source: cnn.com



Base de données YALE

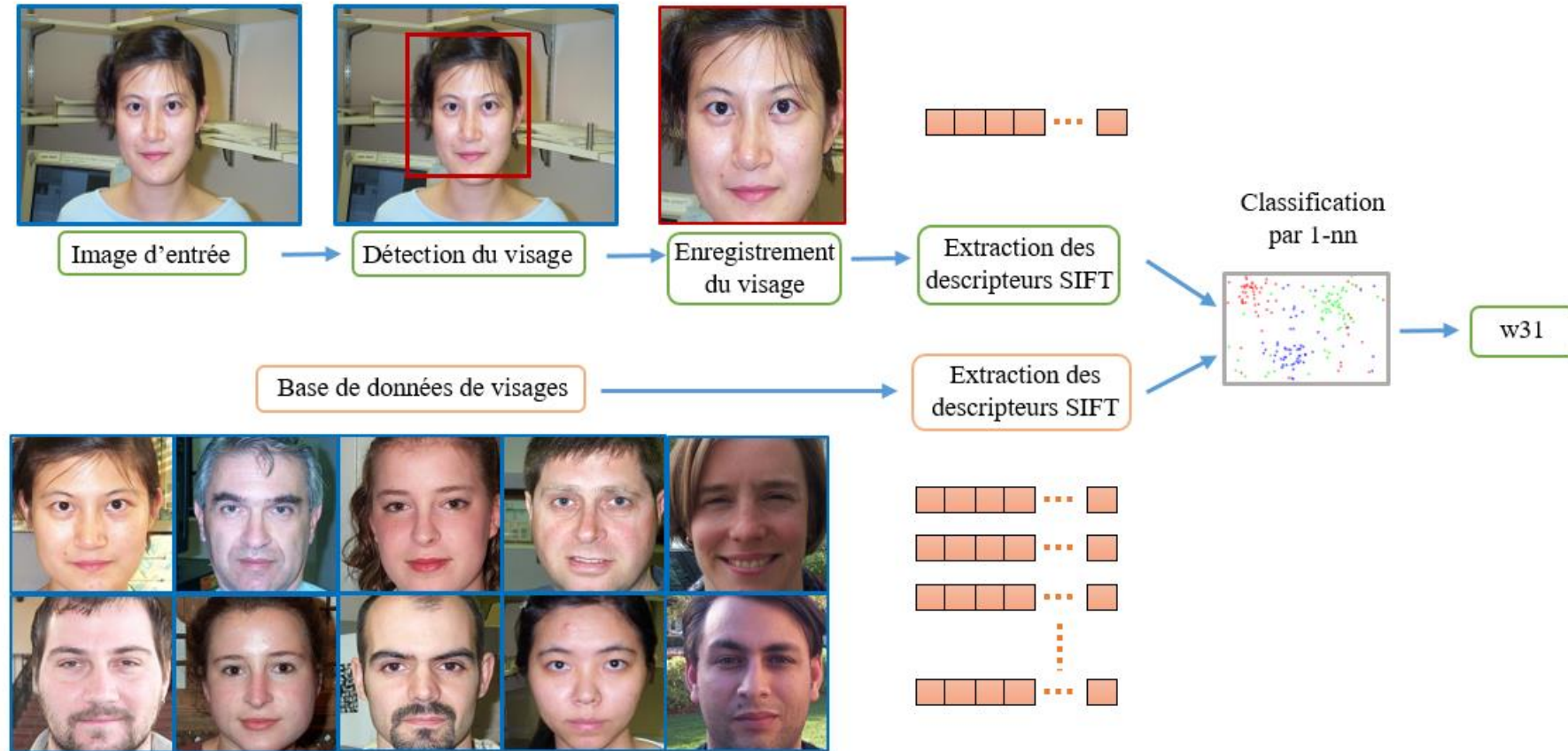
# Contexte

- Bases de données ORL, YALE et Caltech:
  - Parmi les plus utilisées pour évaluer les algorithmes de RF
  - Disponibles gratuitement
  - Toutes les images sont annotées
- Algorithme Viola - Jones pour la détection
  - Détecteur très précis
  - Performance temps réel
- Algorithme de David Lowe pour SIFT
  - Développé en C++
  - Interface Matlab
  - Linux, Windows
- Classification par l'algorithme du plus proche voisin

# Objectifs

- Concevoir un système de reconnaissance faciale basé sur SIFT
  - Détection du visage
  - Extraction des descripteurs
  - Classification par 1nn
- Comparaison des performances avec 2 autres approches: PCA, HOG+SVM

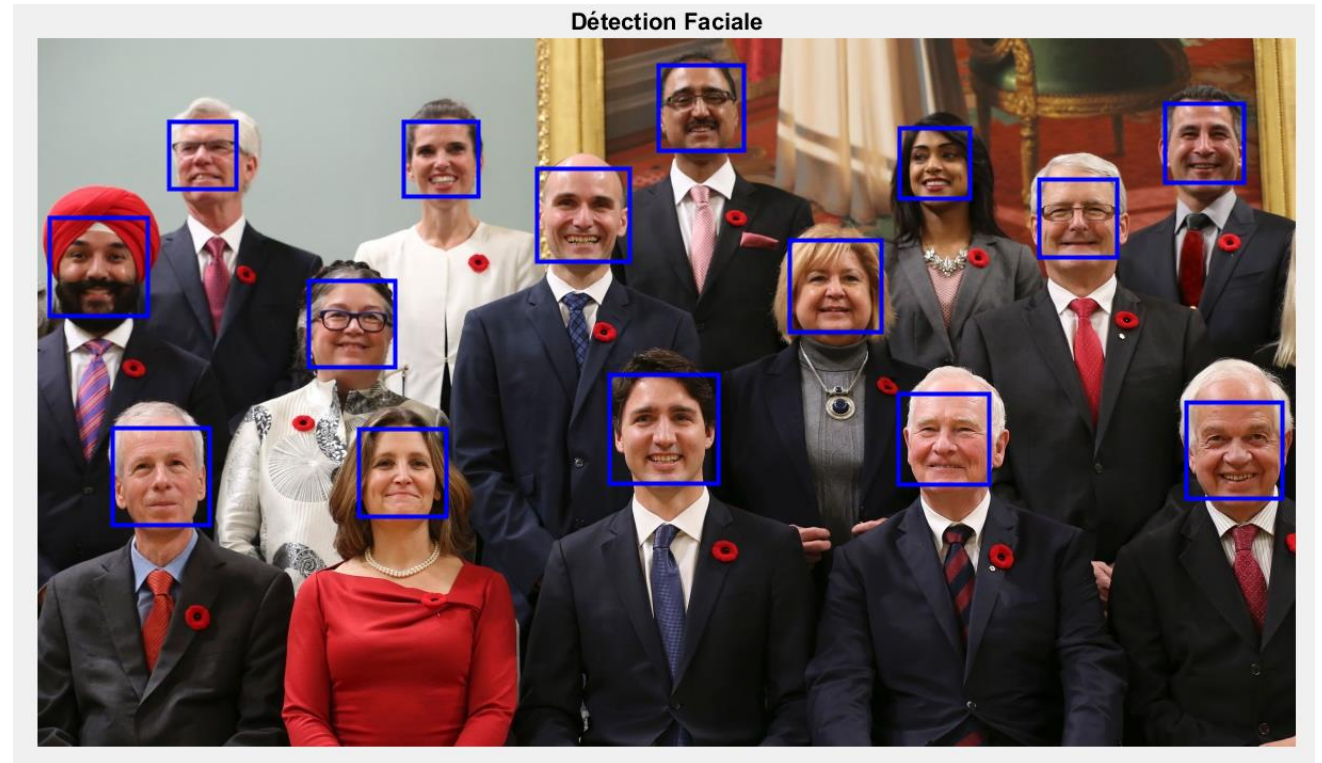
# Méthodes – Schéma du système





# Méthodes – Détection

- Filtre multi-échelles de Haar développé par P. Viola et M. Jones
- Les caractéristiques d'un visage sont décrites dans un fichier xml
- Construites à partir d'un échantillon de quelques centaines d'images tests
- Taux de détection très élevé
- Performance temps réel



Source: cbc.ca



# Méthodes - SIFT

- SIFT extrait un ensemble de points d'intérêt (descripteurs) à partir d'une image
- Construction de l'espace des échelles

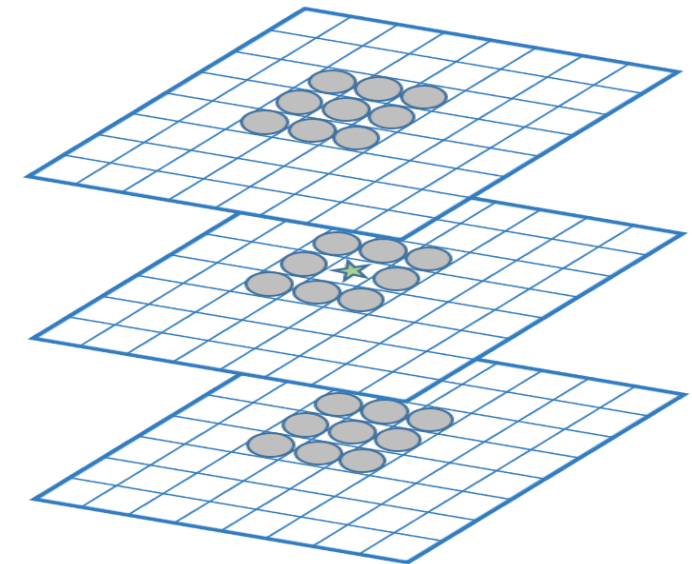
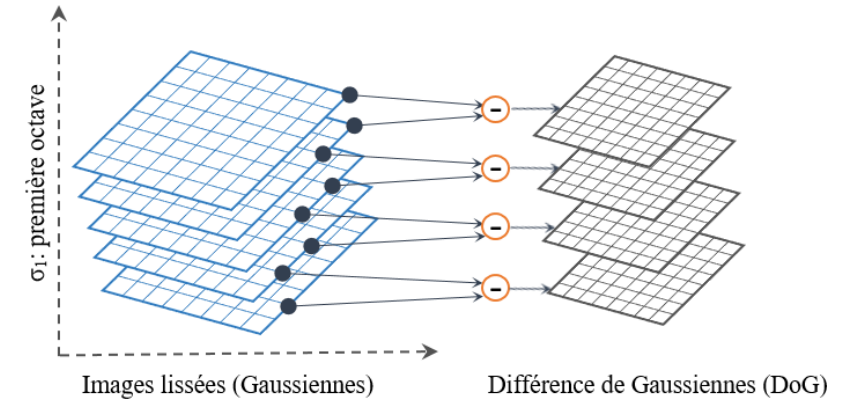
$$L(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} * I(x, y)$$

(x,y) coordonnées de l'image I,  $\sigma$ : facteur d'échelle

- Calcul de la différence de Gaussiennes (DoG)

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

- Détection des extrema dans la DoG
  - Un pixel est sélectionné si sa valeur est plus petite (ou plus grande) que tous les 26 autres qui l'entourent



# Méthodes - SIFT

- Assignment de l'orientation

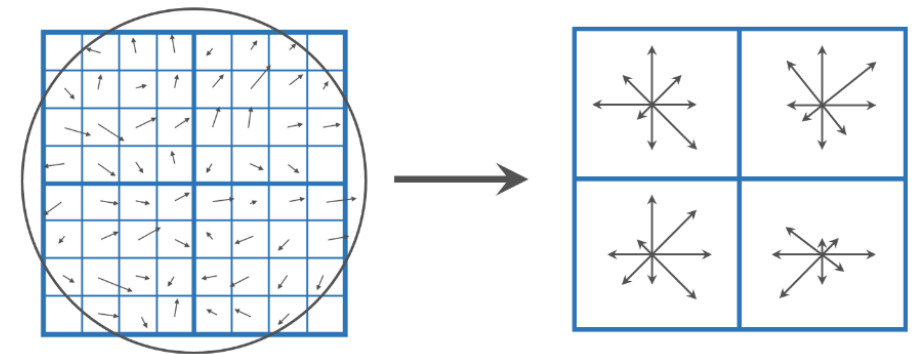
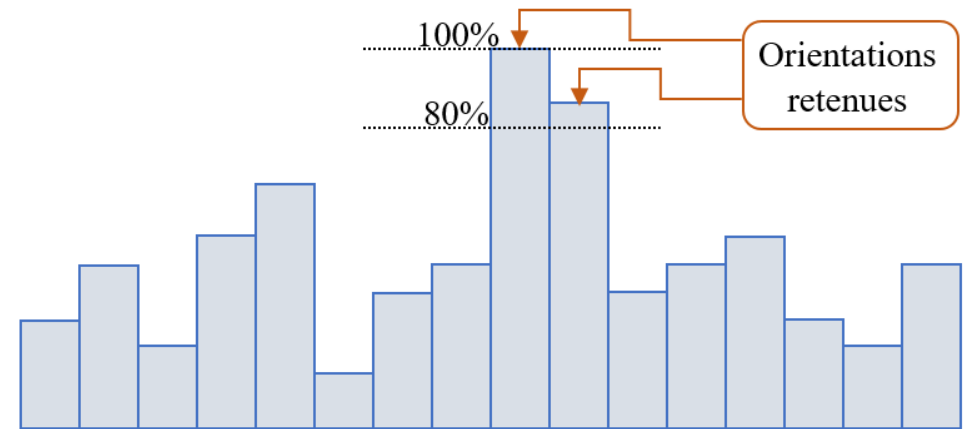
$$m(x, y) = \frac{\sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}}{2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

Un points clé est défini par: (x,y,σ,θ)

- Descripteur local type SIFT

- les descripteurs SIFT sont calculés sur une grille de 16 × 16 points voisins, divisée en 4 blocs de 4 × 4 et des histogrammes de 8 orientations
- Vecteur à 128 composantes (4 \* 8 \* 4)



# Méthodes - Classification par la méthode du plus proche voisin

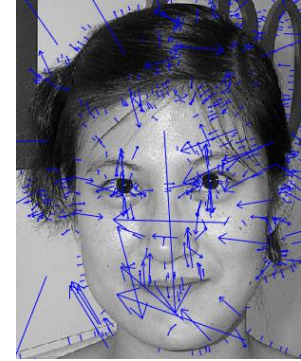
- $I_t$ : image test,  $I_r$ : Image de référence

$$K(I_t) = \{k_1^{I_t}, k_2^{I_t}, \dots, k_P^{I_t}\}$$

$$K(I_r) = \{k_1^{I_r}, k_2^{I_r}, \dots, k_Q^{I_r}\}$$

$$d_i^{min} = \min_j d(k_i^{I_t}, k_j^{I_r}), j = 1, 2, \dots, Q$$

- Distance au plus proche voisin:  $d_1 = d_{min}$
- Distance au prochain plus proche voisin:  $d_2$
- Correspondance entre:  $k_i^{I_t}$  et  $k_j^{I_r}$  ssi  $\frac{d_1}{d_2} < d_s$
- Classification finale de l'image  $I_t$ : classe de l'image  $I_r$  avec le plus grand nombre de correspondances



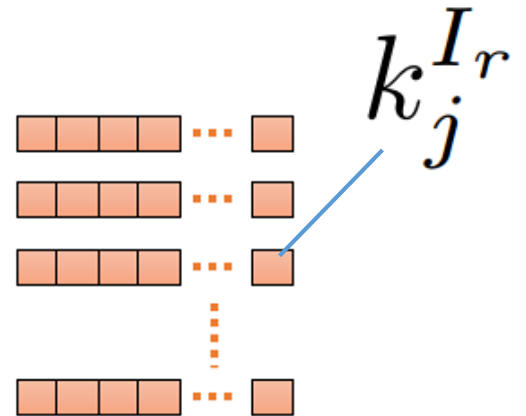
$I_t$



Base de données

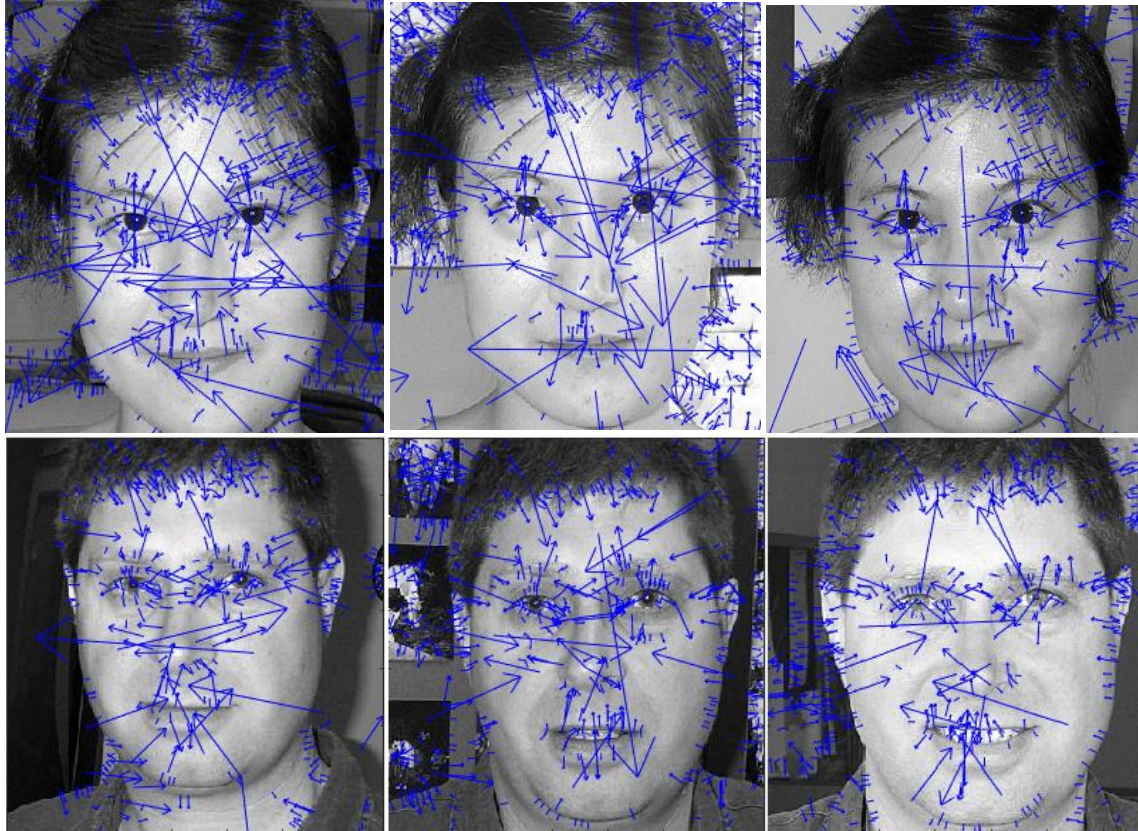


$k_i^{I_t}$

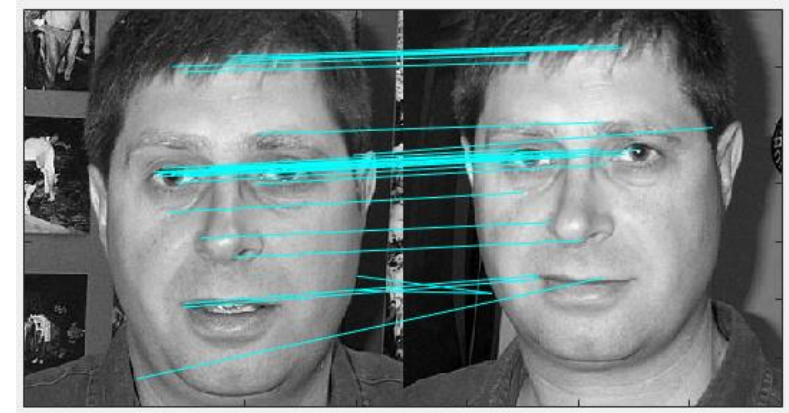


$k_j^{I_r}$

# Méthode – visualisation des descripteurs



Exemple d'images avec les descripteurs SIFT extraits



31 correspondances



1 correspondance

Exemple de correspondances  
entre deux images



# Résultats - Précision

- 3 bases de données: ORL, YALE, Caltech
- Génération aléatoire des sous-ensembles d'apprentissage et de test
- Varier la proportion entres les deux sous-ensembles: de 0.1 à 0.9

$$r = \frac{N_{train}}{N_{total}} = \frac{N_{train}}{N_{train} + N_{test}}$$

- 10 répétitions

$$prec_i = \frac{VP_i}{N_{total}}, i = 1, 2, \dots, 10$$

- Mesurer la précision finale

$$prec = \frac{1}{10} \sum_{i=1}^{10} prec_i$$

- Et la variance, pour chaque valeur de r

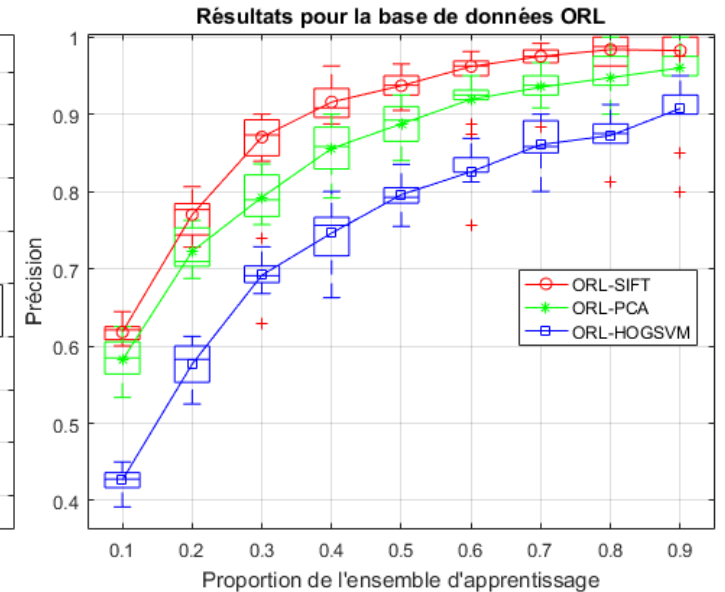
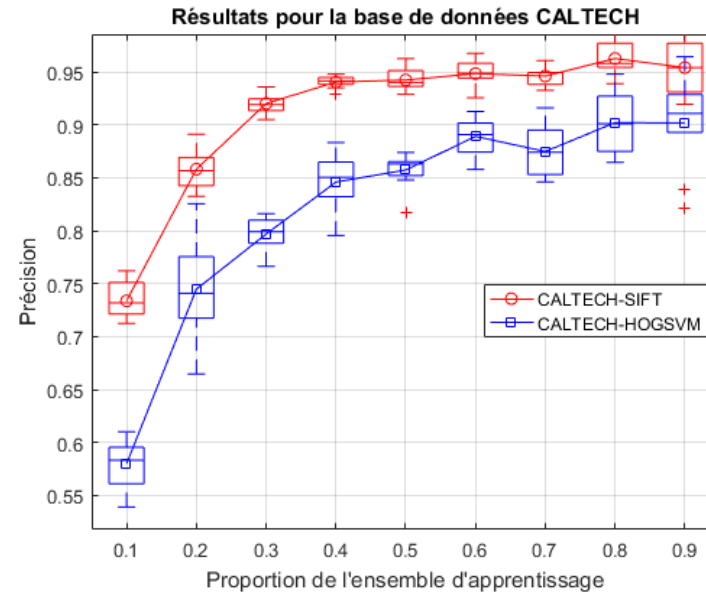


TABLE I  
CARACTÉRISTIQUE DES BASES DE DONNÉES UTILISÉES

BD	$n_p$	$n_i$	$n_{min}$	$n_{max}$	format et taille
ORL	40	400	10	10	PGM, $112 \times 92$
YALE	15	165	11	11	GIF, $116 \times 98$
Caltech	31	450	1	29	JPEG, $896 \times 592$

TABLE II  
RÉSULTATS: PRÉCISION ET VARIANCE

BD	ORL		YALE		CALTECH	
	prec	var	prec	var	prec	var
SIFT	<b>0.98</b>	<b><math>2.8e^{-4}</math></b>	<b>0.85</b>	<b><math>19e^{-4}</math></b>	<b>0.96</b>	<b><math>2.6e^{-4}</math></b>
PCA	0.96	$4e^{-4}$	0.70	$34e^{-4}$	0.31	$27e^{-4}$
HOG+SVM	0.91	$25e^{-4}$	<b>0.85</b>	$21e^{-4}$	0.90	$8e^{-4}$

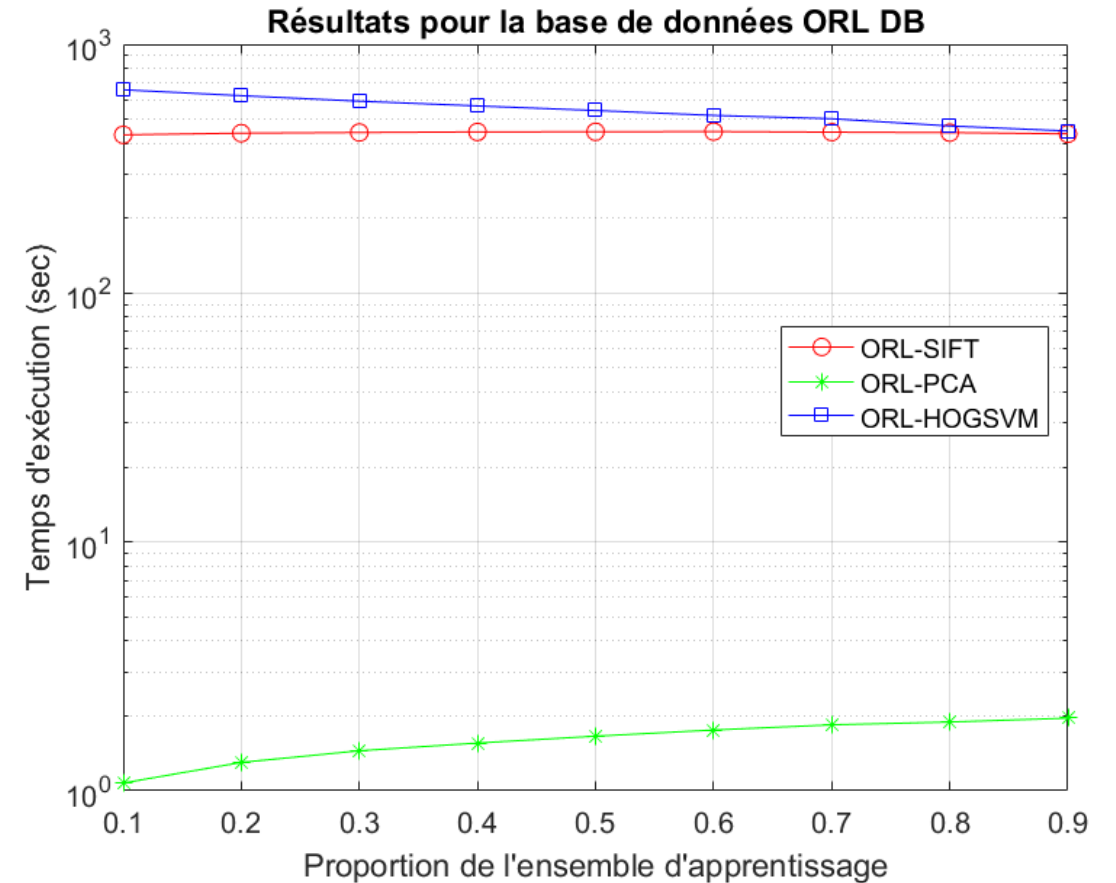
# Résultats – temps de calcul

Le temps de calcul est estimé pour la phase d'apprentissage et de test

TABLE III  
RÉSULTATS: TEMPS DE CALCUL

BD	ORL	YALE	CALTECH
<b>SIFT</b>	441.51	202.02	2329.19
<b>PCA</b>	<b>1.89</b>	<b>0.48</b>	<b>48.36</b>
<b>HOG+SVM</b>	469.43	189.78	519.96

$r = 0.8$



# Discussion

- SIFT présente plusieurs avantages: descripteurs invariants par rapport aux transformations dans l'image: rotation, translation, facteur d'échelle, ...
- C'est le descripteur le plus utilisé, et plusieurs autres sont basés sur SIFT
- Article 1: [3] Cong Geng and X. Jiang, "SIFT features for face recognition,"
  - 2 approches: KPSIFT et PDSIFT
  - Images rognées à 50x57
  - Classification: Similarité Cosinus.
  - Précision: 97% sur ORL
- Article 2: [4] T. Liu, S. H. Kim, H. S. Lee and H. H. Kim, "Face recognition base on a new design of classifier with SIFT keypoints".
  - Descripteur SIFT avec MLP
  - Classification: nombre de descripteurs assignés à une classe
  - Précision: 98% sur ORL pour  $r = 0.8$ , 92% pour  $r = 0.6$
- Article 3 : [5] D. Liu, D. m. Sun and Z. d. Qiu, "Bag-of-Words Vector Quantization Based Face Identification,"
  - Descripteurs SIFT avec un modèle Bag-of-words
  - Classification: SVM
  - Précision: 93%
- Aucun des systèmes n'a atteint 100% de classification
- Systèmes combinant descripteurs locaux avec les descripteurs globaux (texture, couleur, ...) seraient à explorer.



# Conclusion

- Le système de RF basé SIFT est le plus robuste en termes de précision et de stabilité des résultats
- Celui basé sur PCA est le plus rapide en termes de temps de calcul.
- Les résultats pour YALE DB sont moins bons par rapport aux autres bases de données: conditions d'illumination
- Combinaison des méthodes locales et globales donnerait un meilleur résultat

## Perspectives:

- Identification automatique de l'âge et du sexe d'une personne à partir d'une photo de son visage
- Application à la personnalisation de l'écran en fonction de l'utilisateur

# Références

- [1] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol.60, no.2, pp. 91–110, 2004
- [2] Paul Viola, Michael Jones. 2001. "Rapid Object Detection using a Boosted Cascade of Simple Features". Conference On Computer Vision And Pattern Recognition 2001
- [3] Cong Geng and X. Jiang, "SIFT features for face recognition," Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on, Beijing, 2009, pp. 598-602.
- [4] T. Liu, S. H. Kim, H. S. Lee and H. H. Kim, "Face recognition base on a new design of classifier with SIFT keypoints," Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, Shanghai, 2009, pp. 366-370.
- [5] D. Liu, D. m. Sun and Z. d. Qiu, "Bag-of-Words Vector Quantization Based Face Identification," Electronic Commerce and Security, 2009. ISECS '09. Second International Symposium on, Nanchang, 2009, pp. 29-33

Merci de votre attention!

Questions?