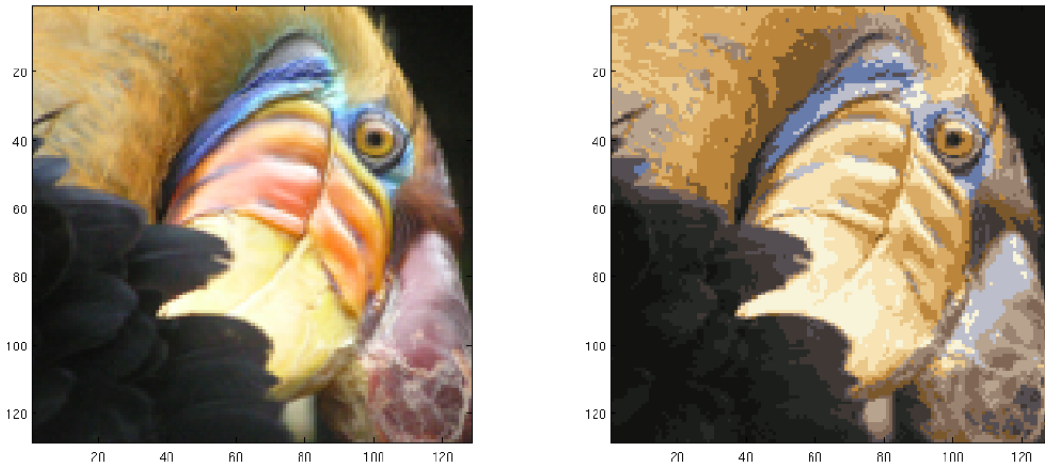


Compression d'image par la méthode de clustering

Dans ce TP nous allons utiliser k-Means comme un méthode de compression l'image. Dans le format de codage des couleurs RGB (Red, Green, Blue), chaque pixel est représenté par trois entiers entre 0..255, qui spécifient l'intensité des couleurs Rouge, Vert et Bleu. Normalement, chaque image est composé de nombreuse couleurs. Une façon de représenter un image de manier efficace est de réduire le nombre des couleurs dans l'image. L'idée est de traiter pixel comme un vecteur de donnée de dimension trois, i.e. chaque pixel est $\mathbf{x} \in \mathbb{R}^3$, et d'utiliser K-means pour grouper les pixels en $K = 16$ clusters de couleurs. Après avoir groupé les pixels, nous allons remplacer chaque pixel dans l'image original par le centre de son cluster. Ceci donne une version compressée de l'image original.



1 Implémentassions de K-Means

L'algorithme K-Means est un méthode pour grouper automatiquement les données similaires. Nous avons un ensemble de données $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}\}$ où $\mathbf{x}^{(i)} \in \mathbf{R}^n$. Notre objectif est de partitionner l'ensemble de données en K groupes "clusters". Chaque cluster C_k pour $k = 1..K$ est représenté par son "centres" dénoté μ_k . L'algorithme K-means est composé de deux phases principales

1. Trouver le centre le plus proche pour chaque point de donnée :

$$c^{(i)} \leftarrow \text{l'indice (entre } 1..K) \text{ du centre le plus proche de } \mathbf{x}^{(i)}$$

où nous utilisons la distance Euclidienne au carré telle que

$$d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) := \sum_{t=1}^n |\mathbf{x}_t^{(i)} - \mathbf{x}_t^{(j)}|^2$$

2. Mettre à jour les centres :

$$\mu_k := \frac{1}{|C_k|} \sum_{j \in C_k} \mathbf{x}^{(j)}$$

$|C_k|$ dénote le nombre des éléments dans cluster C_k . Par exemple si cluster $k = 2$ contient que $\mathbf{x}^{(3)}$ et $\mathbf{x}^{(5)}$ nous calculons μ_2 par $\mu_2 := \frac{1}{2}(\mathbf{x}^{(3)} + \mathbf{x}^{(5)})$.

L'approche est de commencer par K représentants (centres) de clusters $\mu_1, \mu_2, \dots, \mu_K$ choisis aléatoirement, et répéter les deux phase jusque à convergence.

- Téléchargez le support de TP "Support8.zip"
- Complétez et exécutez "compress.sci" en scilab pour compresser l'image "bird_small.png" en $K = 16$ couleurs
- Dans votre implémentations de K-means limiter le nombre des itération à 10

Avant commencer, répondez aux questions suivantes :

1. Quelle est la dimension n de chaque point de donnée $\mathbf{x}^{(i)}$?
2. Quelle est la taille de l'ensemble de données (i.e. le nombre m des elements à clusterer) ?

À rendre

Votre code et la version compressée d'image "bird_small.png" (imprimé en pdf). Le rendu est à envoyer par mail à : azadeh.khaleghi@inria.fr

La date limite de rendu est le mardi 19 mars 2013 à 23 :59