

La compression de données appliquée aux images

Nous manipulons quotidiennement des fichiers compressés sans forcément connaître tous les protocoles qui officient. Il est donc intéressant, en se basant sur le format graphique qu'est l'image, de comprendre comment une information peut être stockée en réduisant sa taille au maximum.

La ville est un espace où l'information est omniprésente, notamment sous la forme d'imagerie de surveillance. Il est donc nécessaire d'optimiser les communications et le stockage des données, d'où l'importance de la compression. Cette compression est effectuée par divers algorithmes, qu'il est intéressant de combiner pour étudier leur optimalité.

Professeur encadrant du candidat

Q. Fortier

Ce TIPE fait l'objet d'un travail de groupe. Listes des membres du groupe

- Maxime LAURENT
- Arsène MALLET
- Sofiane MAMI

Positionnement thématique

- INFORMATIQUE(*Informatique Pratique*)
- MATHEMATIQUES(*Mathématiques Appliquées*)
- MATHEMATIQUES(*Probabilités*)
- (INFORMATIQUE(*Languages*)) (?)

Mots-clés

(français)	(anglais)
Entropie	Entropy
Redondance	Redundancy
Quantification	Quantization
Sans-Perte	Lossless
Codage arithmétique	Arithmetic coding

Bibliographie commentée

La compression de données, procédé consistant à réduire la taille des données tout en conservant l'information qu'elle comporte, est un domaine sujet à de nombreuses études tant il est essentiel au monde du numérique moderne. Basée sur la théorie de l'information^{entropy}, initiée par Harry Nyquist et Ralph Hartley dans les années 1920 et très largement étoffée et formalisée en 1948 par Claude E. Shannon, la compression de données fait notamment intervenir des notions probabilistes, statistiques et informatiques. Les algorithmes de compression se distinguent par leur caractère sans perte ou avec perte, caractère indiquant si de l'information est perdue ou non au cours de la compression. Ainsi, il est intéressant d'étudier des algorithmes avec et sans perte afin de mesurer leur efficacité et leurs potentiels défauts.

L'entropie, telle que définie par Shannon^{entropy} est un concept mathématique essentiel puisqu'il permet de quantifier l'information contenue ou délivrée par une source d'information. C'est par le calcul et l'optimisation de cette grandeur qu'apparaissent des algorithmes de compression sans perte comme le codage de Huffman^{huffman} en 1952, et plus tard le codage arithmétique^{arithmetic-coding}.

Il est mathématiquement démontré qu'il n'existe pas d'algorithme pouvant compresser n'importe quel type d'information sans perte^{compression}. On peut cependant tirer parti du type de données à compresser. C'est pourquoi nous avons choisi de nous focaliser sur le domaine de l'image, un type de données auquel s'appliquent de nombreux procédés de compression^{compressioncode-theory}.

L'algorithme de compression du format JPEG (Joint Photographic Experts Group)^{jpeg}, inventé en 1992, est un procédé de compression d'images avec pertes permettant de réduire entre 3 et 25 fois la taille d'un fichier en fonction de la qualité finale que l'on veut obtenir. Il est aujourd'hui l'un des formats de compression les plus utilisés pour les images.

L'algorithme de compression JPEG est constitué de plusieurs étapes. Il se base sur l'analyse des composantes de couleurs de l'image et de la perception humaine afin d'optimiser le regroupement de l'énergie. Pour ce faire, on traite l'image comme un signal, auquel on applique la Transformée en Cosinus Discrète (DCT)^{jpeg}. Cette transformation permet un portage de l'information essentiellement par les coefficients de basses fréquences. Seul un petit nombre de coefficients sera donc non nul, ce qui permet de réduire le nombre de calculs à effectuer par l'ordinateur.

Vient ensuite la quantification, unique étape responsable de la perte d'information, et donc de la potentielle dégradation de qualité. Cette étape consiste à diviser les composantes formant l'image par des coefficients réducteurs, en fonction de leur importance dans la transmission de l'information^{code-theoryjpeg}. Enfin on code cette information « réduite » au moyen d'un codage comme celui de Huffman.

En effectuant les étapes dans le sens inverse, on peut retrouver l'infor-

mation de départ (potentiellement tronquée), c'est la décompression.

En comparant l'information avant compression et après décompression, nous pouvons quantifier les pertes et les réductions, aussi bien en termes de stockage qu'en termes de temps, et ainsi déterminer l'efficacité des algorithmes en fonction des nécessités.

Problématique Retenue

Il s'agit d'étudier et d'implémenter diverses méthodes de compressions, afin de mesurer leur efficacité et leurs limites, aussi bien théoriques que pratiques.

Objectif du TIPE du premier membre (Maxime)

Objectif du TIPE du second membre (Arsène)

1. Implémentation des différents procédés de compression,
2. Modélisation de leur efficacité par une application concrète, en procédant à la compression d'un grand nombre d'images de différents types et tailles afin d'obtenir des données complètes,
3. Comparaison des différentes méthodes, de leur gains ou pertes temporels et en termes de stockage.

Objectif du TIPE du troisième membre (Sofiane)

Etude théorique des différents modèles de compressions, en se basant grandement sur le modèle du JPEG, afin d'implémenter ces procédés en Python.