

La compression de données appliquées aux images

Nous manipulons quotidiennement des fichiers compressés sans trop savoir ce qui se cache derrière. 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, sous de diverses formes. Il est donc nécessaire d'optimiser les communications, d'où l'importance de la compression. Cette compression est due à divers modèles d'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

<i>(français)</i>	<i>(anglais)</i>
Entropie	Entropy
Redondance	Redundancy
Quantification	Quantization
Sans-Perte	Lossless
Codage Arithmétique	Arithmetic Coding

Bibliographie commentée

La compression de données, procédé consistant en la réduction de 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. Basé sur la théorie de l'information [1], initié par Harry Nyquist and Ralph Hartley dans les années 1920 et très largement étoffé et formalisé en 1948 par Claude E. Shannon, la compression de données fait notamment intervenir des notions probabilistes, statistiques et informatiques. Les algorithmes de compressions se distinguent par leur caractère sans perte ou avec perte, caractère indiquant si de l'information est perdu 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 leur potentiels défauts.

L'entropie, tel que définie par Shannon [1] est un concept mathématique essentiel puisqu'il permet de quantifier l'information contenue ou délivré 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 [2] en 1952, et plus tard le Codage Arithmétique [3].

Bien que certains algorithmes puissent être suffisamment généralisés, il est mathématiquement démontré qu'il n'existe pas d'algorithme pouvant compresser n'importe quel type d'information sans perte [4]. Ainsi lorsqu'il est question de compression, il est nécessaire de l'étudier au travers d'un certain type d'information. C'est pourquoi nous avons choisi de nous focaliser sur le domaine de l'image, un type de données connaissant de nombreux procédés de compression [4] [5].

L'Algorithme de compression du format JPEG (Joint Photographic Experts Group) [6], 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 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) [6]. Cette transformation permet que l'information soit essentiellement portée par les coefficients de basses fréquences. Seul un petit nombre de coefficients seront donc non nuls, 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 potentiel 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 [5] [6]. Enfin on code cette information « réduite » au moyen d'un codage tel qu'évoqué ci-dessus.

En effectuant les étapes dans le sens inverse, on peut retrouver l'information 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.

Références

- [1] C. E. SHANNON. « A Mathematical Theory of Communication ». In : *The Bell System Technical Journal* 27 (juill. 1948), p. 379-423, 623-656.
- [2] David A. HUFFMAN. « A Method for the Construction of Minimum-Redundancy Codes ». In : *Proceedings of the I.R.E.* 40.9 (sept. 1952), p. 1098-1101.
- [3] Guilhem COQ, Olivier ALATA, Christian OLIVIER et Marc ARNAUDON. « Codage arithmétique pour la description d'une distribution ». In : *TAIMA* (mai 2007). hal-00367577f, p. 65-71.
- [4] Matt MAHONEY. *Data Compression Explained*. 2013. URL : <http://mattmahoney.net/dc/dce.html> (visité le 19/01/2022).
- [5] Jean-Guillaume DUMAS, Jean-Louis ROCH, Sébastien VARRETTE et Eric TANNIER. *Théorie des codes*. 3^e éd. Dunod, juin 2018. ISBN : 978-2100781096.
- [6] Gregory K. WALLACE. « THE JPEG STILL PICTURE COMPRESSION STANDARD ». In : *IEEE Transactions on Consumer Electronics* 38.1 (fév. 1992), p. xviii-xxxiv.