

La compression d'images numériques

Principe général de la compression

Les données (texte, son, images, vidéo, programmes) ont besoin d'être compressées pour des raisons d'espace de stockage et parfois de vitesse de transfert. La compression consiste à réduire la taille d'un fichier numérique à l'aide de diverses techniques.

La plupart sont basées sur des formules mathématiques qui permettent de traiter les redondances (c'est-à-dire les éléments qui apparaissent plusieurs fois à l'identique) en les remplaçant par des codes plus courts. Ceci permet, après une étape de décompression, de récupérer le fichier tel qu'il était initialement (compression sans perte, par exemple la compression ZIP utilisable sur tous types de données).

Il existe également des méthodes de compression avec perte, appliquées à des types de données bien particuliers (sons, images ou vidéos). Le principe est d'éliminer les informations les moins perceptibles de manière à obtenir un résultat qui paraisse identique à l'original. Exemple : on ne voit pas les ultra-violets (longueur d'onde < 400 nm).

La compression RLE (Run Length Encoding)

Cette compression est utilisée par les images bitmap (format .bmp notamment)

Elle est basée sur la répétition d'éléments consécutifs. Par exemple, la chaîne « AAAAHHHHHHHHHH » sera compressée en 4A10H soit un gain de compression de $(14-5)/14 = 64,3\%$. En revanche la chaîne REELLEMENT qui deviendrait 1R2E2L1E1M1E1N1T (soit -60% de gain de compression) ne sera pas compressée. En pratique seules les séquences comportant plus de 3 répétitions d'un même élément sont compressées, les autres sont conservées telles quelles, séparées des séquences compressées par un caractère spécial.

La compression RLE n'a donc d'intérêt que pour des images possédant de nombreux éléments consécutifs redondants, notamment les images avec de larges parties uniformes. C'est ce qui explique que les images BMP atteignent fréquemment des tailles importantes.

La compression LZW (Lempel- Ziv - Welch)

Du nom de ses inventeurs, cette méthode de compression est utilisée par les formats d'image GIF et TIFF, mais aussi par les formats d'archive ZIP, ARJ, etc.

v Le principe de base est de repérer des séquences qui apparaissent plusieurs fois, en construisant au fur et à mesure un dictionnaire de séquences, et de les remplacer par leur indice dans le dictionnaire.

v Dans le cas de la compression d'images GIF, elle est particulièrement adaptée pour des images avec de grandes zones de même couleur ou des images géométriques. Elle est limitée à 256 couleurs, et donc peu adaptée aux images photographiques. En revanche c'est l'un des seuls formats d'image (avec le format PNG) qui permette de gérer la transparence, et c'est le seul qui gère des images animées (constituées d'un ensemble d'images affichées l'une après l'autre).

La compression JPEG (Joint Photographic Expert Group)

Méthode de compression avec perte (correspondant au format .jpg)

Elle permet de réduire la taille jusqu'à 1/25e de l'original sans perte notable de qualité

Elle est efficace sur des images photographiques, mais beaucoup moins sur des images géométriques, pour lesquelles la perte d'information est plus souvent visible

Le principe est, à partir de formules mathématiques compliquées (transformation discrète en cosinus), d'atténuer les détails en éliminant certaines composantes de hautes fréquences.

Le taux de compression est variable, ce qui permet de trouver un compromis entre réduction de la taille et niveau de perte de détails acceptable. Le taux minimum est la compression sans perte (réduction de la taille de moitié seulement), utilisée principalement pour la transmission d'images médicales.

En résumé :

Il existe un grand nombre de formats de fichiers. Parmi eux :

Format	Codage des couleurs	Compression	Application et spécificité
BMP	De 1 à 24 bits	Aucune / RLE	Dessins et photos (peu adapté pour internet car trop lourd)
GIF	8 bits	LZW	Adapté pour les dessins au trait et les images avec des plages unies
JPEG	24 bits	JPEG	Adapté pour les photos et les images comportant des dégradés. Très bon taux sans perte notable de qualité