

Les images numériques.

Intervention de Jean Vasseur

La définition des images

Le terme “définition” s’utilise pour qualifier la netteté d’une image numérique de type “bitmap” telle qu’une photographie. Plus le nombre de pixels d’une image “bitmap” est important, plus l’image est potentiellement nette et donc apte à être agrandie.

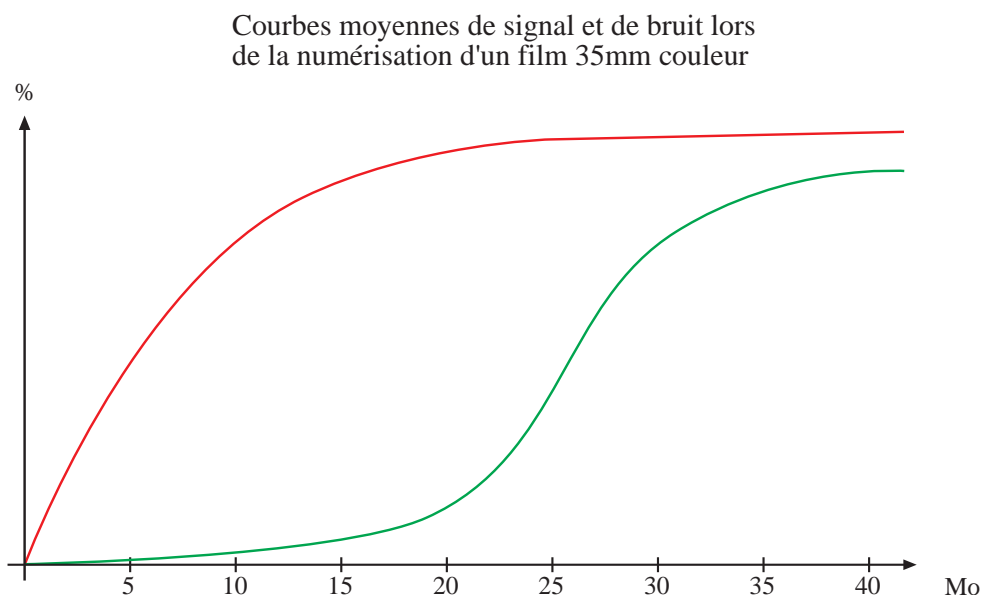
La résolution est un paramètre utilisé pour spécifier à un périphérique (d’acquisition ou de restitution) la précision à laquelle une image de type “bitmap” doit être traitée. Dans le cas d’une acquisition, la résolution va déterminer la définition de l’image.

Avant de numériser (“scanner”) un document, il importe donc de connaître la quantité d’informations contenues dans l’original. En photographie, on appelle “pouvoir séparateur” la capacité d’un support à enregistrer plus ou moins d’informations. L’unité utilisée pour qualifier cette propriété est la paire de lignes / mm. Les émulsions photographiques courantes ont des pouvoirs séparateurs de 120 à 140 paires de lignes / mm pour le noir et blanc et de 45 à 50 paires de lignes / mm pour la couleur.

Taille des images numérisées

Quelle est donc la bonne résolution pour numériser un 24 x 36 ?

Au premier abord, une plus haute résolution fournissant une meilleure définition, il semble préférable d’opter pour la définition la plus haute afin d’obtenir la meilleure qualité lors d’un fort agrandissement. En réalité, deux éléments viennent tempérer cette façon de voir les choses. D’abord, il ne sert à rien d’emmagasiner plus d’informations qu’il n’y en a dans l’original. Ensuite, il faut considérer que toute image comporte à la fois des informations objectives (le signal) et des parasites (le bruit). La qualité de numérisation dépend de ce rapport signal / bruit.



En photographie, le bruit est apporté par le grain de la pellicule. L'augmentation de la résolution permet d'extraire plus de signal mais également plus de bruit. Au delà d'un certain seuil, on atteint le maximum de l'information objective alors que le bruit continue à croître. Le but est de trouver le bon compromis : le maximum d'informations pour un minimum de bruit.

Pour un 24 x 36 couleur, cette zone de bon rapport signal / bruit se situe entre 18 et 25 Mo par image. Plus de définition n'accroît pas la qualité du fichier car le bruit, trop présent, devient une cause de problèmes : pour l'utilisation de filtres améliorant le piqué (USM, accentuation dans Photoshop), pour la correction des couleurs et pour la compression.

Un "gros" fichier ne garantit donc pas systématiquement des résultats optimaux en exploitation. Ainsi, un fichier de 9 Mo issu d'un appareil numérique Canon D30 paraît moins "performant" qu'un fichier de 20 Mo obtenu en numérisant un document argentique. La réalité est inverse, car le fichier du D30 (à 100 / 200 ISO) contient bien moins de bruit que le fichier de l'image numérisée à 20 Mo. Ce fichier de seulement 9 Mo sera utilisable pour l'impression d'une double page dans un magazine avec une qualité identique à celle d'un fichier deux fois plus volumineux issu d'un scanner.

De plus, certains inconvénients annexes interdisent dans la pratique l'usage de trop gros fichiers : par exemple, une image de 72 Mo, réputée de "haute qualité", n'est pas "vendable" car intransportable dans un délai raisonnable, ce qui pose des problèmes face à une concurrence qui assure la "livraison" de fichiers plus légers par Internet. Si, à la place de ce fichier de 72 Mo, on opte pour un fichier de 22 Mo et qu'on accepte une compression JPEG en "haute qualité", on obtient une image d'environ 2 Mo qui devient "livrable" très rapidement et dont la qualité finale imprimée dans un magazine est optimale.

La compression

La compression des fichiers est indispensable pour le stockage et la diffusion des images. Deux méthodes de compression sont couramment utilisées : avec ou sans perte d'informations.

-1- La compression LZW, couramment associée au format Tiff, permet une compression des données sans aucune perte d'informations mais le taux de compression reste faible.

-2- La compression JPEG est, à ce jour, la plus utilisée. Son principe consiste à découper l'image en carrés (ou "matrices") de 8 pixels de côté. On applique alors une formule qui a pour effet de ne conserver que les valeurs les plus typiques de la matrice et de remplacer les autres par la différence qu'elles ont avec ces dernières. De ce fait, la majorité des valeurs de la matrice se trouvent être de très petits nombres.

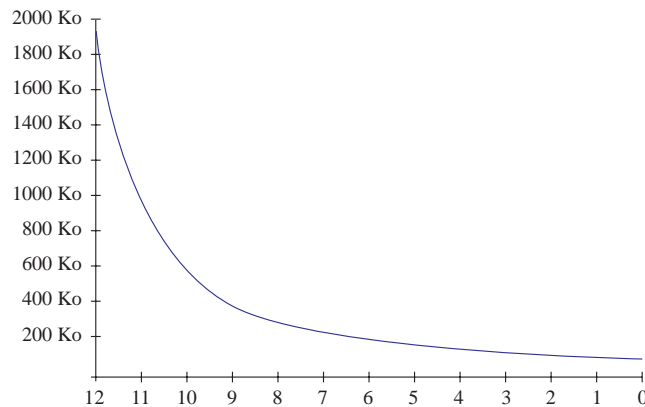
A ce stade de l'opération, on peut affecter un "coefficient d'assimilation" variable. On va ainsi décider que toutes les valeurs d'une matrice n'excédant pas le coefficient choisi sont remplacées par des zéros. Exemple : dans un ciel bleu, chaque pixel est très légèrement différent de son voisin. Dans l'option extrême, la compression JPEG résume les 64 pixels du carré de 8 x 8 pixels à un même bleu moyen.

Il faut donc d'accepter des pertes de détails, mais en limitant ces pertes uniquement à des informations que notre système visuel ne peut percevoir afin de conserver un rendu final sans différences apparentes. Attention, un coefficient de compression excessif produit une perte de détails qui devient visible. Remarque : un fichier qui contient du bruit est moins "compressible" qu'un fichier qui en est exempt. En effet le bruit génère, dans chaque matrice 8 x 8, des pixels ayant des valeurs très éloignées de la moyenne de la matrice.

Dans Photoshop, les coefficients de compression varient de 2 (forte compression / basse qualité) à 12 (compression minimum / qualité maximum).

La compression progresse fortement dans les coefficients élevés pour une qualité stable alors qu'en tendant vers 2, la compression devient bien moins spectaculaire pour une forte dégradation de l'image.

Efficacité de la compression JPEG



En pratique, on peut appliquer les règles admises par les professionnels :

Images sans dégradation : 12

Images d'art (tableaux) : 10

Images de publicité : 9

Images de magazines : 8 (c'est le choix des agences de presse)

Images de presse quotidienne : 7

Images pour la télévision : 6

Éviter les qualités inférieures sauf pour un usage uniquement réservé à Internet, domaine dans lequel la faible taille des fichiers est une quête permanente.

L'ouverture (la lecture) d'un fichier JPEG ne provoque aucune dégradation de l'image. Par contre, il est déconseillé de compresser plusieurs fois de suite la même image car les pertes de qualité se cumulent.

C'est pourtant inévitable : l'appareil photo délivre des fichiers compressés (compression assez faible) puis l'image subit un traitement (réglages, retouches). Elle est ensuite enregistrée et compressée une seconde fois pour être stockée ou expédiée, puis retravaillée par le destinataire et souvent recompressée une troisième fois pour l'envoi à l'impression.

Important : dans les dernières versions de Photoshop (depuis la version 5), seules les zones de l'image ayant subi des modifications sont recompressées, à condition toutefois d'enregistrer le fichier avec le même coefficient de compression que celui qui a été choisi lors du dernier enregistrement. Il est donc utile de connaître ce coefficient. Si l'image provient d'une agence de presse (coefficient 8), il faut recompresser le fichier à 8 au prochain enregistrement. Attention, un choix de coefficient 10, malgré la "haute qualité" présumée, provoquerait une recompression de la totalité du fichier, donc une détérioration inutile de l'image.

Remarque : un coefficient de compression qui produit à l'écran un aspect visuel satisfaisant peut s'avérer excessif à l'impression en faisant apparaître une pixellisation "en damier" typique d'une trop forte compression JPEG. Il ne faut donc pas se fier outre mesure à l'image écran pour juger de la déperdition.

Archivage

Avant le numérique, l'archivage, dans une agence photo, était géré par un documentaliste qui maîtrisait son rangement et il était difficilement remplaçable. Un départ en retraite signifiait la perte d'une très grande quantité d'informations. Par ailleurs, dans les agences d'illustration (Sygma-Corbis, Gamma-Hachette, Sipa...) seule une toute petite partie des images (de 3 à 8%) stockées sont réellement vendues.

L'intérêt du classement en numérique est évident : accès immédiat à l'information, stratégie de stockage, facilité de classement et de recherche. Une des clés essentielles d'un classement numérique efficace est la description textuelle des images. Une légende appliquée à chaque image permet de la retrouver soit par thèmes, soit par mots-clés, soit par n'importe quel terme de sa légende dans le cas de moteurs en texte intégral.

L'**IPTC** (International Press Telecommunication Comitee) est une norme euro-américaine qui a vu le jour en 1991 afin de standardiser la gestion des informations pour la transmission et le stockage des données numériques. Toutes les agences de presse l'utilisent.

Principe

Le format IPTC peut être considéré comme une enveloppe qui comporte toutes les indications nécessaires (expéditeur, destinataire, type de document...) à la distribution du fichier et qui contient :

- 1- **un objet** (fichier) qui peut être un texte, une photo, un film, un graphique ou un son. Chaque type d'objet a un format prédéfini, par exemple, pour la photo, l'objet doit être un fichier JPEG.
- 2- **Un texte** décrivant l'objet ainsi que ses conditions de réalisation. L'objet est ainsi indissociable de sa description : date, heure, ville, pays, auteur, crédit (droit), légende, catégorie (sport, politique...), etc...

Photoshop ne permet pas de créer de véritables fichiers au format IPTC, mais il autorise l'enregistrement des informations textuelles de description quelque soit le format de sauvegarde des images (TIFF, PICT, etc...). Pour produire avec Photoshop un fichier compatible avec la norme IPTC, il faut remplir les champs d'information puis enregistrer les images au format JPEG. Les fichiers ainsi produits seront "compris" par l'immense majorité des bases de données gérées par les agences photographiques et les journaux du monde entier.

Les mots clés doivent être exprimés dans un vocabulaire "réservé" bien défini établi par un documentaliste. Les catégories et sous-catégories de l'IPTC ont été définies pour une utilisation dans la presse. Ces choix, limités dans la variété, peuvent se révéler insuffisants dans d'autres domaines. En octobre 1999, l'IPTC recommandait d'inclure systématiquement un profil colorimétrique dans les images. L'espace choisi est le sRGB, choix qualitatif discutable, mais qui correspond au standard de visualisation sur PC (c'est à dire la majorité des cas).

Consultation

Les images d'agences, même compressées en JPEG s'avèrent trop "lourdes" pour être manipulées directement sur Internet. Il est donc nécessaire, pour la consultation en ligne, de générer des "sous images" (appelées imagerettes ou vignettes) en basse définition. La plupart des agences ont opté pour une définition de 192 x 192 pixels. Ce format carré permet de traiter aussi bien les images verticales qu'horizontales. L'imagerette est généralement codée en milliers de couleurs (16 bits au lieu de 24 bits) et compressée dans des qualités 6 ou 7. Elle présente une qualité suffisante pour une première sélection.

Pour un choix plus qualitatif, une seconde image (image de choix) est disponible, généralement en définition 640 x 640 pixels, en 24 bits (millions de couleurs) et compressée en JPEG qualité 7 ou 8. Insuffisante pour l'impression, cette qualité devient "piratable" pour des utilisations TV, web ou CD-Rom.

La protection des images devient alors indispensable. Deux méthodes de protection des images peuvent être envisagées :

- le watermark : empreinte transparente permanente qui apparaît à l'écran comme à l'impression
- la très forte compression des données qui génère une qualité d'image tolérable en affichage écran mais inacceptable à l'impression.

Gestion des données

Il existe deux types de logiciels permettant la gestion d'images :

-1- Les catalogueurs (Fotostation, par exemple)

Le classement est "traditionnel", et permet d'accéder directement aux différents dossiers dans lesquels sont stockées les images. On voit immédiatement l'analogie avec la boîte de photos placée dans un tiroir, puis dans une armoire, etc... Avec ce classement "analogique", la recherche se fait par les mots inclus dans la légende chaque image. Un catalogueur est un logiciel simple d'utilisation et peu coûteux, mais les temps de recherche peuvent se révéler très élevés quand beaucoup d'images sont consultées simultanément. On peut remédier à ce dernier inconvénient en lui adjoignant un logiciel gestionnaire d'index.

-2-les bases de données (Phraséa, par exemple)

Contrairement aux catalogueurs, ces types de logiciels prennent le contrôle de tous les objets qui leur sont intégrés. Cela permet une administration des données impossible avec un catalogueur. Les données étant centralisées, il est possible d'avoir un index automatiquement actualisé, ce qui réduit considérablement les temps de recherche des images.

Les bases de données peuvent être réparties en deux grandes familles :

- 1- Séquentielles indexées, (File Maker, par exemple)
- 2- En texte intégral comme la plupart des logiciels de recherche sur Internet

Stockage sécurisé

Lors de grands fonds iconographiques, les disques durs s'avèrent rapidement trop limités en capacité. De plus, il faut penser à une stratégie de sauvegarde (en plusieurs lieux distincts). Deux solutions sont actuellement très utilisées : les systèmes "Raid" permettant de gérer de manière sécurisée plusieurs disques durs simultanément et les juke-boxes de CD-Rom ou de DVD.