

A propos de la numérisation des documents...

Aspects pratiques

Rôle d'un scanner (numériseur)

- Traduire sous forme de bits une image en guettant un nombre plus ou moins grand de point par pouce (la taille de l'image peut être exprimée en pouces, 1 pouce = 2,54 cm)
- Distinguer un nombre plus ou moins important de niveaux de gris ou de couleurs (16, 256 niveaux de gris, 256, 65000, 16 millions de couleurs).

Accès au scanner

Un scanner ne se partage pas dans un réseau local, vous devez donc travailler sur le poste auquel est relié le scanner (même si la possibilité technique de numériser depuis un poste distant existait, cela présenterait un piètre avantage puisqu'il faudrait néanmoins déposer et retirer les documents posés sur la vitre du scanner).

Les logiciels de numérisation varient d'un scanner à l'autre. Nombreux sont ceux qui proposent une assistance conviviale, mais certains vont nécessiter une démarche active de l'utilisateur pour obtenir ce qu'il veut ! En effet, le but n'est pas seulement d'obtenir une image numérisée, mais l'image de qualité optimum compte tenu des contraintes de poids (taille en octets)

Nous utiliserons des copies d'écrans d'un processus mené avec un logiciel convivial (ScanWise logiciel livré avec un Snapscan 1212 de chez Agfa) pour illustrer certaines procédures liées à la démarche. Mais ponctuellement un comparatif sera réalisé avec Mirascan (livré avec certains scanners Acer et Guillemot) et un logiciel plus ancien DeskscanII.

1 Lancement du logiciel d'acquisition

Sur le bureau Windows, une icône représentative du scanner permet de provoquer l'exécution du programme de gestion de ce scanner. Voir des exemples ci-contre.

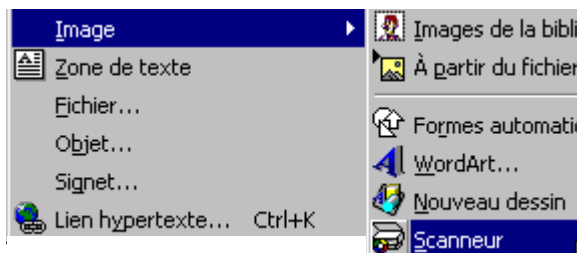


La totalité des nouveaux logiciels de gestion de l'acquisition d'informations via un scanner proposent un accès via les applications clientes ...

du type de PaintShopPro par exemple :



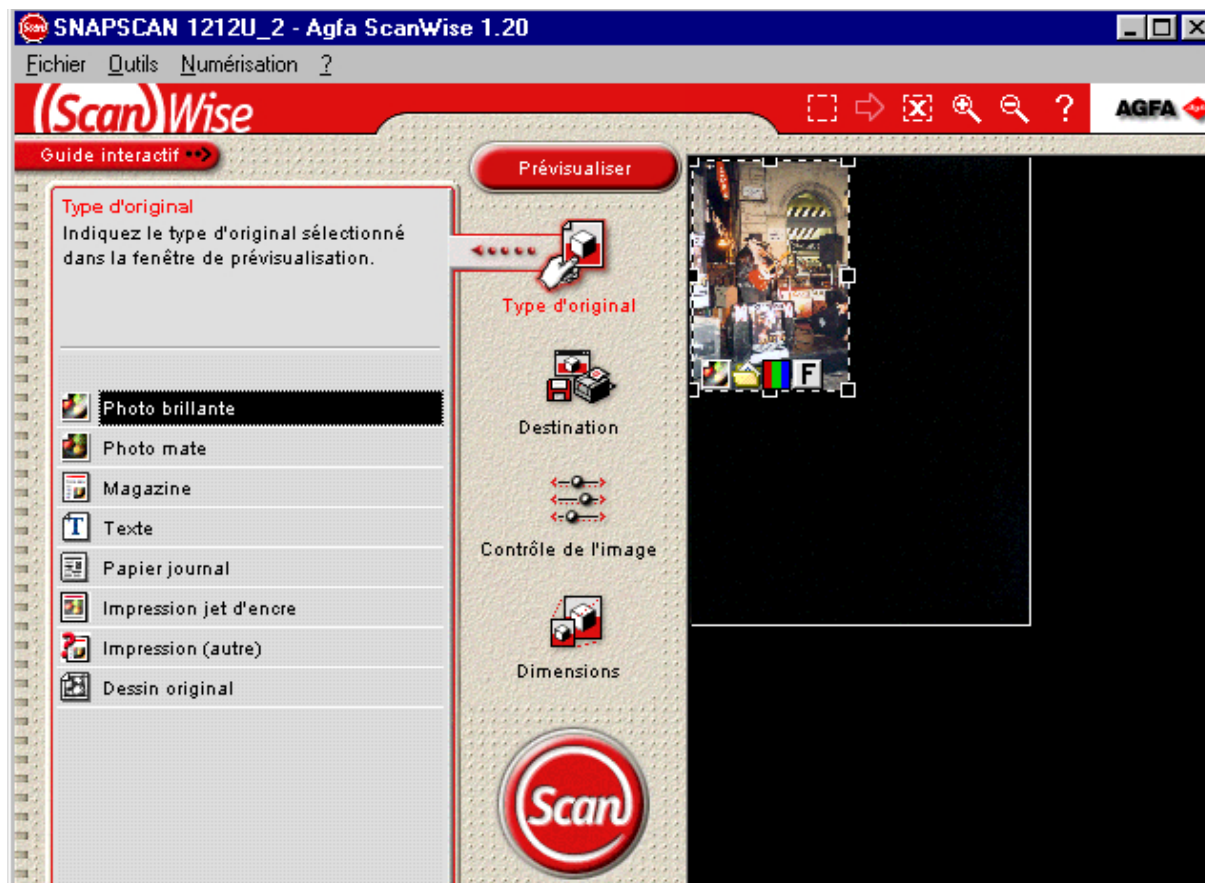
Ou encore depuis Word :



La fenêtre du programme de gestion du scanner s'affiche alors sur l'écran...

2 Choisir le type d'original

Les paramétrages par défaut du logiciel conviennent dans la plupart des cas, il est donc inutile de les modifier (notamment les réglages de luminosité/contraste qui pourront être réalisés ensuite dans un logiciel de retouche d'images tel Photoshop ou PaintshopPro).



Cependant l'image à numériser est bien différente selon qu'il s'agit d'un dessin au trait (un dessin d'élève en noir et blanc ou avec des aplats de couleurs) d'une image noir et blanc comportant un dégradé de niveaux de gris, d'une photographie réalisée avec un appareil photo ou d'une image de magazine ou une photographie présente dans un journal qui a déjà fait l'objet d'un tramage (elle a été reproduite en offset au moyen de lignes de points –chaque point ayant une couleur choisie parmi quatre : cyan, magenta, jaune, noir-) qui ne sont pas tout à fait ronds et qui sont orientés d'une manière bien particulière. Une impression obtenue avec une imprimante jet d'encre présente encore des caractéristiques différentes. **C'est pourquoi il est nécessaire de préciser le type d'original à numériser afin d'indiquer au logiciel le nombre de couleurs à « guetter »**, s'il devra « détarmer » l'image lorsque celle-ci est déjà une image imprimée (Journal : entre 65 et 85 lignes de points par pouce / Magazine : 120 à 133 / magazines de luxe : 175 à 200).

3 Choisir la résolution de numérisation

Il s'agit de choisir le nombre de points que doit « guetter » le scanner (qu'enregistrent les capteurs du scanner) en fonction de la destination de l'image.

Pour une image destinée à être imprimée en taille réelle sur une imprimante couleur, il est souvent souhaitable de numériser à 300 dpi ou ppp (entendez dot per inch ou points par pouce c'est à dire 2,54 cm), mais éventuellement à 600, voire 1200 et 2400 ppp si l'on dispose d'une imprimante qui sait réellement imprimer à cette résolution (les nouvelles imprimantes avec papier photo et cartouche photo permettent d'atteindre un rendu exceptionnel).

Pour une image destinée à être insérée dans un diaporama ou dans une page Web (pour une lecture écran donc) 100 dpi suffisent.

Plutôt que d'inviter l'utilisateur à choisir une résolution, ScanWise propose habilement de choisir la destination de la numérisation.



4 Prévisualiser et cadrer l'image

Un bouton de prévisualisation (Préscan ou Préview suivant les logiciels) permet d'afficher à l'écran une version de l'image en basse résolution (le scanner balaye rapidement l'image).

Il est alors possible de recadrer l'image en définissant un cadre (ou plusieurs suivants les logiciels) autour de la partie de l'image que l'on souhaite effectivement conserver (faire glisser les poignées du cadre de sélection pour le redimensionner). Exemple sous scanWise :



5 Vérifier le « poids » de la sélection et numériser

Les différents logiciels de gestion de scanner permettent de voir le poids (en terme d'encombrement en octets) de la partie d'image sélectionnée.

Ici sous ScanWise :

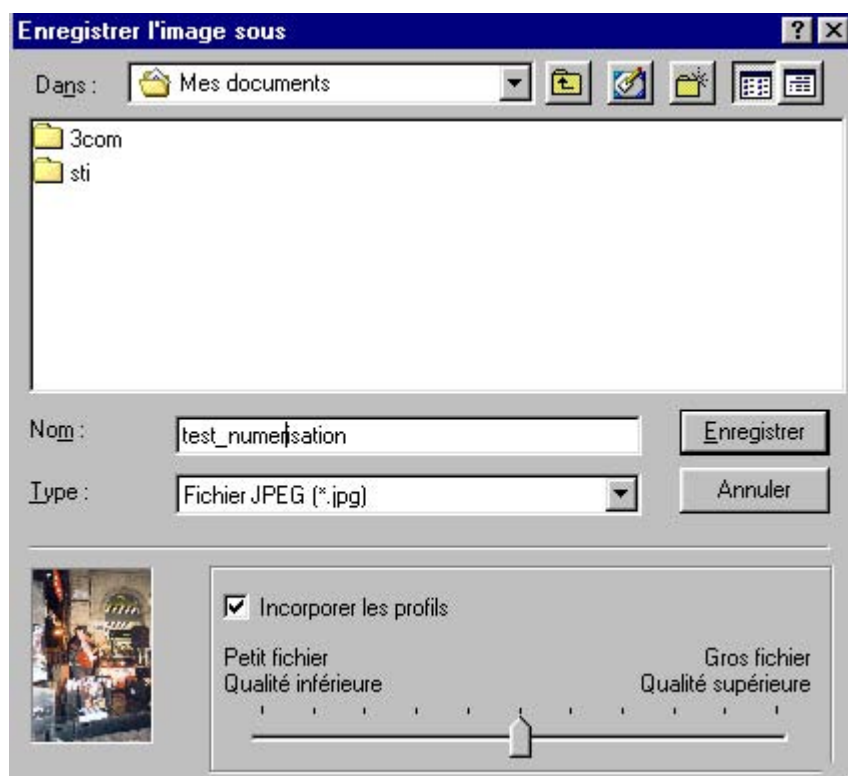
The screenshot shows the ScanWise software interface. On the left, a panel titled 'Dimensions' contains settings for 'Echelle' (Scale) set to 100%, 'Taille' (Size) with width 9.9 and height 14.7, and 'Unités' (Units) set to 'Centimètres'. Below this, it shows 'Résolution d'image : 72 dpi', 'Taille de l'image numérisée : 343,7 Ko', and 'Résolution de numérisation : 72 dpi'. On the right, there's a 'Prévisualiser' button, a preview of the scanned image, and a 'Contrôle de l'image' section with sliders for luminosity and contrast. A red 'Scan' button is at the bottom right. Annotations with arrows point to various elements: one points to the 'Echelle' slider with the text 'Image numérisée pour affichage écran, 72dpi. La taille reste importante. Si l'image doit être diminuée en taille dans le document qui la recevra, il peut être judicieux d'utiliser un taux de réduction.'; another points to the 'Taille' fields; a third points to the 'Résolution d'image' field; a fourth points to the 'Contrôle de l'image' sliders with the text 'Inutile de modifier la luminosité, le contraste, mieux vaut utiliser un logiciel de retouche d'images'; and a fifth points to the 'Scan' button with the text 'Pour lancer la numérisation'.

La boîte de dialogue invitant à l'enregistrement s'affiche (sauf si le logiciel d'acquisition a été ouvert depuis une application hôte comme dans le cas d'une insertion dans Word par exemple)...

Les images sont enregistrées dans un format non compressé si s'agit d'images ayant vocation à être insérées dans une publication papier, ou dans un format compressé GIF ou JPG le plus souvent lorsqu'il s'agit d'images destinées à être affichées sur un écran comme dans le cas des pages Web (ne pas confondre « compressées » avec « réduites en taille », il s'agit ici d'algorithmes qui réduisent le poids de l'image en la dégradant de manière plus ou moins visible à l'œil).

Remarques

- Le dossier qui contiendra les « bruts de scan » sera partagé (au sens du réseau) et donc accessible depuis le poste de l'utilisateur. Afin d'éviter de saturer le disque dur de l'ordinateur « conduisant le scanner », le contenu de ce dossier sera effacé à la fin de chaque séance de travail. Il convient donc de s'assurer d'avoir rapatrié ses documents sur son propre poste de travail.
- Au fait avez-vous pensé à reprendre votre original sur la vitre du scanner ?

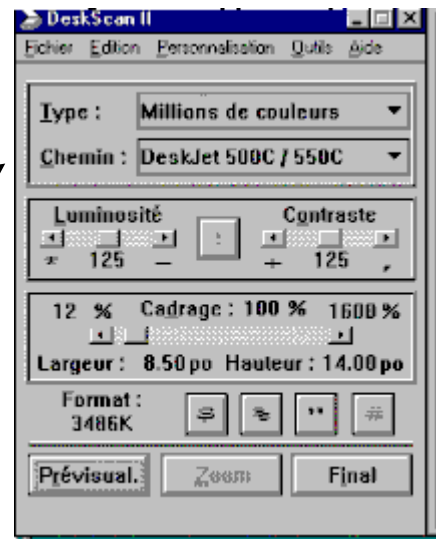


6 Le logiciel d'acquisition peut nécessiter un paramétrage

Attention donc si votre logiciel ne sais pas automatiquement s'adapter au type d'original, il faudra lui spécifier comment il doit travailler.

Avec DeskScan II, il convient de choisir

- le type d'image (dessin noir et blanc, niveaux de gris, millions de couleurs par ex)
- la destination de l'image (imprimante ou écran)



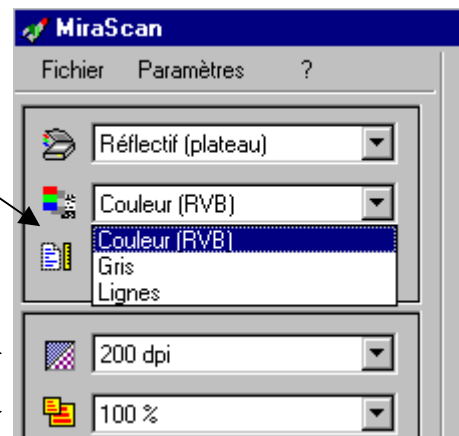
Les options de personnalisation permettent de paramétrer le logiciel finement en prenant en compte le type d'image initial et les possibilités du dispositif d'impression dont dispose l'utilisateur (on définit un « chemin d'impression » à l'aide d'un assistant qui invite l'utilisateur à numériser une photographie puis à l'imprimer et enfin à re-numériser l'impression obtenue ; le logiciel corrige alors les paramètres de numérisation pour minimiser les écarts entre l'original et la copie lors de la prochaine numérisation.

Mirascan quant à lui, nécessite que l'utilisateur précise le type d'image (photographie en couleur, niveaux de gris ou dessin au trait),

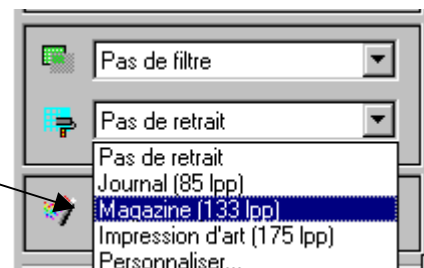
Il convient de spécifier le nombre de points par pouce que l'on souhaite faire « guetter » par le scanner

Il faut préciser l'échelle désirée,

si on numérise une image à 200% elle contiendra quatre fois plus de points qu'une image à 100% car les capteurs retiendront deux fois plus d'informations en largeur et en hauteur.



Si l'image est celle d'un magazine, il faut inviter le logiciel à déramer la photographie pour éviter les effets de moirage



Attention aux agrandissements !

Une image de un pouce de côté numérisée à 100ppp donne un rendu satisfaisant sur un écran. Il n'en va pas de même si l'image doit être agrandie. Si l'image est affichée dans un « carré » de 2 pouces sur deux, il n'y a

pas plus de points pour autant qui définissent l'image, l'image s'affiche à l'écran avec une résolution de 50ppp ce qui est insatisfaisant.

*Inversement, si l'on dispose d'une photographie de 125mm au carré et qu'on souhaite l'imprimer dans un document dans une réserve de 100mm avec une résolution de sortie de 75 ppp, il suffira de numériser à 90 ppp ($100 / 125 * 95 = 75$). Cette image sera donc stockée sur le disque avec moins d'informations sur un format de 125mm et réduite ensuite dans la composition.*

Le nombre de couleurs d'une image

Indépendamment du nombre de points, le poids d'une image (taille du fichier) est proportionnel au nombre de niveaux de gris ou de couleurs identifiés.

Pour identifier la « couleur » d'un point d'une image « en noir et blanc », il suffit d'un bit (2 états : le point est noir ou est blanc). Pour une image en 256 couleurs (ou niveaux de gris), il faut un octet (8 bits donc 256 possibilités possibles de codage), pour identifier 65536 couleurs il faut 2 octets (2^{16} possibilités) et pour 16 777 216 couleurs il est nécessaire de coder la couleur sur 24 bits (2^{24} possibilités).

A propos de la taille des fichiers

Si l'on double la résolution de numérisation, on quadruple la taille du fichier.

Facteurs intervenants :

taille réelle H x L

nombre de niveaux de gris en bit par pixel


(n=1 si Noir et blanc, 8 pour 256 niveaux de gris, 24 pour la couleur)

Taille en Kilo octets = $[(H \times n) \times (L \times n) \times \text{nombre de bits de couleur}] / (8 \text{ bits par octet} \times 1024)$

Une image définie par 600x800 points, (mode d'affichage des écrans les plus courants), occupera en 24 bits (16 millions de couleurs) une place de $=800 \times 600 \times 24 \text{ bits}$ soit 11 520 000 bits soit 1440000 octets.

Il va de soit qu'avec des modems rapatriant les images via une ligne téléphonique à 4ko par seconde (ce qui constitue déjà une performance) , il faudrait 6 minutes environ ($=1440000 / [1024 \times 4 \times 60]$). Les éditeurs de pages html convertissent les images dites Bitmap en images GIF (si elles ont 256 couleurs) et sinon en JPG. Les images sont obtenues avec des algorithmes de compression (taux de compression variable en JPG) et donnent des résultats très satisfaisants pour l'écran.

*Exemple du poids d'une photographie 100*125mm convertie en différents formats. L'image Gif ne contient que 256 couleurs, mais en lecture à l'écran, ces trois images « flattent » autant l'œil l'une que l'autre si on ne cherche pas à les imprimer ou à les agrandir.*

Nom	Taille
 brian.bmp	2 794 Ko
 brian.gif	587 Ko
 brian.jpg	117 Ko

Notes personnelles :

Quelques approfondissements

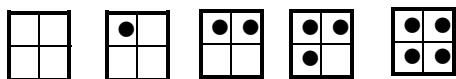
Le nombre de points par pouce retenus pour numériser l'image

Sur un pouce (soit 2,54 cm) on peut « guetter » avec un scanner 100, 200, 300, 600, 1200 points par pouce (ppp) sur une image. Attention, si l'on double le nombre de point « guettés » on quadruple le poids de l'image en nombre de points.

Exemple : images de 5cm*5cm donc 2 pouces de largeur et 2 pouces de hauteur. Si je cherche à identifier 300ppp, l'image sera composée à raison de 300ppp de $(2 \times 300) \times (2 \times 300) = 360\,000$ points. Si la numérisation est faite en 600ppp elle sera composée de $(2 \times 600) \times (2 \times 600) = 1\,440\,000$ points !

Comment une imprimante noir et blanc « réinvente les niveaux de gris » ?

Pour un bon rendu sur une imprimante noir et blanc, il n'est pas nécessaire de numériser plus de 200 points de gris par pouce car l'imprimante ne sachant faire que du noir ou rien, elle doit composer des points de gris en plaçant plus ou moins de points noirs dans une zone donnée.



Si l'on assemble des points noirs sur une matrice carrée occupant deux positions en colonne et deux positions en ligne, on peut simuler 2^2 tons de gris plus le blanc. On dira que l'on crée des points de trame.

L'imprimante ne pourra en imprimer que 300/2 soit 150 sur un pouce de hauteur. On dira que l'image est reproduite avec une linéature¹ de trame de 150ppp (cette fois-ci, ce sont des points de gris). Mais une image à tons continus, (une photographie), donnera l'impression d'être reproduite à tons continus si l'on dispose d'au moins 16 niveaux de gris. Pour cela, il faut utiliser une matrice de points d'impression de 4 sur 4, ce qui permet 16 possibilités de remplissage plus le blanc.

Si l'on utilise une imprimante à 300 ppp, on ne pourra donc obtenir qu'une linéature de $300/4=75$ lpi (lignes par pouce), on dira une trame 75 (75 lpi, lignes par pouce).

Pas de problème a priori, mais ces linéatures de trame doivent être comparées à celles utilisées dans les documents qui nous entourent :

Journal : entre 65 et 85 lpi / Magazine : 120 à 133 lpi / Impression de magazines: 175 à 200 lpi

- D'où l'intérêt d'avoir une imprimante à 1200ppp pour distinguer 64 niveaux de gris (matrice de 8x8) avec une linéature de trame de 150 lpi.
- Les imprimantes couleurs quant à elles, même à 600ppp, permettent un excellent rendu de type magazine moyennant des procédés complexes de répartition des points d'encre, à condition d'utiliser le papier adéquat et des cartouches d'encre de qualité.

Attention aux agrandissements !

Une image de un pouce de côté numérisée à 100ppp donne un rendu satisfaisant sur un écran dont la résolution est en général de 80ppp. Il n'en va pas de même si l'image doit être agrandie. Si l'image est affichée dans un « carré » de 2 pouces sur deux, il n'y a pas plus de points pour autant qui définissent l'image, l'image s'affiche à l'écran avec une résolution de 50ppp ce qui est insatisfaisant.

*Inversement, si l'on dispose d'une photographie de 125mm au carré et qu'on souhaite l'imprimer dans un document dans une réserve de 100mm avec une résolution de sortie de 75 ppp, il suffira de numériser à 90 ppp ($100 / 125 * 95 = 75$). Cette image sera donc stockée sur le disque avec moins d'informations sur un format de 125mm et réduite ensuite dans la composition.*

¹ - la linéature : le nombre de points gris que la machine peut simuler en sachant qu'il faut assembler plusieurs points noirs pour faire un point de gris : Par exemple sur une laser à 300 ppp, une trame de 60 lpi, (matrice 5x5), permettra de rendre compte de $5 \times 5 = 25$ niveaux de gris plus le blanc. Sur une photocomposeuse à 2500ppp, on pourra rendre compte de 256 niveaux de gris avec une matrice 16x16 avec une linéature de 150 lpi ($2500/16$).

A propos des dessins au trait

(dessins en noir et blanc représentant par exemple un tableau avec ses bordures)

Une résolution de numérisation de 600ppp est vivement conseillée (en ne guettant que deux niveaux de gris : le blanc et le noir), mais même dans ce cas, les lignes verticales de schémas ne seront pas parfaites (effets d'escalier), il faudrait numériser à 1200ppp au moins pour avoir un rendu de bonne qualité.

Le nombre de couleurs d'une image

Indépendamment du nombre de points, le poids d'une image (taille du fichier) est proportionnel au nombre de niveaux de gris ou de couleurs identifiés.

Pour identifier la « couleur » d'un point d'une image « en noir et blanc », il suffit d'un bit (2 états : le point est noir ou est blanc). Pour une image en 256 couleurs (ou niveaux de gris), il faut un octet (8 bits donc 256 possibilités possibles de codage), pour identifier 65536 couleurs il faut 2 octets (2^{16} possibilités) et pour 16 777 216 couleurs il est nécessaire de coder la couleur sur 3 octets (2^{24} possibilités).

A propos de la taille des fichiers

Si l'on double la résolution de numérisation, on quadruple la taille du fichier.

Facteurs intervenants :

taille réelle H x L

nombre de niveaux de gris en bit par pixel

(n=1 si Noir et blanc, 8 pour 256 niveaux de gris, 24 pour la couleur)

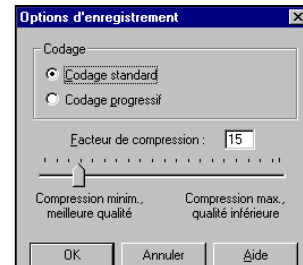
Taille en Kilo octets = $[(H \times n) \times (L \times n) \times \text{nombre de bits de couleur}] / (8 \text{ bits par octet} \times 1024)$

Une image définie par 600x800 points, (mode d'affichage des écrans les plus courants), occupera en 24 bits (16 millions de couleurs) une place de $=800 \times 600 \times 24 \text{ bits}$ soit 11 520 000 bits soit 1440000 octets.

Il va de soit qu'avec des modems rapatriant les images via une ligne téléphonique à 4ko par seconde (ce qui constitue déjà une performance) , il faudrait 6 minutes environ ($=1440000 / [1024 \times 4 \times 60]$). Les éditeurs de pages html convertissent les images dites Bitmap en images GIF (si elles ont 256 couleurs) et sinon en JPG. Les images sont obtenues avec des algorithmes de compression (taux de compression variable en JPG) et donnent des résultats très satisfaisants pour l'écran.

Exemple d'une copie d'écran en 800x600 (écran réglé sur 16 millions de couleurs) : bmp 1407Ko / gif 44 Ko / jpeg 168Ko

Ici l'image a été compressée en JPG avec un facteur de compression donnant une bonne qualité d'image, mais le facteur de compression peut être augmenté, comme par exemple ici sous PaintShop pro 5 :



Remarques diverses

- **Attention aux originaux brillants** : réduire le contraste sur le logiciel de numérisation.
- **Dessin en demi-tons** (image ou photographie déjà tramée, sur un magazine, un livre) : Risques de moirage, pour l'éviter augmenter la résolution de numérisation ou à défaut réduire le contraste. Si la texture des documents est grossière, couvrir l'original d'une feuille de plastique mate.
- Avec une image tramée ou en demi-teintes, on essayera toujours d'avoir **quelques points blancs dans le "noir" le plus foncé de la photo et quelques points noirs dans le blanc afin d'éviter l'empâtement dans les noirs et des visages qui manqueraient de relief dans les blancs**. C'est pourquoi de nombreux logiciels de retouche d'images propose des outils permettant de corriger les teintes extrêmes afin d'avoir 5% de noir dans le blanc et inversement.

Taille, Résolution, Echelle, poids

La numérisation depuis le logiciel fourni avec le numériseur

Image	Taille en mm	Résolution de numérisation (ppp ou dpi)	Echelle	Poids	Remarque
photo	30x30	100	100%	40Ko	Pour une image dont la destination est une page Html
photo	30x30	100	200%	160Ko	
photo	30x30	300	100%	359Ko	Pour une image à imprimer sur papier normal
photo	30x30	300	200%	1 437Ko	
photo	30x30	1200	100%	5 746 Ko	Pour une image à imprimer avec une cartouche photo sur du papier photo

Si l'image finale est affichée avec une échelle inférieure à 1...

Si l'image que l'on souhaite obtenir au final (dans une page Html, sur le papier...) doit avoir une taille moindre que l'original, on peut réduire le nombre de points guettés dans la même proportion (si on utilise par la suite l'image à 50% de sa taille, guetter 150 ppp sera suffisant puisque au final on disposera de 300 points d'informations par pouce).

Si l'on doit par contre afficher l'image finale avec un facteur d'échelle supérieure à 1...

En général, les numériseurs (scanners) ont des qualités optiques qui leur permettent de "guetter" un nombre de points maximum, par exemple 600dpi, pour un numériseur donné. Ils peuvent cependant, par interpolation, recalculer des points (par des algorithmes mathématiques plus ou moins habiles, à l'aide de "moyennes" plus ou moins complexes, ils recalculent par exemple la couleur d'un point entre deux points connus). Il est cependant préférable de laisser faire ce calcul à un logiciel de traitement d'image tel Adobe Photoshop ou Paintshop pro. Ces logiciels disposent de nombreux algorithmes plus "efficaces" (plongez-vous dans l'aide du logiciel pour y interpréter les nombreuses subtilités).

Si l'on souhaite agrandir l'image obtenue dans un traitement de texte par exemple, il faut augmenter d'autant le nombre de points "guettés", mais il serait maladroit de dépasser la résolution optique du numériseur...

Pour cette raison, mais aussi parce qu'il est souvent nécessaire de retravailler l'image obtenue par numérisation (recadrage, modification des couleurs, insertion de texte etc.), en général on numérise à partir d'un tel logiciel. Exemple sous Paintshop pro :

Fichier → importer → Twain → acquérir

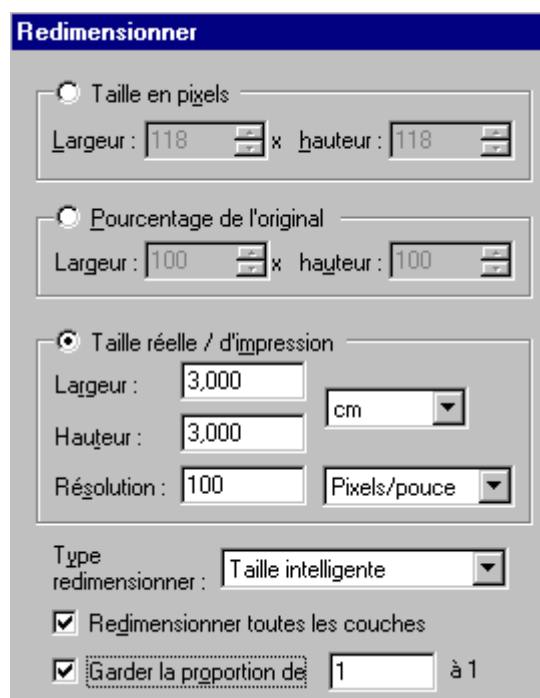
Cette commande permet d'ouvrir le logiciel de numérisation fourni avec le numériseur.

Une fois la numérisation réalisée, l'image est stockée en mémoire et apparaît² dans une fenêtre document de Paintshop pro.

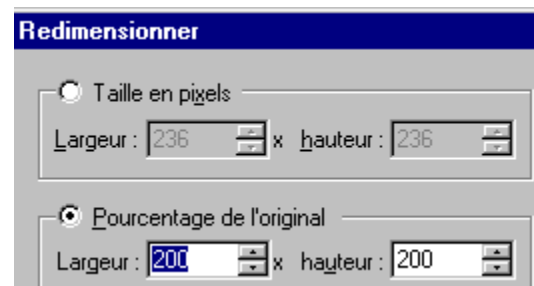
Il est possible de l'agrandir en utilisant la commande : **Image → redimensionner**

Pour une image 30x30mm numérisée à 100ppp :

- L'image comporte 118 pixel au carré
- La taille réelle d'impression peut être modifiée.
- Dans la case Type de redimensionnement, faites défiler la liste pour sélectionner la méthode de redimensionnement que vous désirez utiliser (en fait, en général, si vous ne cherchez pas à optimiser en ayant lu la documentation de Paintshop pro, utilisez l'option "taille intelligente" proposée par défaut !



- Si l'on double la taille de l'image, le nombre de points double en largeur et hauteur, l'image a donc un poids multiplié par quatre...



Attention :

Il convient de distinguer cette manipulation de l'agrandissement autorisé par la **commande Fichier → configuration de page** qui permet de configurer la mise en page et l'échelle d'impression :

Attention, l'action sur cette commande ne fait qu'agrandir la taille de l'image imprimée et ne crée pas de nouveaux points, chaque point d'information sera simplement imprimé sur une surface plus grande de la feuille.

² - Attention, la taille de l'image affichée à l'écran peut apparaître plus grande que la photographie originale. Cela peut provenir d'un facteur de zoom (affichage → zoom) mais aussi simplement parce que Paintshop pro affiche l'image en pixel, ainsi une image de 30mm numérisée à 300ppp, contient 350 points en largeur et hauteur affichés sur 350 pixels, donc occupe presque la moitié de votre écran en largeur si vous disposez d'un écran configuré en 800 pixels de large ! Mais l'aperçu avant impression montre l'image dans sa taille d'impression, et cette taille reste la même quelque soit la résolution de numérisation (c'est la commande de mise en page -voir infra- qui permet d'en changer la taille).