

GUIDE D'UTILISATION DE SCANTAILOR

PAR ALAIN TRENTIN POUR [RETRONIK](#) (CE DOCUMENT EST LARGEMENT INSPIRÉ DU [WIKI](#) DU PROJET)

1 À PROPOS

ScanTailor est un outil interactif de post-traitement des pages numérisées. Utilisé principalement pour numériser des livres et des revues, il permet, entre autres, de séparer des pages ou d'en rogner des parties, de compenser l'angle d'inclinaison et de sélectionner ou pas des parties de contenu et des marges. En partant de scans bruts, vous obtenez des fichiers au format tiff prêts pour l'impression ou l'assemblage dans un fichier PDF ou DjVu

La numérisation, la reconnaissance optique des caractères (OCR) et la création de documents de plusieurs pages ne font pas partie des objectifs du projet.

Des versions du programme existent pour Windows, GNU / Linux et Mac OS X.

2 INSTALLATION ET PREMIER DÉMARRAGE

2.1 POUR WINDOWS

Les versions pour Windows se présentent sous la forme d'un programme d'installation.

Installez le logiciel en suivant le menu qui est intuitif. Les liens de téléchargement valables à ce jour sont ci-dessous

[32-bit installer 0.9.11.1](#)

[64-bit installer 0.9.11.1](#)

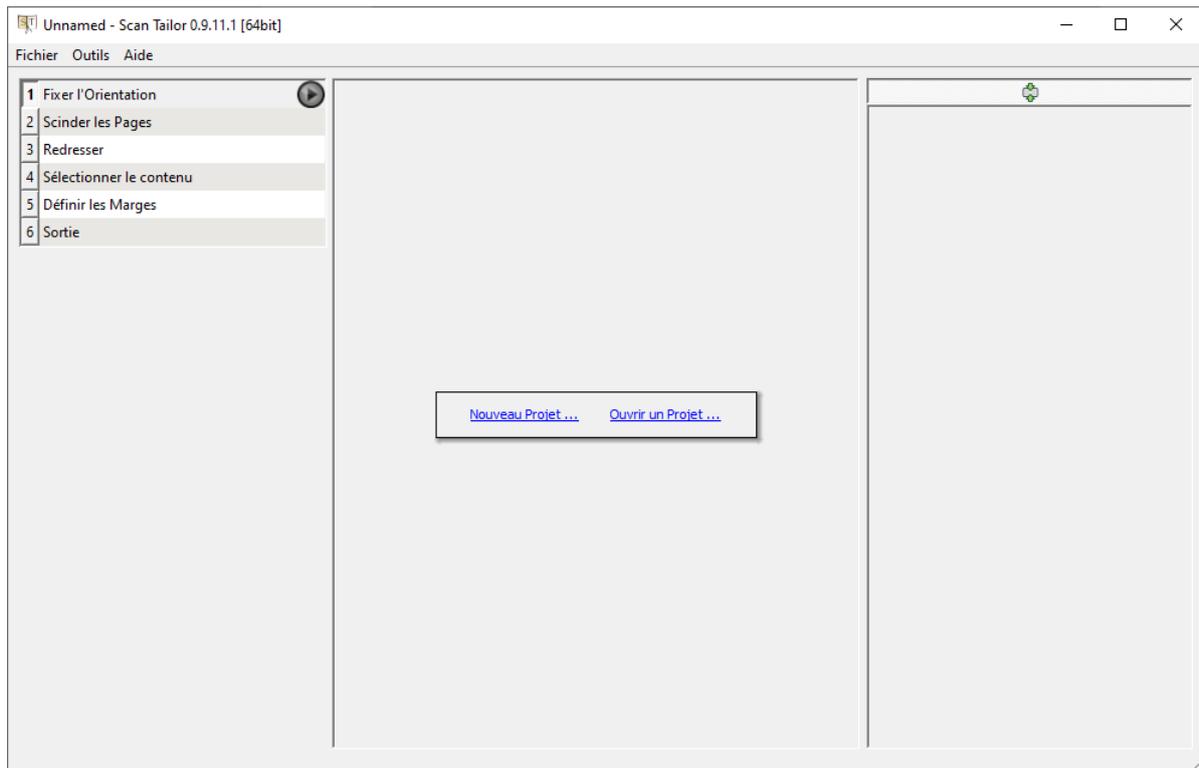
2.2 POUR GNU / LINUX ET MAC OS X

Si votre distribution n'a pas ScanTailor dans ses dépôts ou référentiels, vous devez compiler à partir des sources. Après la compilation, vous pourrez exécuter ScanTailor via Alt + F2, en tapant la commande "scantailor" (sans les guillemets), ou si vous n'avez pas d'environnement Gnome, vous pouvez taper cette même instruction sur la ligne de commande.

Vous pouvez consulter les instructions de compilation en [annexe 1](#)

Sur Mac OS X, le programme doit actuellement être exécuté à partir du terminal après la compilation et l'installation.

Une fois lancée, la fenêtre principale du programme ressemble à ceci :

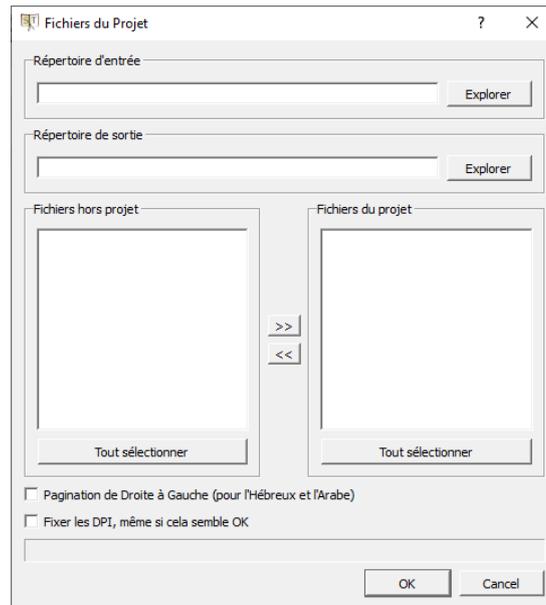


Dans le panneau central, nous voyons les éléments "Nouveau projet ...", "Ouvrir un projet ..." et, le cas échéant, une liste des projets récents.

2.3 CRÉATION D'UN PROJET

Pour créer un premier projet, nous avons besoin de la matière première - ScanTailor nécessite la présence d'au moins un fichier image dans le projet. Ces images peuvent être issues d'une numérisation ou de photographies des pages d'un livre ou d'un magazine. Voyez les conseils de numérisation du [chapitre 6](#) pour obtenir les meilleurs résultats.

Cliquez sur "Nouveau projet ..." – ce qui ouvre la boîte de dialogue "Fichiers de projet".



"Répertoire d'entrée" - le dossier dans lequel se trouvent les scans originaux.

« Répertoire de sortie » - le dossier dans lequel les résultats du traitement sont stockés.

« Fichiers hors projet » - liste des fichiers se trouvant dans le dossier spécifié dans le champ répertoire d'entrée, mais pas encore ajoutés au projet ou fichiers qui ont été supprimés du projet.

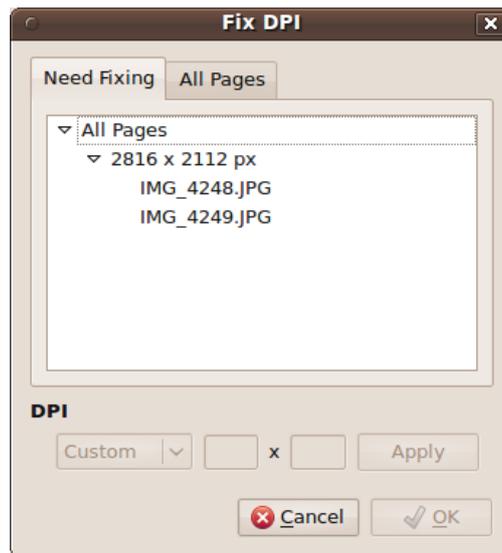
"Fichiers du projet" - fichiers ajoutés au projet,

Nota : Le projet peut prendre des fichiers de différents répertoires. Vous pouvez sélectionner des fichiers dans un répertoire, puis changer de répertoire d'entrée, sélectionner des fichiers à partir de ce dernier, etc.

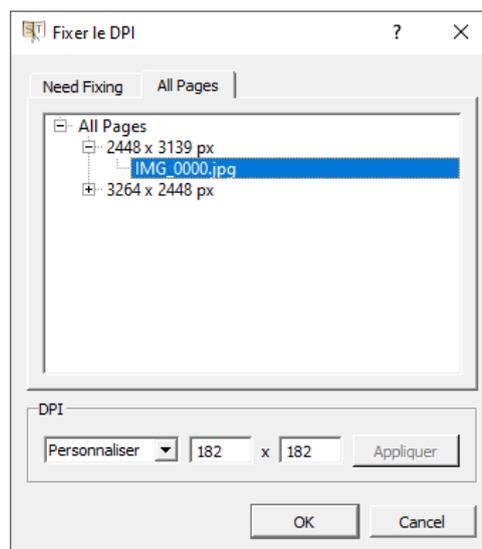
Les boutons ">>" et "<<" entre "Fichiers hors projet" et "Fichiers du projet" déplacent les fichiers sélectionnés d'une liste à une autre. Il est possible de sélectionner tous les fichiers via "Tout sélectionner", ou individuellement - via Ctrl + Clic, ou une plage - par Maj + Clic selon les procédures de sélection standard.

Les types de fichiers pris en charge sont * .tif, * .tiff, * .png, * .jpg, * .jpeg

Une fois que vous avez sélectionné les fichiers à utiliser dans le projet, cliquez sur OK. Habituellement, à ce stade, le processus de création du projet se termine, mais si des fichiers de projet avec un DPI non spécifié ou clairement incorrect ont été sélectionnés, une fenêtre de dialogue "Fix DPI" s'ouvre. Consultez la FAQ en annexe 2 pour en savoir plus sur le concept DPI.



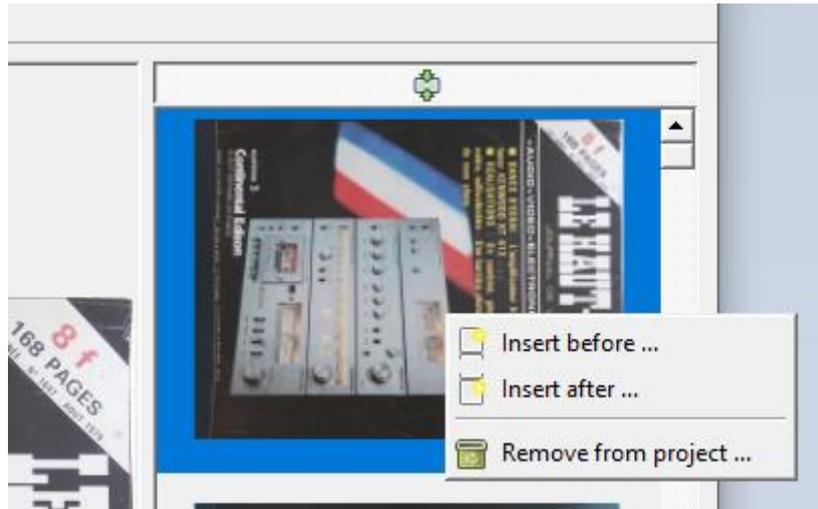
L'onglet "Need Fixing" (Besoin de réparation) répertorie uniquement les fichiers dans lesquels le DPI n'est pas spécifié ou est clairement erroné. Dans l'onglet "All Pages" (Toutes les pages) se trouve la liste de tous les fichiers du projet. On peut également en modifier le DPI.



Si nous savons que tous les fichiers du projet sont de 300 x 300 DPI, alors on peut aller définir le DPI pour tous les fichiers. Pour ce faire, mettez en surbrillance la ligne qui dit "All pages" (Toutes les pages) avec le triangle / flèche à côté (pas l'onglet qui dit Toutes les pages), sélectionnez le PPP correct et cliquez sur Appliquer.

Vous pouvez spécifier le DPI pour des groupes de fichiers avec la même taille de pixel, ainsi que pour des fichiers individuels. Les fichiers corrigés disparaissent de l'onglet "Need fixing". Lorsque l'onglet est complètement vide, cliquez sur OK. Le processus de création du projet sera ainsi terminé.

Nota : À toutes les étapes du traitement que nous verrons par la suite, vous aurez la possibilité d'ajouter / supprimer des fichiers du projet à l'aide d'un menu contextuel, sélectionné via un clic droit de la souris dans la fenêtre d'aperçu (les petites vignettes)

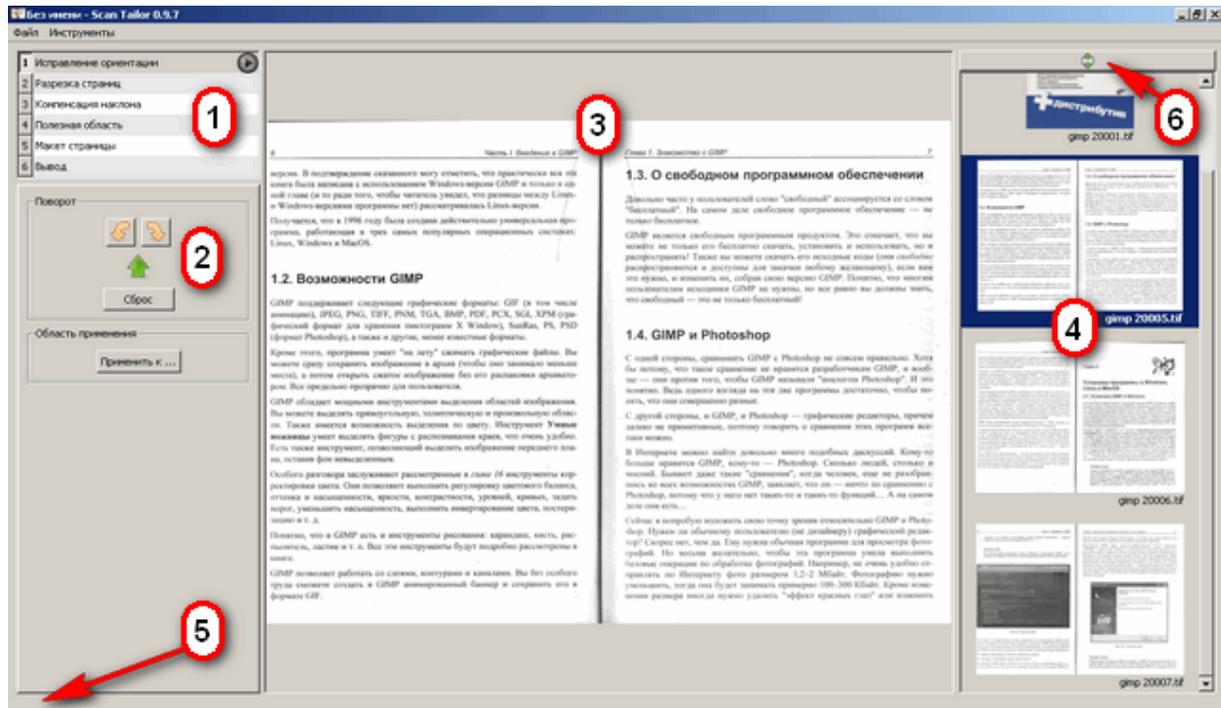


Lorsque vous essayez d'ajouter un fichier avec un DPI non valide, ScanTailor appelle la boîte de dialogue Fix DPI. Si vous souhaitez modifier le DPI d'un fichier dans le projet déjà créé, appelez Fix DPI via le menu « outils ». Les paramètres PPP que vous spécifiez ne sont pas écrits dans les fichiers - ils sont uniquement stockés dans le fichier de projet ScanTailor.

3 LE CONCEPT DE TRAITEMENT DES SCANS DANS SCANTAILOR

Tout d'abord, vous devez comprendre le concept général du programme. Le traitement des pages numérisées dans ScanTailor est comme un convoyeur d'usine. Il y a des étapes, dont chacune traite d'une manipulation spécifique de l'image pour produire le produit final.

Regardons la fenêtre principale du programme:



(1) Liste des étapes,

(2) Options pour l'étape de traitement en cours,

(3) Zone de travail principale qui affiche l'image actuelle, ainsi que des outils pour la manipuler,

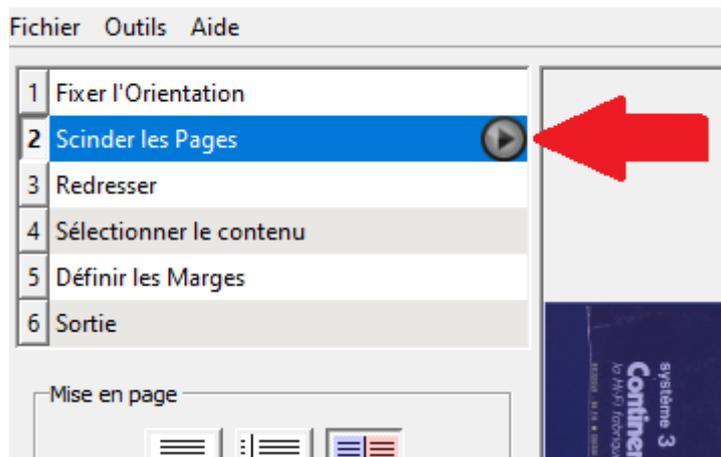
(4) Bande de prévisualisation,

(5) Info-bulles (ici vous pouvez voir des invites pour savoir comment utiliser l'outil ou l'élément sous le pointeur de la souris),

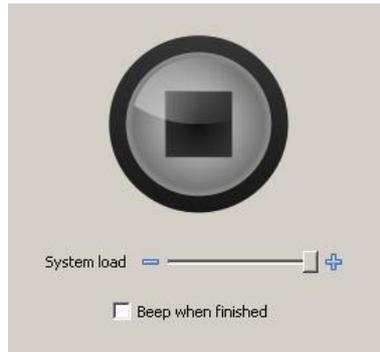
(6) Bouton "Garder la page active dans le champ de vision" (le fait d'appuyer sur ce bouton amène le programme à afficher la page active sur la bande de prévisualisation).

La bande de prévisualisation peut être considérée comme un convoyeur. Il est important de comprendre que, comme sur une chaîne d'assemblage classique, si le produit a atteint un certain stade, toutes les étapes précédentes sont censées avoir été réalisées dans un ordre strict. Par exemple, si vous entrez dans l'étape "Select Content" (Sélectionner le contenu) immédiatement après le lancement du programme, la page sur laquelle vous travaillez sera traitée (de manière invisible) en passant par toutes les étapes précédentes (c'est-à-dire «Diviser les pages» et «Redresser»). À tout moment, vous pouvez revenir à l'une des étapes précédentes pour voir et corriger l'opération qui a été effectuée sur l'image à ce stade.

Le bouton rond "Play" à droite du nom de chaque étape exécute ce traitement par lots (pour toutes les pages), comme pour démarrer le convoyeur décrit ci-dessus.



En appuyant sur le bouton "Play", une par une les pages, passent par toutes les étapes du traitement, jusqu'à et y compris l'étape en cours. Passer par toutes les étapes du traitement ne signifie pas nécessairement être traité. Par exemple, si l'image numérisée avait déjà été scindée, lorsqu'elle reviendra au stade « scinder » elle ne sera pas retraitée à moins que vous n'ayez changé quelque chose d'important tel que son orientation. Vous pouvez arrêter le traitement par lots à n'importe quel stade en utilisant le gros bouton rond "stop" de la zone de travail principale (3) pendant le traitement.



Aucune des étapes, à l'exception de la dernière nommée "Sortie", n'enregistre réellement l'image résultante sur le disque. Elles sont purement analytiques et ne sont mémorisées que dans le fichier projet.

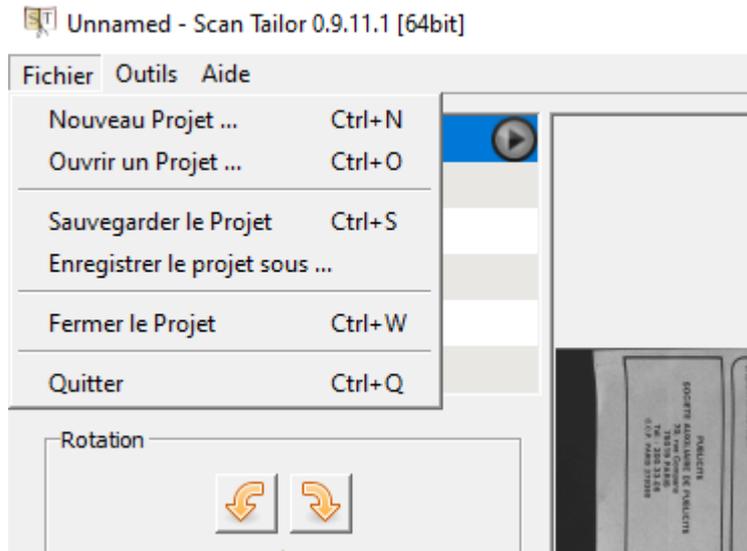
4 MENUS DU PROGRAMME

4.1 MENU FICHER

Le menu fichier est assez évident et ne nécessite aucune explication, à l'exception peut-être de l'item « Fermer le projet ». Cette commande ferme le projet, mais pas tout le programme : La fenêtre principale sera affichée comme immédiatement après le démarrage.

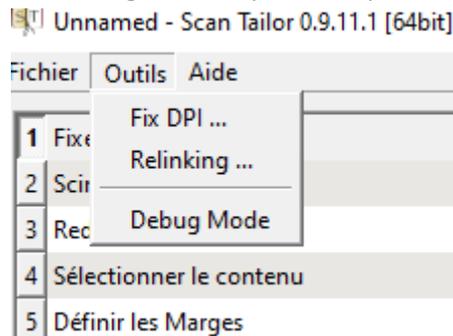
Les raccourcis clavier sont indiqués à droite des commandes de menu.

Autrement dit, par exemple appuyer sur la combinaison de touches CTRL + S Sauvegardera le projet.

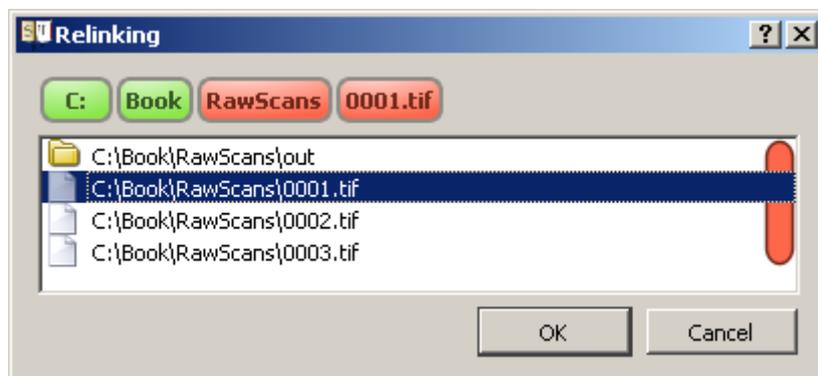


4.2 MENU OUTILS

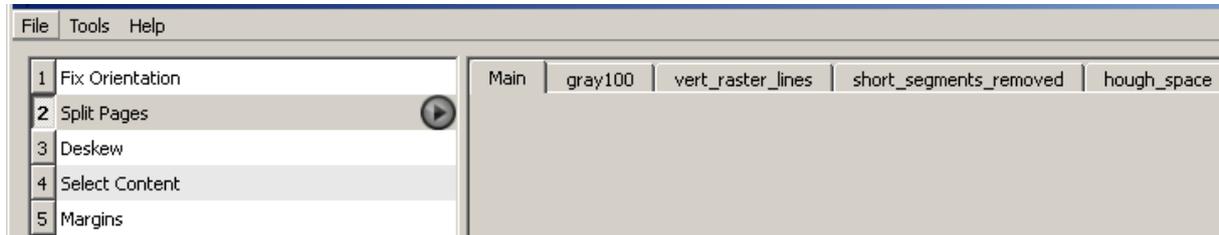
La commande Fix DPI appelle la boîte de dialogue Fix DPI présentée précédemment



La commande Relinking vous permet de réassocier des chemins des répertoires où se trouvent les images après la copie d'un projet d'un ordinateur à un autre (ou de son déplacement d'un disque dur à un autre). Si un dossier n'est pas reconnu sur le chemin proposé, il est mis en surbrillance en rouge. Vous pouvez cliquer sur cette zone rouge et choisir un nouveau chemin pour le répertoire.



La commande Debug Mode affiche les résultats du traitement intermédiaire sur des onglets spéciaux. Ces onglets seront affichés en haut de la zone centrale après le traitement d'une page. Ce mode est destiné aux développeurs.



4.3 MENU D'AIDE

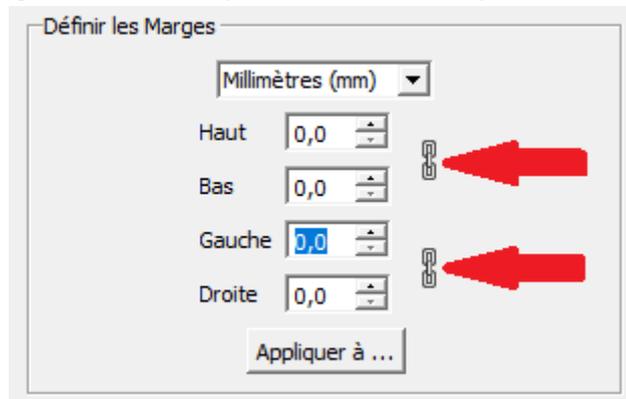
Dans le menu Aide, vous ne pouvez appeler que la boîte de dialogue "À propos", qui contient des informations sur la version du programme et ses développeurs



5 PRINCIPE DE TRAVAIL

1. Ouvrez ScanTailor et sélectionnez "Nouveau projet" pour commencer avec un nouvel ensemble d'images.
2. Parcourez et sélectionnez le répertoire contenant les images que vous souhaitez traiter. Sélectionnez les images souhaitées, si nécessaire, définissez leur résolution correcte (DPI).
3. ScanTailor utilisera le même répertoire pour les images traitées et la sortie, créant le sous-répertoire "out" par défaut. Mais à cette étape, vous pouvez maintenant choisir un autre répertoire de sortie.
4. Pour les images issues de photos, vous devrez peut-être ajuster la valeur DPI. (Correction de la commande DPI dans le menu Outils.) Cliquez sur « Toutes les pages » et sélectionnez une valeur dans la liste (par exemple, 300 * 300) ou définissez votre propre valeur « personnalisée ». Pour estimer le DPI réel, vous pouvez diviser la taille de l'image d'origine en pixels par la taille physique réelle du livre en pouces.
5. Lors de la première étape du traitement "Fixer l'orientation", vous pouvez faire pivoter les images de 90 degrés du portrait au paysage. Si toutes les numérisations doivent pivoter de la même manière, ajustez l'orientation de la première image, cliquez sur « Appliquer à » et appliquez cette option à toutes les autres images.
6. Cliquez sur la deuxième étape de traitement "Scinder les pages". Si l'image représente deux pages d'un livre double, ScanTailor tentera de la diviser en deux pages. Cliquez sur le "bouton de lecture" (une flèche en forme de triangle à droite du nom de la scène) pour exécuter le traitement par lots des images restantes. Si la détection automatique du type de page était incorrecte, sélectionnez le type prédominant, appliquez-le à toutes les pages, puis modifiez-le pour des pages individuelles. Une fois le traitement par lots terminé, vous pouvez corriger manuellement l'emplacement des lignes de coupe dans la fenêtre centrale. Il est pratique de vérifier les résultats avec le panneau de vignettes. Cliquer sur la vignette ouvrira cette page dans le panneau central, où vous pourrez ajuster son type et ses lignes de coupe.
7. À la 3^{ème} étape «Redresser », ScanTailor tentera de compenser l'angle d'inclinaison de la page. Si les images sont déjà alignées, vous pouvez appliquer un angle nul à toutes les images. Utilisez Ctrl + Molette pour faire pivoter des pages individuelles si l'angle détecté automatiquement est incorrect. Encore une fois, utilisez la flèche triangulaire pour effectuer des lots et le panneau des miniatures pour vérifier rapidement les pages.
8. Passez à la 4^{ème} étape "Sélectionner le contenu" et lancez le traitement par lots. Attendez la fin du processus. Ajustez les zones de contenu détectées en faisant glisser leurs bords dans la fenêtre centrale. Si la zone de contenu n'a pas été créée, vous pouvez la créer manuellement en cliquant avec le bouton droit de la souris dans la fenêtre. Les zones inutiles sur une page vierge peuvent également être supprimées en cliquant avec le bouton droit de la souris. Si l'erreur de détection a été causée par une compensation d'inclinaison incorrecte à l'étape précédente, revenez en arrière d'une étape et corrigez le problème, puis revenez à «Sélectionner le contenu» et assurez-vous que le cadre a été correctement défini cette fois. Il est pratique d'inspecter les pages en les triant par largeur et hauteur de la zone de contenu (le type d'ordre est indiqué dans la liste déroulante en bas du panneau des miniatures).
9. Passez à la 5^{ème} étape "Définir les marges". Sélectionnez la page la plus typique du livre (souvent la page de garde, parfois la couverture pour les revues) et définissez ses marges au plus près des marges réelles du livre. Appliquez ce paramètre à toutes les pages. Pour la couverture du livre, définissez des marges nulles et désactivez la case à cocher "Faire correspondre la taille avec les autres pages". Faites ensuite attention à la distance entre les lignes pointillées et pleines. S'il est suffisamment grand, le projet a des pages plus grandes que les pages réelles du livre. Soit leurs zones de contenu, soit leurs marges ont été définies de manière incorrecte. Pour trouver de telles pages, il est pratique de trier les vignettes par largeur ou par hauteur de la page. Pour modifier la zone de contenu incorrecte, vous devez revenir à l'étape précédente. Dans les modes «Trier par hauteur croissante» et «Trier par

largeur croissante», observez les pages les plus basses et les plus étroites en haut du panneau des miniatures. Ces pages sont généralement incomplètes et nécessitent des changements d'alignement (haut / bas / etc.) ou une personnalisation unique des marges. (N'oubliez pas de rompre l'association entre les paires de marges lors d'une telle personnalisation en cliquant sur les chaînes.)



10.

11. Cliquez sur la 6ème étape "Sortie". Le mode "Noir et blanc" est sélectionné par défaut pour toutes les pages. Si certaines lettres semblent "manquantes", essayez d'augmenter l'épaisseur du texte en déplaçant le curseur du seuil de binarisation vers la droite. Sélectionnez le mode «Mixte» pour les pages contenant des images et des photos. Il est recommandé de traiter la sortie avec une résolution de 600 DPI. C'est la valeur par défaut, car elle fournit de jolies lettres lisses.
12. Démarrez le traitement par lots. Lorsque vous avez terminé, vérifiez le résultat. Si les zones de texte ou d'illustration n'ont pas été détectées correctement sur certaines pages en mode mixte, vous pouvez les définir manuellement dans l'onglet "Zones d'image" du panneau central. Redémarrez le traitement par lots (cette fois, ce sera beaucoup plus rapide que la première fois.)
13. Les fichiers traités seront écrits dans le répertoire spécifié (par défaut - dans le sous-dossier "out") au format tiff. Ces images, si nécessaire, peuvent être encore améliorées dans les éditeurs graphiques ou utilisées pour assembler des fichiers de livre.

6 CONSEILS POUR LA NUMÉRISATION

Pour obtenir un bon résultat avec ScanTailor et minimiser les erreurs dans les processus automatiques

- Ne numérisez pas en mode noir et blanc - utilisez des niveaux de gris ou, si nécessaire, des couleurs.
- Ne numérisez pas à une résolution inférieure à 300 DPI, numérisez généralement à 300 ou 600 DPI.
- Ne pas numériser au format JPEG - le format est avec perte et la conversion vers d'autres formats n'entraînera pas une meilleure qualité. (Les utilisateurs d'appareils photo devront probablement utiliser le mode JPG mais utiliser la meilleure résolution et les meilleurs paramètres de qualité JPEG possibles).
- Lors de la numérisation, utilisez généralement le format TIFF, mais soyez prudent car TIFF peut être utilisé pour les algorithmes de compression jpeg.
- Si vous avez besoin de compresser l'image, choisissez LZW - un format sans perte
- Si vous ne pouvez pas utiliser TIFF, numérisez vers PNG qui garantit d'utiliser des algorithmes de compression sans perte, ou, dans les cas extrêmes, vers BMP mais la taille du fichier sera décompressée et donc énorme, il est donc préférable de se servir d'un format qui utilise la compression sans perte.
- Évitez le mode de numérisation "Document", et essayez généralement de désactiver toutes les options censées améliorer les numérisations.

7 EXPLICATIONS DES COMMANDES :

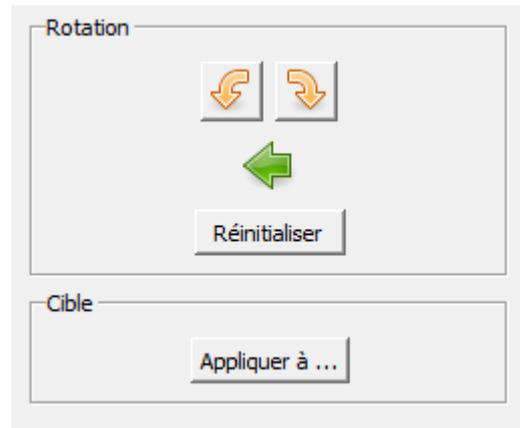
7.1 FIXER L'ORIENTATION

Cette étape permet de tourner les scans par multiples de 90 degrés. C'est-à-dire pour corriger les scans latéraux ou à l'envers.

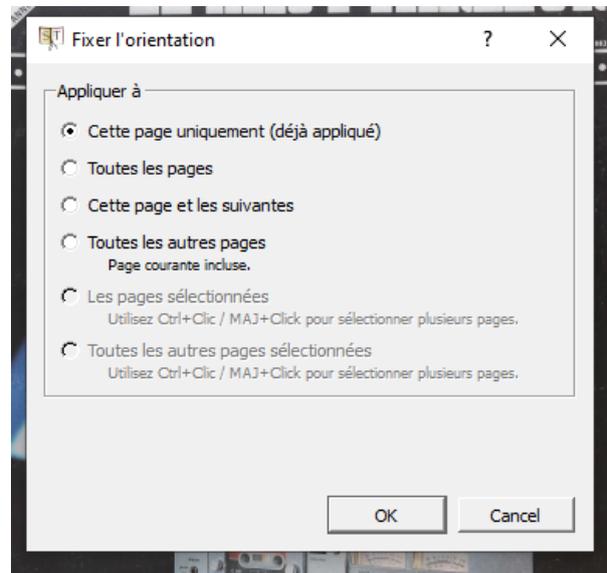
Il s'agit d'une étape manuelle car le programme ne sait pas comment déterminer l'orientation correcte des numérisations - l'utilisateur doit le faire. Cela signifie également que l'utilisation du traitement par lots à ce stade est inutile, par contre il est possible d'appliquer la rotation à toutes les pages grâce au bouton Apply to (voir plus bas).

Le panneau Paramètres a cet aspect :

- Rotate : Les boutons avec des flèches jaunes font pivoter la numérisation de 90 degrés dans un sens ou dans l'autre. Un pointeur vert indique dans quelle direction le balayage est orienté à un moment donné. Le bouton "reset" ramène le scan à la position initiale - le pointeur vert indiquera vers le haut.
- Scope : permet de tourner non seulement cette page (ce qui se fait automatiquement), mais aussi d'autres. Cette boîte de dialogue s'ouvre en cliquant sur le bouton "Appliquer à ..." et offre les options suivantes :



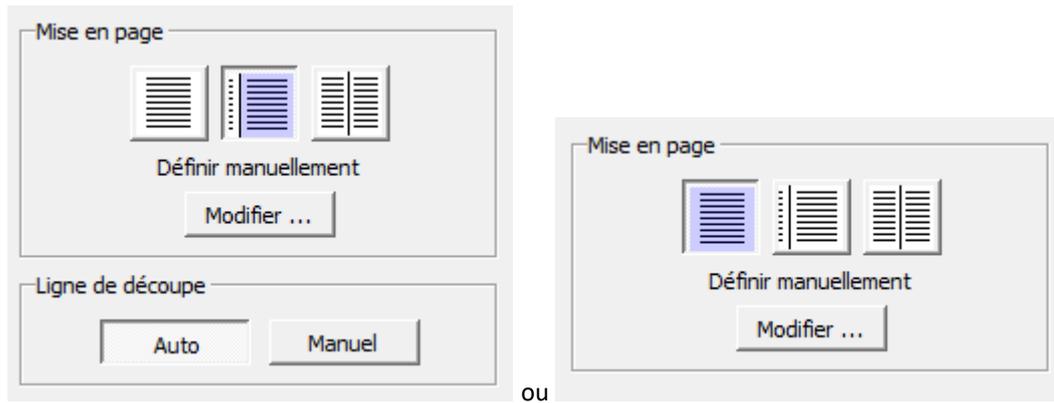
- Cette page uniquement.
- Toutes les pages du projet
- Cette page et les suivantes. L'opération qui a été appliquée à la page en cours sera aussi appliquée aux pages qui suivent.
- Une page sur deux : appliquera la rotation à chaque page paire ou à chaque page impaire selon que la page actuelle est paire ou impaire.
- Les deux dernières options ne sont activées que si deux ou plusieurs pages sont sélectionnées dans le volet d'aperçu.
- Pour utiliser Toutes les autres pages sélectionnées, les pages sélectionnées doivent être continues. Nota : Pour sélectionner des pages continues, il est souvent plus pratique d'utiliser Maj + clic.



7.2 DIVISER LES PAGES

Cette étape détermine si vous souhaitez diviser la ou les pages.

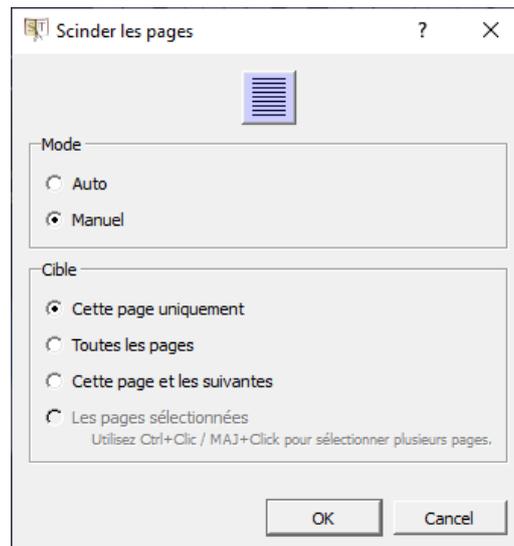
La boîte de contrôle des paramètres se présente ainsi:



Mise en page :

1. icone de gauche : Un scan complet, sans aucune partie de la page suivante ou précédente. Ces scans sont généralement obtenus à partir de scanners de livres ou de photographies spécialisés.
2. icone du centre : Un scan, qui comporte une partie de la page suivante ou précédente.
3. icone de droite : Scan double page.

Le type de division est déterminé automatiquement, mais peut être défini manuellement. Cliquer sur l'icone qui vous convient et utilisez le bouton "Modifier..." pour le redéfinir manuellement. Le type de division peut être appliqué à toutes les pages à la fois ou à des pages individuelles.



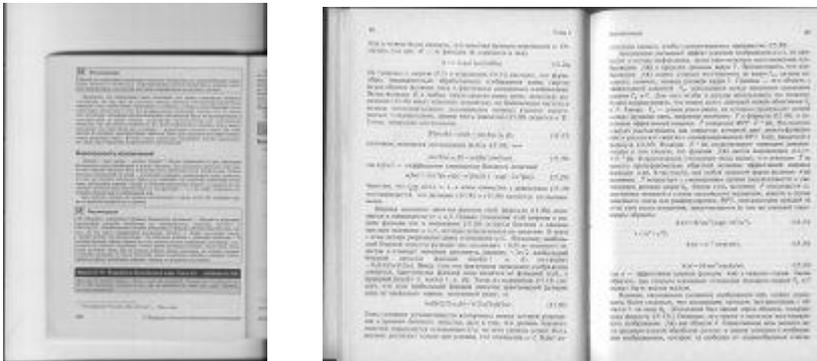
Le bouton auto peut être utilisé pour sélectionner automatiquement le type de division.

La ligne de séparation peut également être déterminée / déplacée automatiquement ou spécifiée manuellement, mais elle ne peut pas être appliquée à d'autres pages. Il est utile de consulter rapidement le volet d'aperçu de chaque page pour s'assurer que les fractionnements de page ont été correctement appliqués - parfois, les images dans les pages peuvent affecter l'opération de fractionnement. Les point de départ et d'arrivée de la ligne de séparation peuvent être déplacés avec la souris et les bulles les matérialisant.

Pour augmenter les chances de détermination automatique correcte du type d'incision et de la ligne de séparation, essayez de suivre ces règles lors de la numérisation :

- Si vous effectuez une numérisation unique, faites-le en orientation portrait plutôt qu'en orientation paysage.
- Lors de la numérisation, choisissez le mode le plus brut, c'est-à-dire celui dans lequel la machine n'essaiera pas d'améliorer ou de compresser quoi que ce soit.

Exemples de bons scans :



Exemple de scan défectueux (logiciel de scanner prétraité) :

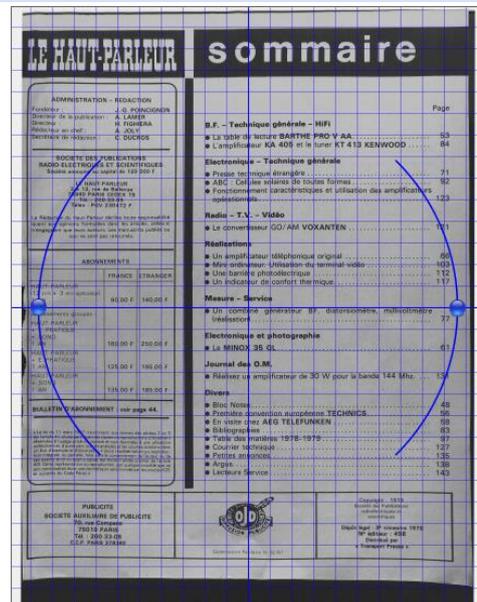


7.3 REDRESSER

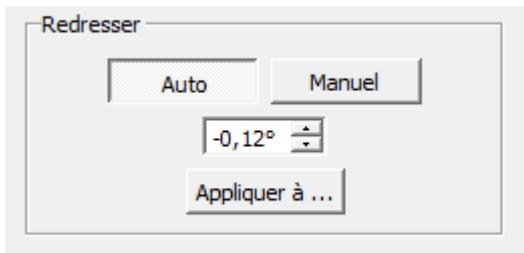
On peut maintenant déterminer l'angle dont la page doit être tournée pour que le texte soit bien horizontal. La compensation étant une simple rotation, des distorsions telles que la distorsion trapézoïdale ou le curling ne peuvent pas être corrigées à ce stade. L'angle de rotation est déterminé automatiquement, mais vous pouvez également le définir manuellement.

Voici ci-contre la zone de travail dans ce mode :

Les images peuvent être tournées en faisant glisser les poignées rondes sur les bords.



Voici à quoi ressemble le panneau d'options:



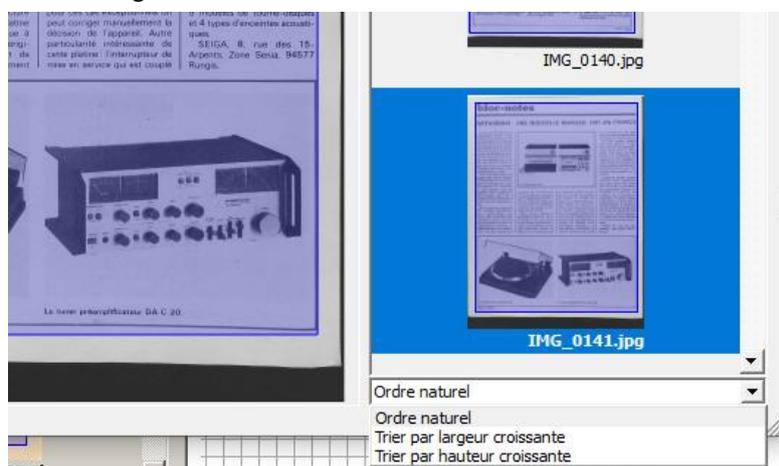
Vous pouvez également spécifier explicitement l'angle de rotation en degrés. Les angles positifs feront pivoter l'image dans le sens horaire, négatif dans le sens antihoraire. Pour un réglage précis de l'angle, il peut être pratique de cliquer avec la souris sur la partie texte du coin du champ de saisie, puis de déplacer la molette de la souris pour le corriger.

7.4 SÉLECTIONNEZ LE CONTENU

Cette étape détermine la région rectangulaire contenant la partie «utile» ou utilisable (surlignée en couleur). Pourquoi définir cette zone? Tout d'abord afin de déterminer la taille de la page à la sortie. Le contenu sera ajouté à la zone de marge totale et la limite extérieure de ces marges affecte la taille du fichier de sortie. Deuxièmement, pour que les images finales ne montrent pas la ligne de pli ou d'autres défauts des bords. À proprement parler, le fait que les défauts tombent dans la marge à l'étape de sortie dépend du mode. Dans la plupart des modes, la marge est remplie en blanc.

Si des zones sont identifiées de manière incorrecte, vous pouvez modifier manuellement des pages individuelles en plaçant le pointeur de la souris sur le bord, en cliquant et en faisant glisser si nécessaire. Parfois, ScanTailor peut trouver du contenu inexistant ou, au contraire, ne pas sélectionner le contenu là où il le devrait. Dans ce cas, vous pouvez créer / supprimer manuellement une région en cliquant avec le bouton droit de la souris sur l'image et en sélectionnant l'élément de menu approprié.

Il est très pratique de rechercher des zones de contenu mal détectées en réorganisant les pages par largeur / hauteur. Activez-le avec la liste déroulante sous le panneau des miniatures :

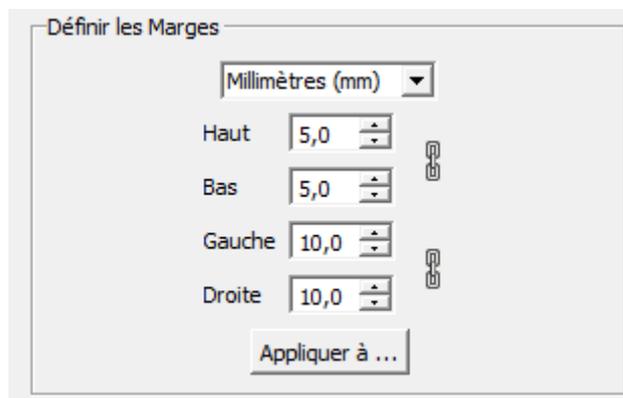


7.5 MISE EN PAGE



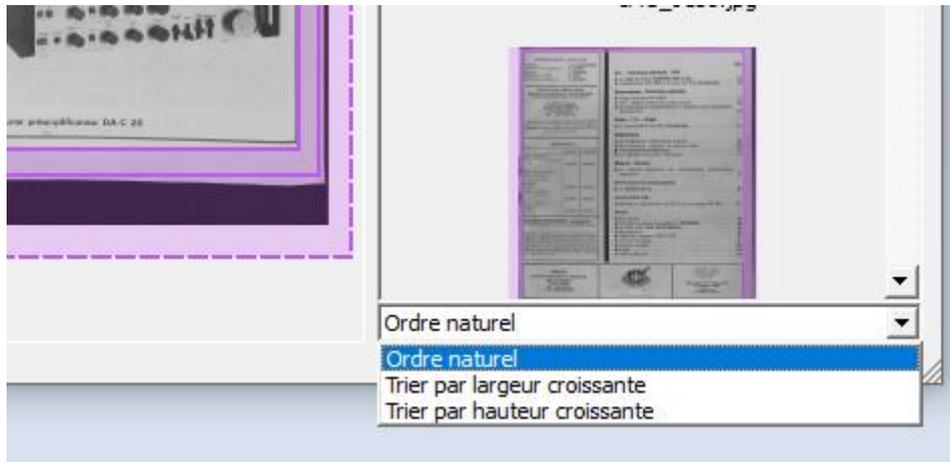
À ce stade, vous pouvez ajuster les marges ajoutées à la zone de contenu. Il existe deux types de marges : dures et douces.

La marge dure est l'espace entre les lignes pleines. Ces marges sont définies par l'utilisateur. Vous pouvez déplacer n'importe quelle ligne pleine - externe ou interne, ou définir la taille des marges avec leurs valeurs numériques:

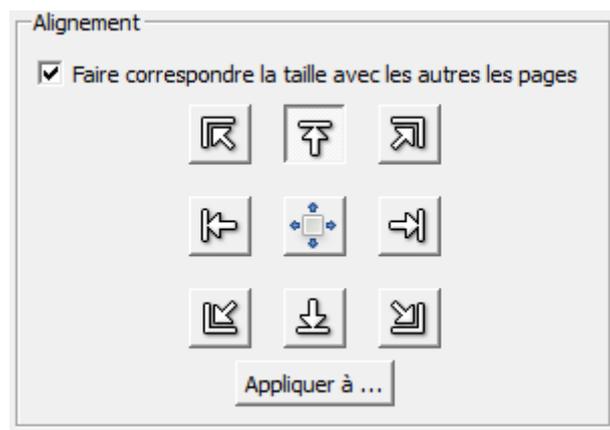


La marge douce est l'espace entre le trait plein et les lignes pointillées. Ces marges sont automatiquement ajoutées pour amener la taille de la page à la taille des autres pages. Si vous voyez une ligne en pointillé, cela

signifie que quelque part dans le projet, il y a au moins une page avec cette largeur (zone de contenu + marge rigide) et peut-être une autre avec cette hauteur. Vous pouvez trouver ces plus grandes pages en utilisant le tri des pages par largeur / hauteur:



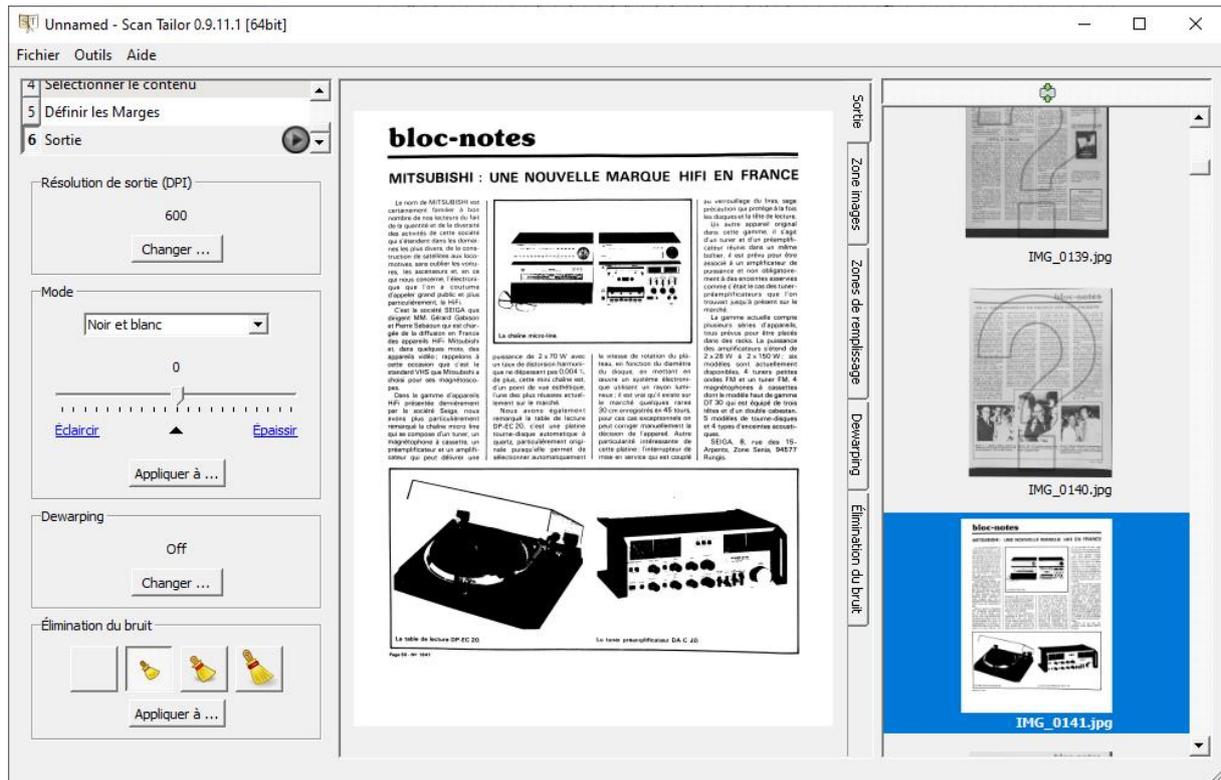
Autrement dit, une grande page fait apparaître des marges douces sur toutes les autres pages, à moins que l'option «faire correspondre la taille avec les autres pages» ne soit désactivée pour elles. Les paramètres d'alignement déterminent simplement s'il faut ajouter des marges douces et à partir de quels côtés ajouter:



7.6 PRODUCTION DU DOCUMENT FINAL

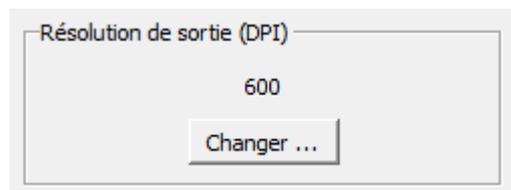
7.6.1 ÉTAPE "SORTIE"

C'est à ce stade que les fichiers de sortie sont créés à partir des images et écrits sur le disque. Les images résultantes apparaissent également dans la fenêtre centrale du programme.

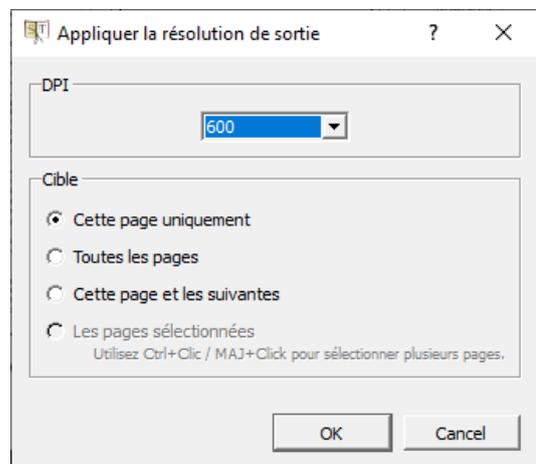


Contrairement aux autres étapes, l'étape «Sortie» n'est disponible qu'après que toutes les pages ont passé les étapes «Sélectionner le contenu» et «Définir les marges». Cela est dû au fait que la taille des pages dans la sortie dépend les unes des autres. . Par exemple, si une grande page est trouvée, tous les autres champs augmentent (pour en savoir plus, consultez la documentation sur l'étape "Mise en page"). Il est donc important de connaître la taille finale des pages, et cela ne peut se faire qu'à travers les étapes de "Sélectionner le contenu" et "Mise en page". Pourquoi l'étape "Sélectionner le contenu" est-elle nécessaire ? Parce que toutes les options de l'étape "Mise en page" sont définies manuellement ou prennent des valeurs par défaut.

Configurez les éléments suivants:

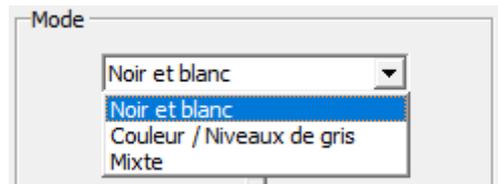


Résolution de sortie (DPI) - vous pouvez spécifier manuellement une résolution pour les fichiers de sortie:



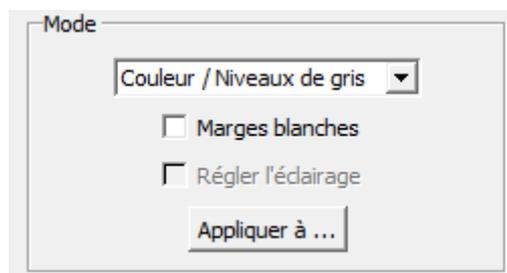
La valeur par défaut est 600 DPI. Dans certains cas, il est de 300.

7.6.2 MODE - SÉLECTIONNE LE MODE DE SORTIE DES PAGES:



Le mode noir et blanc ne nécessite guère d'explication. Il n'y a clairement pas de teintes grises disponibles, ce qui ne conviendrait donc pas à des images et à certains dessins. Il existe une option pour « supprimer les taches » et pour augmenter ou diminuer l'épaisseur de la ligne (c'est-à-dire du texte). En général, il est préférable de ne pas supprimer les taches si l'image est raisonnablement propre, car la suppression des taches peut entraîner la perte de certaines parties de texte. Cela peut être compensé dans une certaine mesure en augmentant l'épaisseur de la ligne, mais il est probablement important d'expérimenter sur quelques pages avant de l'appliquer à l'ensemble du projet.

L'option Couleur / Niveaux de gris:



Les marges peuvent être remplies de blanc ou laissées telles quelles. Si les marges sont remplies de blanc, l'option d'égalisation de l'éclairage devient également disponible. Cette option normalise la couleur d'arrière-plan, l'amène au blanc et normalise le contraste, en l'augmentant dans les zones ombrées.

Le mode mixte est utilisé pour les projets dans lesquels il y a des numérisations à partir d'images en demi-teintes (niveaux de gris ou couleur). Les images seront automatiquement détectées et affichées telles quelles, tout comme dans "Couleur / Gris" avec l'éclairage égalisé inclus. Le reste de la page s'affiche en noir et blanc.

L'image automatique fonctionne assez bien, mais si l'image se fond facilement dans l'arrière-plan, le résultat peut être insatisfaisant. Dans ce cas, vous devez créer et configurer les images de la zone d'image. Il est important de noter que la création de zones d'images n'est possible qu'en mode mixte.

7.6.3 LE FORMAT DES FICHIERS DE SORTIE

La sortie est au format TIFF. Nous utilisons LZW pour garantir des images compressées sans perte.

7.7 ONGLETS DE SORTIE: ZONES D'IMAGE

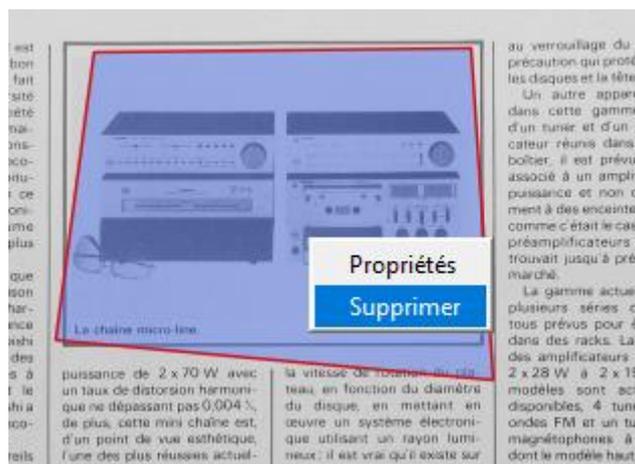
Zones d'image

Les zones d'image sont disponibles en mode de sortie mixte. Ces zones sont affichées sur l'onglet dans le coin supérieur droit de la fenêtre centrale:

Après le traitement de sortie, vous verrez le calque automatique (images détectées automatiquement surlignées en bleu) sur cet onglet:



Si la détection est erronée, vous devez soit ajouter une zone d'image, soit exclure une partie de l'image automatiquement reconnue. Pour ce faire, vous devez créer une zone - un polygone fermé de forme arbitraire. Cette zone peut avoir des propriétés d'ajout / de soustraction de sa zone du calque automatique. Pour créer un tel polygone, cliquez sur le bouton gauche de la souris en conséquence sur les points de sommets potentiels. Le dernier clic doit être fait sur le premier sommet - cela fermera le polygone de zone. Pour supprimer un polygone non fermé non terminé, appuyez simplement sur la touche Échap.



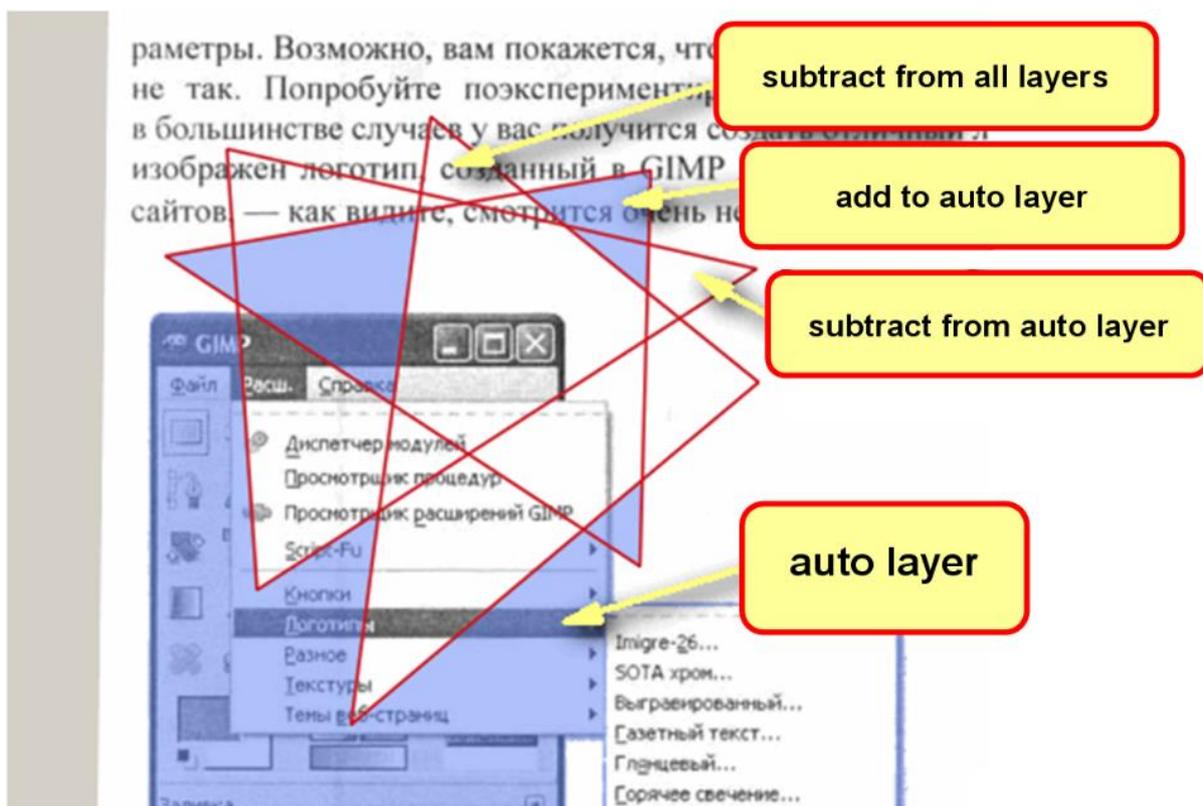
Lorsqu'une zone est finalement créée, ses sommets peuvent être décalés. Faites-les glisser avec la souris. Pour créer un nouveau sommet, cliquez simplement sur le bord de la zone. Pour supprimer un sommet, faites glisser le sommet à supprimer sur l'un de ses deux voisins. Lorsque vous cliquez avec le bouton droit de la souris sur la zone, un menu contextuel contenant les éléments Propriétés et Supprimer apparaît. La sélection de «Supprimer» supprime toute la zone. La sélection de «Propriétés» ouvre la fenêtre «Propriétés de la zone»:



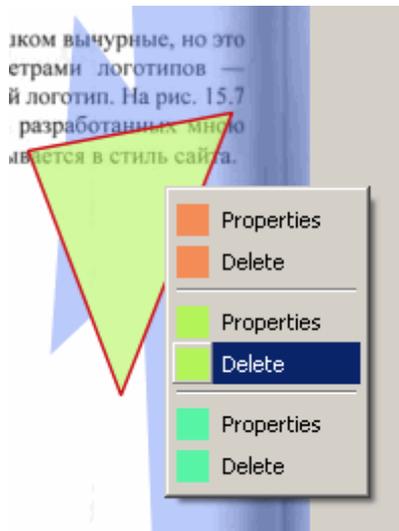
"Soustraire de toutes les couches" effectue la binarisation du contenu de cette zone - rend cette zone en noir et blanc - indépendamment des autres zones (manuelles ou automatiques) chevauchant cette zone. "Ajouter au calque automatique" rend cette zone en couleur après la sortie. "Soustraire du calque automatique" effectue la binarisation du contenu de cette zone là où il n'est pas recouvert de zones ajoutées au calque automatique. En superposant des zones et en modifiant leurs propriétés, vous pouvez définir la couleur et les zones n / b de la sortie pour correspondre à la page d'origine. Tenez compte des règles de priorité suivantes:

1. "Soustraire de toutes les couches" - zone avec la priorité la plus élevée
2. "Ajouter au calque automatique"
3. "Soustraire du calque automatique" - n'est pas appliqué à la zone de priorité 2, mais affecte la zone de priorité 4
4. Couche automatique - zone avec la priorité la plus basse

La figure ci-dessous montre la contribution de chaque type de zone dans l'image résultante:



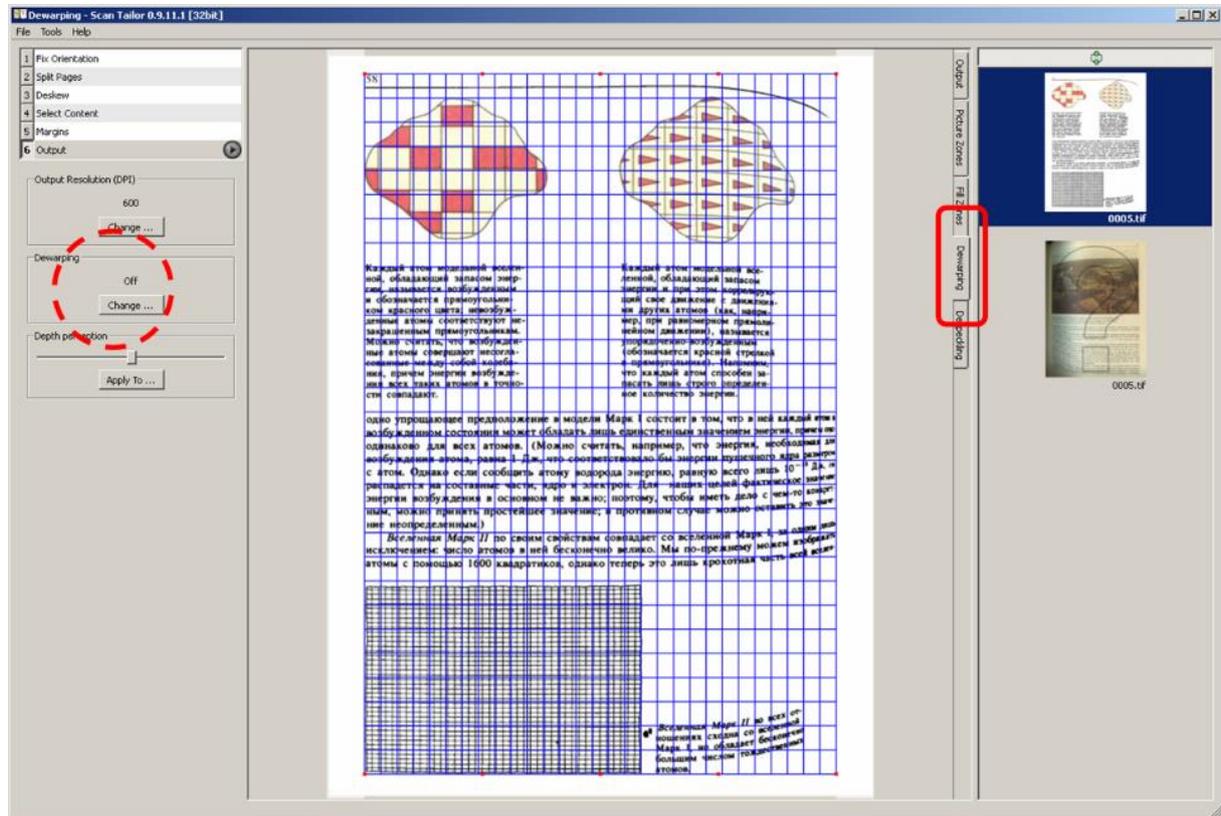
Il est très pratique qu'après l'étape "Output", les zones de l'image ne soient pas seulement surlignées en bleu mais clignotent également. Ce clignotement garantit que ces zones sont bien distinguées sur l'image de n'importe quelle couleur. Tout changement de type de zone fait que le choix devient le type par défaut pour les prochaines zones créées. Pour déterminer la zone active parmi les zones qui se chevauchent, vous devez sélectionner les commandes appropriées dans le menu contextuel. Il y a une paire d'éléments "Propriétés" et "Supprimer" colorés différemment pour chaque zone de chevauchement dans le menu. Lorsque vous les survolez avec la souris, la zone correspondante commence à être mise en évidence avec cette couleur :



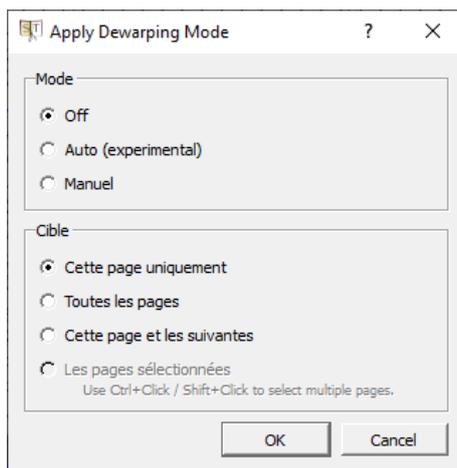
Pour voir l'effet final des zones créées, cliquez sur l'onglet "Sortie" dans le coin supérieur droit de l'espace de travail.

7.8 ONGLETS DE SORTIE: DEWARPING

La grille de déformation de la page traitée en cours est affichée sur l'onglet du même nom dans le coin supérieur droit de la fenêtre centrale. Si le mode de déformation est désactivé, la grille est rectangulaire avec des lignes droites:

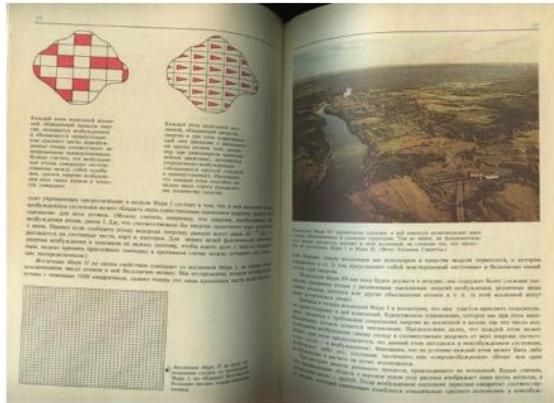


En appuyant sur "Changer", vous verrez les deux principaux modes de déformation ST - Auto et Manuel:

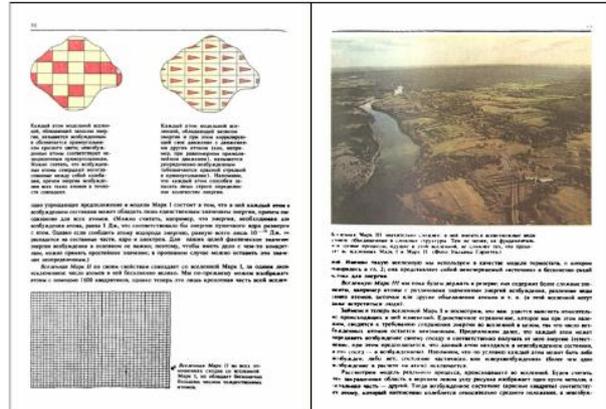


La déformation automatique facilite l'obtention de très bons résultats de sortie pour une entrée courbe:

without dewarping

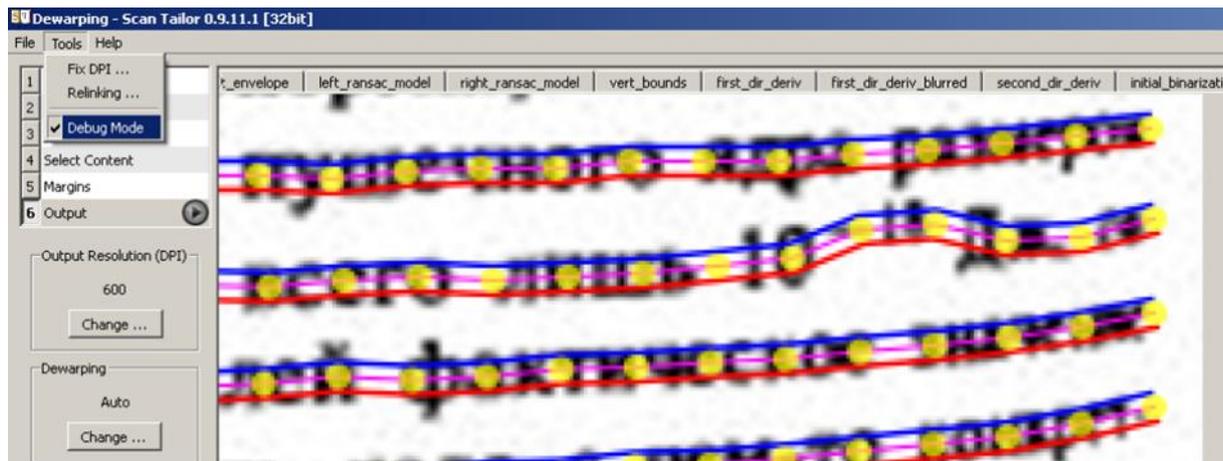


with dewarping



L'algorithme consiste à placer deux limites verticales et deux courbes représentant de longues lignes de texte horizontales (perpendiculaires aux limites et étant éloignées ; elles n'ont pas besoin d'être les lignes les plus hautes et les plus basses). Ce modèle de distorsion suffit à déformer toute la page:

Toutes les étapes de la détection sont visibles en mode débogage :



Les principales étapes sont: 1. Détection des limites verticales; 2. Traçage grossier des lignes de texte, obtention de polygones; 3. Amélioration des polygones avec des «serpents couplés»; 4. Ajuster les X-splines aux polygones dans le but d'éditer ces splines (et notre grille de déformation) plus tard manuellement; 5. Sélection de la meilleure paire de splines (plus la distance entre elles est grande, mieux c'est) et construction du modèle de distorsion.

Il existe également un traceur d'arête intégré à l'algorithme de déformation. Le cas échéant, il trace les bords supérieur et inférieur de la page, détectant leur bonne position. Ces bords complètent les lignes de texte et seront traités de la même manière. Après tout, l'algorithme n'a pas besoin de lignes de texte spécifiques, mais de tout ce qui est long et horizontal sur la page plate.

Les zones d'image ne sont pas prises en compte pour la déformation. Donc, cette mise en page est l'un des cas les plus désespérés d'amélioration automatique. Pas de bords de livre, pas de longues lignes de texte, et chacun d'entre eux commence ou se termine par une image:

Opening Highlights



Alexey Shirov

Shirov was the hardest-working man in chess business in recent months. Among many other feats, we saw a great tournament victory in Shanghai and a spectacular game against Baadur Jobava at the Khanty-Mansiysk Olympiad. In an **Advanced Caro-Kann** Shirov, true to form, won brilliantly with the ultra-sharp 8.g4. In his Survey on page 97, Van de Walle investigates whether the sharpest move is really the best.



Fabiano Caruana

The winner of this year's exceptionally strong Biel Young Masters analyzed his decisive game against Wesley So for us. In a **Semi-Slav**, the young Italo-American surprised the still younger Philippine with the audacious 16.f2f5. As Paul Blazynski demonstrates in his Survey on page 146, Black does not really need to fear this thrust theoretically, but facing it in a crucial game is a different kettle of fish!



Romain Edouard

He seemed to be cruising towards the French title. Romain Edouard got very close but no cigar as Laurent Fressinet pipped him at the post in a bitz playoff. The ambitious young grandmaster played some brilliant games though. He joins our ranks with a Survey on Annaparashvili's move 7.Wd4 in the **Main Line Benoni**, a "courtesy surprise" that may be a good way to stymie Black's kerkuz Benoni piece play. See page 229.



Loek van Wely

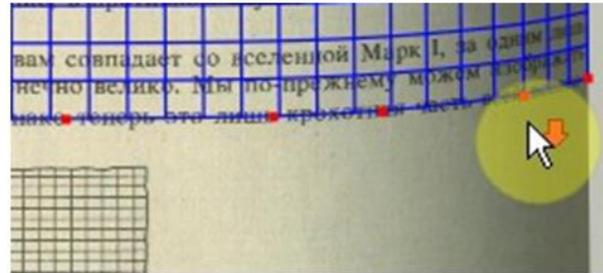
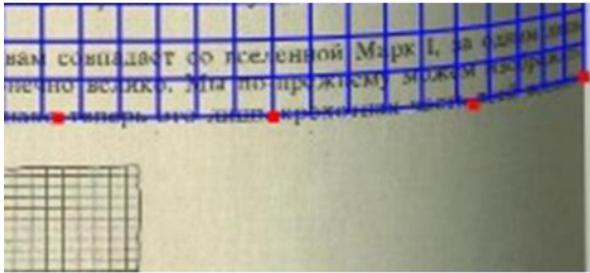
With his principled approach to openings the Dutchman sometimes suffers little "accidents" – against Hikaru Nakamura in the NH Tournament it was more like a frontal collision. After confusing two moves from his own New In Chess analysis Van Wely was blown away in a **Najdorf Sicilian** by the fast-and-furious American. Roud Reviz Dithof's Survey on page 36 and also Daniel Stellwagen's comment in our FORUM section on page 11.

Si la grille détectée automatiquement n'a pas été définie correctement, l'utilisateur peut l'ajuster manuellement. (Pour appliquer ces modifications, basculez le mode de déformation sur MANUEL.) Tout d'abord, notez la relation entre les limites verticales de la grille et les limites de la zone de contenu. Rien ne sera ajouté (1) ni perdu (2). (Le mode de sortie ici est "Couleur / Niveaux de gris" avec "Marges blanches".) Vous pouvez donc éditer les bords gauche et droit de la grille sans précaution. Le contenu de la zone de contenu ne disparaîtra pas.

Deuxièmement, notez qu'avec le modèle de distorsion de notre exemple initial, la ligne du bas a besoin de plus de précision car ce n'est pas la ligne de texte la plus basse.

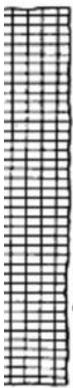
Vous pouvez améliorer les résultats de déformation dans les zones supérieure et inférieure en ajoutant plus de nouveaux points de contrôle (points rouges) et en faisant correspondre une meilleure ligne avec les mouvements des points existants. Les points rouges non dans les coins peuvent ne pas se trouver sur la ligne de texte, mais la spline générée par eux devrait l'être. Les points de contrôle inutiles peuvent être supprimés avec les touches D ou DEL:

Bottom area improvement with more control points above

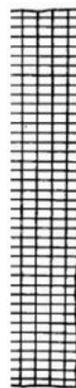


» велико. Мы по-прежнему можем изображать теперь это лишь крохотная часть всей вселен-

» велико. Мы по-прежнему можем изображать теперь это лишь крохотная часть всей вселен-



• Вселенная Марк II во всех отношениях сходна со вселенной Марк I, но обладает бесконечно большим числом тождественных атомов.



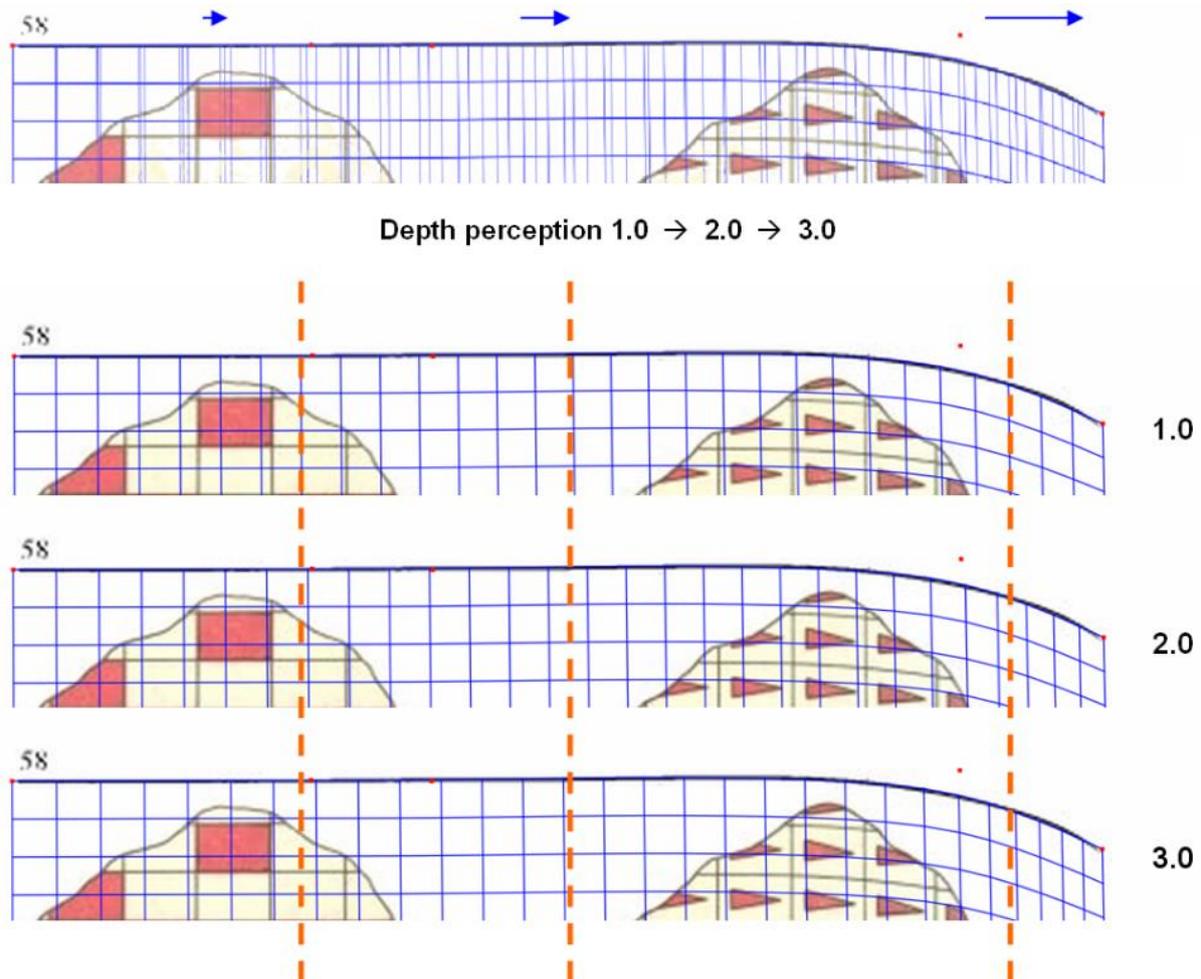
• Вселенная Марк II во всех отношениях сходна со вселенной Марк I, но обладает бесконечно большим числом тождественных атомов.

Les points de contrôle rouges du milieu sont configurés comme des points d'approximation : la X-spline n'a pas à les traverser, mais sera attirée vers eux comme si elle était tirée par des ressorts. Ces points de contrôle n'ont pas d'importance en eux-mêmes - ils aident uniquement l'utilisateur à contrôler la forme des courbes du haut et du bas. Si vous supprimez tous les points médians et que ces courbes seront des lignes droites, vous obtiendrez une déformation sans déformation.

Enfin, considérez l'effet de la commande «Perception de la profondeur», qui est réglable en modes de déformation AUTO et MANUEL:



Cette commande ajuste la distance entre la caméra et la page, mais en unités normalisées. Sa valeur passe de 1 à 3 par pas de 0,1. ST n'a pas pu déduire cette distance de l'image elle-même. Ses modifications affectent la compression horizontale des lettres au niveau de la reliure:



Une fois la grille de déformation définie sur la page (le modèle de distorsion est construit), l'image de sortie sera transformée de sorte que cette grille devienne rectangulaire avec des lignes droites et équidistantes.

Les références:

1. Cao H., Ding X., Liu C., "Rectifying the bound document image captured by the camera: A model based approach." Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on. IEEE, 2003 [[download](#)]
2. Bukhari S., Shafait F., Breuel T., "Coupled snakelets for curled text-line segmentation from warped document images" // International Journal on Document Analysis and Recognition, vol 16(1), March 2011, pp 1-21 [[download](#)]
3. Blanc C., Schlick C., "X-splines: A spline model designed for the end-user." Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. ACM, 1995. [[download](#)]

7.9 ONGLETS DE SORTIE: ZONES DE DÉPOUSSIÉRAGE ET DE REMPLISSAGE

7.9.1 ÉLIMINATION DU BRUIT

Sur l'onglet Dépoussiérage, vous pouvez voir les résultats du nettoyage de l'image des ordures sales (suppression des points supplémentaires de l'image). Le dépoussiérage ne peut pas être effectué en mode Couleur / Niveaux de gris:

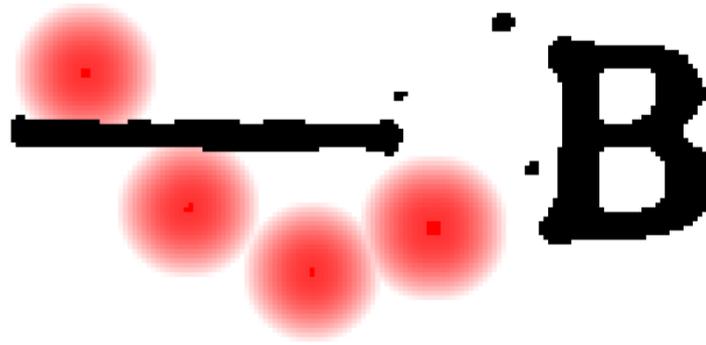


Plus le balai est gros, plus le nettoyage est intensif:

Attention, après avoir changé le balai, vous devez appuyer sur l'onglet Sortie, pour que le résultat soit écrit sur le disque dur, pas seulement indiqué. Faites attention, qu'après de tels changements, le point d'interrogation apparaît sur la vignette de la page dans la barre d'aperçu. Deux méthodes sont suggérées:

1. Effectuez la sortie en mode batch -> passez à l'onglet Élimination du bruit -> parcourez toutes les pages en changeant le balai manuellement si nécessaire -> exécutez le traitement par lots pour la deuxième fois (cela se fera plus rapidement qu'à la première fois).
2. Chaque fois que vous avez modifié le niveau de suppression des taches, passez à l'onglet Sortie pour écrire le résultat sur le disque.

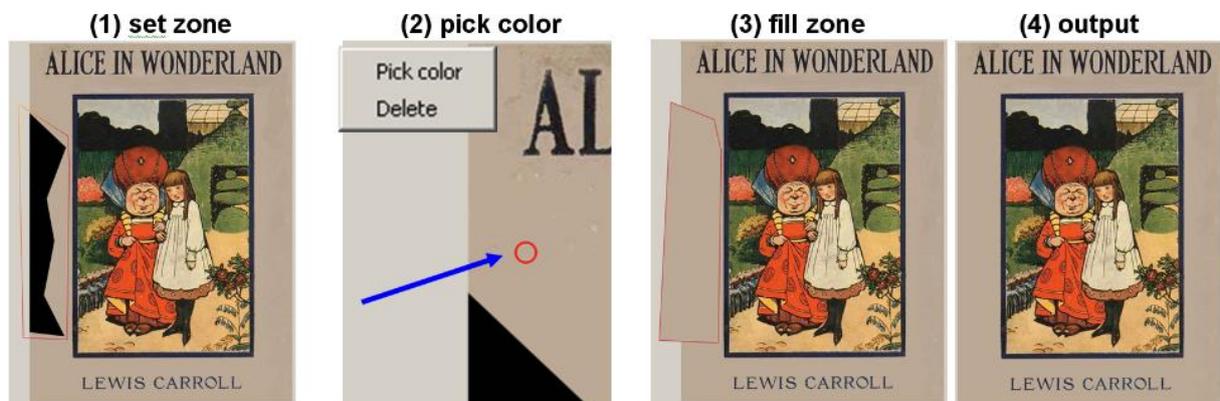
Le balai n'est pas seulement un filtre par taille. Le modèle est plus compliqué. Il existe une limite supérieure de taille et les objets plus volumineux ne seront pas supprimés. Les lettres dépassent cette limite, mais pas les signes de ponctuation. De même que pour les corps célestes: un objet peut contenir un autre objet s'il n'est ni trop grand ni trop éloigné. Cet objet, à son tour, peut également contenir d'autres objets et ainsi de suite. Pour chaque objet, ST vérifie s'il y a un porte-objet dans la chaîne qui a atteint la limite supérieure de taille.



7.9.2 REMPLIR LES ZONES

Dans l'onglet Remplir les zones, vous pouvez sélectionner manuellement une zone d'une page. Vous pouvez le remplir avec une couleur de votre choix, en utilisant le même outil de création de zone que les zones d'image. Habituellement, nous utilisons des zones de remplissage comme gomme pour les grands objets tels que les tampons et les inscriptions. Il est également pratique de pouvoir effectuer des suppressions mineures dans ScanTailor, sans utiliser aucun autre logiciel:

Lorsque vous cliquez avec le bouton droit de la souris sur la zone, un menu contextuel avec les éléments Choisir la couleur et Supprimer apparaît. La sélection de «Supprimer» supprime la zone de remplissage. La sélection de "Choisir la couleur" affiche un cercle rouge, qui doit viser la couleur cible. Cet outil permet souvent de combler les trous sur les couvertures colorées:



8 ANNEXE 1 BUILDING FROM SOURCE CODE ON LINUX AND MAC OS X

8.1 COMPILATION LINUX

Conditions préalables

Toutes les conditions requises doivent être disponibles pour votre distribution. Les plus importantes sont:

- Les outils de base de création de programme. Sur Ubuntu et Debian, ceux-ci sont disponibles dans le méta-paquet build-essential. Sur d'autres distributions, vous devrez peut-être installer le compilateur GNU C ++ séparément. Le paquet aura c ++ ou g ++ dans son nom. Si vous ne trouvez pas l'équivalent du méta-paquet build-essential pour votre distribution, assurez-vous d'installer également l'utilitaire "make".
- CMake, le système de construction multiplateforme.
- Qt version 4.4.0 ou supérieure (mais Qt version 5.x ne fonctionnera pas). Si KDE 4.1 ou version ultérieure est installé, vous aurez les bibliothèques d'exécution, mais vous devrez peut-être installer les fichiers d'en-tête et les utilitaires. Ceux-ci se trouvent dans le paquet libqt4-dev sur Ubuntu et Debian. En général, les packages de développement ont le suffixe -dev ou -devel.
- Plusieurs autres packages de développement:

Ce sont les noms de paquets sur Debian et Ubuntu, trouvez les noms similaires pour votre distribution. Pour Fedora, utilisez "-devel" au lieu de "-dev" comme suffixe et supprimez les numéros de version.

Package	Version	Possible Package Names
libjpeg	any	libjpeg62-dev Fedora: libjpeg-turbo-devel
zlib	any	zlib1g-dev
libpng	any	libpng12-dev
libtiff	any	libtiff4-dev
libboost	>= 1.35	libboost1.40-all-dev boost-libs (if that fails, try: libboost1.55-all-dev)
libxrender	any	libxrender-dev libXrender-devel

Configuration

Ouvrez une fenêtre de console et accédez au répertoire qui contient les sources scantailor. À partir de là, exécutez ce qui suit (notez le point à la fin):

```
cmake .
```

Il y aura beaucoup de messages de cmake se terminant par:

```
- Configuring done
```

```
- Generating done
```

- Si vous avez des dépendances manquantes, vous recevrez un message d'erreur vous indiquant ce qui manque. Vous pouvez ensuite rechercher le package manquant et l'installer.
- Si vous avez une bibliothèque et des fichiers d'en-tête installés dans un emplacement non standard, cmake ne les trouvera pas. Dans ce cas, vous pouvez exécuter le programme interactif "'ccmake' " qui vous permet de spécifier les chemins vers les bibliothèques et les fichiers d'en-tête.

Compilation et installation

```
make
sudo make install
```

8.2 COMPILATION POUR MAC OS X

8.2.1 MACPORTS

MacPorts a déjà scantailor en tant que [package](#) (les prérequis Xcode pour MacPorts s'appliquent)

```
sudo port selfupdate
sudo port install scantailor
```

La construction sur Mac OS X est très similaire à la construction sur Linux, mais les noms de paquetages sont différents et le système d'empaquetage MacPorts doit être installé afin d'obtenir les paquets nécessaires. L'emplacement non standard dans lequel MacPorts installe qmake doit également être spécifié.

Prérequis

- Xcode. Installez-le à partir de votre support d'installation Mac OS X ou supérieur, téléchargez la dernière version sur <http://developer.apple.com/TOOLS/Xcode/>
- MacPorts. Choisissez et installez la bonne version de dmg pour votre version de Mac OS X à partir de <http://www.macports.org/install.php>, puis ouvrez une fenêtre Terminal et exécutez:

```
sudo port selfupdate
```

- Ensuite, installez le reste des packages nécessaires avec

```
sudo port install cmake qt4-mac-devel boost xrender
```

Cela installe:

- CMake, le système de construction multiplateforme.
- Qt version 4.6.0 beta actuellement
- Boost version 1.40.0 avec libboost
- libxrender

Les autres packages nécessaires sont installés en tant que dépendances de ceux-ci.

Configuration

Téléchargez le package de code source scantailor, puis rétablissez la source en saisissant ce qui suit dans le répertoire où réside le package source:

```
tar xf scantailor-0.9.8.1.tar.gz
```

Accédez ensuite au répertoire qui contient les sources scantailor. À partir de là, exécutez ce qui suit (une seule ligne et faites attention au point à la fin):

```
cmake -DCMAKE_LIBRARY_PATH=/opt/local/lib -DCMAKE_INCLUDE_PATH=/opt/local/include
-DQT_QMAKE_EXECUTABLE=/opt/local/libexec/qt4-mac-devel/bin/qmake .
```

Il y aura beaucoup de messages de cmake se terminant par

```
- Configuring done  
- Generating done
```

- Si vous avez des dépendances manquantes, vous recevrez un message d'erreur vous indiquant ce qui manque. Vous pouvez ensuite rechercher le package manquant et l'installer.
- Si vous avez une bibliothèque et des fichiers d'en-tête installés dans un emplacement non standard, cmake ne les trouvera pas. Dans ce cas, vous pouvez les spécifier comme ci-dessus pour qmake ou exécuter le programme interactif " 'ccmake' " qui vous permet de spécifier les chemins vers les bibliothèques et les fichiers d'en-tête.

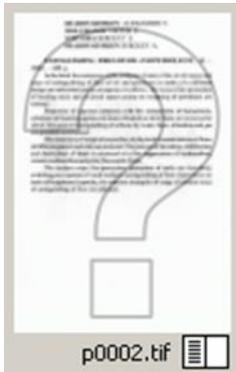
Compilation et installation

```
make  
sudo make install
```

9 ANNEXE 2 FAQ

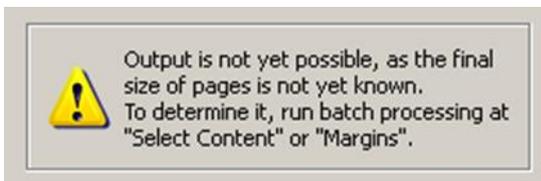
FAQ

9.1 QUE SIGNIFIE LE POINT D'INTERROGATION SUR UNE VIGNETTE DANS LE PANNEAU D'APERÇU ?



Cela signifie que cette page n'a pas encore dépassé l'étape actuelle et que, par conséquent, ST ne peut pas l'afficher comme étant en cours de traitement à ce stade. Par exemple, si l'étape actuelle est « Redresser », les pages non traitées seront affichées sans compensation d'inclinaison et avec un point d'interrogation. Il est également possible que la page ait déjà passé l'étape actuelle mais, depuis, certaines modifications ont été apportées qui nécessitent le retraitement de cette étape. Par exemple, la page avait passé l'étape « scinder les pages », mais après cela, vous avez apporté des modifications à « Fixer l'orientation ».

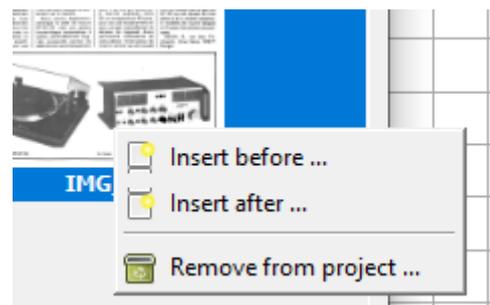
9.2 JE NE SUIS PAS AUTORISÉ À FAIRE OUTPUT, EN DISANT QUE JE DOIS TERMINER LES ÉTAPES PRÉCÉDENTES - POURTANT JE LES AI DÉJÀ RÉUSSIES !



Cela signifie qu'après avoir traité toutes les étapes précédentes, vous avez fait plusieurs pas en arrière et avez changé quelque chose. Dans ce cas, vous devrez recommencer le traitement par lots aux étapes "Sélection du contenu" ou "Mise en page". Ne vous inquiétez pas, le retraitement sera terminé beaucoup plus rapidement que la première fois, car il n'affectera probablement que les pages modifiées.

9.3 J'AI COMMENCÉ À TRAVAILLER SUR UN PROJET SCANTAILOR, MAIS J'AI OUBLIÉ QUELQUES PAGES. PUIS-JE LES AJOUTER AU PROJET EXISTANT ?

À toutes les étapes du traitement, il est possible d'ajouter / supprimer des fichiers du projet à l'aide du menu contextuel, qui est appelé en cliquant avec le bouton droit de la souris sur une vignette de page. Les nouveaux fichiers ne doivent pas nécessairement se trouver dans le même répertoire. En fait, ils ne sont même pas tenus d'avoir des noms distincts de ceux des fichiers qui se trouvent déjà dans le projet. Si vous essayez d'ajouter un fichier avec un DPI explicitement incorrect, le programme appellera la boîte de dialogue Fix DPI.



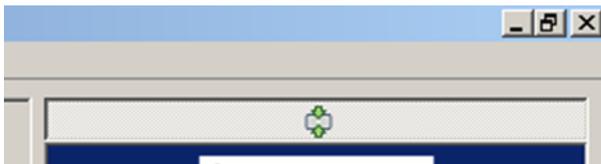
9.4 LE MODE AUTOMATIQUE DE SCANTAILOR FAIT SOUVENT ÉCHOUER MES TRAITEMENTS. POURQUOI ?

Si des erreurs se produisent très souvent, cela signifie que votre matériel source n'est pas en accord avec les présuppositions faites par ScanTailor. Voici quelques exemples des attendus de ScanTailor:

- il y a des marges autour du contenu, et plus ils sont grands, mieux c'est. Tout ce qui touche le bord est supposé être du rebus ;
- il y a au moins deux lignes de texte sur la page ;
- le DPI écrit dans les images ou spécifié manuellement est proche du vrai.
- les lettres ne sont pas trop petites (le texte est lisible, non dégradé, non flou), ce qui arrive parfois pour les prises de vue ;

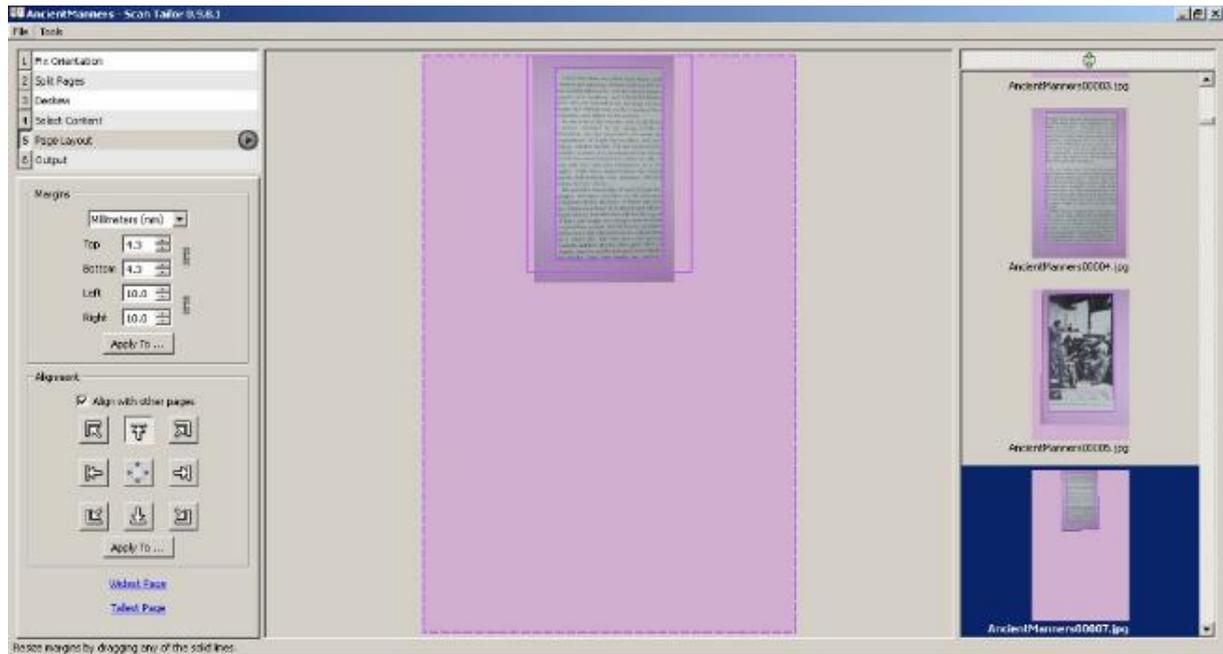
Des problèmes surviennent fréquemment lorsque le logiciel du scanner prétraite (améliore) les numérisations brutes. Ce prétraitement est une source inépuisable de problèmes. En général, un bon travail sur de telles analyses doit être perçu comme un miracle, et un mauvais - comme une réalité. Sauf si vous avez une bonne raison, fournissez à ScanTailor avec des scans / photos bruts. ScanTailor aime que son entrée soit aussi brute que possible, c'est-à-dire moins "d'améliorations" faites par le logiciel du scanner, mieux c'est.

9.5 À QUOI SERT LE BOUTON EN HAUT À DROITE (SITUÉ AU-DESSUS DU PANNEAU DES MINIATURES)? À QUOI CELA SERT-IL?



1. Il a une info-bulle dans la ligne d'état tout en bas de la fenêtre du logiciel "Maintenir l'affichage de la page courante". Ce bouton fait défiler les miniatures, ramenant la page actuelle (ouverte également dans la fenêtre centrale) dans votre vue au milieu du panneau des miniatures. Scénarios d'utilisation: Vous pouvez inspecter des miniatures arbitraires sur le panneau pendant le traitement par lots (à la recherche d'erreurs). Après avoir fait défiler manuellement les vignettes, le bouton passe en position non enfoncée. Si vous appuyez sur le bouton, la bande de vignettes reviendra à l'image de traitement actuelle. Vous pourrez ainsi regarder à tout moment combien sont déjà traitées et combien il en reste et revenir ou vous en étiez. Donc, en mode batch, ce bouton fonctionne comme un commutateur de mode de défilement automatique.
2. Ce bouton est également très utile si vous avez ordonné les vignettes en augmentant la largeur ou en augmentant la hauteur et que vous effectuez des corrections. Si tel est le cas, vous perdrez votre page hors de vue après que ses marges ou sa zone de contenu aient changé dans la fenêtre centrale. Appuyer sur ce bouton le ramène à la vue. Donc en mode manuel ce bouton fonctionne comme une commande "aller à la page courante".

9.6 SUR CERTAINES PAGES, LE CONTENU EST RÉDUIT À UNE FRACTION DE SA TAILLE D'ORIGINE. EXISTE-T-IL UNE SOLUTION DE CONTOURNEMENT À CE PROBLÈME?

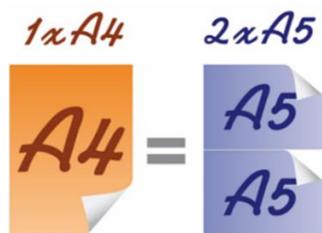


Vous devez dans ce cas vérifier s'il existe une incohérence dans les paramètres ppp (points par pouces) pour cette page. Le DPI actuel fourni par le fichier lui-même ou par vous manuellement est beaucoup plus élevé que la valeur réelle. Appelez la commande "Fix DPI" dans le menu et trouvez cette page dans la liste. Ne le faites pas dans l'onglet "Need fixing", car il ne contient que des pages manifestement erronées. Faites-le plutôt dans l'onglet "All pages" en sélectionnant cette page ou en vérifiant la cohérence des autres. En inspectant les DPI pour chaque page dans la boîte de dialogue «Fix DPI», vous devriez être en mesure de localiser et de corriger ces fichiers. Notez qu'il ne modifie pas les fichiers eux-mêmes, il enregistre simplement les valeurs correctes dans le fichier de projet de ScanTailor.

9.7 JE NE COMPRENDS PAS LE CONCEPT DPI. COMMENT ESTIMER LE DPI INCONNU?

DPI signifie «points par pouce» qui définit la correspondance entre les dimensions physiques (pouces, centimètres) et la taille en pixels d'une image. Les pixels sont ici les points qui composent l'image numérique. Les composants horizontaux et verticaux du DPI sont généralement les mêmes.

Par exemple, vous avez des numérisations d'une taille de 2816 x 2112 pixels. Si nous supposons que leur DPI est de 300 (300 horizontal et 300 vertical), alors les dimensions physiques sont : (2816/300) x (2112/300) = 9,4 x 7 pouces = 23,8 x 17,8 cm. La résolution est déterminée par les paramètres du scanner et le logiciel de numérisation produit généralement un DPI correct. Cependant, lors de manipulations ultérieures, il peut être perdu. Dans ce cas, vous devez restaurer sa valeur normale. Les DPI les plus courants sont 300 et 600. Si vous connaissez approximativement la taille du support et la taille des pixels de l'image exactement, le tableau suivant avec A4 et la moitié des formats A4 à différents DPI peut vous aider à évaluer la résolution de numérisation:



Resolution (DPI)	A4 pixel size	A5 pixel size
100	827*1169	585*827
150	1240*1754	877*1240
200	1654*2339	1170*1654
300	2480*3508	1754*2480
400	3307*4677	2339*3307
500	4134*5846	2923*4134
600	4961*7016	3508*4961

Pour les images d'appareils photo numériques, vous devez toujours corriger manuellement le DPI. Une caméra ne peut pas vous donner le DPI correct, car elle ne connaît ni la taille réelle de l'objet, ni sa distance. Vous devrez peut-être mesurer les dimensions physiques de la zone photographiée, puis diviser le nombre de pixels par la longueur en pouces pour obtenir le DPI. Gardez cela à l'esprit avant de vous séparer d'un livre que vous avez numérisé, bien que vous puissiez souvent trouver les dimensions physiques du livre en ligne. Au cas où vous n'auriez plus accès au livre physique pour le mesurer, il existe toujours un moyen d'estimer approximativement le DPI. Vous le faites en chargeant une page dans un éditeur graphique (Gimp par exemple fera l'affaire) et en utilisant un outil de sélection rectangulaire pour sélectionner 6 lignes de texte. Gimp affiche la taille en pixels de votre sélection pendant que vous la sélectionnez. La hauteur de cette sélection est à peu près votre DPI. C'est parce que la plupart des livres sont imprimés de telle manière que 6 lignes de texte s'inscrivent verticalement dans un pouce. Bien sûr, personne n'oblige les éditeurs de livres à insérer exactement 6 lignes dans un pouce. Pourtant, c'est une estimation raisonnable qui devrait produire de bons résultats avec ScanTailor. Ainsi, pour les livres avec une petite taille de police, il peut être avantageux de faire en sorte que ScanTailor pense qu'il est plus grand qu'il ne l'est en spécifiant un DPI inférieur.

Pour changer le DPI dans la boîte de dialogue «Fix DPI», vous devez travailler sur l'onglet «All pages» plutôt que sur «Need fixing», car certaines de ces images sont normales pour ScanTailor.

9.8 TOUT EST OK AVEC LE DPI DANS MES SCANS. POURQUOI SCANTAILOR ME DEMANDE-T-IL DE RÉSOUDRE CE PROBLÈME ?

Lorsque ST déclare que la résolution de l'image (dpi) nécessite une correction, il vérifie les éléments suivants horizontalement et verticalement :

1. La résolution est comprise entre [150-9999]. Si $ppp < 150$, alors il est trop petit pour obtenir des résultats acceptables. Si $ppp > 9999$, c'est probablement faux.
2. La taille de l'image physique pour le DPI donné ne dépasse pas 50 cm. ($500 < 25,4 * pixel_size / dpi$)
Donc, dans ce cas, ce n'est pas le DPI qui est faux, mais la taille de l'image qui est trop grande. Cela peut vraiment signifier que vous travaillez avec des supports trop volumineux, mais le plus souvent cela indique probablement un DPI incorrect (trop faible). Un DPI sous-estimé peut facilement conduire à un manque de mémoire, en particulier dans un environnement PC 32 bits.

Astuce: Dans ce cas, vous pouvez essayer de « bluffer » ScanTailor en entrant de fausses valeurs de dpi plus élevées dans la boîte de dialogue Fix DPI ; mais vous devez également l'augmenter proportionnellement pour les autres pages.

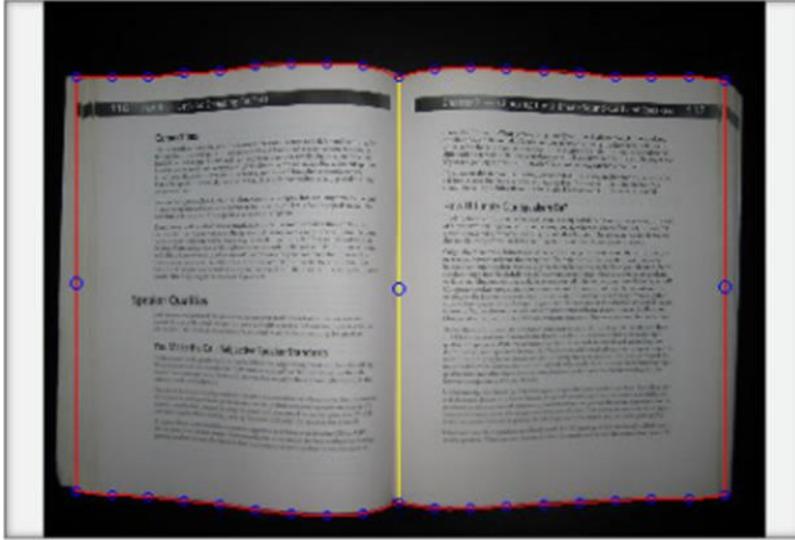
9.9 IL Y A UNE EXPOSITION À LA LUMIÈRE ÉTRANGE SUR LES NUMÉRISATIONS BLANC / NOIR. LES IMAGES DE COUVERTURE NE SONT PAS TOUJOURS BIEN PRODUITES NON PLUS. POURQUOI ? COMMENT TRAITER CES PAGES ?



ScanTailor n'a pas été conçu pour fonctionner avec du texte clair sur un fond sombre. Sortez cette page de texte en mode Niveaux de gris / Couleur, puis binarisez le texte dans Gimp / Photoshop ou un autre logiciel de

traitement d'images. La binarisation de ST ne fonctionne pas vraiment bien pour les images couleur non texte. Il est également suggéré de sortir cette page de garde en mode "Couleur / Niveaux de gris". N'activez pas l'égalisation de l'éclairage à cet endroit.

9.10 COMMENT CORRIGER LES DISTORSIONS GÉOMÉTRIQUES DE LA PAGE DE LA PHOTO CI-DESSOUS?



Tout d'abord, si vous avez deux pages sur une seule image de sortie, vos paramètres "Split Pages" sont incorrects. La déformation ne fonctionnera pas sur deux pages à la fois. Cependant, votre cas est trop difficile à gérer pour la correction automatique de la déformation, vous devrez donc ajuster la grille de déformation manuellement.

9.11 COMMENT SCANTAILOR PEUT-IL ÊTRE AUTOMATISÉ ? J'AI PROGRAMMÉ UN SCRIPT...[EN APPELANT D'AUTRES DANS UN LOT]. EXISTE-T-IL UNE VERSION EN LIGNE DE COMMANDE DE SCANTAILOR ?

La branche CLI (ScanTailor Command Line Interface) était en fait incluse dans la version principale il y a quelque temps. Donc, si vous êtes sous Windows, vous pouvez simplement télécharger le package de distribution ST et vous trouverez scantailor-cli.exe à l'intérieur. La version CLI est capable de traiter les images à partir de la ligne de commande. Vous pouvez également lui donner un fichier de projet existant. Si output_directory est spécifié comme dernier argument, il écrase celui du fichier projet.

```
scantailor-cli [options] <project_file> [output_directory]
```

```
scantailor-cli [options] <image, image, ...> <output_directory>
```

9.12 COMMENT CRÉER ST POUR WINDOWS À PARTIR DES FICHIERS SOURCE?

Il existe un fichier README dans l'archive source. Il contient les instructions de compilation.

9.13 COMMENT LE TYPE TIFF DE SORTIE EST-IL CORRÉLÉ AU TYPE D'ENTRÉE ? LE TIFF 16 BITS ENVOYÉ À LA ST PRODUIRA-T-IL DE MEILLEURS RÉSULTATS OU N'EST-IL PAS NÉCESSAIRE ?

Les TIFF de sortie ScanTailor sont de 1 bit pour le mode de sortie N / B, pour le mode Couleur/Niveaux de gris - RVB ou niveaux de gris, selon l'original, c'est soit 8, 24 ou même 32 bits si l'image d'origine avait de la transparence. Cependant, cela ne dépasse jamais 8 bits par canal. En mode mixte, ST produit des images en

niveaux de gris ou en couleur et ne produit pas d'images en noir et blanc, même pour les pages sans aucune image. C'est un détail de mise en œuvre. C'est toujours la même chose que le mode "Couleur / Niveaux de gris", c'est-à-dire en niveaux de gris 8 bits ou RVB 8 + 8 + 8 bits ou RVB 8 + 8 + 8 + 8 bits. Il n'y a aucun avantage à exporter des images 16 bits pour les saisir dans ScanTailor. Comme tout le traitement interne est effectué en 8 bits par canal, la toute première chose qui est faite à une image d'entrée est de la convertir dans ce format.

9.14 EST-IL POSSIBLE D'OBTENIR UNE SORTIE SÉPARÉE POUR LES CALQUES DE MASQUE ET D'ARRIÈRE-PLAN DANS ST AU LIEU D'ASSIGNER LA TÂCHE DE LES DIVISER À L'ENCODEUR DJVU?

La sortie séparée des images d'image et des images de texte dans ST n'est pas prévue, bien qu'elle ne soit pas difficile à organiser, car cette séparation se produit toujours en mode mixte. (POUR RÉFÉRENCE : une telle exportation de texte / images séparés existe dans la branche "Featured".) La sortie mixte de ST et en fait tous les autres modes de sortie, suivent une certaine convention pour faciliter la séparation du contenu N / B des images. La convention est de ne jamais utiliser de noir pur et de blanc pur dans les images. Les couleurs pures sont réservées aux zones de texte uniquement. Pour les images grises 8 bits dans les zones d'image, la couleur 0x00 est remplacée par la couleur 0x01 et la couleur 0xFF par 0xFE. Pour les zones d'image RGB32 et ARGB32, la couleur 0x00000000 passe à 0xFF010101 et la couleur 0x00FFFFFF à 0xFFFEFEFE. Ce comportement permet de détecter ultérieurement ces zones d'image et de les traiter différemment, par exemple en les encodant comme couche d'arrière-plan au format DjVu. C'est ainsi que des utilitaires externes tels que PDFMaker et djvubind séparent le texte et les images, en utilisant la sortie ST.

10 TABLE DES MATIÈRES

1	À propos	1
2	Installation et premier démarrage	1
2.1	Pour Windows	1
2.2	Pour GNU / Linux et Mac OS X.....	1
2.3	Création d'un projet	2
3	Le concept de traitement des scans dans ScanTailor.....	5
4	Menus du programme	8
4.1	Menu Fichier.....	8
4.2	Menu Outils	8
4.3	Menu d'aide.....	9
5	Principe de travail	10
6	Conseils pour la numérisation.....	11
7	Explications des commandes :	11
7.1	Fixer l'orientation	12
7.2	Diviser les pages	12
7.3	Redresser	14
7.4	Sélectionnez le contenu.....	15
7.5	Mise en page	16
7.6	Production du document final.....	17
7.6.1	Étape "Sortie"	17
7.6.2	Mode - Sélectionne le mode de sortie des pages:	19
7.6.3	Le format des fichiers de sortie.....	19
7.7	Onglets de sortie: zones d'image.....	19
7.8	Onglets de sortie: Dewarping	23
7.9	Onglets de sortie: zones de dépoussiérage et de remplissage.....	28
7.9.1	Élimination du bruit.....	28
7.9.2	Remplir les zones.....	29
8	Annexe 1 Building from Source Code on Linux and Mac OS X	30
8.1	Compilation Linux	30
8.2	Compilation pour Mac OS X.....	31
8.2.1	MacPorts	31
9	Annexe 2 FAQ.....	33
9.1	Que signifie le point d'interrogation sur une vignette dans le panneau d'aperçu ?	33
9.2	Je ne suis pas autorisé à faire Output, en disant que je dois terminer les étapes précédentes - pourtant je les ai déjà RÉUSSIES !	33

9.3	J'ai commencé à travailler sur un projet scantailor, mais j'ai oublié quelques pages. Puis-je les ajouter au projet existant?	33
9.4	Le mode automatique de ScanTailor fait souvent échouer mes traitements. Pourquoi ?	34
9.5	À quoi sert le bouton en haut à droite (situé au-dessus du panneau des miniatures)? À quoi cela sert-il?	34
9.6	Sur certaines pages, le contenu est réduit à une fraction de sa taille d'origine. Existe-t-il une solution de contournement à ce problème?	34
9.7	Je ne comprends pas le concept DPI. Comment estimer le DPI inconnu?	35
9.8	Tout est OK avec le DPI dans mes scans. Pourquoi ScanTailor me demande-t-il de résoudre ce problème?	36
9.9	Il y a une exposition à la lumière étrange sur les numérisations blanc / noir. Les images de couverture ne sont pas toujours bien produites non plus. POURQUOI ? Comment traiter ces PAGES ?	36
9.10	Comment corriger les distorsions géométriques de la page de la photo ci-dessous?	37
9.11	Comment ScanTailor peut-il être AUTOMATISÉ ? J'ai programmé un script...[en appelant d'autres dans un lot]. Existe-t-il une version en ligne de commande de SCANTAILOR ?	37
9.12	Comment créer ST pour Windows à partir des fichiers source?	37
9.13	Comment le type TIFF de sortie est-il corrélé au type D'ENTRÉE ? Le TIFF 16 bits envoyé à la ST produira-t-il de meilleurs résultats ou n'est-il pas NÉCESSAIRE ?	37
9.14	Est-il possible d'obtenir une sortie séparée pour les calques de masque et d'arrière-plan dans ST au lieu d'assigner la tâche de les diviser à l'encodeur DJVU?	38