

Numérisation de documents bombés : simulation de la mise à plat par *shape from shading*

Digitization of convex documents :
simulation of the setting flat by shape from shading

**Frédéric Courteille, Alain Crouzil, Jean-Denis Durou
et Pierre Gurdjos**

IRIT, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9,
{courteille,crouzil,durou,gurdjos}@irit.fr

Manuscrit reçu le 10 novembre 2005

Résumé et mots clés

Nous proposons une méthode originale permettant la numérisation rapide d'un document bombé (par exemple une page d'un livre ouvert) à partir d'une photographie de celui-ci. Cette méthode réalise la « mise à plat » du document grâce à une formulation réaliste du problème en terme de *shape from shading*, prenant en compte notamment la projection perspective. Des résultats sont présentés sur images réelles.

Numérisation, documents, mise à plat, vision par ordinateur, *shape from shading*.

Abstract and key words

We describe a novel method for the fast digitization of a skew document (e.g., a page of an open book), from one image taken by a camera. Our method carries out the setting flat of the document thanks to a realistic formulation of the problem in terms of shape from shading, involving perspective projection. Results are shown on real images.

Digitization, documents, flattening, computer vision, shape from shading.

Remerciements

Ce travail a été partiellement financé par EGIDE, dans le cadre du projet Galilée « PLATONOV » (Photocopies de Livres Anciens par Techniques d'Optimisation Numérique et sOlutions de Viscosité).

1. Introduction

La numérisation de documents connaît actuellement un essor considérable, du fait du développement d'Internet qui autorise la consultation à distance. Notre travail s'inscrit dans la lignée de nombreux programmes de recherche publics et privés sur la réalisation de bibliothèques numériques à base d'images (pro-

grammes DLI¹ pour «Digital Library Initiative» aux États-Unis, HUMI² pour «HUmanities Media Interface» au Japon, «Mémoire du Monde»³ de l'UNESCO, «Gutenberg Digital»⁴

1. <http://www.dli2.nsf.gov>

2. <http://www.humi.keio.ac.jp>

3. http://www.unesco.org/webworld/mdm/fr/index_mdm.html

4. <http://www.gutenbergdigital.de>

en Allemagne, etc.). Si le procédé classique, à l'aide d'un scanner à balayage, est satisfaisant pour les situations courantes, il présente deux faiblesses : lorsqu'il s'agit d'un livre, la manipulation répétitive consistant à tourner une page puis à retourner le livre afin de le plaquer sur la vitre du scanner peut devenir relativement fastidieuse ; par ailleurs, l'image numérisée comportera des défauts (caractères flous ou déformés, effets de parallaxe, papier non encré de niveau de gris non uniforme) si le document est « bombé », comme c'est le cas pour un livre ouvert. Pour pallier ces deux faiblesses, un certain nombre de systèmes spécifiques ont été développés, mais bien entendu il ne s'agit plus alors de matériel « grand public ».

Une alternative consiste à simuler la mise à plat des documents bombés, en remplaçant le scanner, dont le système optique est difficile à modéliser, par un appareil photographique numérique à visée réflex. Notre travail se place dans ce cadre. Pour produire une image corrigée géométriquement et photométriquement, il est nécessaire de calculer le relief du document à partir de l'image initiale, puis d'utiliser ce relief pour simuler l'image du document « mis à plat ». Plusieurs techniques classiques de reconstruction tridimensionnelle peuvent alors être utilisées. Dans le cadre de ce travail, nous utilisons une technique bien connue, le *shape from shading* (SFS), qui présente l'avantage de ne pas nécessiter de matériel sophistiqué et de pouvoir s'appliquer quasiment à tous les types de documents. L'utilisation simultanée de plusieurs techniques de reconstruction tridimensionnelle permettrait sans doute d'améliorer les performances, mais nous montrons ici que le seul SFS permet d'obtenir des résultats très satisfaisants sur images réelles.

L'engouement connu par le SFS dans les années 1980-1990 s'est régulièrement effrité entre 1996 et 2000, probablement parce que les résultats obtenus sur images réelles étaient décevants, comme l'ont montré deux articles d'état de l'art récents [ZTCS99, DFS04]. Depuis 2001, le SFS a connu un nouvel élan lorsque trois équipes ont proposé une modélisation plus réaliste [PF03, TSY03, CCDG04a], grâce à laquelle des applications ont pu être envisagées : reconstruction de visages [Pra04] ; reconstruction d'organes à partir d'images médicales [TSY04] ; simulation de la mise à plat de documents gauches [CCDG04a, CCDG04b] qui constitue notre propos. L'intérêt de scènes possédant un relief aussi simple qu'un livre ouvert posé sur une table est que le niveau de gris peut être directement interprété en termes d'orientation de la surface, sans aucune ambiguïté. En particulier, l'ambiguïté classique concave/convexe du SFS, qui est à l'origine d'illusions d'optique, est évitée. Dès lors, le SFS peut devenir une méthode de reconstruction tridimensionnelle de précision, mais encore faut-il que la relation entre le niveau de gris et l'orientation de la surface soit fiable. Nous nous attachons donc à reformuler le SFS en remettant en question le modèle de projection le plus souvent utilisé.

Dans le paragraphe 2, nous passons en revue les différentes techniques qu'il est possible d'utiliser pour la numérisation de documents et montrons en quoi l'utilisation du SFS nous semble appropriée. Dans le paragraphe 3, nous remettons en question la

modélisation classique du SFS, ce qui nous amène à une nouvelle formulation du problème. Dans le paragraphe 4, nous proposons une méthodologie originale de numérisation de documents bombés. Les résultats expérimentaux de cette application sont présentés dans le paragraphe 5. Le paragraphe 6 tire les conclusions de notre étude et dégage quelques perspectives.

2. Présentation de la problématique

2.1. Systèmes d'acquisition

Il existe plusieurs types de capteurs optiques numériques, parmi lesquels les scanners et les appareils photographiques, qui sont fondés sur le même principe mais sont conçus pour des situations bien différentes.

2.1.1. Scanners

Le scanner « classique » est spécialement conçu pour numériser les documents plans. Son capteur unidimensionnel balaie toute la surface du document afin de produire son image, ce qui en fait un capteur relativement lent. Mais si le document est parfaitement plan, alors l'image est une reproduction parfaite de l'original, c'est-à-dire qu'il n'y a ni défaut géométrique (l'original et sa copie numérique sont parfaitement superposables à un facteur d'échelle près), ni flou (il y a stigmatisme parfait), ni défaut photométrique (les niveaux de gris de la copie sont proportionnels à ceux de l'original). La qualité de l'image ainsi obtenue est seulement limitée par la résolution, qui peut quand même dépasser 1 200 *dpi* pour les scanners les plus perfectionnés. Si on doit numériser un livre de cent pages au format A4, il faut répéter cinquante fois la même suite d'opérations consistant à soulever le livre, tourner une page, reposer le livre en prenant soin de l'aplatir convenablement sur la vitre, attendre la fin de la numérisation, ce qui prend au total dix minutes environ. Mais, puisque la copie sera quasiment indiscernable de l'original, on est prêt à accepter la relative lenteur du procédé.

Lorsqu'on sort de cette utilisation « normale » du scanner, c'est-à-dire si le document n'est pas plan mais bombé, comme par exemple pour un livre épais, alors trois légers défauts apparaissent simultanément près de la reliure (*cf.* l'image I_0^1 sur la figure 1.a) : défaut géométrique (les lignes de texte ne sont plus parallèles entre elles et les caractères sont déformés), défaut de flou et défaut photométrique (le papier non encré n'a plus un niveau de gris uniforme). Nous reviendrons plus loin en détail sur les traitements qui permettent de corriger ces défauts *a posteriori*, mais intéressons-nous au préalable aux systèmes de numérisation qui ont été spécifiquement développés pour les corriger *a priori* et dont le principe est de rétablir le contact entre le



(a)



(b)

Figure 1. Page de gauche d'un livre : (a) image scannée I_0^1 et (b) photographie I_1^1 . Lors de la numérisation : en (a), le document est quasi-plan ; en (b), il est bombé.

document et la vitre du scanner. Une première solution consiste à réduire la taille du capteur, que l'on doit passer sur le document comme on le ferait avec un fer à repasser, ce qui explique que l'on appelle ces systèmes des « scanners à main ». Ce procédé permet effectivement de gommer les défauts qui ont été mentionnés plus haut. Néanmoins, il risque de devenir fastidieux si l'on doit numériser un livre en entier et, qui plus est, on ne pourra certainement pas l'utiliser dans le cas d'un document ancien, que l'on risquerait d'endommager.

D'autres systèmes, les « scanners de livres » ou « scanners à livre ouvert » ont été spécifiquement développés pour numériser les livres anciens : le livre est posé sur le dos ; il n'est plus forcément ouvert à 180 degrés, ce qui évite d'endommager la reliure ; une vitre se pose sur chaque page afin de l'aplatir sans pression excessive (« mise à plat »). Il s'agit d'un procédé relativement lent, puisque la tête du scanner doit être relevée à chaque page, mais qui fournit des résultats de très bonne qualité. Le système « Parc Book Scanner »⁵, fondé sur ce principe, a été retenu dans le cadre du programme Digital Library Initiative dont nous avons déjà parlé. Notons également l'existence d'un grand nombre de sociétés proposant des scanners fondés sur ce même principe, comme par exemple I2S⁶. Remarquons que ce genre de systèmes spécifiques s'apparentent plus à un appareil photographique numérique qu'à un scanner traditionnel puisque, très souvent, ils sont équipés d'un capteur CCD bidimensionnel. Le problème de ces systèmes est leur coût très élevé. Nous nous orientons donc préférentiellement vers les solutions consistant à

corriger les défauts des images obtenues avec des capteurs grand public.

2.1.2. Appareils photographiques numériques

Les appareils photographiques numériques comportent un capteur CCD ou CMOS bidimensionnel et leur système optique n'est plus uniquement prévu pour des scènes planes à distance fixe. En quoi un capteur non spécifique peut-il être intéressant pour numériser des documents si des capteurs spécifiques (les scanners) existent par ailleurs ? Dans le cas de documents plans, la rapidité de la numérisation est grandement accrue, comme cela a été démontré par le projet NumMobile⁷ du Centre de Recherche en Histoire des Sciences et des Techniques, mais la fidélité de l'image à l'original n'est pas aussi bonne qu'avec un scanner, à cause des « aberrations » de l'objectif. En revanche, pour des documents bombés, l'image numérique comportera les mêmes types de défauts qu'avec un scanner. Sur la figure 1, on a représenté en (a) l'image scannée I_0^1 et en (b) la photographie numérique I_1^1 d'un même document. Il apparaît que les défauts géométriques présents sur l'image I_1^1 sont nettement plus forts que ceux de I_0^1 , car le livre est posé sur le dos, alors que dans le cas de I_0^1 , il est plaqué contre la vitre du scanner, ce qui fait que le relief de la page de I_0^1 est nettement plus faible que celui de I_1^1 . Malgré cela, l'image I_1^1 est bien nette, car on a pu augmenter la profondeur de champ en diminuant l'ouverture du diaphragme. Quant au défaut photométrique, il est lui aussi beau-

5. http://www2.parc.com/eml/members/ready/parc_bookscanner.htm

6. <http://www.i2s-bookscanner.com/fr/>

7. <http://www.crhst.cnrs.fr/nm>

coup plus important sur I_1^1 que sur I_0^1 , car les pentes y sont beaucoup plus fortes. Remarquons que ce dernier défaut pourrait être évité en éclairant la page non plus au flash, mais avec un éclairage diffus, mais, comme pour les images scannées, les photographies numériques de documents bombés peuvent être corrigées *a posteriori* et nous verrons par la suite que ce « défaut » photométrique sera en fait très utile pour ce faire.

2.2. Techniques de simulation de la mise à plat

Notre ambition est la suivante : proposer un système de numérisation de documents bombés qui n'utilise pas de matériel spécifique et qui atteigne des performances voisines de celles des scanners à livre ouvert, où une réelle mise à plat des documents est effectuée. La seule possibilité semble bien être de numériser les documents avec un scanner classique ou un appareil photographique, puis d'appliquer un traitement aux images qui permette de simuler la mise à plat (l'idée ayant été originellement proposée par Wada *et al.* dans [WM92]). Ces deux étapes étant totalement indépendantes, il est envisageable de numériser un ouvrage dans sa totalité, puis d'effectuer la simulation de la mise à plat lorsque les ressources de calcul le permettent. On peut même imaginer que la simulation soit effectuée en direct, à la demande de l'internaute, s'il juge par exemple que le texte est difficilement déchiffirable près de la reliure.

2.2.1. Transformations bidimensionnelles

Il est envisageable d'effectuer un traitement purement bidimensionnel de l'image, consistant à déformer (par interpolation) l'image tout en tentant de corriger ses défauts par rapport à un modèle *a priori* d'image du document mis à plat. Dans [ZT01], l'orientation des caractères est évaluée, afin de les redresser, mais les résultats sont plutôt décevants car, s'il semble relativement aisé d'imposer que les lignes de texte soient parallèles, on voit plus difficilement comment corriger l'effet de raccourcissement des caractères situés près de la reliure. Dans [TB04], une déformation bidimensionnelle judicieuse est proposée, qui rend la page rectangulaire. Les résultats sont bons, mais l'effet de raccourcissement est corrigé en glissant une mire derrière le document. Toutes les autres méthodes proposées calculent le relief du document au moment où il a été numérisé, à l'aide des techniques classiques de la vision par ordinateur que sont : l'utilisation de deux photographies prises sous deux angles différents ou avec deux sources lumineuses différentes ; la projection d'une lumière structurée ; l'étude de la déformation des lignes de texte ; l'étude de la déformation des contours ; enfin, l'analyse des niveaux de gris des zones non encrées.

2.2.2. Utilisation de deux photographies prises sous deux angles différents

La stéréovision est une technique bien connue en vision par ordinateur, qui consiste à reconstruire le relief à l'aide de deux

photographies (ou plus) prises sous deux angles différents. La phase délicate de cette méthode est la phase d'appariement, qui consiste à chercher pour chaque pixel d'une image son correspondant dans l'autre image représentant le même point de la scène. Des mesures de corrélation de plus en plus fiables sont régulièrement proposées [CC03], et permettent d'obtenir une estimation du relief de très bonne qualité. Cependant, ce gain en qualité se fait au prix d'une augmentation considérable du temps de calcul. Ainsi dans [YKKM04], pas moins de six heures de calcul sont nécessaires pour corriger une image de taille 2048×1360 (dans [CHT04], les temps de calcul semblent moins élevés, cependant aucun résultat n'étant présenté, il est difficile d'évaluer la qualité de la méthode proposée). Par ailleurs, la méthode ne peut fonctionner que si l'on sait estimer avec précision les positions de l'appareil photographique au moment des deux prises de vue.

2.2.3. Utilisation de deux photographies prises avec deux sources lumineuses différentes

Cette technique de reconstruction tridimensionnelle, appelée « stéréophotométrie », utilise deux photographies (ou plus) prises sous le même angle, mais avec deux sources lumineuses différentes. Son principe consiste à utiliser un modèle liant le niveau de gris de l'image à la pente de la surface analysée. Elle a été proposée par Cho *et al.* [CSO97]. Les résultats présentés sont de qualité moyenne, ce qui peut être dû aux deux faiblesses suivantes : la modélisation utilisée n'est pas réaliste, car les photographies étant prises à courte distance, il est indispensable, comme nous le verrons plus loin, de prendre en compte la perspective ; la simulation de la mise à plat utilise une méthode d'interpolation linéaire.

2.2.4. Projection d'une lumière structurée

Cette technique consiste également à prendre deux photographies avec deux éclairages différents, sachant que pour une des deux photographies, un motif lumineux de structure connue est projeté sur le document [DBQ97, BS01]. En analysant la déformation du motif sur cette image, il est possible de reconstruire le relief. La deuxième photographie est nécessaire si l'on souhaite que l'image simulée ne comporte pas les artéfacts dus au maillage lumineux. Pour analyser correctement la déformation du motif, on doit segmenter précisément les zones de lumière. Dans [DBQ97], cette tâche, qui est réalisée par une transformée en ondelettes, semble constituer la principale difficulté de la méthode. Notons également l'existence de l'appareil photographique Minolta 3D1500, qui permet de projeter sur la scène des bandes de lumière colorée et qui calcule automatiquement le relief de la scène. C'est cet appareil photographique qui a été utilisé par Brown et Seales [BS00] pour la numérisation de manuscrits très abîmés, dans le cadre du projet « Digital Atheneum »⁸.

8. <http://www.digitalatheneum.org>

2.2.5. Étude de la déformation des lignes de texte

Une connaissance *a priori* sur la page consiste à supposer que le texte est formé de lignes parallèles et éventuellement équidistantes, ce qui est vrai pour la plupart des documents. Bien entendu, pour un document ne comportant aucun texte ou comportant un texte ne suivant pas des lignes horizontales, comme dans l'exemple du calligramme de Guillaume Apollinaire représenté sur la figure 2, un tel traitement serait inexploitable. Cette technique, récemment exploitée par Cao *et al.* [CDL03], fonctionne rapidement (temps de calcul de l'ordre de quelques secondes) et fournit des résultats de qualité convenable. L'étape cruciale consiste à extraire les lignes de texte, puis à calculer le relief par une analyse de la texture. La principale faiblesse de cette méthode est qu'elle repose sur une hypothèse risquée.

Une technique voisine consisterait à étudier la déformation des caractères. Si l'on suppose que la fonte ne change pas sur une même page, il semble effectivement possible, en comparant les différentes occurrences d'un même caractère situées en différents endroits de la page, d'estimer la pente locale du document. Cela suppose bien sûr que l'on soit capable de reconnaître les différentes occurrences d'un même caractère. Cette difficulté, qui relève de la reconnaissance optique de caractères, explique sans doute pourquoi l'idée n'a jamais été publiée.

2.2.6. Étude de la déformation des contours

Il est également possible d'utiliser les contours du document pour estimer son relief, si l'on suppose qu'il est rectangulaire (un document ancien risque de ne pas respecter cette hypothèse). Cette technique a été expérimentée par Kashimura *et al.* [KNO⁺99] dans le cadre du projet HUMI. La technique ne nécessitant la présence que d'un seul bord du document (bord supérieur ou bord inférieur), cela permet de ne pas trop dégrader la résolution. Cette technique a été également exploitée par Tsoi et Brown [TB04], mais les deux bords (supérieur et inférieur) de la page doivent être visibles sur l'image, ce qui est dommage car les bords d'un document contiennent généralement peu d'information.

2.2.7. Analyse des niveaux de gris des zones non encrées par *shape from shading*

Les plus anciens travaux sur la simulation de la mise à plat [WM92] utilisent cette technique, où elle est appliquée à des images provenant d'un scanner classique, alors que toutes les autres techniques utilisent des photographies numériques. Wada *et al.* ont eu l'idée d'utiliser le « défaut » photométrique d'images telles que celle de la figure 1 a), qui est en réalité lié à la pente de la surface du document aux points où le papier n'a pas été encré, pour calculer le relief du document, ce qui permet ensuite de corriger les défauts géométrique et photométrique (*cf.* également [WUM97]). L'idée a été récemment reprise et améliorée par Zhang *et al.* [ZTF04], qui obtiennent des résultats plutôt convaincants, mais à qui l'on peut faire les deux mêmes

reproches qu'à Cho *et al.* dans [CSO97] : modélisation peu réaliste et utilisation d'une méthode d'interpolation linéaire.

La méthode de simulation de la mise à plat que nous proposons dans cet article reprend la même idée, mais en essayant d'éviter les écueils que nous venons de mentionner. En premier lieu, nous optons délibérément pour l'utilisation d'un appareil photographique numérique au lieu d'un scanner, car : le système optique d'un appareil photographique est plus facile à modéliser que celui d'un scanner ; ce choix permet de numériser plus vite et sans contact direct avec le document. Par ailleurs, nous proposons une nouvelle modélisation du SFS qui prend en compte la perspective. Enfin, nous choisissons la méthode d'interpolation la plus performante parmi plusieurs méthodes existantes. Des versions moins abouties de notre méthode ont déjà été publiées dans [CCDG04a, CCDG04b].

Bien entendu, l'utilisation conjointe de plusieurs techniques permettrait sans doute d'améliorer les résultats, mais notre but est ici de montrer que le seul SFS permet de réaliser des simulations de mise à plat de photographies numériques de très bonne qualité. En outre, c'est une technique qui permet de traiter des images ne contenant pas les contours. Qui plus est, elle peut également fonctionner pour des documents ne contenant pas de texte, ou contenant du texte ne suivant pas des lignes horizontales, comme sur l'exemple de la figure 2. Ce type de document, s'il est bombé, est particulièrement difficile à numériser. En effet, comme il est impossible d'utiliser la moindre connaissance *a priori* sur le contenu (le texte n'est pas rectiligne, de multiples fontes sont mélangées), quelques-unes des techniques que nous avons décrites seraient incapables de fournir un résultat satisfaisant.

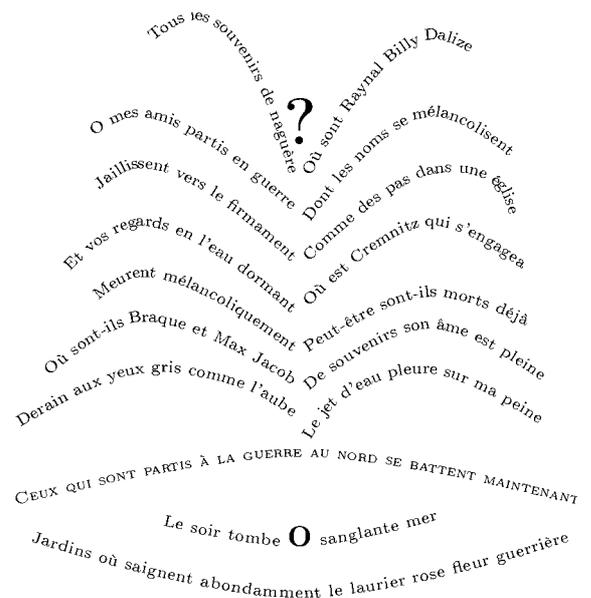


Figure 2. Calligramme de Guillaume Apollinaire : « la colombe poignardée et le jet d'eau ».

3. Modélisation du *shape from shading* en conditions réalistes d'acquisition photographique

Le *shape from shading* consiste à reconstruire le relief 3D d'une scène à partir d'une seule image de la scène, grâce à l'analyse de ses niveaux de gris.

3.1. Description générale d'un montage photographique

La figure 3 positionne les trois entités fondamentales qui doivent être mises en présence pour prendre une photographie : la scène ; la source lumineuse ; l'appareil photographique, constitué de l'objectif et du récepteur photosensible. Nous choisissons un repère tridimensionnel $\mathcal{R} = (Cxyz)$ lié à l'appareil photographique, dont l'origine est le centre optique C et tel que l'axe Cz coïncide avec l'axe optique. Il est bien connu que tout plan objet orthogonal à l'axe optique, d'équation $z = -d$, est le conjugué d'un plan orthogonal à l'axe optique, d'équation $z = d'$, tel que les distances d et d' vérifient la relation de Descartes :

$$\frac{1}{d'} + \frac{1}{d} = \frac{1}{f'} \tag{1}$$

où f' désigne la distance focale image de l'objectif (qui est positive). Le récepteur photosensible de l'appareil photographique,

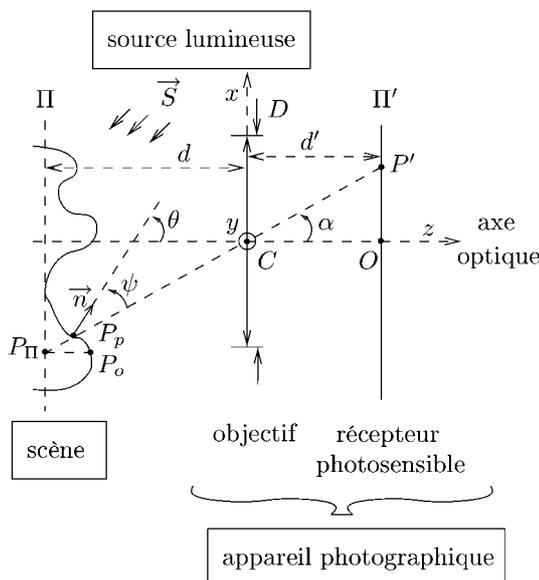


Figure 3. Schéma d'un montage photographique.

situé sur un plan Π' orthogonal à l'axe optique d'équation $z = d'$, est donc le conjugué d'un plan Π orthogonal à l'axe optique, d'équation $z = d$, appelé « plan de mise au point », qui contient les points objets dont l'image est parfaitement nette.

3.2. Équation de luminance

Le problème du SFS étant relativement complexe, il est d'usage de le simplifier grâce aux hypothèses suivantes [Dan00] :

- H1 - La réflectance de la scène, qui décrit la manière dont la lumière est réémise, est connue.
 - H2 - Les réflexions secondaires de la lumière sont négligeables.
 - H3 - L'image est nette et les éventuelles aberrations du système optique sont négligeables.
 - H4 - Le récepteur photosensible est linéaire.
 - H5 - L'angle de champ α (cf. figure 3) reste faible.
 - H6 - Le matériau est lambertien.
 - H7 - La scène est éclairée par une source lumineuse unique, émettant un flux lumineux parallèle et uniforme qui peut être décrit, en direction et en densité, par un vecteur \vec{S} .
 - H8 - La surface ne comporte ni arête, ni partie cachée, donc la normale unitaire sortante est définie en tout point.
- Sous les hypothèses H1, H2 et H3, le SFS est modélisé par une équation dite « équation de luminance » (*image irradiance equation*, cf. [HS79]) :

$$t \frac{\pi}{4} \cos^4 \alpha \frac{D^2}{d^2} L(P, \vec{PC}) = E(P'), \tag{2}$$

où t désigne le coefficient de transmission des lentilles de l'objectif, D le diamètre de sa pupille d'entrée, $L(P, \vec{PC})$ la luminance de la surface de la scène en un point objet P , dans la direction du centre optique C , et $E(P')$ l'éclairement du point image P' conjugué de P . L'hypothèse H4 consiste à supposer qu'il existe un coefficient r , appelé « facteur de réponse », tel que le niveau de gris $I(P')$ du point P' ait pour valeur :

$$I(P') = r E(P'). \tag{3}$$

L'hypothèse H5 permet de négliger l'effet du facteur $\cos^4 \alpha$. Les hypothèses H6, H7 et H8 permettent de modéliser la luminance sous une forme analytique particulièrement simple, indépendante de la direction d'émission⁹ :

$$L(P, \vec{PC}) = -\frac{\rho}{\pi} \vec{S} \cdot \vec{n}(P), \tag{4}$$

où ρ désigne l'albédo de la surface et $\vec{n}(P)$ la normale unitaire sortante au point P . En rajoutant toutes ces hypothèses, l'équation (2) peut donc être réécrite :

$$-r t \frac{\rho}{4} \frac{D^2}{d^2} \vec{S} \cdot \vec{n}(P) = I(P'). \tag{5}$$

9. L'écriture exacte de (4) est $L(P, \vec{PC}) = \max\{-\frac{\rho}{\pi} \vec{S} \cdot \vec{n}(P), 0\}$, car un point non éclairé a une luminance nulle.

La résolution de cette dernière équation n'est possible que si l'on connaît la relation entre P et P' , c'est-à-dire si l'on précise le modèle de projection. Nous allons envisager successivement la projection orthogonale, puis la projection perspective. Auparavant, nous allons rajouter aux hypothèses H1 à H8 une neuvième hypothèse, qui est très souvent rencontrée dans la littérature du SFS et qui permet d'alléger les notations :

H9 - L'éclairage est « frontal », c'est-à-dire que $\vec{S} = (0, 0, -S)$. Précisons néanmoins que les équations qui vont être introduites par la suite seraient faciles à généraliser au cas d'une direction d'éclairage quelconque.

3.3. Équation de l'eikonale

La projection orthogonale consiste à associer au point image P' le point objet P_o dont le projeté orthogonal P_{Π} sur le plan Π a comme image P' , par projection centrale de centre C sur le plan Π' (cf. figure 3). La donnée du problème est la fonction i telle que $i(x, y) = I(P')$, P' étant un point image de coordonnées (x, y, d') . L'inconnue du problème est la fonction u_o telle que P_o ait comme cote $u_o(x, y)$. Comme les coordonnées de P_o sont $(x/g, y/g, u_o(x, y))$, où $g = -d'/d$ est le grandissement transversal, $\vec{n}(P_o)$ peut facilement s'exprimer en fonction des composantes de $\vec{\nabla}u_o(x, y)$:

$$\vec{n}(P_o) = \frac{(-g \partial_x u_o(x, y), -g \partial_y u_o(x, y), 1)}{\sqrt{g^2 \|\vec{\nabla}u_o(x, y)\|^2 + 1}}. \quad (6)$$

Grâce à cette expression et à l'hypothèse H9, l'équation (5) se réécrit :

$$g^2 \|\vec{\nabla}u_o(x, y)\|^2 = \left[\frac{r t \rho D^2 S}{4 d'^2 i(x, y)} \right]^2 - 1. \quad (7)$$

Cette équation, dite « équation de l'eikonale », est la plus couramment rencontrée dans la littérature du SFS.

C'est une équation aux dérivées partielles du premier ordre (équation de Hamilton-Jacobi du premier ordre) non linéaire. Le premier membre est positif ou nul et s'annule pour les points P_o de la scène où la normale $\vec{n}(P_o)$ est parallèle à Cz , appelés « points singuliers ». Pour chacun de ces points, le niveau de gris atteint sa valeur maximale i_{\max} :

$$i_{\max} = \frac{r t \rho D^2 S}{4 d'^2}. \quad (8)$$

L'équation de l'eikonale (7) peut donc être réécrite :

$$g^2 \|\vec{\nabla}u_o(x, y)\|^2 = \frac{i_{\max}^2}{i(x, y)^2} - 1. \quad (9)$$

Le seul paramètre de cette équation est i_{\max} , qui peut être mesuré directement sur l'image si celle-ci contient au moins un point singulier.

3.4. Équation de l'eikonale perspective

La prise en compte de la projection perspective en SFS n'avait fait l'objet, jusqu'à très récemment, que de très peu de travaux [Pen89, LK94, HT96]. En outre, ces travaux préliminaires n'avaient modélisé le SFS dans le cas perspectif qu'en supposant la surface de la scène constituée de facettes triangulaires. Récemment, trois groupes d'auteurs ont établi simultanément, et indépendamment les uns des autres, une modélisation perspective générale du SFS [PF03, TSY03, CCDG04b]. Le conjugué du point image P' est maintenant le point objet P_p qui a comme image P' par projection centrale, de centre C , sur le plan Π' (cf. figure 3). Le point P_p a comme coordonnées $(x u_p(x, y)/d', y u_p(x, y)/d', u_p(x, y))$, la fonction u_p étant la nouvelle inconnue du problème. Un calcul de quelques lignes permet d'obtenir les coordonnées du vecteur $\vec{n}(P_p)$:

$$\vec{n}(P_p) = \frac{(-\hat{g}(x, y) \partial_x u_p(x, y), -\hat{g}(x, y) \partial_y u_p(x, y), 1)}{\sqrt{\hat{g}(x, y)^2 \|\vec{\nabla}u_p(x, y)\|^2 + 1}}, \quad (10)$$

où $\hat{g}(x, y) = d'/\hat{u}_p(x, y)$, avec :

$$\hat{u}_p(x, y) = u_p(x, y) + x \partial_x u_p(x, y) + y \partial_y u_p(x, y). \quad (11)$$

L'équation (5) se réécrit donc :

$$\hat{g}(x, y)^2 \|\vec{\nabla}u_p(x, y)\|^2 = \frac{i_{\max}^2}{i(x, y)^2} - 1. \quad (12)$$

Cette équation, que nous appelons « équation de l'eikonale perspective », est très similaire à (9). En particulier, il s'agit encore d'une équation de Hamilton-Jacobi du premier ordre, non linéaire. En réalité, l'équation (12) n'est pas valide pour les points P' non éclairés. En éclairage frontal, ce cas de figure ne peut pas se produire sous l'hypothèse de la projection orthogonale, mais il peut se produire sous l'hypothèse de la projection perspective, comme le montre l'exemple de la figure 4.

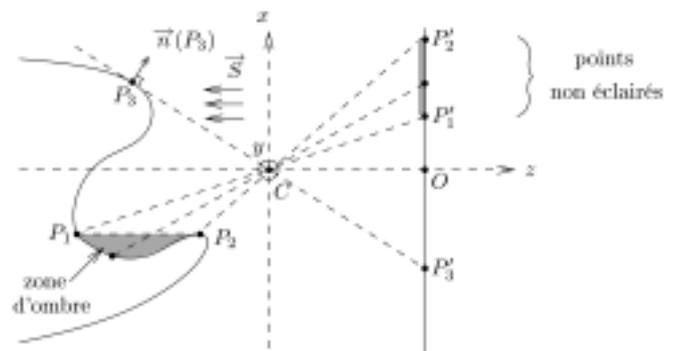


Figure 4. Existence de points non éclairés, en éclairage frontal, sous l'hypothèse de la projection perspective.

La seule différence, en apparence, entre les deux équations de l'eikonale (9) et (12), vient de ce que la constante g , qui peut être mesurée par un calibrage géométrique préalable de l'appareil photographique, est remplacée par l'expression $\hat{g}(x,y)$, mais ceci a plusieurs conséquences notables. Alors que l'inconnue u_o n'apparaît dans (9) qu'au travers de son gradient, il n'en va plus de même pour l'inconnue u_p dans (12), qui apparaît explicitement (mais indirectement) dans l'expression (11) de $\hat{u}_p(x,y)$. À partir d'une solution u_o de (9), on peut construire une infinité de solutions $u_o + K$, K étant un réel quelconque, ce qui n'est plus le cas pour l'équation (12). En revanche, on peut remarquer que l'équation (12) est homogène, ce qui signifie qu'à partir d'une solution u_p de (12), on peut construire une infinité de solutions $K' u_p$, K' étant un réel quelconque. En d'autres termes, la résolution de l'équation de l'eikonale (9) ne peut être faite qu'à une constante additive près, alors que la résolution de l'équation de l'eikonale perspective (12) ne peut être faite qu'à une constante multiplicative près.

Il existe une autre différence, tout aussi fondamentale, entre ces deux équations. Dans (9), les variables x et y apparaissent implicitement, alors qu'elles apparaissent explicitement dans (12), comme on peut le vérifier en observant l'expression de $\hat{u}_p(x,y)$. Pour la résolution de cette dernière équation, il est donc nécessaire de connaître la position du « point principal » O , qui est l'intersection de l'axe optique et du plan Π' . Les coordonnées du point principal peuvent bien sûr être déterminées par calibrage géométrique de l'appareil photographique, mais il faut garder à l'esprit que si celui-ci n'est pas fait avec précision, alors la résolution de (12) ne pourra pas non plus être effectuée correctement.

On peut craindre que la résolution de (12) ne soit plus difficile que celle de (9). Dans [PF03], (12) est résolue par le biais de la recherche de ses solutions de viscosité. Un schéma numérique d'approximation de ces solutions, ainsi que la preuve de convergence de ce schéma, sont tirées de la généralisation des résultats de [LRT93]. Nous abordons la résolution des équations (9) et (12) sous un angle relativement différent.

3.5. Pseudo-équation de l'eikonale

Comme de nombreux auteurs [HB86], nous rajoutons comme inconnue au problème du SFS (orthogonal ou perspectif) la normale $\vec{n}(P)$. Ce vecteur unitaire ayant deux degrés de liberté, il est équivalent d'introduire deux fonctions inconnues scalaires, par exemple p et q telles que :

$$\vec{n}(P) = \frac{(-p(x,y), -q(x,y), 1)}{\sqrt{p(x,y)^2 + q(x,y)^2 + 1}}. \quad (13)$$

Remarquons que de cette façon, les fonctions p et q sont bien définies sans ambiguïté, puisque d'après l'hypothèse H8, la normale unitaire sortante est définie sans ambiguïté.

L'équation (5) se réécrit alors :

$$p(x,y)^2 + q(x,y)^2 = \frac{i_{\max}^2}{i(x,y)^2} - 1. \quad (14)$$

Contrairement à (9) et à (12), cette équation n'est pas une équation aux dérivées partielles. Pour cette raison, nous proposons de l'appeler « pseudo-équation de l'eikonale ». Insistons sur le fait qu'elle est valable pour les deux types de projection, orthogonale et perspective. Sa résolution est évidemment un problème mal posé, puisqu'en chaque pixel (x,y) il y a une équation pour deux inconnues $p(x,y)$ et $q(x,y)$, mais différentes stratégies¹⁰ peuvent être envisagées pour le rendre bien posé [LK94].

Une fois p et q calculées, il reste ensuite à calculer le relief à partir de p et q , via une étape connue sous le nom « d'intégration ». Sous l'hypothèse de la projection orthogonale, on déduit de (6) et (13) que :

$$\vec{\nabla} u_o(x,y) = \left(\frac{p(x,y)}{g}, \frac{q(x,y)}{g} \right). \quad (15)$$

En suivant un raisonnement similaire, sous l'hypothèse de la projection perspective, on déduit de (15) et (13) que :

$$\vec{\nabla} u_p(x,y) = \left(\frac{p(x,y)}{\hat{g}(x,y)}, \frac{q(x,y)}{\hat{g}(x,y)} \right). \quad (16)$$

Cette égalité équivaut au système linéaire suivant (les dépendances en (x,y) ont été omises pour alléger les notations) :

$$\begin{cases} [x p - d'] \partial_x u_p + y p \partial_y u_p = -p u_p, \\ x q \partial_x u_p + [y q - d'] \partial_y u_p = -q u_p. \end{cases} \quad (17)$$

Le déterminant δ de ce système d'équations linéaires en $\partial_x u_p$ et $\partial_y u_p$ vaut $d'(d' - x p - y q)$. Or, en utilisant les notations de la figure 3, le produit scalaire $\vec{n}(P_p) \cdot \vec{C P}'$ vaut : d'une part, $(d' - x p - y q) \cos \theta$, en utilisant (13) et en remarquant que la troisième coordonnée de $\vec{n}(P_p)$ vaut $\cos \theta$; d'autre part, $d' \cos \psi / \cos \alpha$, en utilisant le fait que $\|\vec{C P}'\| = d' / \cos \alpha$ et en introduisant l'angle $\psi = (\vec{C P}', \vec{n}(P_p))$. Par conséquent :

$$\delta = \frac{d'^2 \cos \psi}{\cos \alpha \cos \theta}, \quad (18)$$

qui ne s'annule que si $\psi = \pi/2$. Sinon, c'est-à-dire si P' n'est pas situé sur un « contour occultant » (en guise d'illustration, le point P'_3 de la figure 4 est situé sur un contour occultant), alors la solution de (17) se trouve facilement :

$$\vec{\nabla} u_p(x,y) = \frac{u_p(x,y)}{d' - x p(x,y) - y q(x,y)} (p(x,y), q(x,y)). \quad (19)$$

10. Dans le cadre de notre application, nous verrons que des connaissances *a priori* sur la surface de la scène contraignent suffisamment le problème pour nous permettre de calculer p et q sans ambiguïté.

L'égalité (15) est un système de deux équations aux dérivées partielles en u_o , de même que l'égalité (19) est un système de deux équations aux dérivées partielles en u_p , mais contrairement aux équations de l'eikonale (9) et (12), il s'agit ici d'équations linéaires, qui seront donc beaucoup plus simples à résoudre. On peut observer les mêmes différences entre (15) et (19) qu'on avait déjà faites entre (9) et (12), à savoir: la résolution de (15) ne peut être faite qu'à une constante additive près, alors que celle de (19) ne peut être faite qu'à une constante multiplicative près; la résolution de (19) requiert la connaissance des coordonnées du point principal, car les variables x et y apparaissent explicitement au dénominateur du membre droit, ce qui n'est pas le cas pour la résolution de (15).

4. Application à la correction de photographies de documents gauches

Maintenant que nous avons défini une modélisation réaliste du SFS, nous pouvons considérer l'application qui nous intéresse, en l'occurrence la simulation de la mise à plat de documents bombés.

4.1. Montage photographique utilisé et discussion des hypothèses

La figure 5 montre le montage photographique utilisé. Le document est posé sur le dos, sur une table horizontale qui a été préalablement recouverte d'un tissu noir. Ceci permet de limiter les éventuelles réflexions secondaires de la scène qui pourraient considérablement dégrader le relief reconstruit par SFS, et de ce fait la qualité de la mise à plat. Nous avons utilisé un appareil numérique à visée réflex (Canon EOS 300D, capteur CMOS de 6,5 mégapixels), fixé sous un pied à une distance d'environ 400 mm de la table, et orienté (par calibrage) de telle sorte que les bords gauche et droit du document coïncident avec des colonnes de l'image (pour lesquelles la coordonnée x est constante). L'utilisation d'un appareil photographique permet, en outre, une vitesse de numérisation élevée, puisque notre système peut fonctionner à la cadence de quarante pages par minute environ.

Comme nous l'avons vu précédemment, si nous voulons utiliser l'équation (19), il est nécessaire d'estimer les paramètres internes de l'appareil photographique, (la distance d' et la position du point principal O). Pour cela, nous avons utilisé la méthode de calibrage géométrique classique de Bouguet¹¹. Par



Figure 5. Montage photographique utilisé.

ailleurs, nous utilisons comme source lumineuse le flash intégré de l'appareil photographique. De cette façon le montage optique est extrêmement facile à réaliser. L'acquisition des images est réalisée page par page, en prenant soin de cacher une des deux pages par du tissu noir. L'intérêt de procéder ainsi, et non pas réellement avec un livre ouvert, est d'éviter les réflexions secondaires près de la reliure, conformément à l'hypothèse H2. Il existe bien des techniques permettant de prendre en compte ce phénomène, mais ayant testé la méthode décrite dans [NIK91] et constaté que le temps de calcul était très élevé, nous préférons éluder ce problème pour le moment.

Afin de vérifier l'hypothèse H3, l'objectif est ouvert le moins possible, ce qui minimise le flou, et la distorsion est estimée lors du calibrage. En revanche, les hypothèses H4, H5, H6 et H7 ne sont pas réalistes. Une façon de pallier ce problème (et aussi de prendre en compte l'aberration de vignettage) consiste à réaliser un calibrage photométrique (*cf.* le prochain paragraphe).

4.2. Reconstruction du relief et élimination des ombres propres

La figure 6 décrit la chaîne de traitement telle qu'elle est mise en œuvre. La donnée est l'image I_1 d'origine (prise avec le système optique décrit précédemment), et il s'agit donc de produire comme résultat une image I_6 qui simule la mise à plat. Pour cela, la première étape consiste à appliquer à l'image I_1 une correction photométrique élémentaire, réalisée en photographiant une page blanche (plus grande que le livre ouvert), posée à plat sur la table, puis à diviser I_1 par cette image de référence I_{ref} (corrigée de la distorsion), ce qui fournit une nouvelle image I_2 .

11. <http://www.vision.caltech.edu/bouguetj>

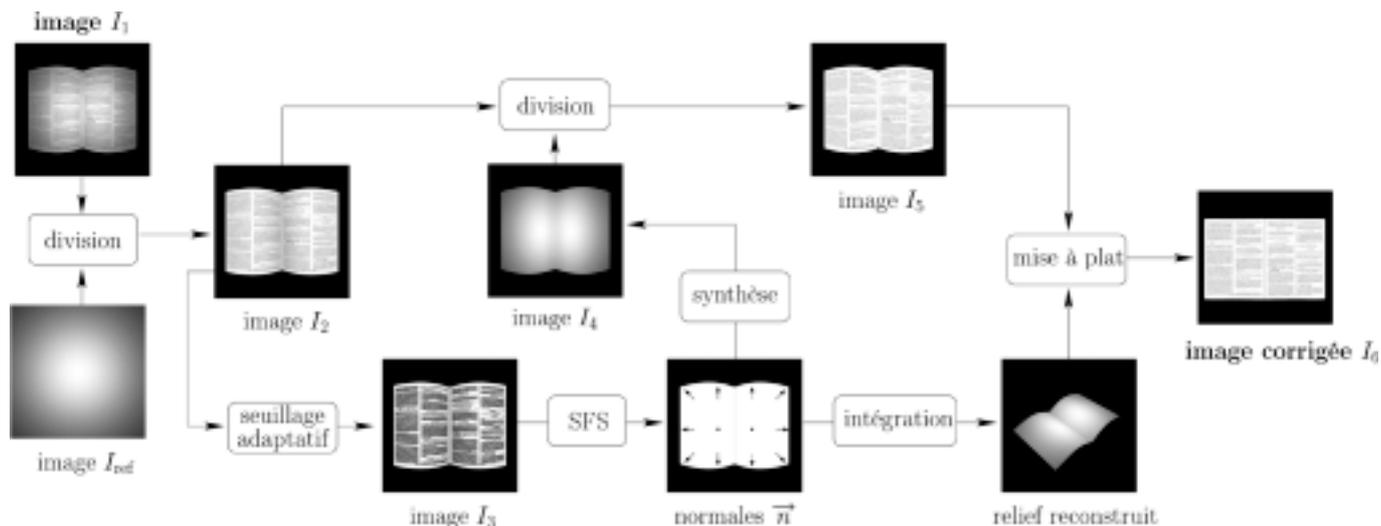


Figure 6. Détail de la chaîne de traitement.

Cette procédure est primordiale puisque, comme nous pouvons le constater facilement, l'image I_{ref} n'a pas un niveau de gris uniforme, comme cela devrait être le cas si les hypothèses H1 à H9 étaient parfaitement respectées. Par ailleurs, les équations du SFS sont fondées sur l'hypothèse implicite d'un albédo uniforme. Or, il est patent que l'albédo n'est pas uniforme : le papier non encré a un albédo ρ_0 quasiment égal à 1, alors que les zones encrées ont des albédos qui peuvent prendre toutes les valeurs de l'intervalle $[0, \rho_0[$. Il est donc nécessaire de différencier les zones encrées des zones non encrées. On cherche donc à calculer une image I_3 sur laquelle toutes les zones encrées sont masquées. Pour réaliser cette tâche, nous appliquons un seuillage adaptatif à l'image I_2 , dont chaque seuil est calculé localement sur une fenêtre de 5×5 pixels.

À présent, nous pouvons calculer le relief du document par SFS. Comme cela avait été annoncé, le calcul des fonctions p et q est facilité par le fait que la surface du document est « réglée développable », c'est-à-dire que deux points d'une même colonne de l'image correspondent à deux points de la scène de même altitude, ce qui implique que q est nulle et que les fonctions p et u_p ne dépendent que de la variable d'espace x . L'équation (13) permet donc de calculer $p(x)^2$ en fonction de la moyenne en y de $i(x, y)$ sur les pixels non encrés, notée $\bar{i}(x)$:

$$p(x)^2 = \frac{\bar{i}_{max}^2}{\bar{i}(x)^2} - 1, \tag{20}$$

où \bar{i}_{max} est la valeur maximale de $\bar{i}(x)$, atteinte en $x = x_{max}$. De plus, la surface étant convexe, le signe de p est positif si $x \geq x_{max}$ et négatif si $x \leq x_{max}$. Par la suite, nous calculons u_p en intégrant l'équation (19). La connaissance du relief du document est utile à double titre : elle doit permettre, bien sûr, de simuler la photographie du document à plat, mais elle est également utile pour calculer l'image I_5 des albédos. Pour ce faire, nous calculons, grâce à (12), l'image I_4 des ombres propres cor-

respondant au relief estimé et nous obtenons I_5 en divisant I_2 par I_4 . Chaque point (x, y) correspond donc à un albédo et à un point de la surface du document, de coordonnées $(x, u_p(x, y)/d', y, u_p(x, y)/d', u_p(x, y))$. Pour simuler la photographie I_6 du document mis à plat, il suffit alors de « déplier » la surface gauche. Pour cela, plusieurs techniques sont envisageables.

4.3. Mise à plat du document

Contrairement à ce qui a été fait par certains auteurs [CSO97, KNO⁺99] qui ont développé eux-mêmes des méthodes d'interpolation linéaires, nous avons testé trois méthodes d'interpolation classiques. La première de ces techniques est fondée sur l'algorithme de Smythe [Smy90]. Cette technique, qui consiste à appliquer à l'image deux interpolations unidimensionnelles successives (suivant les lignes et suivant les colonnes), a donné naissance au logiciel libre *xmorph*, permettant entre autres de faire de la déformation et du plaquage d'images. La seconde technique consiste à utiliser la fonction `griddata` de Matlab, qui effectue une triangulation de Delaunay puis une interpolation sur l'image grâce aux coordonnées barycentriques de chacun des points discrétisant la scène. La troisième technique consiste à utiliser la puissance matérielle par le biais d'OpenGL (interface de programmation vers le matériel graphique). Dès lors que l'on connaît le relief du document, il devient très facile, grâce aux fonctions de plaquage de texture et au lancer de rayons, de simuler la mise à plat.

Le tableau 1 synthétise les avantages et inconvénients de ces différentes méthodes. Visuellement, nous avons pu constater que les résultats obtenus avec ces trois techniques sont similaires. Donc, pour des raisons de rapidité, de fonctionnalité et de facilité d'intégration, nous avons retenu OpenGL.

Tableau 1. Comparaison des trois méthodes d'interpolation testées.

	xmorph	griddata	OpenGL
Avantages	<ul style="list-style-type: none"> • rapide (environ 10 s) • logiciel libre • utilisé depuis longtemps 	<ul style="list-style-type: none"> • très précis 	<ul style="list-style-type: none"> • très rapide (environ 1 s) • mise en œuvre facile • puissance matérielle
Inconvénients	<ul style="list-style-type: none"> • précision en-deçà du pixel • difficulté d'utilisation 	<ul style="list-style-type: none"> • lent (environ 1 min) • logiciel non libre • triangulation inutile 	<ul style="list-style-type: none"> • affichage du résultat à l'écran inévitable

5. Résultats expérimentaux

5.1. Analyse qualitative des résultats

En nous référant à la figure 7, nous pouvons voir le résultat I_6^1 fourni par notre algorithme à partir de l'image I_1^1 . Nous rappelons au lecteur que seules les zones non encrées de I_1^1 sont analysées. Ainsi, le quasi-parallélisme des lignes de texte dans I_6^1 n'est pas utilisé comme connaissance *a priori* sur le document analysé. De la même façon, le cadre au bas de l'image I_6^1 a retrouvé sa forme rectangulaire originelle. Notre technique de mise à plat semble donc plutôt satisfaisante. Du point de vue de la correction photométrique, il apparaît que certains caractères de l'image corrigée, en particulier ceux qui sont situés au voisinage de la zone de pente nulle, sont légèrement moins contras-

tés que d'autres. Ceci provient probablement du fait que les hypothèses du SFS ne sont pas parfaitement vérifiées, et que la correction photométrique (qui est censée gommer les écarts aux hypothèses H_4 , H_5 , H_6 et H_7) est moins performante en ces points. Il serait intéressant de modifier l'éclairage, en utilisant par exemple un flash annulaire, ou de proposer une modélisation plus fine de l'éclairage.

Comme on peut le voir sur la figure 8, qui représente en trois dimensions le document reconstruit (a) avant et (b) après mise à plat, le relief estimé par SFS présente de fortes pentes, ce qui souligne la bonne qualité du résultat.

La figure 9 montre deux imagerie extraites des images I_1^1 et I_6^1 , correspondant à une même zone de texte. À l'œil nu, il semble que la correction soit de très bonne qualité.

La figure 10 représente deux occurrences du même mot « capitaines » extraites de différentes positions. Le mot M_6 est le résultat de la mise à plat de M_1 , qui se situe près de la reliure (la déformation élevée de M_1 est due à la non planarité de la page à cet endroit). Le mot N_6 est le résultat de la mise à plat de N_1 , qui se situe dans une zone où la pente de la page est relativement



(a)



(b)

Figure 7. Exemple 1 : (a) photographie I_1^1 et (b) image corrigée I_6^1 .



Figure 8. Vue 3D de la surface reconstruite de la page : (a) avant et (b) après mise à plat.

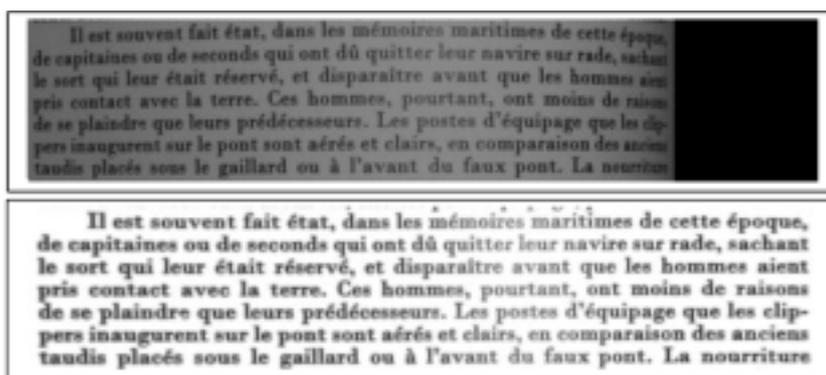


Figure 9. Extraits des images I_1^1 et I_6^1 .

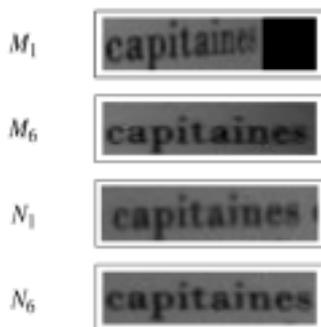


Figure 10. Quatre occurrences du mot « capitaines » extraites des images I_1^1 et I_6^1 : M_1 devient M_6 après mise à plat ; N_1 devient N_6 après mise à plat.

faible. On peut donc conclure que la correction est de bonne qualité, puisque toutes les lettres de M_6 et N_6 sont alignées horizon-

talement, qu'elles ont toutes la même taille et que deux occurrences du même mot, qui apparaissent très différentes dans l'image d'origine I_1^1 , deviennent semblables après mise à plat.

La figure 11 montre un autre exemple. Le document photographié est un plan de mat de navire. On peut remarquer sur l'image I_6^2 que la correction apportée a permis de restaurer les lignes droites.

Les deux derniers exemples illustrent clairement le fait que notre algorithme n'utilise que les zones non encrées du document, indépendamment de sa structure géométrique, de sa mise en forme ou des éventuelles figures qu'il comporte. L'exemple de la figure 2 représente les images du calligramme « la colombe poignardée et le jet d'eau » : (a) scannée, (b) photographiée et (c) corrigée.

On peut remarquer que la photographie comporte bien les défauts géométrique et photométrique déjà mentionnés. Notamment, le mot « Jardins » (premier mot de la dernière ligne de texte) est complètement déformé et presque illisible, alors

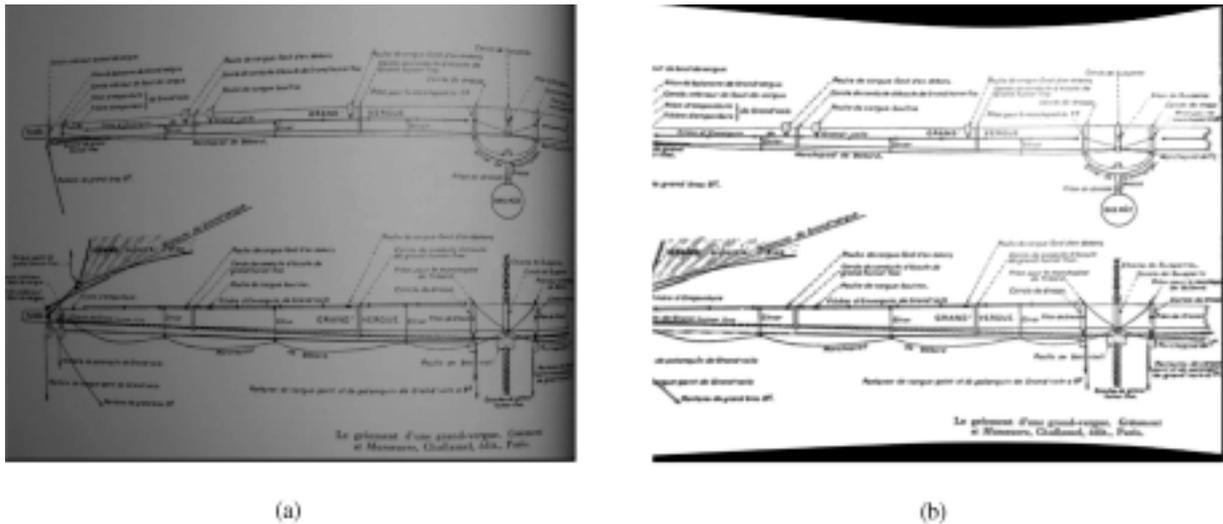


Figure 11. Exemple 2 : (a) photographie I_1^2 et (b) image corrigée I_6^2 .

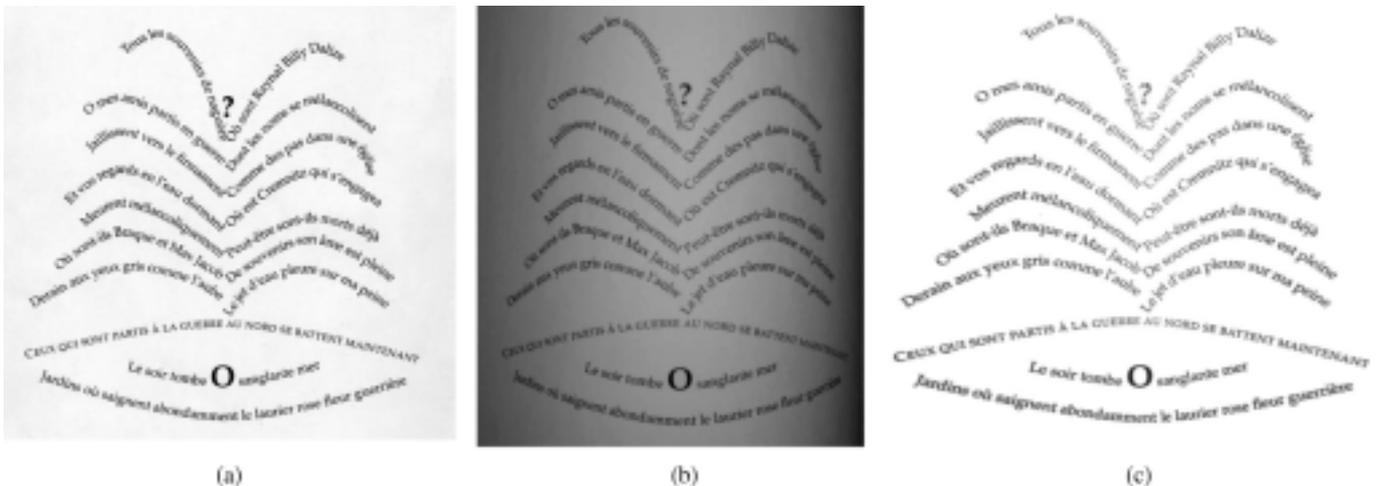


Figure 12. Exemple 3 : (a) image scannée I_0^3 , (b) photographie I_1^3 et (c) image corrigée I_6^3 .

que le même mot, une fois mis à plat, est parfaitement lisible, tant au niveau de sa géométrie que de son contraste par rapport au fond blanc. Cependant, le résultat n'est pas parfait, puisque si l'on compare l'image du document scannée à plat et l'image corrigée obtenue par notre méthode, on peut s'apercevoir que, d'une part, le texte est légèrement déformé (différentes occurrences d'un même caractère n'ayant pas la même taille), et que d'autre part, le défaut photométrique déjà présent sur l'exemple de la figure 7 est encore visible.

La figure 13 montre un dernier exemple, qui cette fois-ci ne comporte que des dessins. La mise à plat d'un tel document serait difficilement réalisable par des méthodes n'utilisant pas le SFS. On peut observer que les deux figurines situées sur les bords gauche et droit de l'image de la figure 13.b (avant mise à plat) sont très dégradées par rapport à l'original (cf. figure 13.a). Néanmoins, la mise à plat est de très bonne qualité : les trois figurines sont bien restaurées, tant au niveau de la forme que de

la couleur (cf. figure 13.c). Ce résultat n'est cependant pas parfait, puisque le contraste n'est pas le même sur l'image scannée I_0^4 et sur l'image corrigée I_6^4 . Insistons sur le fait que, dans le cadre de ce travail, tous les résultats qui ont été présentés ont été obtenus avec la chaîne de traitement de la figure 6.

5.2. Analyse quantitative des résultats

Afin d'évaluer quantitativement notre méthode de mise à plat, il semble valide de comparer les performances d'un logiciel de reconnaissance de caractères (OCR) appliqué aux trois images suivantes d'un même document comportant du texte : image scannée I_0 , photographie I_1 et image corrigée I_6 .

Nous avons appliqué l'OCR libre `gocr` aux trois images I_0^1 (cf. figure 1.a), I_1^1 et I_6^1 . Les résultats, reportés dans le tableau 2, montrent que le pourcentage de caractères reconnus avec succès

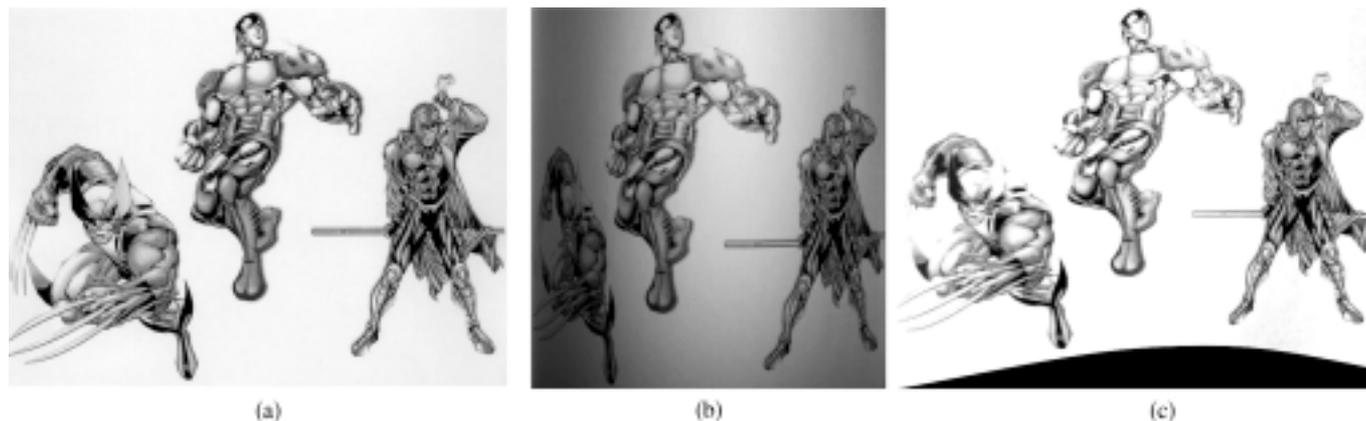


Figure 13. Exemple 4 : (a) image scannée I_0^4 , (b) photographie I_1^4 et (c) image corrigée I_6^4 .

Tableau 2. Résultats de l'OCR sur les zones de texte des images I_0^1 , I_1^1 et I_6^1 .

	I_0^1	I_1^1	I_6^1
Taux de succès	92 %	2 %	61 %
Taux d'échec	6 %	70 %	24 %
Taux d'abstention	2 %	28 %	15 %

a été amélioré d'un facteur 30 environ entre I_1^1 et I_6^1 (61 % au lieu de 2 %). Le pourcentage d'échecs a été réduit d'un facteur 3. Le taux d'abstention a été, quant à lui, divisé d'un facteur 2, ce qui nous permet de conclure que les corrections géométrique et photométrique effectuées permettent à l'OCR de décider plus souvent et de manière plus fiable. Cependant, nous sommes encore loin du taux de succès (92 %) obtenu sur l'image scannée I_0^1 , ce qui s'explique en partie par le fait que la résolution est meilleure pour les images scannées que pour les photographies numériques: l'image I_0^1 comporte 2264×2581 pixels, contre 1024×1360 pixels pour la photographie I_1^1 .

5.3. Limites de la méthode proposée

Nous venons de voir que, si les hypothèses du SFS sont respectées, les résultats optenus sont de bonne qualité. Cependant, notre méthode n'est plus utilisable dès lors qu'une des hypothèses du SFS n'est plus valide, comme par exemple pour les deux photographies de la figure 14.

En effet, la photographie de la figure 14.a a été prise sans aucun contrôle sur la source lumineuse (en direction et en intensité). Ainsi, les hypothèses H2 et H7 ne sont pas vérifiées. Par

ailleurs, le document dont la photographie est représentée sur la figure 14.b n'est pas lambertien et son albédo est inconnu. De ce fait, les hypothèses H1 et H6 ne sont pas vérifiées.

6. Conclusion

Nous avons proposé dans cet article une chaîne complète de traitement permettant la numérisation rapide et peu coûteuse de documents bombés. La méthode proposée consiste à utiliser un appareil photographique numérique grand public pour acquérir les images, qui comportent des défauts géométrique et photométrique, puis à corriger ces défauts dans un deuxième temps, en simulant la mise à plat. Les résultats obtenus sur images réelles prouvent le potentiel de cette méthode. Même si le taux de reconnaissance de caractères est meilleur sur une image scannée à plat, il n'en reste pas moins que notre méthode présente, par rapport aux images scannées à plat, les avantages d'être plus rapide et non intrusive.

De nombreuses améliorations, qui ont déjà été évoquées dans l'article, doivent être envisagées, notamment :

- La méthode de résolution du SFS que nous avons utilisée doit être généralisée aux documents de surface quelconque, c'est-à-dire pour lesquels l'hypothèse $q = 0$ n'est plus valide. Cette amélioration permettrait par exemple de numériser des documents anciens.
- L'utilisation conjointe d'une technique utilisant les contours du document, supposé rectangulaire, permettra de mieux contraindre le problème du calcul du relief par SFS.
- Le système d'éclairage pourrait être amélioré en remplaçant le flash interne par un flash annulaire.
- La méthode de calibrage photométrique relativement naïve que nous avons proposée présente l'avantage de gommer d'un coup les écarts aux hypothèses H4, H5, H6 et H7, mais on doit encore pouvoir l'améliorer en modélisant plus finement, par exemple, la réflectance de la scène.

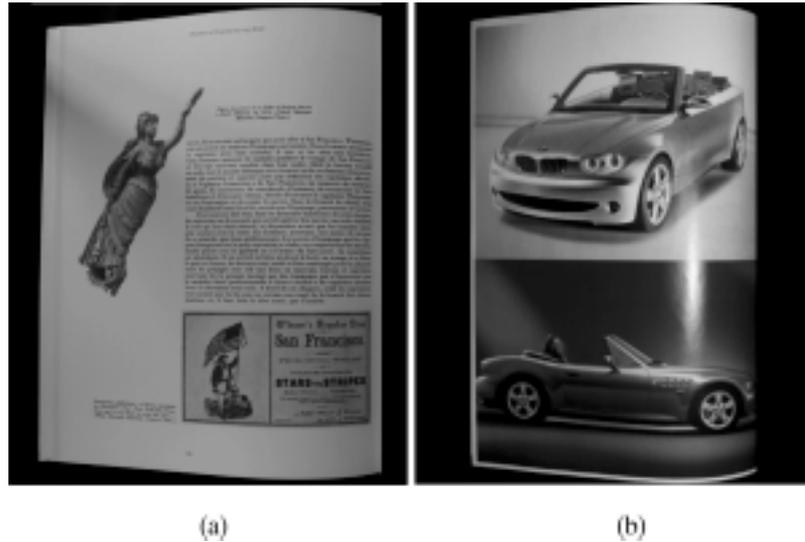


Figure 14. Exemples d'images sur lesquelles notre méthode ne peut être utilisée :
 (a) photographie prise dans des conditions ne vérifiant pas les hypothèses H2 et H7 ;
 (b) photographie d'un document ne vérifiant pas les hypothèses H1 et H6.

• La prise en compte des réflexions secondaires permettrait de numériser les deux pages d'un livre simultanément. À l'heure actuelle, notre méthode ne fonctionne réellement que si l'une des deux pages est recouverte d'un tissu noir.

À terme, nous envisageons d'ajouter aux fonctionnalités de l'appareil photographique la simulation de la mise à plat de documents, comme cela est déjà le cas pour certains traitements tels que la création d'images panoramiques.

Références

- [BS00] BROWN M.S., SEALES W.B., «Beyond 2D Images: Effective 3D Imaging for Library Materials». Dans *Proceedings of the 5th ACM Conference on Digital Libraries*, pages 27-39, San Antonio, Texas, États-Unis, juin 2000.
- [BS01] BROWN M.S., SEALES W.B., «Document Restoration Using 3D Shape: A General Deskewing Algorithm for Arbitrarily Warped Documents». Dans *Proceedings of the 8th IEEE International Conference on Computer Vision (volume I)*, pages 365-375, Vancouver, Canada, juillet 2001.
- [CC03] CHAMBON S., CROUZIL A., «Dense matching using correlation: new measures that are robust near occlusions », Dans *Proceedings of the British Machine Vision Conference*, pages 143-152, Norwich, Royaume-Uni, septembre 2003,
- [CCDG04a] COURTEILLE F., CROUZIL A., DUROU J.-D., GURDJOS P., «Shape from shading en conditions réalistes d'acquisition photographique». Dans *Actes du 14^{ème} Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (volume II)*, pages 925-934, Toulouse, France, janvier 2004.
- [CCDG04a] COURTEILLE F., CROUZIL A., DUROU J.-D., GURDJOS P., «Towards shape from shading under realistic photographic conditions». Dans *Proceedings of the 17th International Conference on Pattern Recognition (volume II)*, pages 277-280, Cambridge, Royaume-Uni, août 2004.
- [CDL03] CAO H., DING X., LIU C., «A Cylindrical Surface Model to Rectify the Bound Document Image». Dans *Proceedings of the 9th IEEE International Conference on Computer Vision (volume I)*, pages 228-233, Nice, France, octobre 2003.
- [CHT04] CUBAUD P., HAAS J.-F., TOPOL A., «Numérisation 3D de documents par photogrammétrie». Dans *Actes du Colloque International Francophone sur l'Écrit et le Document*, pages 228-233, La Rochelle, France, juin 2004.
- [CSO97] CHO S. I., SAITO H., OZAWA S., «A Divide-and-conquer Strategy in Shape from Shading Problem». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 413-419, San Juan, Porto Rico, juin 1997.
- [Dan00] DANIEL P., «Peut-on extraire le relief d'une seule image ? ». Thèse de doctorat, Université Paul Sabatier, Toulouse, France, janvier 2000.
- [DBQ97] DONCESCU A., BOUJU A., QUILLET V., «Former books digital processing: image warping». Dans *Proceedings of the IEEE Workshop on Document Image Analysis*, pages 5-9, San Juan, Porto Rico, juin 1997.
- [DFS04] DUROU J.-D., FALCONE M., SAGONA M., «A Survey of Numerical Methods for Shape from Shading». Rapport de Recherche 2004-2-R, Institut de Recherche en Informatique de Toulouse, Toulouse, France, janvier 2004.
- [HB86] HORN B. K. P., BROOKS M. J., «The Variational Approach to Shape From Shading». *Computer Vision, Graphics, and Image Processing*, 33(2):174-208, février 1986.
- [HS79] HORN B. K. P., SJOBERG R. W., «Calculating the Reflectance Map». *Applied Optics*, 18(11):1770-1779, juin 1979.
- [HT96] HASEGAWA J. K., TOZZI C. L., «Shape from Shading with Perspective Projection and Camera Calibration». *Computers and Graphics*, 20(3):351-364, juin 1996.

- [KNO*99] KASHIMURA M., NAKAJIMA T., ONDA N., SAITO H., OZAWA S., «Practical Introduction of Image Processing Technology to Digital Archiving of Rare Books». Dans *Proceedings of the International Conference on Signal Processing Application Technology*, Toronto, Canada, octobre 1999.
- [LK94] LEE K. M., KUO C.-C. J., «Shape from Shading with Perspective Projection». *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(2):202-212, mars 1994.
- [LRT93] LIONS P.-L., ROUY E., TOURIN A., «Shape-from-Shading, viscosity solutions and edges». *Numerische Mathematik*, 64(3):323-353, mars 1993.
- [NIK91] NAYAR S. K., IKEUCHI K., KANADE T., «Shape from Interreflections». *International Journal of Computer Vision*, 6(3):173-195, août 1991.
- [Pen89] PENNA M. A., «A Shape from Shading Analysis for a Single Perspective Image of a Polyhedron». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):545-554, juin 1989.
- [PF03] PRADOS E., FAUGERAS O., «“Perspective Shape from Shading” and Viscosity Solutions». Dans *Proceedings of the 9th IEEE International Conference on Computer Vision (volume II)*, pages 826-831, Nice, France, octobre 2003.
- [Pra04] PRADOS E., «*Application of the theory of the viscosity solutions to the Shape From Shading problem*». Thèse de doctorat, Université de Nice - Sophia Antipolis, Nice, France, octobre 2004.
- [Smy90] SMYTHE D. B., «A two-pass mesh warping algorithm for object transformation and image interpolation». ILM Technical Memo 1030, Computer Graphics Department, 1990, Lucasfilm Ltd.
- [TB04] TSOI Y.-C., BROWN M. S., «Geometric and Shading Correction for Images of Printed Materials: A Unified Approach Using Boundary». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (volume I)*, pages 240-246, Washington, D.C., États-Unis, juin 2004.
- [TSY03] TANKUS A., SOCHEN N., YESHURUN Y., «Reconstruction of Medical Images by Perspective Shape-from-Shading». Dans *Proceedings of the 9th IEEE International Conference on Computer Vision (volume II)*, pages 862-869, Nice, France, octobre 2003.
- [TSY04] TANKUS A., SOCHEN N., YESHURUN Y., «A New Perspective [on] Shape-from-Shading». Dans *Proceedings of the 17th International Conference on Pattern Recognition (volume III)*, pages 778-781, Cambridge, Royaume-Uni, août 2004.
- [WM92] WADA T., UKIDA H., MATSUYAMA T., «Shape from Shading with Interreflections Under a Proximal Light Source: Distortion-Free Copying of an Unfolded Book». Dans *Proceedings of the IAPR Workshop on Machine Vision and Applications*, pages 591-594, Tokyo, Japon, decembre 1992.
- [WM92] WADA T., MATSUYAMA T., «Shape from Shading on Textured Cylindrical Surface Restoring Distorted Scanner Images of Unfolded Book Surfaces». *International Journal of Computer Vision*, 24(2):125-135, septembre 1997.
- [YKKM04] YAMASHITA A., KAWARAGO A., KANEKO T., MIURA K. T., «Shape Reconstruction and Image Restoration for Non-Flat Surfaces of Documents with a Stereo Vision System». Dans *Proceedings of the 17th International Conference on Pattern Recognition (volume I)*, pages 482-485, Cambridge, Royaume-Uni, août 2004.
- [ZT01] ZHANG Z., TAN C. L., «Recovery of Distorted Document Images from Bound Volumes». Dans *Proceedings of the International Conference on Document Analysis and Recognition*, pages 429-433, septembre 2001.
- [ZTCS99] ZHANG R., TSAI P.-S., CRYER J. E., SHAH M., «Shape from Shading: A Survey». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690-706, août 1999.
- [ZTF04] ZHANG Z., TAN C. L., FAN L., «Restoration of Curved Document Images through 3D Shape Modeling». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (volume I)*, pages 10-15, Washington, D.C. États-Unis, juin 2004.



Frédéric **Courteille**

Frédéric Courteille a obtenu son DEA en Informatique de l'Image et du Langage à l'Université Paul Sabatier de Toulouse en 2003. Son stage portait sur la prise en compte de la perspective pour le *Shape From Shading*. Il termine actuellement une thèse sur la prise en compte des défauts des images réelles pour le *Shape From Shading* à l'IRIT (Institut de Recherche en Informatique de Toulouse).



Jean-Denis **Durou**

Jean-Denis Durou a obtenu son doctorat en informatique à l'Université Paris XI d'Orsay en 1993. Il est, depuis 1994, Maître de Conférences à l'Université Paul Sabatier et membre de l'équipe TCI (Traitement et Compréhension d'Images) de l'IRIT (Institut de Recherche en Informatique de Toulouse). Ses recherches portent sur la reconstruction tridimensionnelle en vision par ordinateur, et tout particulièrement sur la technique du *Shape From Shading*.



Alain **Crouzil**

Alain Crouzil a obtenu son doctorat en informatique à l'Université Paul Sabatier de Toulouse en 1997. Il est, depuis 1998, Maître de Conférences à l'Université Paul Sabatier et membre de l'équipe TCI (Traitement et Compréhension d'Images) de l'IRIT (Institut de Recherche en Informatique de Toulouse). Ses recherches portent sur la stéréovision, l'analyse du mouvement et le *Shape From Shading*.



Pierre **Gurdjos**

Pierre Gurdjos est ingénieur d'études de l'INPT (Institut National Polytechnique de Toulouse) depuis 1986. Il est actuellement membre de l'équipe TCI (Traitement et Compréhension d'Images) de l'IRIT (Institut de Recherche en Informatique de Toulouse). Ses activités de recherche sont liées aux problèmes géométriques spécifiques à la vision par ordinateur et au *Shape From Shading*.

S

