

# Mise à plat de documents par *shape from shading*

## Setting flat of documents by *shape from shading*

Frédéric Courteille

Alain Couzil

Jean-Denis Durou

Pierre Gurdjos

IRIT, Université Paul Sabatier

118 route de Narbonne

31062 TOULOUSE Cedex 4

{courteille,crouzil,durou,gurdjos}@irit.fr

### Résumé

*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 un modèle sténopé de l'appareil photographique. Des résultats sont présentés sur images réelles.*

### Mots Clef

*Shape from shading*, documents, numérisation.

### Abstract

*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 a pinhole model of the camera. Results are shown on real images.*

### Keywords

*Shape from shading*, documents, digitization.

## 1 Introduction

La numérisation de documents connaît actuellement un essor considérable, du fait du développement d'Internet qui autorise la consultation à distance. Or, 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 non uniformément blanc) si le document est « bombé », comme c'est le cas pour un livre épais ouvert,

voire pour un parchemin déformé. Pour pallier ces deux faiblesses, un certain nombre de systèmes spécifiques ont été développés, comme par exemple les « scanners à livre ouvert », systèmes où le livre est posé sur le dos et n'est plus forcément ouvert à 180 degrés, ce qui évite d'endommager la reliure, et où 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 (deux pages par minute environ), puisque la tête du scanner doit être relevée à chaque page et, de plus, son coût est très élevé. Nous nous orientons donc préférentiellement vers une solution consistant à simuler la mise à plat des documents bombés grâce aux techniques de vision par ordinateur, en remplaçant le scanner, dont le système optique est difficile à modéliser, par un appareil photographique numérique à visée réflex. Comme nous le verrons, plusieurs techniques de vision par ordinateur peuvent alors être considérées. Dans le cadre de ce travail, nous utiliserons une technique bien connue : le *shape from shading* (SFS). L'utilisation simultanée de plusieurs techniques de vision par ordinateur permettrait sans doute d'améliorer les performances, mais nous montrons ici que le seul SFS permet d'obtenir des résultats extrêmement satisfaisants sur images réelles. D'un point de vue théorique, nous remettons en cause certains aspects du SFS. 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 est évitée. Dès lors, le SFS peut devenir une méthode de reconstruction tridimensionnelle de précision.

Dans le paragraphe 2, nous passons en revue les différentes techniques qu'il serait possible d'utiliser pour la numérisation de documents et montrons en quoi l'utilisation du SFS nous semble appropriée. Dans le paragraphe 3, nous présentons une modélisation perspective du SFS, indispensable dans le cas qui est le nôtre. Dans le paragraphe 4, nous proposons une méthodologie originale de numérisa-

tion 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 des perspectives.

## 2 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. Il est envisageable d'effectuer un traitement purement bidimensionnel de l'image, consistant à déformer l'image en tentant de corriger les défauts par rapport à un modèle *a priori* d'image du document mis à plat. Mais s'il semble relativement aisé d'imposer que les lignes de texte soient parallèles, on voit plus difficilement comment faire pour que les caractères situés près de la reliure soient correctement étirés. Et c'est sans doute le côté empirique d'un tel procédé qui explique qu'à notre connaissance, aucune technique de simulation de la mise à plat basée sur un traitement purement bidimensionnel n'ait été publiée. Intéressons-nous donc aux différentes techniques de vision par ordinateur qui ont déjà été utilisées pour simuler la mise à plat de documents.

**Utilisation de deux photographies prises sous deux angles différents.** La stéréovision est une technique 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 et permettent d'obtenir une estimation du relief de très bonne qualité. Cependant, comme on peut le voir dans les travaux de Yamashita *et al.* [21], ce gain en qualité se fait au prix d'une augmentation considérable du temps de calcul.

**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 des 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. Les résultats proposés par Cho *et al.* [2] présentent les deux faiblesses suivantes : la modélisation de la relation entre niveau de gris et pente n'est pas réaliste, car la perspective n'est pas prise en compte ; la simulation de la mise à plat utilise une méthode d'interpolation naïve.

**Projection d'une lumière structurée.** Cette technique a été mise en œuvre par Doncescu *et al.* [5]. Elle 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. 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. La faiblesse de cette méthode réside dans l'étape délicate de segmentation du motif.

**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 de la figure 7(a), un tel traitement serait inexploitable. Cette technique, récemment proposée par Cao *et al.* [1], permet d'obtenir des résultats de qualité convenable, mais repose sur une hypothèse risquée.

**Étude de la déformation des contours.** Il est également possible d'utiliser les contours d'un document pour estimer son relief, si l'on suppose qu'il est rectangulaire (un livre ancien risque de ne pas respecter cette hypothèse). Cette technique a été expérimentée par Kashimura *et al.* [9] dans le cadre du projet HUMI<sup>1</sup>. Elle a également été exploitée par Tsoi et Brown [18], mais les deux contours (supérieur et inférieur) de la page doivent être visibles sur l'image, ce qui est dommage car les zones d'un document situées près des contours contiennent généralement peu d'information.

**Analyse des niveaux de gris par *shape from shading*.** Les travaux les plus anciens sur la simulation de la mise à plat sont dus à Wada *et al.* [19] et appliquent la technique du SFS à des images provenant d'un scanner classique, alors que toutes les autres techniques utilisent des photographies numériques. Wada *et al.* prennent en compte le défaut photométrique d'images telles que celle de la figure 4(a) pour calculer le relief du document, ce qui permet ensuite de corriger les défauts géométrique et photométrique. L'idée a été récemment reprise et améliorée par Zhang *et al.* [22], qui obtiennent des résultats plutôt convaincants, mais à qui l'on peut faire les deux mêmes reproches qu'à Cho *et al.* [2] : modélisation peu réaliste ; utilisation d'une méthode d'interpolation naïve.

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 papier. 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 [3]. Notre technique permet de traiter des images ne contenant pas les contours du document. 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.

1. HUMANITIES MEDIA INTERFACE.

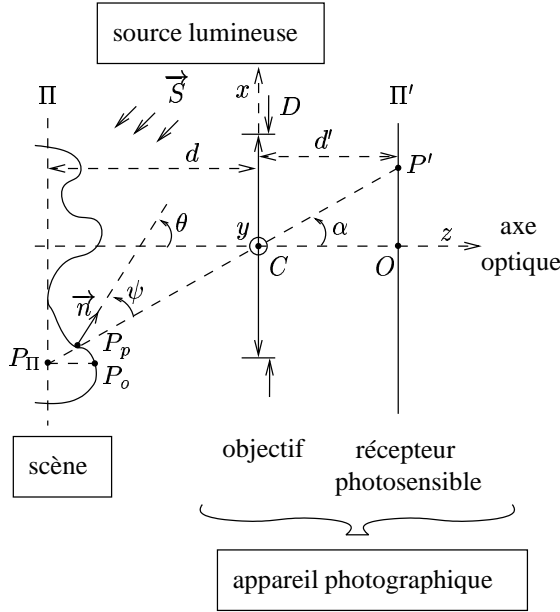


FIG. 1 – Schéma du montage photographique.

### 3 Modélisation perspective du *shape from shading*

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 du montage photographique

La figure 1 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 :

$$\frac{1}{d'} + \frac{1}{d} = \frac{1}{f'} \quad (1)$$

où  $f'$  désigne la distance focale image de l'objectif (qui est positive). Le récepteur photosensible de l'appareil photographique, situé sur un plan  $\Pi'$  orthogonal à l'axe optique, d'équation  $z = d'$ , est donc le conjugué d'un plan  $\Pi$  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 [4] :  
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  $\alpha$  (cf. figure 1) 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. [8]) :

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

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 :

$$I(P') = r E(P'). \quad (3)$$

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'émission<sup>2</sup> :

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

où  $\rho$  désigne l'albédo de la surface et  $\vec{n}(P)$  la normale unitaire sortante au point  $P$  (qui est définie sans ambiguïté sous l'hypothèse H8). En rajoutant toutes ces hypothèses, l'équation (2) peut donc être réécrite :

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

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.

2. 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.

### 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  $P_{\Pi}$  sur le plan  $\Pi$  a comme image  $P'$ , par projection centrale de centre  $C$  sur le plan  $\Pi'$  (cf. figure 1). 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, qui s'appelle « 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 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 [13, 10, 6], et aucun de ces travaux n'avait proposé de nouvelle modélisation du SFS. Trois groupes d'auteurs ont établi simultanément, et indépendamment les uns des autres, cette nouvelle modélisation [15, 17, 3]. 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  $\Pi'$  (cf. figure 1). 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  $\hat{g}(x,y) = d'/\widehat{u}_p(x,y)$ , avec :

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

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. La seule différence, en apparence, entre les deux équations (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  $\widehat{u}_p(x,y)$ . À partir d'une solution  $u_o$  de (9), on peut donc 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 (11) de  $\widehat{u}_p(x,y)$ . 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  $\Pi'$ . 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 [14], (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 [11]. Nous abordons la résolution de (12) sous un angle relativement différent.

### 3.5 Pseudo-équation de l'eikonale

Comme de nombreux auteurs [7], nous rajoutons comme inconnue 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 :

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

Cette équation, qui ressemble à (9) et à (12), n'est pas une équation aux dérivées partielles. Pour cette raison, nous proposons de l'appeler « pseudo-équation de l'eikonale ». Sa résolution est évidemment un problème mal posé, puisqu'en chaque  $(x,y)$  il y a une équation pour deux inconnues  $p(x,y)$  et  $q(x,y)$ , mais différentes stratégies<sup>3</sup> peuvent être envisagées pour le rendre bien posé [10].

Une fois  $p$  et  $q$  calculées, il reste ensuite à calculer le relief à partir de  $p$  et  $q$ , via une étape appelée « 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 (10) 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) :

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

Le déterminant  $\delta$  de ce système d'équations linéaires en  $\partial_x u_p$  et  $\partial_y u_p$  vaut  $d'(d' - xp - yq)$ . Or, en utilisant les notations de la figure 1, le produit scalaire  $\vec{n}(P_p) \cdot \vec{CP}'$  vaut : d'une part,  $(d' - xp - yq) \cos \theta$ , 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  $\|\vec{CP}'\| = d' / \cos \alpha$  et en introduisant l'angle  $\psi = (\vec{CP}', \vec{n}(P_p))$ . Par conséquent :

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

3. 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.

qui ne s'annule que si  $\psi = \pi/2$ . Sinon, c'est-à-dire si  $P'$  n'est pas sur la silhouette, alors la solution de (17) se trouve facilement :

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

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, 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 Discussion des hypothèses



FIG. 2 – Photographie du montage utilisé.

La figure 2 est une photographie du montage optique 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 (hypothèse H2) 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 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 Bouguet<sup>4</sup>. Par ailleurs, nous utilisons comme source lumineuse le flash intégré de l'appareil photographique. 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. Il existe bien des techniques permettant de prendre en compte ce phénomène, mais après avoir testé la méthode décrite dans [12] 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 le moins ouvert 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 consiste à réaliser un calibrage photométrique (cf. ci-dessous).

## 4.2 Reconstruction du relief

La figure 3 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 du même papier que celui du document, posée à plat sur la table, puis à diviser  $I_1$  par cette image de référence  $I_{\text{ref}}$  (corrigée de la distorsion), ce qui fournit une nouvelle image  $I_2$ . Cette procédure est primordiale puisque, comme nous pouvons le constater facilement, l'image  $I_{\text{ref}}$  n'est pas uniformément blanche, alors qu'elle devrait l'être si les hypothèses H1 à H9 étaient parfaitement respectées. De plus, les équations du SFS sont basées 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  $\rho_0$  quasiment égal à 1, alors que les zones encrées (là où le papier a été imprimé) 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. Pour réaliser cette tâche, nous appliquons un simple seuillage adaptatif à l'image  $I_2$ . Le résultat de ce traitement est l'image  $I_3$ , sur laquelle toutes les zones encrées ont été masquées. 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 (14) permet donc de calculer  $p(x)^2$  en fonction de la moyenne en  $y$  de  $i(x, y)$ , notée  $j(x)$  :

$$p(x)^2 = \frac{j_{\text{max}}^2}{j(x)^2} - 1, \quad (20)$$

où  $j_{\text{max}}$  est la valeur maximale de  $j(x)$ , atteinte en  $x = x_{\text{max}}$ . De plus, la surface étant convexe, le signe de  $p$  est positif si  $x \geq x_{\text{max}}$  et négatif si  $x \leq x_{\text{max}}$ . Par la suite, nous calculons  $u_p$  en utilisant la méthode d'intégration de Wu et Li [20], sous l'hypothèse de la projection perspective. 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 correspondant 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 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 [2, 9] qui utilisent des méthodes d'interpolation relativement naïves, nous avons testé trois méthodes qui ont déjà fait leurs preuves. La première de ces techniques est basée sur l'algorithme de Smythe [16]. Cette technique, qui consiste à appliquer à l'image deux interpolations unidimensionnelles successives (suivant les lignes et 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 d'`OpenGL` (interface de programmation vers le matériel graphique). 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 préférons utiliser `OpenGL`.

## 5 Résultats expérimentaux

Sur la figure 4, nous pouvons voir un exemple de résultat  $I_6^1$  fourni par notre algorithme à partir d'une image  $I_1^1$ . Nous rappelons au lecteur que seules les zones non encrées de

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

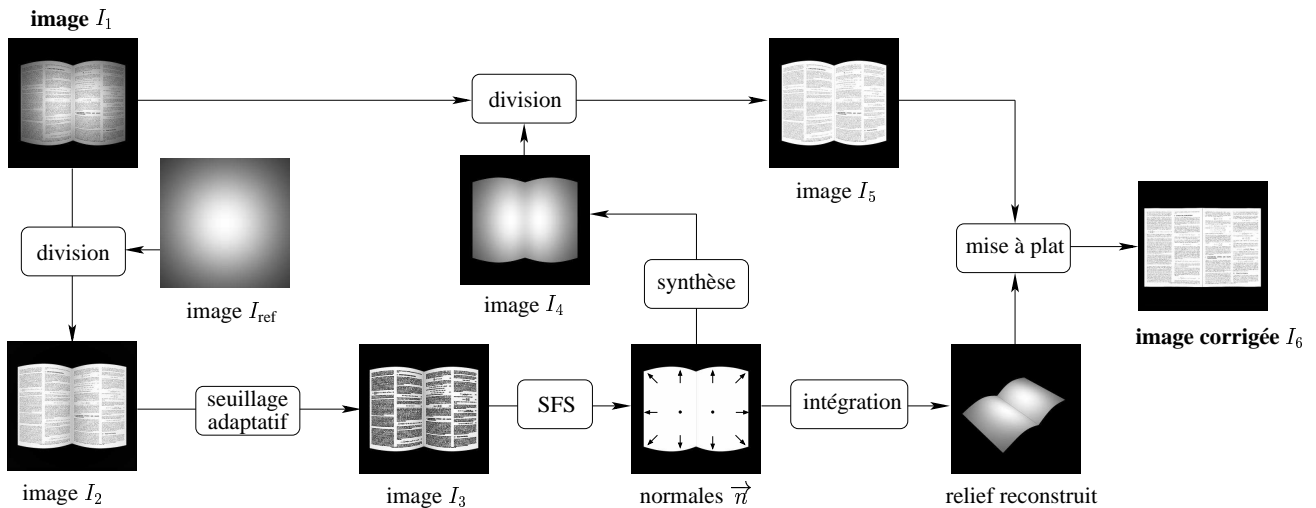


FIG. 3 – Détail de la chaîne de traitement.

$I_1^1$  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 originale. 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 moins « contrastés » 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 H4, H5, H6 et H7) est moins performante en ces points. Cependant, ce défaut est relativement léger et il existe de nombreux opérateurs de traitement d'images qui seraient capables d'uniformiser le contraste de l'image  $I_6^1$ . La figure 5 montre trois imagerie extraites des images  $I_0^1$ ,  $I_1^1$  et  $I_6^1$ , qui correspondent à une même zone de texte. À l'œil nu, il semble que la correction soit de très bonne qualité. La figure 6 représente deux occurrences  $M$  et  $N$  du même mot « capitaines » extraites de différentes positions. Le mot noté  $M_6^1$  est le résultat de la mise à plat de  $M_1^1$ , qui se situe près de la reliure (en effet, la déformation élevée de  $M_1^1$  est due à la non planarité de la page à cet endroit). Le dernier mot, noté  $N_6^1$ , a été extrait de l'image corrigée  $I_6^1$ , dans une zone où la pente du livre était relativement faible. On peut donc conclure que la correction est de bonne qualité, puisque toutes les lettres sont alignées horizontalement, qu'elles ont toutes la même taille et que, d'autre part, deux mots identiques, dont les positions dans l'image sont éloignées, deviennent semblables après mise à plat. Pour valider plus rigoureusement notre résultat, nous avons effectué des tests de reconnaissance de caractères. Pour cela, nous avons utilisé l'OCR libre `gocr` sur les trois imagerie de la figure 5. Les résultats, reportés dans le tableau 1, montrent que le pourcentage de caractères re-

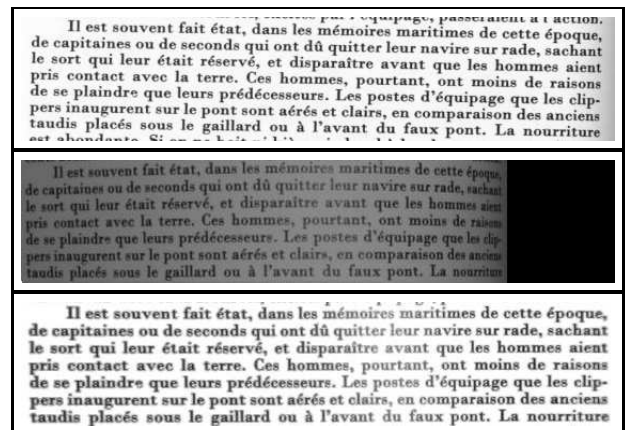


FIG. 5 – Extraits des images  $I_0^1$ ,  $I_1^1$  et  $I_6^1$ .

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

TAB. 1 – Résultats de l'OCR.

connus avec succès a été amélioré d'un facteur 30 environ (61 % au lieu de 2 %), que le pourcentage d'échecs a été réduit d'un facteur 3, et que le taux d'abstention a été, quant à lui, divisé d'un facteur 2, ce qui nous permet de conclure que la rectification effectuée permet à l'OCR de décider plus souvent et de manière plus fiable.

Les autres exemples vont illustrer 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 7 représente les images du calligramme « la colombe poignardée et le jet d'eau » de

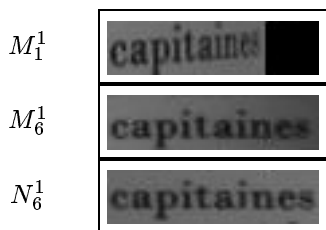


FIG. 6 – Deux occurrences du mot « capitaines » : une première extraite des images  $I_1^1$  et  $I_6^1$  ( $M_1^1$  devient  $M_6^1$  après mise à plat) ; une seconde extraite de l'image  $I_6^1$  ( $N_6^1$ ).

Guillaume Apollinaire : (a) scannée, (b) photographiée et (c) après correction. On peut remarquer que la photographie comporte les défauts géométrique et photométrique. Notamment, le mot « Jardins » (premier mot de la dernière ligne du texte) est complètement déformé et presque illisible, alors que le même mot, une fois mis à plat, est tout à fait 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 1 est encore visible. La figure 8 est un autre exemple qui, cette fois-ci, est un document en couleurs ne comportant que des dessins, et dont le papier est légèrement glacé. La mise à plat d'un tel document serait difficilement réalisable par des méthodes n'utilisant pas le SFS. On peut remarquer que, même si les hypothèses du SFS ne sont pas complètement satisfaites (la scène n'étant pas lambertienne), le résultat obtenu est tout à fait satisfaisant. Si l'on observe de plus près les deux personnages situés sur les bords gauche et droit de l'image, la mise à plat est de très bonne qualité : les trois figurines sont très proches des originaux, tant au niveau de la forme qu'au niveau de la couleur. Cependant, le résultat n'est pas parfait puisque, comme on peut le voir sur la figure 8(c), la tête du personnage central a été dégradée. Le papier n'étant pas lambertien, le flash a manifestement produit un artefact à cet endroit.

## 6 Conclusion

Nous avons proposé dans cet article une chaîne complète de traitement permettant une 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 du document. Les résultats très satisfaisants qui ont été obtenus sur images réelles prouvent le potentiel de cette méthode. De nombreuses améliorations, qui ont déjà été évoquées

dans l'article, doivent être envisagées, notamment : la généralisation de la méthode à des documents de surfaces quelconques, c'est-à-dire pour lesquels l'hypothèse  $q = 0$  n'est plus valide ; la prise en compte des contours du document pour mieux contraindre le problème du calcul du relief par SFS ; la prise en compte des réflexions secondaires, relativement importantes au voisinage de la reliure.

## 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).

## Références

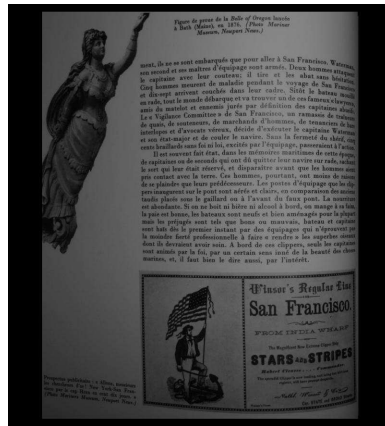
- [1] H. CAO, X. DING et C. LIU. « A Cylindrical Surface Model to Rectify the Bound Document Image ». Dans *Proceedings of the 9<sup>th</sup> IEEE International Conference on Computer Vision*, pages 228–233, Nice, France, octobre 2003.
- [2] S. I. CHO, H. SAITO et S. OZAWA. « A Divide-and-conquer Strategy in Shape from Shading Problem ». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 413–419, Porto Rico, juin 1997.
- [3] F. COURTEILLE, A. CROUZIL, J.-D. DUROU et P. GURDJOS. « Towards shape from shading under realistic photographic conditions ». Dans *Proceedings of the 17<sup>th</sup> International Conference on Pattern Recognition (volume II)*, pages 277–280, Cambridge, Royaume-Uni, août 2004.
- [4] P. DANIEL. « Peut-on extraire le relief d'une seule image ? ». Thèse de doctorat, Université Paul Sabatier, Toulouse, France, janvier 2000. (en français).
- [5] A. DONCESCU, A. BOUJU et V. QUILLET. « Former books digital processing: Image warping ». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5–9, Porto Rico, juin 1997.
- [6] J. K. HASEGAWA et C. L. TOZZI. « Shape from Shading with Perspective Projection and Camera Calibration ». *Computers and Graphics*, 20(3):351–364, mai-juin 1996.
- [7] B. K. P. HORN et M. J. BROOKS. « The Variational Approach to Shape From Shading ». *Computer Vision, Graphics, and Image Processing*, 33(2):174–208, février 1986.
- [8] B. K. P. HORN et R. W. SJOBERG. « Calculating the Reflectance Map ». *Applied Optics*, 18(11):1770–1779, juin 1979.
- [9] M. KASHIMURA, T. NAKAJIMA, N. ONDA, H. SAITO et S. OZAWA. « 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.
- [10] K. M. LEE et C.-C. J. KUO. « Shape from Shading with Perspective Projection ». *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(2):202–212, mars 1994.
- [11] P.-L. LIONS, E. ROUY et A. TOURIN. « Shape-from-Shading, viscosity solutions and edges ». *Numerische Mathematik*, 64(3):323–353, mars 1993.
- [12] S. K. NAYAR, K. IKEUCHI et T. KANADE. « Shape from Interreflections ». *International Journal of Computer Vision*, 6(3):173–195, 1991.
- [13] M. A. PENNA. « A Shape from Shading Analysis for a Single Perspective Image of a Polyhedron ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):545–554, juin 1989.
- [14] E. PRADOS. « *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.
- [15] E. PRADOS et O. FAUGERAS. « “Perspective Shape from Shading” and Viscosity Solutions ». Dans *Proceedings of the 9<sup>th</sup> IEEE International Conference on Computer Vision (volume II)*, pages 826–831, Nice, France, octobre 2003.
- [16] D. B. SMYTHE. « A two-pass mesh warping algorithm for object transformation and image interpolation ». ILM Technical Memo 1030, Computer Graphics Department, 1990. Lucasfilm Ltd.
- [17] A. TANKUS, N. SOCHEN et Y. YESHURUN. « A New Perspective [on] Shape-from-Shading ». Dans *Proceedings of the 9<sup>th</sup> IEEE International Conference on Computer Vision (volume II)*, pages 862–869, Nice, France, octobre 2003.
- [18] Y. C. TSOI et M. S. BROWN. « Geometric and Shading Correction of Imaged Print Materials: A Unified Approach Using Boundary ». Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, D.C., États-Unis, juin 2004.
- [19] T. WADA, H. UKIDA et T. MATSUYAMA. « Shape from Shading on Textured Cylindrical Surface Restoring Distorted Scanner Images of Unfolded Book Surfaces ». Dans *Proceedings of the IAPR Workshop on Machine Vision and Applications*, pages 591–594, décembre 1992.
- [20] Z. WU et L. LI. « A Line-Integration Based Method for Depth Recovery from Surface Normals ». *Computer Vision, Graphics, and Image Processing*, 43(1):53–66, juillet 1988.
- [21] A. YAMASHITA, A. KAWARAGO, T. KANEKO et K. T. MIURA. « Shape Reconstruction and Image Restoration for Non-Flat Surfaces of Documents with a Stereo Vision System ». Dans *Proceedings of the 17<sup>th</sup> International Conference on Pattern Recognition (volume I)*, pages 482–485, Cambridge, Royaume-Uni, août 2004.
- [22] Z. ZHANG, C. L. TAN et L. FAN. « 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.



(a)

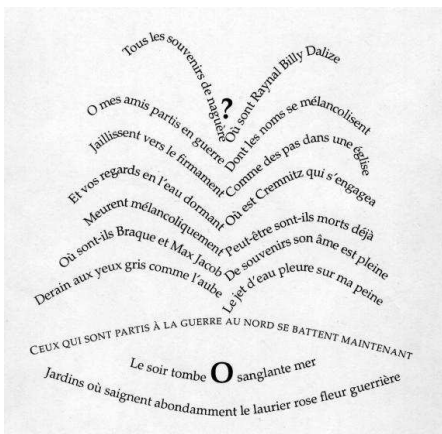


(b)

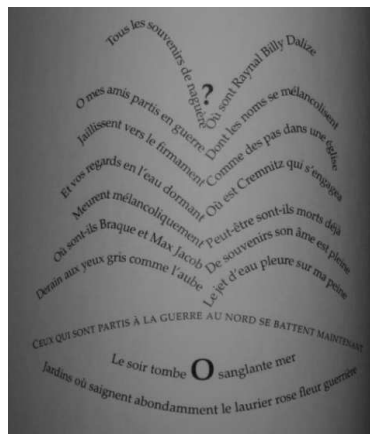


(c)

FIG. 4 – Exemple 1 : (a) image scannée  $I_0^1$ , (b) photographie  $I_1^1$  et (c) image corrigée  $I_6^1$ .



(a)



(b)



(c)

FIG. 7 – Exemple 2 : (a) image scannée  $I_0^2$ , (b) photographie  $I_1^2$  et (c) image corrigée  $I_6^2$ .



(a)



(b)



(c)

FIG. 8 – Exemple 3 : (a) image scannée  $I_0^3$ , (b) photographie  $I_1^3$  et (c) image corrigée  $I_6^3$ .