



HAL
open science

Shape-from-texture revisité : reconstruction 3D et auto-calibrage à partir d'une seule image

Jean-Denis Durou, Adrien Bartoli, Pierre Gurdjos

► **To cite this version:**

Jean-Denis Durou, Adrien Bartoli, Pierre Gurdjos. Shape-from-texture revisité : reconstruction 3D et auto-calibrage à partir d'une seule image. Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA 2010), AFRIF; AFIA, Jan 2010, Caen, France. (support électronique). hal-04584912

HAL Id: hal-04584912

<https://hal.science/hal-04584912>

Submitted on 24 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Shape-from-texture revisité : reconstruction 3D et auto-calibrage à partir d'une seule image

Shape-from-Texture Revisited : 3D-reconstruction and Self-calibration from a Single Image

Jean-Denis DUROU^{1,2}

Adrien BARTOLI³

Pierre GURDJOS²

¹CMLA, ENS Cachan, CNRS, Universud

²IRIT, Université Paul Sabatier, Toulouse

³Université d'Auvergne, Clermont-Ferrand

durou@cmla.ens-cachan.fr

adrien.bartoli@gmail.com

gurdjos@irit.fr

Résumé

Dans cet article, nous revisitons le problème du shape-from-texture. Notre première contribution est d'ordre méthodologique. Nous montrons que l'estimation de la profondeur et de la normale est équivalente à un calcul de pose. Notre deuxième contribution consiste à utiliser la redondance d'information entre la profondeur et la normale pour auto-calibrer la caméra. Nous validons ces deux contributions sur images de synthèse et sur image réelle.

Mots Clef

Reconstruction 3D, shape-from-texture, textons, calibrage de caméras.

Abstract

This paper revisits the shape-from-texture problem. Our first contribution is a methodological one : we show that the depth and normal vector estimation boils down to a plane-based pose computation. Our second contribution is to show that the redundancy of information makes it possible to compute the camera's internal parameters. We validate both these contributions on simulated and real images.

Keywords

3D-reconstruction, shape-from-texture, textons, camera calibration.

1 Introduction

Le shape-from-texture est une technique classique de reconstruction 3D qui utilise généralement une seule image [9, 4]. Récemment, plusieurs articles ont proposé de nouvelles approches du problème. Dans [6], Loh et Hartley font très peu d'hypothèses sur la scène : en particulier, le

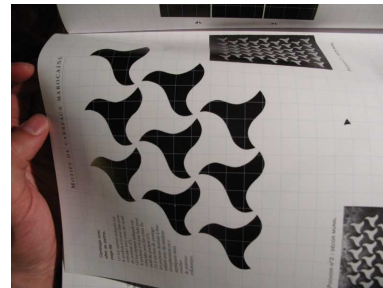


FIGURE 1 – Exemple de motif dissymétrique utilisable en shape-from-texture.

motif de base de la texture n'est pas supposé connu. Dans [5], Lobay et Forsyth proposent une technique de détection automatique de textons (ou texels), et proposent une nouvelle méthode de shape-from-texture adaptée aux caméras affines.

Contrairement à l'ensemble des travaux précédents, nous abordons le problème du shape-from-texture sous l'angle d'un calcul de pose [8, 10] que nous effectuons pour chaque texton. Si la surface \mathcal{S} de la scène est localement assimilable à une portion de plan, et qu'un motif plan de forme connue (ou patron), est imprimé en différents endroits de \mathcal{S} (voir figure 1), alors les différentes occurrences du patron se projettent dans l'image en textons, qui sont significatifs de la profondeur et de l'orientation locale de la scène. Le premier objectif de cet article consiste à savoir si cette relation constitue un problème bien posé, c'est-à-dire si chaque texton permet de retrouver sans ambiguïté les caractéristiques géométriques locales de la surface \mathcal{S} . Un patron fortement symétrique, comme par exemple un cercle, fournira probablement moins d'informations qu'un

patron dissymétrique tel que celui de la figure 1. Dans ce premier travail, nous étudions en détail le cas d'un patron ayant la forme d'un triangle rectangle isocèle. Nous montrons qu'en chaque lieu de l'image où un texton a pu être détecté, il est possible de calculer la profondeur sans ambiguïté et la normale à une ambiguïté près, si la caméra est calibrée. L'ambiguïté sur la normale est facile à lever car la profondeur et la normale sont deux informations redondantes. Nous tirons également profit de cette redondance pour auto-calibrre la caméra. Nous proposons donc une nouvelle méthode de reconstruction 3D mono-image non calibrée, que nous testons sur images de synthèse et sur image réelle.

Dans la section 2, nous montrons comment calculer la profondeur et la normale de la scène à partir de la détection d'un texton. Dans la section 3, nous montrons comment estimer les paramètres de calibrage de la caméra. La section 4 est consacrée aux tests. La section 5 conclut ce travail et ouvre quelques perspectives.

2 Reconstruction 3D par les textons

Bien que notre travail soit directement inspiré du calcul de pose d'une caméra [8, 10], nous avons privilégié dans cet article la clarté de l'exposé. C'est la raison pour laquelle nous n'utilisons pas le formalisme de la géométrie projective.

2.1 Notations

Nous notons \mathbf{n} le vecteur unitaire normal au plan contenant le patron. Comme il existe deux tels vecteurs égaux et opposés, nous choisissons celui qui pointe vers le centre optique C . Nous numérotons les sommets d'une occurrence du patron de telle sorte que l'angle $\widehat{P_1P_0P_2}$ soit égal à $\pi/2$, si le plan est orienté par \mathbf{n} (voir figure 2). Si Q_0, Q_1 et Q_2 sont les images des sommets P_0, P_1 et P_2 , alors on voit que l'angle $\widehat{Q_1Q_0Q_2}$ est nécessairement compris dans l'intervalle $[0, \pi]$.

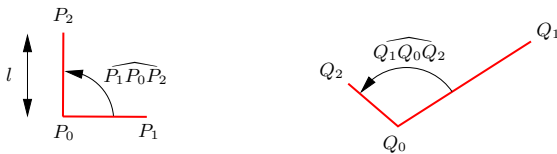


FIGURE 2 – Si les sommets du triangle sont numérotés de telle sorte que $\widehat{P_1P_0P_2} = \pi/2$, alors $\widehat{Q_1Q_0Q_2} \in [0, \pi]$.

Les coordonnées cartésiennes de \mathbf{n} relativement au repère objet $\mathcal{R}_o = Cxyz$ peuvent être exprimées en fonction de ses coordonnées sphériques θ et ϕ (voir figure 3) :

$$\mathbf{n} = [\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta]^T. \quad (1)$$

Le produit vectoriel normalisé de \mathbf{n} par $\mathbf{z} = [0, 0, 1]^T$ donne :

$$\mathbf{n}_1 = [\sin \phi, -\cos \phi, 0]^T. \quad (2)$$

En calculant le produit vectoriel de \mathbf{n} par \mathbf{n}_1 , nous obtenons :

$$\mathbf{n}_2 = [\cos \theta \cos \phi, \cos \theta \sin \phi, -\sin \theta]^T, \quad (3)$$

de telle sorte que $(\mathbf{n}_1, \mathbf{n}_2, \mathbf{n})$ forme une base orthonormée directe de \mathbb{R}^3 . Comme l'angle $\widehat{P_1P_0P_2}$ est égal à $\pi/2$, il existe un angle ψ tel que :

$$\overrightarrow{P_0P_1} = l (\cos \psi \mathbf{n}_1 + \sin \psi \mathbf{n}_2), \quad (4a)$$

$$\overrightarrow{P_0P_2} = l (-\sin \psi \mathbf{n}_1 + \cos \psi \mathbf{n}_2), \quad (4b)$$

où l désigne la taille des deux côtés égaux du triangle (voir figure 2). L'angle ψ semble arbitraire, dans la mesure où il dépend de la base choisie, mais il disparaît bien sûr dans le produit vectoriel :

$$\overrightarrow{P_0P_1} \wedge \overrightarrow{P_0P_2} = l^2 \mathbf{n}. \quad (5)$$

Nous choisissons d'exprimer les coordonnées des vecteurs $\overrightarrow{Q_0Q_1}$ et $\overrightarrow{Q_0Q_2}$ dans le repère image $\mathcal{R}_i = Oxy$:

$$\overrightarrow{Q_0Q_1} = [a, b]^T, \quad (6a)$$

$$\overrightarrow{Q_0Q_2} = [c, d]^T. \quad (6b)$$

Si l'appareil photographique est calibré, ce que nous supposons dans un premier temps, alors les coordonnées du point Q_0 dans le repère \mathcal{R}_i sont connues, puisque la position du point principal O l'est :

$$Q_0 = [x, y]^T. \quad (7)$$

Notons enfin f la distance focale, qui elle aussi est connue par hypothèse.

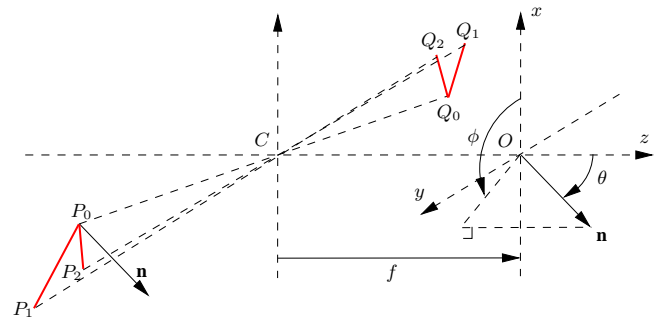


FIGURE 3 – Coordonnées sphériques (θ, ϕ) de \mathbf{n} .

2.2 Système d'équations à résoudre

Les inconnues du problème sont les coordonnées cartésiennes (X, Y, Z) du point P_0 , ainsi que les coordonnées sphériques θ et ϕ de \mathbf{n} . À cela s'ajoute l'angle ψ , qui n'a pas de signification géométrique intrinsèque. Ces six inconnues sont liées par un système de six équations qui traduisent la relation de conjugaison entre point objet et point

image. Pour la paire (P_0, Q_0) , cette relation donne les deux équations :

$$x = f \frac{X}{Z}, \quad (8a)$$

$$y = f \frac{Y}{Z}. \quad (8b)$$

Les équations de la paire (P_1, Q_1) s'obtiennent en utilisant (2), (3), (4a), (6a) et (7) :

$$x + a = f \frac{X + l(\cos \psi \sin \phi + \sin \psi \cos \theta \cos \phi)}{Z - l \sin \psi \sin \theta}, \quad (9a)$$

$$y + b = f \frac{Y + l(-\cos \psi \cos \phi + \sin \psi \cos \theta \sin \phi)}{Z - l \sin \psi \sin \theta}. \quad (9b)$$

Enfin, celles de la paire (P_2, Q_2) s'obtiennent en utilisant (2), (3), (4b), (6b) et (7) :

$$x + c = f \frac{X + l(-\sin \psi \sin \phi + \cos \psi \cos \theta \cos \phi)}{Z - l \sin \psi \sin \theta}, \quad (10a)$$

$$y + d = f \frac{Y + l(\sin \psi \cos \phi + \cos \psi \cos \theta \sin \phi)}{Z - l \sin \psi \sin \theta}. \quad (10b)$$

Les inconnues $(X, Y, Z, \theta, \phi, \psi)$ sont donc solutions du système \mathcal{S}_1 formé par les équations (8a), (8b), (9a), (9b), (10a) et (10b). La difficulté du problème vient bien sûr de ce que ces équations sont non linéaires.

Il est facile d'éliminer du système \mathcal{S}_1 les inconnues X et Y . En effet, en retranchant membre à membre (8a) de (9a), (8b) de (9b), (8a) de (10a) et (8b) de (10b), il vient :

$$\begin{aligned} a &= \frac{fl}{Z - l \sin \psi \sin \theta} (\cos \psi \sin \phi + \sin \psi \cos \theta \cos \phi \\ &\quad + \frac{x}{f} \sin \psi \sin \theta), \\ b &= \frac{fl}{Z - l \sin \psi \sin \theta} (-\cos \psi \cos \phi + \sin \psi \cos \theta \sin \phi \\ &\quad + \frac{y}{f} \sin \psi \sin \theta), \\ c &= \frac{fl}{Z - l \cos \psi \sin \theta} (-\sin \psi \sin \phi + \cos \psi \cos \theta \cos \phi \\ &\quad + \frac{x}{f} \cos \psi \sin \theta), \\ d &= \frac{fl}{Z - l \cos \psi \sin \theta} (\sin \psi \cos \phi + \cos \psi \cos \theta \sin \phi \\ &\quad + \frac{y}{f} \cos \psi \sin \theta). \end{aligned} \quad (11)$$

Nous obtenons ainsi un deuxième système \mathcal{S}_2 de quatre équations à quatre inconnues (Z, θ, ϕ, ψ) .

Un moyen de simplifier le système \mathcal{S}_2 consiste à supposer que la taille l du triangle est négligeable devant $|Z|$. Cela permet d'approcher le dénominateur des quatre équations

de \mathcal{S}_2 par Z . Il en résulte le nouveau système \mathcal{S}_3 suivant :

$$\alpha Z = \cos \psi \sin \phi + \sin \psi \cos \theta \cos \phi + \frac{x}{f} \sin \psi \sin \theta, \quad (12a)$$

$$\beta Z = -\cos \psi \cos \phi + \sin \psi \cos \theta \sin \phi + \frac{y}{f} \sin \psi \sin \theta, \quad (12b)$$

$$\gamma Z = -\sin \psi \sin \phi + \cos \psi \cos \theta \cos \phi + \frac{x}{f} \cos \psi \sin \theta, \quad (12c)$$

$$\delta Z = \sin \psi \cos \phi + \cos \psi \cos \theta \sin \phi + \frac{y}{f} \cos \psi \sin \theta, \quad (12d)$$

où α, β, γ et δ désignent les quotients respectifs de a, b, c et d par fl .

Parmi les quatre inconnues du système \mathcal{S}_3 , seules Z, θ et ϕ nous intéressent réellement. Sachant que le produit vectoriel $\overrightarrow{P_0 P_1} \wedge \overrightarrow{P_0 P_2}$ permet d'éliminer l'inconnue ψ , il pourrait en aller de même de $\overrightarrow{Q_0 Q_1} \wedge \overrightarrow{Q_0 Q_2}$. En effet, en calculant (12a)×(12d)−(12b)×(12c), nous obtenons après simplification :

$$(\alpha\delta - \beta\gamma)Z^2 = \cos \theta + \frac{\sin \theta}{f}(x \cos \phi + y \sin \phi). \quad (13)$$

En calculant (12a)²+(12c)² et (12b)²+(12d)², nous obtenons deux autres équations qui ne font plus intervenir ψ :

$$\begin{aligned} (\alpha^2 + \gamma^2)Z^2 &= \sin^2 \phi + \cos^2 \theta \cos^2 \phi + \frac{x^2}{f^2} \sin^2 \theta \\ &\quad + \frac{2x}{f} \sin \theta \cos \theta \cos \phi, \end{aligned} \quad (14)$$

$$\begin{aligned} (\beta^2 + \delta^2)Z^2 &= \cos^2 \phi + \cos^2 \theta \sin^2 \phi + \frac{y^2}{f^2} \sin^2 \theta \\ &\quad + \frac{2y}{f} \sin \theta \cos \theta \sin \phi. \end{aligned} \quad (15)$$

Il nous faut donc maintenant résoudre le système \mathcal{S}_4 formé par les équations (13), (14) et (15).

2.3 Résolution du problème dans les conditions de Gauss

Les conditions de Gauss de l'optique géométrique reviennent à supposer que $|x|$ et $|y|$ sont négligeables devant f . Sous cette hypothèse, les équations du système \mathcal{S}_4 donnent le système simplifié \mathcal{S}_5 suivant :

$$(\alpha\delta - \beta\gamma)Z^2 = \cos \theta, \quad (16a)$$

$$(\alpha^2 + \gamma^2)Z^2 = \sin^2 \phi + \cos^2 \theta \cos^2 \phi, \quad (16b)$$

$$(\beta^2 + \delta^2)Z^2 = \cos^2 \phi + \cos^2 \theta \sin^2 \phi. \quad (16c)$$

Pour résoudre \mathcal{S}_5 , nous pouvons par exemple calculer (16b)+(16c) :

$$(\alpha^2 + \beta^2 + \gamma^2 + \delta^2)Z^2 = 1 + \cos^2 \theta. \quad (17)$$

Des équations (16a) et (17), nous tirons l'équation du second degré suivante en $\cos \theta$ (nous pouvons simplifier par Z^2 car $Z \neq 0$ par hypothèse, voir figure 3) :

$$(\alpha^2 + \beta^2 + \gamma^2 + \delta^2) \cos \theta = (\alpha\delta - \beta\gamma)(1 + \cos^2 \theta). \quad (18)$$

Le cas dégénéré où $\alpha\delta - \beta\gamma = 0$ se produit lorsque les vecteurs $\overrightarrow{QQ_1}$ et $\overrightarrow{QQ_2}$ sont colinéaires, ce qui ne peut arriver que si \mathbf{n} est orthogonal à la direction d'observation, c'est-à-dire si $\theta = \pi/2$. Cette prédiction est cohérente avec l'équation (18), puisque $\cos \pi/2 = 0$, mais le cas n'est pas très pertinent, car un tel texton sera difficile à détecter dans la pratique.

Dans le cas général, nous trouvons l'expression suivante pour le discriminant de l'équation (18) :

$$\Delta = (\alpha^2 + \beta^2 - \gamma^2 - \delta^2)^2 + 4(\alpha\gamma + \beta\delta)^2, \quad (19)$$

Il est clair que cette expression est toujours positive ou nulle. Comme le produit des racines de l'équation (18) vaut 1, et que l'inconnue $\cos \theta$ est dans l'intervalle $[-1, 1]$, nous en déduisons que la solution est égale à la plus petite racine en valeur absolue, à savoir :

$$\cos \theta = \frac{\alpha^2 + \beta^2 + \gamma^2 + \delta^2 - \sqrt{(\alpha^2 + \beta^2 - \gamma^2 - \delta^2)^2 + 4(\alpha\gamma + \beta\delta)^2}}{2(\alpha\delta - \beta\gamma)}. \quad (20)$$

C'est la normale pointant vers le centre optique C qui nous intéresse, ce qui contraint l'angle θ à être dans l'intervalle $[0, \pi/2]$, donc $\cos \theta$ à être positif ou nul. Or, nous avons déjà signalé au paragraphe 2.1 que l'angle $\widehat{Q_1 Q_0 Q_2}$ est compris dans l'intervalle $[0, \pi]$. Par conséquent, le dénominateur de (20), qui est du même signe que le sinus de cet angle, est positif. On vérifie facilement qu'il en va de même du numérateur. La solution en θ est donc unique et vaut :

$$\theta^* = \arccos \left\{ \frac{\alpha^2 + \beta^2 + \gamma^2 + \delta^2 - \sqrt{(\alpha^2 + \beta^2 - \gamma^2 - \delta^2)^2 + 4(\alpha\gamma + \beta\delta)^2}}{2(\alpha\delta - \beta\gamma)} \right\}. \quad (21)$$

De (16a) et (20), nous tirons la solution en Z (le signe – vient de ce que $Z < 0$, conformément au choix du repère objet $\mathcal{R}_o = Cxyz$, voir figure 3) :

$$Z^* = - \frac{\sqrt{\alpha^2 + \beta^2 + \gamma^2 + \delta^2 - \sqrt{(\alpha^2 + \beta^2 - \gamma^2 - \delta^2)^2 + 4(\alpha\gamma + \beta\delta)^2}}}{\sqrt{2}(\alpha\delta - \beta\gamma)}. \quad (22)$$

Nous pouvons réécrire l'expression (22) uniquement à l'aide des vecteurs $\overrightarrow{Q_0 Q_1}$ et $\overrightarrow{Q_0 Q_2}$ et de fl :

$$\frac{Z^*}{fl} = - \frac{\sqrt{Q_0 Q_1^2 + Q_0 Q_2^2 - \sqrt{(Q_0 Q_1^2 - Q_0 Q_2^2)^2 + 4(Q_0 Q_1 \cdot Q_0 Q_2)^2}}}{\sqrt{2} \|Q_0 Q_1 \wedge Q_0 Q_2\|}. \quad (23)$$

Enfin, nous tirons de (16b) :

$$\sin^2 \theta \sin^2 \phi = (\alpha^2 + \gamma^2) Z^2 - \cos^2 \theta. \quad (24)$$

Cette équation en ϕ dégénère lorsque $\theta = 0$, ce qui est sans importance puisque l'angle sphérique ϕ est indéterminé dans ce cas, mais elle admet quatre solutions en ϕ

lorsque $\theta \neq 0$. Pour tenter de lever cette ambiguïté, il est judicieux de revenir au système \mathcal{S}_3 . En effet, en calculant (12a)– $\cos \theta \times$ (12d) et (12b)+ $\cos \theta \times$ (12c), nous obtenons :

$$(\alpha - \delta \cos \theta) Z = \cos \psi \sin \phi \sin^2 \theta, \quad (25a)$$

$$(\beta + \gamma \cos \theta) Z = -\cos \psi \cos \phi \sin^2 \theta. \quad (25b)$$

En divisant (25a) par (25b) membre à membre, nous obtenons l'équation suivante :

$$\tan \phi = - \frac{\alpha - \delta \cos \theta}{\beta + \gamma \cos \theta}, \quad (26)$$

qui permet de calculer ϕ à π près, y compris lorsque son dénominateur s'annule. Il semble donc y avoir une ambiguïté sur ϕ , alors que θ et Z sont calculables sans ambiguïté. Effectivement, il existe deux orientations du patron qui donnent le même texton si $\theta \neq 0$ (voir figure 4).

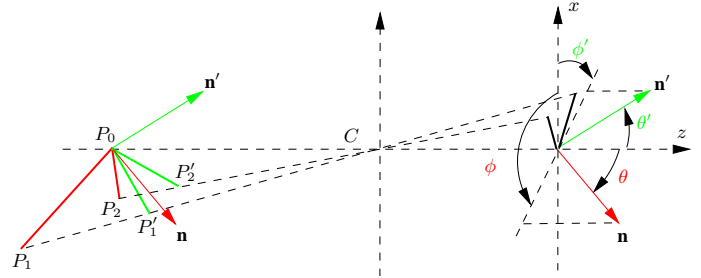


FIGURE 4 – Il existe généralement deux orientations du patron qui donnent le même texton. Les deux normales \mathbf{n} et \mathbf{n}' sont telles que $\theta' = \theta$ et $\phi' = \phi + \pi$.

2.4 Choix de la meilleure normale

La méthode de reconstruction 3D qui vient d'être décrite permet de calculer indépendamment la profondeur (sans ambiguïté) et le champ de normales (à une ambiguïté près) d'une scène. Dans [2], la situation est apparemment analogue, mais profondeur et normale sont recherchées de front, et aucune des deux ne saurait être calculée séparément de l'autre. Nous nous trouvons donc dans un cas de figure inédit, pour lequel nous disposons d'informations sur la scène fortement redondantes. Nous pouvons utiliser cette redondance pour lever l'ambiguïté sur la normale. En effet, supposons que nous connaissions les profondeurs Z_1 , Z_2 et Z_3 en trois coins voisins $Q_{0,1}$, $Q_{0,2}$ et $Q_{0,3}$ (voir figure 5). Des formules (8a) et (8b), nous déduisons la position 3D de leurs points conjugués $P_{0,1}$, $P_{0,2}$ et $P_{0,3}$. Une façon très simple de choisir entre les deux normales possibles en $P_{0,1}$ consiste donc à comparer les angles qu'elles forment avec le vecteur $\overrightarrow{P_{0,1} P_{0,2}} \wedge \overrightarrow{P_{0,1} P_{0,3}}$, qui constitue une estimation certes grossière de la normale en $P_{0,1}$, mais suffisamment discriminante en pratique pour lever l'ambiguïté. C'est cette technique qui est utilisée dans les tests.

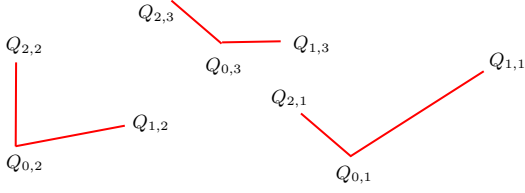


FIGURE 5 – Utilisation de trois coins voisins pour lever l’ambiguïté sur la normale.

3 Calibrage à partir de textons

3.1 Intégration d’un champ de normales

Lorsqu’on connaît la profondeur d’une scène, on peut calculer le champ de normales correspondant par dérivation. Inversement, la connaissance du champ de normales permet de calculer la profondeur par intégration (à une constante près). En réalité, ces opérations ne sont vraiment fiables que si les données sont denses. À ce titre, les techniques de reconstruction 3D photométriques que sont le shape-from-shading et la stéréophotométrie sont remarquables, car elles fournissent généralement des informations denses sur le relief de la scène inconnue, c’est-à-dire une valeur en chaque point de l’image.

Le problème de l’intégration d’un champ de normales a été largement exploré (voir [7, 3] pour de bons états de l’art). Récemment, une nouvelle méthode d’intégration inspirée de [2] a été proposée, qui prend en compte la perspective et se dispense de toute connaissance a priori sur la profondeur [1]. Les équations de l’intégration perspective d’un champ de normales $\mathbf{n} = [n_X, n_Y, n_Z]^T$ s’écrivent [1] :

$$\nabla \mathcal{Z} = [r, s]^T, \quad (27)$$

où les notations suivantes sont utilisées :

$$\mathcal{Z} = \ln |Z|, \quad (28a)$$

$$r = -\frac{n_X}{x n_X + y n_Y + f n_Z}, \quad (28b)$$

$$s = -\frac{n_Y}{x n_X + y n_Y + f n_Z}. \quad (28c)$$

Les définitions de r et s font intervenir non seulement la distance focale f (explicitement), mais également la position du point principal O (implicitement), puisque les coordonnées x et y sont exprimées dans un repère ayant O pour origine. Il est donc tentant d’utiliser la redondance d’information entre profondeur et normales pour estimer ces trois paramètres, c’est-à-dire d’utiliser les équations (27) autrement que pour les résoudre en \mathcal{Z} au sens des moindres carrés, ce qui ramène le problème à une équation de Poisson [1].

Nous introduisons donc le critère suivant, qui dépend de la distance focale f et des coordonnées (x_0, y_0) du point

principal :

$$\mathcal{C}(f, x_0, y_0) = \sum \left[\partial_x \mathcal{Z} + \frac{n_X}{(x-x_0)n_X + (y-y_0)n_Y + f n_Z} \right]^2 + \left[\partial_y \mathcal{Z} + \frac{n_Y}{(x-x_0)n_X + (y-y_0)n_Y + f n_Z} \right]^2, \quad (29)$$

où $\partial_x \mathcal{Z}$ et $\partial_y \mathcal{Z}$ désignent les coordonnées de $\nabla \mathcal{Z}$, et où la sommation est effectuée sur l’ensemble des textons. Comme la profondeur n’est calculable que là où des textons ont pu être détectés, elle ne constitue pas une information dense, mais l’estimation de $\partial_x \mathcal{Z}$ et $\partial_y \mathcal{Z}$ en un texton $Q_{0,1}$ peut quand même être effectuée grâce aux développements limités suivants (voir les notations de la figure 5) :

$$\mathcal{Z}_2 = \mathcal{Z}_1 + \nabla \mathcal{Z}(Q_{0,1}) \cdot \overrightarrow{Q_{0,1}Q_{0,2}}, \quad (30a)$$

$$\mathcal{Z}_3 = \mathcal{Z}_1 + \nabla \mathcal{Z}(Q_{0,1}) \cdot \overrightarrow{Q_{0,1}Q_{0,3}}, \quad (30b)$$

où $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3$ désignent les valeurs de \mathcal{Z} en $Q_{0,1}, Q_{0,2}$ et $Q_{0,3}$. Si les vecteurs $\overrightarrow{Q_{0,1}Q_{0,2}}$ et $\overrightarrow{Q_{0,1}Q_{0,3}}$ ne sont pas parallèles, alors les équations (30a) et (30b) forment un système de Cramer facile à résoudre en $\partial_x \mathcal{Z}(Q_{0,1})$ et $\partial_y \mathcal{Z}(Q_{0,1})$.

À titre d’exemple, la courbe d’évolution du critère \mathcal{C} en fonction de f , pour l’image réelle qui sera testée dans le paragraphe 4.2, est représentée sur le graphique de la figure 14. Il apparaît que ce critère est suffisamment discriminant pour estimer la focale. Par manque de place, nous ne montrons pas de résultat équivalent pour x_0 et y_0 , mais les rôles très symétriques des trois paramètres de calibrage dans l’expression (29) de \mathcal{C} montrent clairement que nous obtiendrions des courbes analogues.

4 Tests

4.1 Tests sur images de synthèse

Nous plaquons sur un cylindre à section circulaire une série de grilles comportant $N \times N$ mailles carrées, pour $N \in [3, 50]$. La taille des mailles est ajustée de telle sorte que les grilles aient toutes la même taille globale. Trois de ces grilles sont visualisées sur la figure 6, et les champs de normales réels leur sont superposés, sous la forme de petites flèches. Nous calculons les images de ces grilles, en utilisant comme distance focale $f = 200$, étant entendu que toutes les distances sont exprimées en nombres de pixels. Les images sont calculées pour différentes valeurs de la distance de mise au point d , qui est la distance du centre optique C au plan moyen de la scène. Les images des trois grilles de la figure 6 sont simulées à courte distance (voir figure 7) et à longue distance (voir figure 8). Il apparaît que l’effet de la perspective est prononcé à courte distance mais disparaît à longue distance. Seuls les coins des textons sont visualisés.

Nous nous plaçons tout d’abord dans la situation idéale où la distance focale réelle est connue et les données ne sont pas bruitées, afin de valider les formules de reconstruction (21), (23) et (29). Le relief reconstruit en appliquant

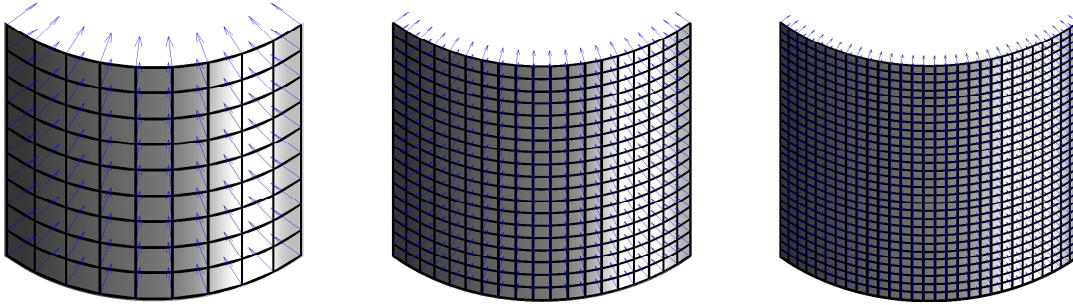


FIGURE 6 – Grilles de 10×10 , 20×20 et 30×30 mailles carrées, plaquées sur un cylindre à section circulaire.

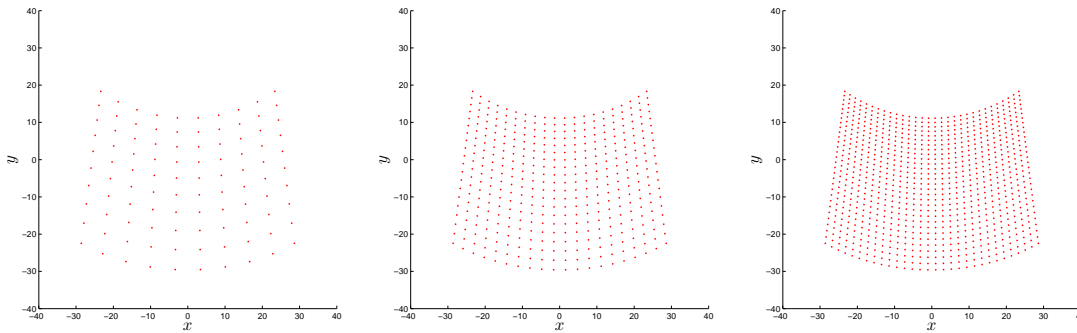


FIGURE 7 – Images des grilles de la figure 6 à courte distance ($d/f = 2, 5$) : l’effet de la perspective est prononcé.

la formule (23) à l’image du centre de la figure 7 est représenté à gauche de la figure 9. Le champ de normales estimé en utilisant les formules (21) et (29), puis en levant l’ambiguïté sur l’angle ϕ suivant la procédure décrite dans le paragraphe 2.4, est superposé au relief. Visuellement, il semble bien que le relief et le champ de normales reconstruits soient identiques à ceux de la scène d’origine (comparer les reliefs de gauche des figures 6 et 9).

Les coordonnées des coins peuvent être parfaitement connues en utilisant des données de synthèse, mais dans la pratique, la détection de coins n’est pas si simple. C’est pourquoi il nous a semblé utile de simuler d’éventuelles erreurs de localisation. Nous appliquons pour cela, à chaque coordonnée de chaque coin, un bruit additif gaussien d’écart-type σ modulable. L’effet de cette dégradation des données est illustré sur les deux autres reconstructions de la figure 9, où σ est égal à 0,02 % et à 0,04 % de la taille de l’image, soit 0,1 et 0,2 pixels pour une image de taille 512×512 .

Le relief étant reconstruit à un facteur près, il conviendrait d’ajuster ce facteur pour évaluer numériquement la précision des reconstructions, mais il est plus facile de mesurer la précision du champ de normales reconstruit, qui est indépendant du facteur d’échelle. L’écart entre le champ de normales estimé et le champ de normales réel peut être mesuré, par exemple, par l’angle entre ces vecteurs. Nous utilisons donc comme mesure de précision des reconstructions la racine carrée de la moyenne des carrés de ces

angles, exprimée en degrés et notée RMSE. L’évolution de la précision de la reconstruction en fonction du nombre de textons N , pour différentes valeurs de la distance de mise au point d , est représentée sur le graphique de la figure 10. Nous constatons que la précision croît avec le nombre de textons, ce qui est en conformité avec l’approximation du système (11) par les équations (12a)-(12d), puisque la taille du motif décroît lorsque N croît. D’autre part, la précision de la reconstruction croît avec d , ce qui est conforme aux simplifications du problème effectuées dans le paragraphe 2.3 puisque, pour une scène donnée, les conditions de Gauss sont d’autant mieux vérifiées que le rapport d/f est grand.

Enfin, nous testons la méthode d’estimation de la focale décrite dans la section 3. L’évolution de la valeur estimée de f , en fonction du nombre de textons N , est représentée sur le graphique de la figure 11, pour différentes valeurs de la distance de mise au point d . Nous constatons que la valeur estimée se stabilise très rapidement lorsque N croît. Par ailleurs, l’estimation de la focale est d’autant plus précise que la distance de mise au point d est élevée. Cela est normal puisque la précision de la profondeur influe sur celle du critère \mathcal{C} , et que la profondeur est d’autant plus précise que d est plus élevée (voir figure 10).

4.2 Tests sur image réelle

Nous avons également effectué des tests sur l’image réelle de la figure 12. Les coins ont été relevés manuellement,

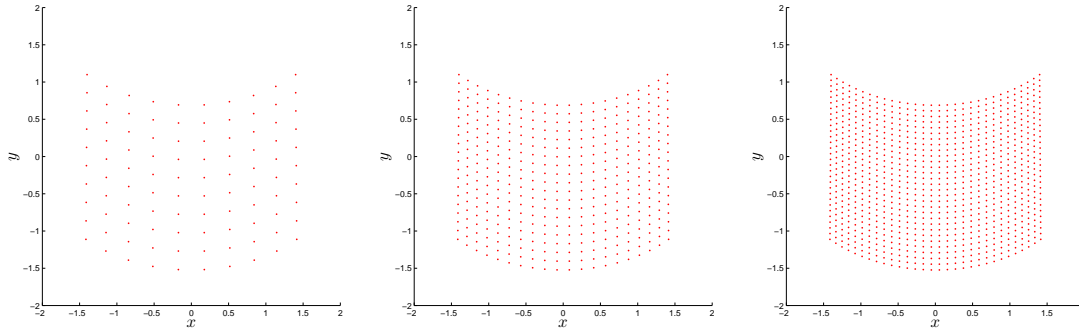


FIGURE 8 – Images des grilles de la figure 6 à longue distance ($d/f = 50$) : l’effet de la perspective disparaît.

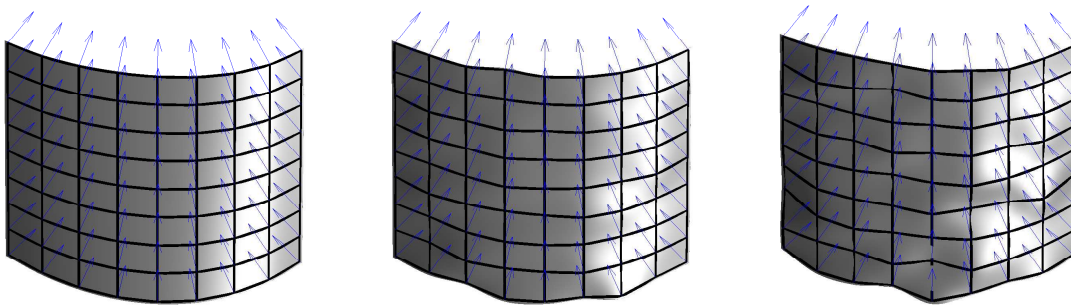


FIGURE 9 – À gauche : reliefs et champs de normales calculés à partir de l’image de gauche de la figure 7. Au centre et à droite : reliefs et champs de normales calculés à partir de la même image bruitée à l’aide d’un bruit additif gaussien d’écart-type σ égal à 0,02 % (au centre) et à 0,04 % (à droite) de la taille de l’image.

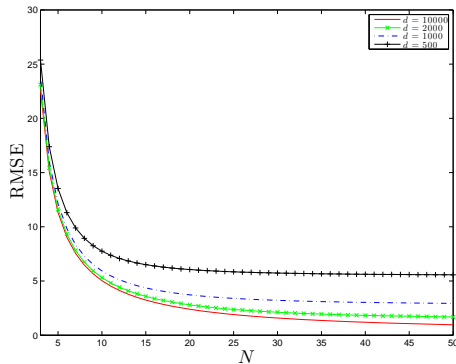


FIGURE 10 – Précision de la reconstruction 3D en fonction du nombre N de textons, pour différentes valeurs de la distance de mise au point d .

donc leurs coordonnées sont probablement entachées d’erreurs importantes. Néanmoins, la reconstruction obtenue en utilisant les paramètres de calibrage estimés par la méthode de Bouguet¹ semble cohérente avec notre intuition de ce relief.

Enfin, nous avons testé notre méthode d’auto-calibrage sur

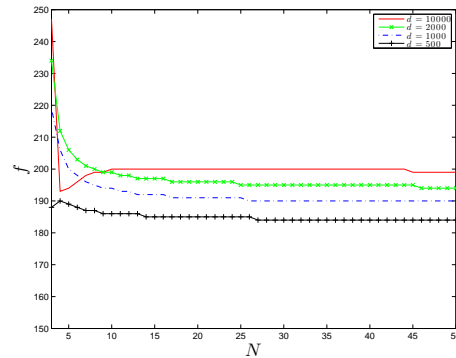


FIGURE 11 – Estimation de la focale f en fonction du nombre N de textons, pour différentes valeurs de la distance de mise au point d (la valeur effective de f est 200).

cette image. La courbe d’évolution du critère \mathcal{C} en fonction de f est représentée sur la figure 14. L’estimation de la focale qui en résulte donne $f^* = 1660$, alors que la méthode de Bouguet indique $f = 1274,99$. Cette imprécision provient probablement de l’imprécision sur les coordonnées des coins, et aussi de ce que le rapport d/f semble plutôt faible.

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

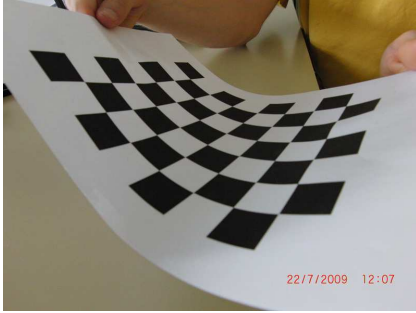


FIGURE 12 – Photographie d'une grille à mailles carrées.

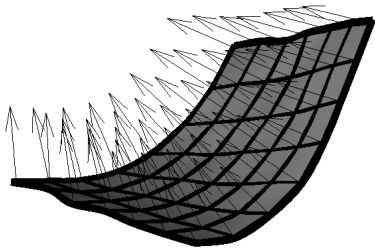


FIGURE 13 – Relief et normales obtenus à partir de la photographie de la figure 12.

5 Conclusion et perspectives

Les contributions théoriques de cet article sont au nombre de deux. D'une part, nous utilisons un parallèle entre la reconstruction 3D par shape-from-texture et le calcul de pose de caméra, dans le cas d'un motif plan de forme connue. D'autre part, nous montrons que la redondance entre profondeur et normales permet d'estimer les paramètres internes de la caméra. Ces deux contributions sont testées sur images de synthèse et sur image réelle. Les premiers résultats obtenus sont très encourageants.

Une première perspective vise à généraliser ce travail au cas d'un motif quelconque, et à déterminer les cas dégénérés provenant de motifs fortement symétriques, comme par un exemple un cercle. Une deuxième perspective s'inspire du fait que les paramètres internes de la caméra peuvent être estimés en même temps que la pose. Chaque texton peut donc fournir une estimation des paramètres internes, et il convient de s'interroger sur la façon de combiner ces différentes estimations.

Références

[1] J.-D. DUROU et F. COURTEILLE. « Integration of a Normal Field without Boundary Condition ». Dans *Proc. ICCV'07, Workshop on Photometric Analysis*

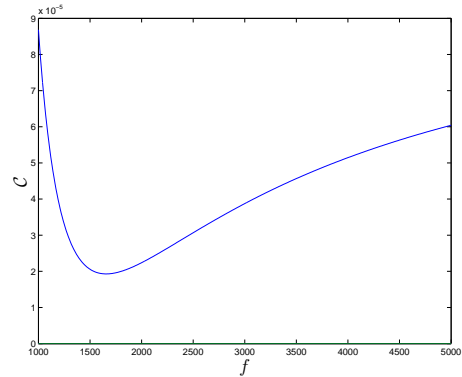


FIGURE 14 – Évolution du critère C en fonction de f , calculé sur l'image de la figure 12.

for Computer Vision, Rio de Janeiro, Brésil, octobre 2007. 5

- [2] B. K. P. HORN. « Height and Gradient from Shading ». *IJCV*, 5(1) :37–75, août 1990. 4, 5
- [3] I. HOROWITZ et N. KIRYATI. « Depth from Gradient Fields and Control Points : Bias Correction in Photometric Stereo ». *IVC*, 22(9) :681–694, août 2004. 5
- [4] K. IKEUCHI. « Shape from Regular Patterns ». *AI*, 22(1) :49–75, janvier 1984. 1
- [5] A. LOBAY et D. A. FORSYTH. « Shape from texture without boundaries ». *IJCV*, 67(1) :71–91, avril 2006. 1
- [6] A. M. LOH et R. HARTLEY. « Shape from non-homogeneous, non-stationary, anisotropic, perspective texture ». Dans *Proc. BMVC'05*, pages 69–78, Oxford, Royaume-Uni, septembre 2005. 1
- [7] K. SCHLÜNS et R. KLETTE. « Local and global integration of discrete vector fields ». Dans *Advances in Computer Vision, Advances in Computing Science*, pages 149–158, Springer, 1997. 5
- [8] P. STURM. « Algorithms for Plane-Based Pose Estimation ». Dans *Proc. CVPR'00*, pages 1010–1017, Hilton Head, États-Unis, juin 2000. 1, 2
- [9] A. P. WITKIN. « Recovering Surface Shape and Orientation from Texture ». *AI*, 17(1–3) :17–45, août 1981. 1
- [10] Z. ZHANG. « Camera Calibration By Viewing a Plane From Unknown Orientations ». Dans *Proc. ICCV'99*, pages 666–673, Corfou, Grèce, septembre 1999. 1, 2