

Capteurs et dimensionnement physique en vision

Jean Louchet

TELIN/IPI - INWE

Universiteit Gent

St. Pietersnieuwstraat 41, B-9000 Gent, Belgique

jean.louchet@gmail.com

Résumé

Le dimensionnement physique des capteurs en amont d'un système de vision artificielle, est un point souvent négligé ou sous-estimé, conduisant parfois à des efforts ou raffinements algorithmiques coûteux, voire des échecs qui auraient pu être évités si les caractéristiques précises du capteur avaient été étudiées dès la conception du système. Dans ce tutoriel, nous examinerons non les aspects électroniques (codage, cartes d'acquisition etc.) mais seulement les aspects proprement physiques et optiques de l'acquisition d'images, les limitations associées et leurs conséquences sur le choix judicieux de caméras et d'optiques en fonction de l'application.

Mots Clef

Stéreo vision, caméra, optique, géométrie, vision robotique.

Abstract

The choice and dimensioning of imaging sensors in machine vision systems is often overlooked, possibly resulting into unnecessary algorithmic refinement or costly failures which could have been avoided if the specification of the imaging sensors had been more carefully thought out from the very start of the project. In this tutorial we shall not examine the electronic aspects (coding, interfaces etc.) but focus on the physical and optical aspects of image capture, their limitations and their consequences on the choice of cameras and lenses depending on the application.

Keywords

Stereo vision, camera, optics, geometry, robot vision.

1 Introduction

Dans ce tutoriel, nous aborderons les bases d'optique et de géométrie appliquées au choix et au dimensionnement des imageurs pour les systèmes de vision artificielle. Pour cela, nous parcourrons

- les lois physiques et techniques sur les systèmes optiques focaux: loi de Descartes, angle de champ aberrations et correction optique, focales variables, mise au point, ouverture absolue et relative
- flou de mise au point, modélisation (loi de Se-tälä), hyperfocale

- diffraction, ouverture relative et limites physiques en lien avec les propriétés du capteur
- disparité, hyperfocale stéréo et dimensionnement d'un système stéréo.

Nous terminerons cet exposé par des exemples réels, et par la présentation de quelques axes de recherche possibles.

Nous adopterons les notations suivantes:

- f la distance focale
- t le tirage (égal à f lorsqu'on met au point à l'infini)
- z la distance (ou profondeur) de l'objet
- δ le diamètre de pupille, ou ouverture absolue
- ω le nombre d'ouverture, ou inverse de l'ouverture relative, ou diaphragme: $\omega = f / \delta$
- Δ le diamètre du cercle de confusion dû à l'erreur de mise au point
- λ la longueur d'onde de la lumière
- p le pas interpixels
- h l'hyperfocale,

et en stéréo:

- b la base stéréo
- h_s l'hyperfocale stéréo.

2 Conception optique

"Plus un objectif a de lentilles plus il a de défauts optiques". Il faut tordre le cou à cette contre-vérité, entendue parfois même dans le milieu du traitement d'images. Il est hors de propos ici d'entrer dans les schémas de conception des optiques, mais l'optique la plus simple, le ménisque, a pour seule vertu que son modèle au 1er ordre est une optique "parfaite" vérifiant la loi de Descartes ci-dessous. En réalité, plus on veut approcher de cette loi idéale, plus il sera nécessaire de complexifier l'optique. Ceci est spécialement vrai pour les grandes ouvertures et les grands angles de champ. Par bonheur, nous (traiteurs d'images) pouvons nous appuyer sur la technologie actuelle pour avoir des optiques quasi parfaites – c'est-à-dire dont le modèle mathématique soit simple, et entre notamment dans le cadre de la géométrie projective. Il reste à savoir quantifier leurs caractéristiques.

2.1 Loi de Descartes

Une optique idéale met en correspondance parfaite tout point en avant de l'objectif avec un point en arrière – on dit que ces points sont *caustiques*. Pour une optique donnée, la somme des inverses des profondeurs de deux points caustiques est une constante. On appelle distance focale l'inverse de cette constante, d'où la loi

$$\frac{1}{f} = \frac{1}{t} + \frac{1}{z}$$

où z désigne la profondeur de l'objet (en avant de l'objectif) et t celle de son image (en arrière). Le réglage de distance sur une caméra (aussi appelé "mise au point" ou MAP) se fait donc en positionnant le capteur sur un plan à la profondeur t , correspondant à la profondeur z des objets dont on veut que l'image soit nette. ajustant le tirage t , le plus souvent par rotation d'une bague qui agit à l'aide d'une rampe hélicoïdale¹. Si le tirage est réglé sur la valeur t , alors les objets à la profondeur z seront nets, mais tous les autres seront flous, comme nous verrons plus loin.



Figure 1 - Graduations de profondeur de champ sur un Voigtlander Nokton 1.5/50. On remarque sur l'objectif: les graduations de diaphragme (en haut: 2.8; 4; 5.6; 8 etc. en puissances de $\sqrt{2}$; les graduations de distance (au milieu: 1.5; 2; 3; 5; 10; ∞) et celles de profondeur de champ (les traits disposés symétriquement).

On règle donc le tirage selon la formule $\frac{1}{t} = \frac{z-f}{zf}$, et comme en général z (plusieurs mètres) est très grand devant f (quelques millimètres ou centimètres), on peut écrire $t - f = \frac{-f^2}{z-f} \approx \frac{-f^2}{z}$. Au repos (objectif réglé à l'infini)

¹ "Mettre au point" se dit "focus" en Anglais, d'où une confusion fréquente chez les francophones: *lorsqu'on met au point, il ne s'agit pas de régler la focale, ce qui est le propre des objectifs à focale réglable (dits "zoom"), mais de régler le tirage, la focale restant inchangée. Les objectifs à focale variable présentent un faible intérêt en vision, leurs propriétés géométriques médiocres et leur calibration problématique.*

on a $t = f$, et cette formule montre que le supplément de tirage (l'angle dont il faut tourner la bague des distances) est proportionnel à $\frac{1}{z}$. Autrement dit, la bague des distances d'un objectif sera graduée linéairement en inverses de distances

2.2 Angle de champ

Cette expression recouvre deux notions. La première est l'*angle de champ couvert* par un objectif, qui désigne l'angle au sommet du cône de vision de l'objectif. Si l'on projette l'image par exemple sur une grande feuille de papier blanc (qui prendra la place du capteur), le cône de vision se projette sur le papier sous la forme d'un cercle: l'image est claire et nette à l'intérieur de ce cercle, et elle se dégrade (flou) ou s'assombrit fortement (vignettage) à l'extérieur.

On appelle vignettage cet assombrissement des images loin du centre, dû à la conception de l'optique. Le vignettage propre à l'optique est aggravé par les capteurs électroniques, moins sensibles aux rayons d'incidence oblique. Le vignettage de l'optique est habituellement plus fort à pleine ouverture; il est notable surtout avec les optiques à très grande ouverture et certains grands angulaires. Le vignettage du capteur, généralement plus fort avec les objectifs à grand angle, ne dépend pas que de la focale de l'optique mais aussi de sa conception. Certains capteurs récemment développés (Kodak-Leica) comportent des micropismes de compensation, mais cette technologie n'est actuellement pas appliquée aux capteurs vidéo.

On appelle *champ couvert* par l'objectif, le diamètre du cercle à l'intérieur duquel le vignettage est jugé acceptable. Normalement, on choisit un capteur qui tienne entièrement dans ce cercle, donc un format dont la diagonale sera inférieure au champ couvert de l'objectif. Cette notion est intrinsèque à chaque objectif. La deuxième notion est l'*angle de champ d'une combinaison appareil-objectif*: cette fois, c'est l'angle limité par les bords du capteur. On le mesure soit par rapport à la largeur du champ, soit par rapport à sa diagonale. Si t est le tirage, et que l'on appelle θ le demi-angle de champ, l'angle de champ total 2θ est ainsi égal à $2 \operatorname{Arctan} \frac{\text{demi largeur}}{t}$, ou à $2 \operatorname{Arctan} \frac{\text{demi diagonale}}{t}$. Ainsi, avec un objectif de focale 50 mm réglé sur l'infini ($t = 50$ mm) en format 24×36 mm, en largeur $2\theta = 2 \operatorname{Arctan} \frac{18}{50} = 39.6^\circ$, et en diagonale $2\theta = 2 \operatorname{Arctan} \frac{\sqrt{18^2+12^2}}{50} = 46.8^\circ$. Lorsqu'on augmente le tirage (pour mettre au point plus près), on diminue l'angle de champ.

La *focale*, pour un format donné de capteur, détermine l'angle de champ de la caméra. Plus la focale est grande, plus l'angle de champ est faible, et le grossissement (à dis-

tance donnée) sera fort. A l'inverse, pour obtenir un grand angle de champ il faut une focale courte, ce qui est techniquement plus difficile à réaliser. Pour un objectif à perspective standard (*homographique*) l'angle de champ est nécessairement inférieur à 180 degrés; en pratique le grand-angulaire homographique (ou rectilinéaire) le plus extrême du commerce, le 12 mm de Voigtländer, atteint 122 degrés en diagonale (112 degrés en largeur).



Figure 2 - L'un des ancêtres de la prise de vue cylindrique: Horizont panoramique 24x58. Le film est enroulé sur un cylindre dont l'axe est l'axe de rotation de l'objectif, un simple 2.8/28 mm qui n'est pas optiquement un grand-angle car sa diagonale de champ est seulement 24 mm.

Pour aller au-delà de ces valeurs angulaires, il faut abandonner la géométrie homographique – ce qui signifie sacrifier la rectitude des lignes droites, ce qui n'est pas anodin, car on perd alors la puissance de l'outil homographique et de ses invariants (birapports). Trois approches sont possibles:

- La technique la plus répandue est celle du *fish-eye* (oeil de poisson), objectif spécial permettant d'atteindre un angle de champ jusqu'à 180 degrés, parfois plus
- les systèmes mécaniques tournants. Utilisés en photographie (Widelux, Noblex ou Horizont russe) ils sont maintenant transposés en numérique, faisant appel à une barrette linéaire de photodétecteurs (Panoscan). Le résultat est une perspective cylindrique. L'inconvénient majeur est le temps de balayage (de l'ordre de la seconde ou de la minute)
- les systèmes à miroir panoramique, par exem-

ple conique, parabolique ou hyperbolique.

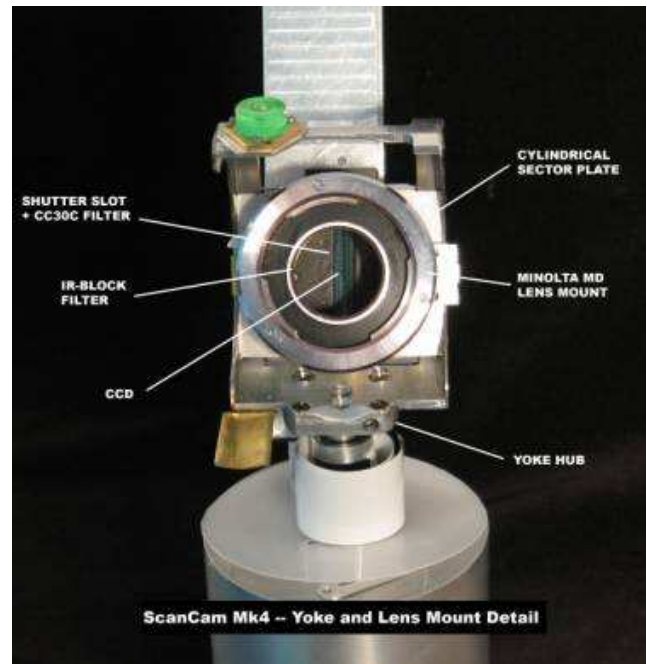


Figure 3 - Un appareil panoramique de T.K. Sharpless, construit à partir des composants d'un scanner.

Il ne faut pas confondre les systèmes panoramiques tournants à géométrie cylindrique, avec les appareils panoramiques rectilinéaires comme le Linhof Technorama, le Fuji 617 ou le numérique à balayage Seitz 617 qui utilisent un grand-angulaire homographique, sont donc rectilinéaires et ne peuvent pas atteindre des angles extrêmes.

Sur le marché actuel, outre les optiques fish-eye, on trouve les produits suivants:

- caméras à barrette tournante – ces caméras sont habituellement à très haute résolution (393 MPixels pour la Panoscan) mais à acquisition lente, ce qui les rend souvent impropres aux applications en vision:
 - Betterlight - Panoscan MK-3
 - SLiVR (Panoptic Vision)
 - Spheron VR panocam
 - Eyescan MM1
- miroirs panoramiques (la solution la mieux adaptée à la saisie en cadence vidéo):
 - Kaidan, EyeSee360. La production de Kaidan semble actuellement interrompue (septembre 2011)
 - 0-360 Panoramic, complément optique qui s'adapte à l'avant d'un objectif existant
 - Egg 360Lens



Figure 4 - EyeSee360 de Kaidan

3 Ecarts à la mise au point de Descartes: le bokeh et le théorème de Setälä

3.1 Théorème de Setälä

Il faut distinguer l'ouverture absolue ou diamètre de pupille d'une part, notée ici δ , de l'ouverture relative ou nombre d'ouverture (souvent appelé diaphragme) d'autre part, ici appelée ω . Si f est la distance focale et t le tirage, on a normalement: $\omega = \frac{t}{\delta} \approx -\frac{f}{\delta}$. Ainsi, un objectif de 50 mm d'ouverture maximum 2 (on écrit souvent "un 2/50" ou un "50 mm f:2") aura un diamètre de pupille de 25 mm. Le mécanisme du diaphragme, habituellement à iris, permet de réduire le diamètre de pupille par rapport au diamètre nominal, et donc de régler l'ouverture. Le même objectif "diaphragmé sur 8" aura alors un diamètre utile de pupille (ou ouverture absolue) égal à $50/8 = 6,25$ mm. On dit qu'on "ferme d'un diaph" lorsqu'on multiplie ω par $\sqrt{2}$, et qu'on "ouvre d'un diaph" dans le cas inverse. Aux très courtes distances, le tirage t devient sensiblement plus grand que la focale f et il convient alors de redéfinir l'ouverture relative comme $\omega = \frac{t}{\delta}$.

Le théorème de Setälä², qui décrit le flou de mise au point, est le suivant: Δ étant le diamètre de la tache de confusion sur le capteur, δ l'ouverture absolue (diamètre de pupille), f la focale, z la profondeur de l'objet et d la profondeur sur laquelle est réglée la caméra,

$$\Delta \approx \delta f \left| \frac{1}{z} - \frac{1}{d} \right|$$

qui peut s'écrire avec le nombre d'ouverture $\omega = f/\delta$:

$$\Delta \approx \frac{f^2}{\omega} \left| \frac{1}{z} - \frac{1}{d} \right|$$

Cette formule n'est valable qu'aux grandes distances. Aux courtes distances, il faut utiliser la formule exacte (non simplifiée): $\Delta = \delta \times f \times \left| \frac{d-z}{z(d-f)} \right|$. L'approximation de Setälä (utilisée dans les échelles de profondeur de champ gravées sur les objectifs) n'est valable que si $f \ll z$ et $f \ll d$, ce qui est généralement le cas. Comme toujours, le tirage t est lié à f et à d par la relation de Descartes $1/f = 1/d + 1/t$. En mise au point à l'infini, on a $t = f$.

3.2 Le Bokeh

Bokeh est un mot d'origine japonaise, qui désigne le rendu du flou de mise au point sur un objectif donné, ou la forme de la tache de flou donnée en photographiant un point lumineux sur fond noir, avec une erreur de mise au point donnée – la "réponse impulsionnelle"³. Notons que l'image floue (par défaut de mise au point) d'un objet situé à une profondeur donnée, est la convoluée de l'image nette du même objet par l'image de la pupille (généralement circulaire). Cette convolution a un effet atténuateur sur les hautes fréquences spatiales mais ces hautes fréquences spatiales restent presque toutes présentes sur l'image (cf. la transformée de Fourier d'un créneau...) ce qui autorise les opérations de défouage. Il est donc faux de dire que le flou de mise au point soit équivalent à un filtre passe-bas ou à la convolution par une gaussienne, deuxième contre-vérité, sauf dans le cas exceptionnel où les lamelles de diaphragme sont remplacées p.ex. par un verre dégradé de façon adéquate.

C'est ici qu'intervient l'expertise des meilleurs concepteurs d'objectifs. Sur l'objectif théorique, le bokeh est un disque uniforme; les objectifs de prise de vue générale de haut de gamme sont habituellement conçus pour adoucir les bords du disque et rendre visuellement plus agréables les flous. A l'opposé, la plupart des objectifs à focale variable ont un bokeh où le pourtour du cercle de confusion est rehaussé: cela donne souvent une sensation désagréable de dédoublement des contours qui sont hors de la mise au point. Le sommet du désagrément (c'est subjectif bien sûr) est atteint avec les objectifs catadioptriques ("Cassegrain") où la réponse impulsionnelle est de forme annulaire (torique). Désagréable ne veut pas forcément dire inutile (cf. la section sur la "déconvolution de bokeh").

² Vilho Suonio Setälä (Helsinki , 4 septembre 1892 – Espoo , 11 avril 1985) fut un journaliste, photographe, scientifique et écrivain finlandais. Fils du conseiller d'Etat E. N. Setälä et de la romancière Helmi Krohn, diplômé en physique, chimie et mathématiques, esprit universel, il travailla à partir de 1922 pour une maison d'édition, coordonna la création d'une encyclopédie et fut le grand promoteur de l'Espéranto en Finlande. Leicaïste de la première heure, il écrivit un manuel de photographie. On lui doit les calculs de profondeur de champ et l'invention des échelles de profondeur de champ.

³ Les photographes de portrait connaissent (ou connaissent...) bien ce problème qui a justifié la fabrication d'objectifs spéciaux pour le portrait (comme le Hektor de Leitz) où la "douceur" d'image, c'est-à-dire l'absence de hautes fréquences spatiales, qui ne peut pas être obtenue par flou de mise au point, est obtenue par construction de l'objectif, et donne une sensation visuelle tout-à-fait différente de la sensation de flou.



Figure 6 - Bokeh théorique (à gauche), bokeh doux (à droite)



Figure 7 - Bokeh dur (à gauche), bokeh torique (à droite)

Voici les effets simulés sur l'image-test de Lena (ici en 128x128), avec la même taille de disque de confusion:



bokeh théorique



bokeh doux



bokeh dur



bokeh annulaire
(catadioptrique)

L'ouverture absolue (diamètre δ de pupille) a évidemment une influence sur l'éclairement de la surface sensible: normalement chaque point de la surface sensible voit la pupille d'entrée sous l'angle solide $\pi\delta^2/t^2$, où t est le tirage (si la distance de mise au point d est grande devant f , alors t est proche de f), et avec une luminance égale à la luminance du point de la scène correspondant, aux pertes par absorption près – celles-ci sont très faibles sur les optiques modernes traitées multicouches. A un point de la scène de luminance L correspondra donc un point image recevant l'éclairement:

$$E = kL\delta^2/t^2$$

soit

$$E \approx kL\delta^2/f^2$$

dans l'hypothèse de l'“approximation des petits angles”, supposant que le point considéré soit proche de l'axe optique et que le rapport δ/t (ouverture relative) soit faible. En pratique cette formule (théoriquement affectée d'un coefficient $\cos^4\theta$ lorsqu'on s'éloigne de l'axe optique) est maintenue presque vraie sur l'ensemble de la surface sensible, même à des ouvertures relativement élevées, au prix d'une complexification de conception de l'objectif compensant la perte latérale d'éclairement encore présente sur nombre de dispositifs simples. Encore une fois, la complexité optique vise à obtenir ou à rétablir des effets mathématiquement simples.

Dans la pratique, les ouvertures relatives les plus usitées sont de l'ordre de 2 à 10, et peuvent en fait varier de 1 ou 0,5 (en particulier en infrarouge où l'on possède des matériaux à très fort indice: les limitations sont alors liées à la taille des lentilles et à la correction des aberrations), jusqu'à des valeurs très élevées (100, 200...) où le facteur limitant est la perte de résolution due à la diffraction.



Figure 8 - Optique catadioptrique dit “Cassegrain”, Lentar 5.6/250 mm. La lentille frontale est annulaire, le miroir primaire (annulaire) est au fond, le miroir secondaire est circulaire et placé au milieu de la lentille frontale, le groupe de sortie est placé au fond, au centre du miroir primaire. La tache de bokeh est homothétique de la lentille frontale. Légères et compactes, ces optiques atteignent rarement la qualité et le contraste des optiques traditionnelles, et sont limitées en angle de champ.



Figure 9 - Détail d'une image prise avec une optique catadioptrique, montrant les effets du flou de mise au point. Le bokeh Cassegrain est très caractéristique.

3.3 Profondeur de champ et hyperfocale

La notion de profondeur de champ dérive directement des formules de Setälä. On appelle profondeur de champ, la fourchette de distances entre lesquelles les objets forment une image raisonnablement nette sur le capteur.

Le tout est de savoir ce qu'on appelle "raisonnablement". Au bon vieux temps de la pellicule, on disait, assez arbitrairement, que l'image d'un point était nette si son image formait une tache de moins de 50 microns sur la pellicule (soit un vingtième de millimètre, en termes à la mode on dirait 500 points par pouce). Sur une image traditionnelle 24x36, cela signifie que l'on considère la netteté comme encore acceptable tant que la tache de flou d'un point est inférieure à 1/720ème de la largeur de l'image. En numérique, les choses sont plus claires: on peut garder cette même tolérance – qui est à peine, voire pas du tout atteinte par les appareils bon marché ou les bridge coûteux en raison de la diffraction – mais ici, on connaît parfaitement bien la distance entre deux pixels voisins, et on peut donc définir ce qu'est une image numérique "parfaitement nette". c'est à dire ne gaspillant aucun pixel⁴. Il existe donc deux manières de faire une table de profondeurs de champ: la traditionnelle, basée sur le 1/720ème de la largeur de l'image, et la "rigoureuse", basée sur la taille du pixel et souvent plus exigeante.

Reprenons la formule de Setälä aux longues distances: Δ étant le diamètre de la tache de confusion sur le capteur, δ l'ouverture absolue (diamètre de pupille), f la focale, z la profondeur de l'objet et d la profondeur sur laquelle est réglée la caméra,

$$\Delta = \delta f \left| \frac{1}{z} - \frac{1}{d} \right|$$

On peut retourner cette formule, ce qui donne pour la mise au point d , les profondeurs z_{min} et z_{max} entre lesquelles les objets seront nets:

$$z_{min} = \frac{d\delta f}{\Delta d + \delta f}; \quad z_{max} = \frac{d\delta f}{-\Delta d + \delta f}$$

L'échelle de distance sur la bague de réglage des objectifs est linéaire en $1/z$: cela découle de la loi de Descartes. C'est ainsi que la graduation 10 est située exactement à mi-distance entre 5 et ∞ , que 5 est à mi-chemin entre 2.5 et 10, etc. Regardons donc en inverses de distances, comment sont représentés $1/z_{min}$ et $1/z_{max}$.

$$\frac{1}{z_{min}} = \frac{\delta f + \Delta d}{d\delta f} = \frac{1}{d} + \frac{\Delta}{\delta f}$$

$$\frac{1}{z_{max}} = \frac{\delta f - \Delta d}{d\delta f} = \frac{1}{d} - \frac{\Delta}{\delta f}$$

Lisons de près ces formules: elles signifient que $1/d$ correspondant au réglage de la bague des distances, si l'on tra-

ce un trait de chaque côté du zéro, décalé de $\Delta/\delta f$ alors les objets dont la profondeur est comprise entre ces deux traits seront nets. Et par chance, la position de ces traits ne dépend pas de d donc on peut les tracer "en dur" sur l'objectif: un trait pour chaque valeur de δ . Ce sont les graduations de profondeur de champ inventées par Setälä! (fig. 1)

La valeur de profondeur de champ elle-même – l'épaisseur de la tranche d'espace qui sera nette sur la photo – est:

$$z_{max} - z_{min} = \frac{d^2\delta\Delta f}{\delta^2f^2 - \Delta^2d^2} = \frac{2\delta f\Delta d^2}{\delta^2f^2 - \Delta^2d^2}$$

Ceci constitue la base de calcul des tables de profondeur de champ.

On définit alors l'*hyperfocale* comme la distance minimale en dessous de laquelle les objets ne sont plus nets lorsque l'objectif est réglé sur l'infini. La valeur de l'hyperfocale est $hyp = \frac{f\delta}{\Delta} = \frac{f^2}{\Delta \times \text{ouverture}}$. Si l'on règle un objectif sur son hyperfocale (qui dépend de l'ouverture), alors la profondeur de champ ira de la moitié de l'hyperfocale jusqu'à l'infini, ce qui est pratique en photographie. En vision, où l'on s'intéresse rarement aux objets lointains, il est souvent plus pratique pour une application donnée, de retourner à la formule de Setälä pour trouver le compromis qui donnera une netteté d'image suffisante sur toutes les profondeurs concernées de la scène à analyser.

4 Diffraction et ouverture relative

La diffraction est une propriété de la propagation d'un rayon lumineux à travers une fente ou un trou, aboutissant à une dispersion de la lumière d'autant plus large que le trou est petit par rapport à la longueur d'onde.

La diffraction est régie par la règle pratique suivante: si on note θ (en radians) l'angle de dispersion par diffraction de la lumière à la sortie d'une pupille de diamètre δ , on a (ceci est une approximation d'ingénieur):

$$\theta = 1,22 \frac{\lambda}{\delta}$$

Comme longueur d'onde λ de la lumière visible on peut prendre en pratique $0,55 \mu\text{m}$, qui correspond à peu près au centre du spectre visible.

Le diamètre du lobe principal de la tache de diffraction⁵ sur l'image, située à la distance t de la pupille, est donc (en microns)

$$Diam_{diff} = \frac{1,22 \times 0,55 \times t}{\delta} \approx 0,67 \frac{f}{\delta} = 0,67 \omega$$

Dans le rouge ($\lambda = 0,68$) le diamètre de tache de diffraction devient $0,83 \omega$.

Ainsi, pour un objectif ouvrant à $\omega = 2$, le diamètre de la tache de diffraction sera égal à environ 1,5 micron. Plus le

⁴ Les distances interpixels sont, selon la caméra, entre 1 et 10 microns, et en pratique aucune caméra dont la distance interpixels est égale ou inférieure à 2 microns n'est capable de faire une image parfaitement nette, quelles que soient les conditions de prise de vue.

⁵ car en réalité la tache de diffraction possède une forme plus complexe, formée d'anneaux concentriques. Ici on ne considère que le coeur de la tache.

diaphragme sera fermé, plus le diamètre de diffraction sera grand: par exemple, environ 5 microns au diaphragme 8. Ceci est une limite physique, infranchissable.

Ici, le constructeur d'optiques doit faire un compromis. Il est coûteux de concevoir et produire un objectif à très faibles aberrations optiques, et ces aberrations décroissent à mesure que l'on ferme le diaphragme, de manière à peu près proportionnelle. Tentons donc un modèle simple du flou combiné, résultant de la diffraction et des aberrations optiques. Si l'on admet que le flou dû aux aberrations optiques est de la forme A/ω , alors le diamètre de la tache de flou résultant de la combinaison de la diffraction et des aberrations sera égal à: $0.75 \omega + \frac{A}{\omega}$. Ce diamètre sera minimum pour l'ouverture ω telle que les aberrations optiques soient de même amplitude que la diffraction.

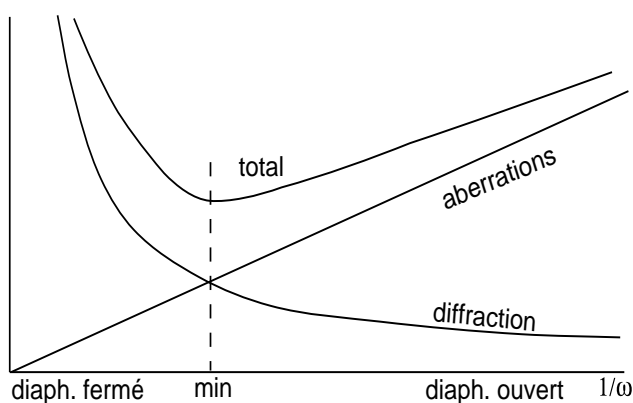


Figure 10 - Diamètre du cercle de confusion (en ordonnée) en fonction de l'ouverture.

Comme il est coûteux de réduire les aberrations optiques, la plupart des constructeurs s'arrangent pour qu'à l'ouverture la plus utilisée (souvent 1 à 2 valeurs de diaphragme en dessous de l'ouverture maximum, quelquefois plutôt vers l'ouverture maximum mais uniquement pour les optiques plus chères) les deux valeurs soient égales, et à cette valeur d'ouverture le flou sera le double de celui uniquement dû à la diffraction. On peut donc dire qu'en moyenne le diamètre de la tache de flou sera au moins égal à 1,5 fois le nombre d'ouverture maximum ω . Soyons optimistes et admettons que la qualité de construction étant exceptionnellement bonne, le diamètre de la tache de flou soit égal à ω .

Comment calculer pratiquement le nombre de mégapixels utiles d'un appareil photographique ou d'une caméra numérique? Il faut d'abord estimer la taille réelle du capteur, qui est rarement indiquée en clair sur les appareils grand public. On utilise pour cela la règle de trois:

$$\text{largeur capteur} \approx \frac{\text{focale minimum}}{\text{focale équivalente}} \times 35 \text{ mm}$$

Ensuite, le nombre de pixels par ligne d'image vaut:

$$\frac{\text{largeur capteur}}{\text{tache de flou}} = \frac{35 \times \text{focale minimum}}{\text{focale équivalente} \times \omega}$$

Une colonne ayant une longueur égale aux 3/4 d'une ligne, on a:

$$\text{Nb. de megapixels utiles} \approx \frac{3}{4} \left(\frac{35}{\text{focale équivalente}} \times \frac{\text{focale minimum}}{\omega} \right)^2$$

5 Décentrement, bascule et principe de Scheimpflug

Il y a confusion fréquente entre ces deux notions, confusion souvent entretenue par ceux qui voudraient la clarifier. La raison en est... mécanique.

La manière la plus simple de faire du décentrement est de prendre une photo avec un appareil normal (si possible muni d'un objectif grand angulaire), puis de découper la photo obtenue pour n'en garder qu'un morceau décentré par rapport à la photo initiale. Si je photographie *avec un bon grand-angle* le tympan de ma cathédrale préférée par exemple, je peux soit viser vers le haut – dans ce cas je verrai les lignes verticales converger désagréablement vers le ciel – soit viser à l'horizontale, et à ce moment les verticales seront parallèles, mais la moitié inférieure de la photo sera occupée par l'image des pavés. Qu'à cela ne tienne: un coup de ciseaux, et je jette la moitié inférieure de la photo. J'ai fait du décentrement.

Pour faire du décentrement sans post-manipulation, il suffit d'acheter dans le commerce un objectif à décentrement comme celui illustré fig. 11. Plutôt que de penser la vis de décentrement comme une vis qui décentre l'objectif par rapport au boîtier, il est conceptuellement plus clair de dire que la vis décentre le capteur par rapport à l'axe de l'objectif. Autrement dit, cet objectif forme une très grande image, mais on place le capteur sur la petite zone qui nous intéresse à l'intérieur de cette grande image. On a fait la même chose qu'avec la paire de ciseaux (mais pour plus cher).

Bref, le décentrement est un outil pour corriger les perspectives. Cet outil intéressant est en train de passer de mode, car il est devenu tellement plus facile de faire la même chose par calcul!



Figure 11 - Optique Nikon PC "perspective correction" à décentrement: ce n'est autre qu'une optique couvrant le moyen format, que l'on promène grâce à une vis micrométrique devant une surface sensible de petit format.

La bascule est bien autre chose. Il s'agit d'un principe breveté en 1904 par le capitaine Theodor Scheimpflug, de l'armée autrichienne. Il s'agit de **faire pivoter le capteur par rapport à sa position normale**. L'idée est simple: normalement, le capteur est placé dans un plan perpendiculaire à l'axe optique, et les objets nets seront tous contenus dans un plan lui aussi perpendiculaire à l'axe optique. Ceci n'est pas toujours souhaitable: on peut vouloir par exemple que les points en bas du champ de l'image soient nets à 1 mètre et ceux en haut du champ, nets à l'infini. Pour cela, il suffit de faire en sorte que le plan de la surface sensible ne soit plus orthogonal à l'axe optique. La profondeur de chaque point de la surface sensible en arrière du centre optique détermine la profondeur de son point caustique en avant de l'optique. Si la surface sensible est plane, alors sa surface caustique est plane car les deux sont reliés par une transformation homographique, et les homographies respectent la relation de co-planarité. Un élégant raisonnement homographique démontre sans calcul que l'intersection de deux plans caustiques est une droite contenue dans le "plan de l'objectif", autrement dit, dans le plan orthogonal à l'axe optique passant par le centre optique. D'où le schéma suivant qui illustre le principe de Scheimpflug:

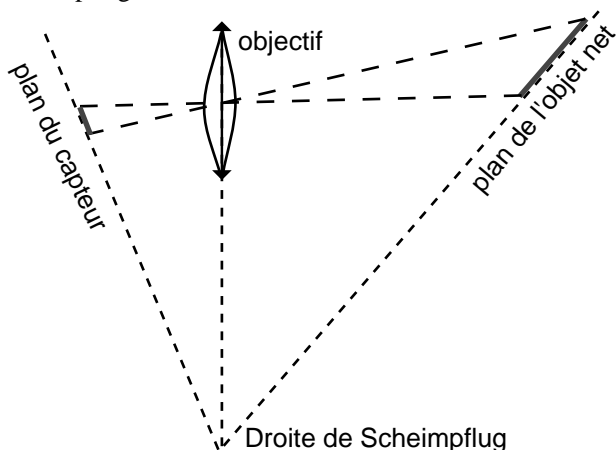


Figure 12 - Bascule: le principe de Scheimpflug

Gagner en profondeur de champ grâce au principe de Scheimpflug permet d'utiliser une ouverture optique plus grande et donc des temps d'intégration plus courts. On pourrait généraliser le principe de Scheimpflug en utilisant des surfaces sensibles gauches pour des applications particulières, ce qui à ma connaissance n'a jamais été fait, mais c'est probablement plus difficile surtout en numérique.

Il ne faut pas confondre la bascule de Scheimpflug avec le décentrement. En toute rigueur le décentrement sert à "redresser les perspectives" alors que la bascule sert à contrôler (augmenter ou parfois diminuer) la profondeur de champ. Dans la pratique une grande confusion de vocabulaire s'est installée, car *il est mécaniquement plus facile de faire basculer un objectif autour de son centre que de faire basculer la surface sensible autour du sien*: en conséquence **ce que les photographes appellent "bascule" est très souvent en réalité une combinaison de bascule et de dé-**

centrement, ce qui leur fait affirmer à faux que la bascule permet de corriger les perspectives. Ce n'est pas le cas par exemple sur le Mamiyapress S23 qui permet une **bascule pure** grâce au dos à soufflet et à ses 4 tiges guides. Ce genre de dispositif, commun sur les chambres grand format et assez peu répandu sur les appareils moyen format, est proposé par Canon et par Nikon en petit format, sous la forme d'une optique à bascule pour appareil reflex: dans ce dernier cas, le mécanisme de bascule étant incorporé dans l'objectif et non le dos du boîtier, une fois de plus on n'est pas en présence d'une bascule pure mais d'une *combinaison bascule-décentrement*. ("tilt-shift").



Figure 13 - "Mini chambre" photographique Mamiyapress avec bascule du dos à soufflet à 3 degrés de liberté: une bascule **pure sans décentrement**, contrairement aux mécanismes de bascule souvent placés dans l'objectif. Plus utile pour les applications en vision, il existe un adaptateur à bascule destiné au système Micro-Four-Thirds.

6 Dimensionnement d'un système stéréo

6.1 Disparité et hyperfocale stéréo

On suppose les caméras en disposition standard (deux caméras identiques et parallèles). L'hyperfocale stéréo est la profondeur pour laquelle la disparité vaut 1 pixel – intuitivement, le point le plus lointain que le système peut distinguer de l'infini. L'homothéticité des triangles (fig. 14) donne la relation entre profondeur z et disparité d :

$$\frac{f}{d} = \frac{z}{b}$$

soit

$$z = \frac{fb}{d}$$

D'où, dans le cas particulier où la disparité vaut l'écart interpixels p , la formule donnant l'hyperfocale stéréo:

$$h_s = \frac{f \cdot b}{p}$$

Une disparité de n pixels (soit $d = np$) va donc correspondre à une profondeur $z = h_s/n$. Les profondeurs corres-

pendant aux disparités 0, 1, 2, 3 pixels etc. seront donc: $\infty, h_s, h_s/2, h_s/3$ etc. Une grande hyperfocale stéréo donnera une meilleure précision en profondeur. A la profondeur $z = h_s/n$, la résolution en profondeur vaudra:

$$\frac{h_s}{n} - \frac{h_s}{n+1} = \frac{h_s}{n(n+1)} \approx \frac{z}{n} = \frac{z^2}{h_s}$$

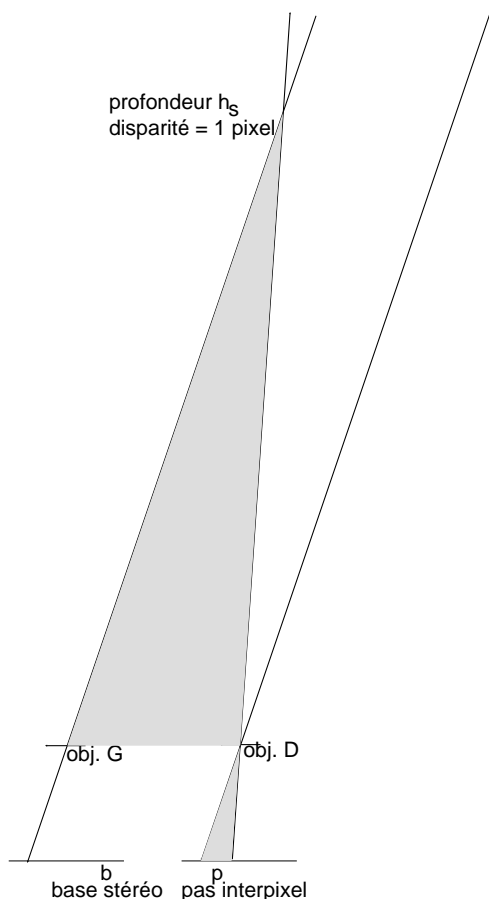


Figure 14 - Hyperfocale stéréo. L'homothéticité des triangles donne $\frac{f}{p} = \frac{h_s}{b}$

6.2 Exemples

La stéréovision s'effectue habituellement à l'aide de deux caméras. Cependant il est possible, lorsqu'on n'a pas besoin d'une très grande base, de faire appel à une optique stéréo qui permet de créer la paire stéréo sur une seule caméra, simplifiant ainsi les problèmes d'acquisition, de synchronisation et de calibration. Nous les prendrons ici comme exemples.

Les trois objectifs stéréo les plus courants sur le marché neuf sont:

- en format Micro-Four-Thirds (format 13.5 x 18 mm)
 - Panasonic focale 12.5 mm, ouverture 12, base 10 mm;
 - Loreo, focale 40 mm, ouverture 11,

base 90 mm et convergence réglée automatiquement sur la mise au point;

- en format APS-C (format Nikon DX):
 - Loreo, focale 38 mm, ouverture 11, base 90 mm.



Figure 15 - Optique stéréo (dite "3D") LOREO, en monture Micro-Four-Thirds.

En matériel plus ancien, il existe notamment:

- l'adaptateur KMZ (de fabrication russe) qui est un complément optique se vissant à l'avant d'un objectif d'angle de champ normal (par exemple 50 mm en 24x36) – ce qui présente le grand avantage d'être *indépendant du format du capteur*, pourvu que la caméra soit munie d'un objectif d'angle de champ équivalent. Ceci en fait un bon candidat pour les applications en vision. En pratique sa calibration est assez délicate, en raison des aléas du montage et du contrôle qualité, mais la qualité optique est excellente;
- l'optique cinéma stéréo Kern pour cinéma 16 mm (format 7.49x10.26 mm) ou super 16 (simple perforation, 7.41x12.52 mm, soit une diagonale de 14,6 mm), dont la qualité tant optique que mécanique est remarquable. Focale 12.5 mm, ouverture 2.8, base stéréo 64 mm, mise au point fixe. Ce modèle se prête bien aux applications robotiques; le capteur commercial le mieux adapté semble être le Nikon-1 (format CX), de diagonale 16 mm, commercialisé en octobre 2011. La base stéréo de 64 mm correspond à une valeur moyenne pour le système visuel humain.

Attention, avec un objectif stéréo on n'a pas droit à un capteur de taille inférieure au champ couvert total, car sinon les champs couverts par chacun des objectifs ne se recouvriraient pas suffisamment pour rendre le système utilisable.



Figure 16 - Optique Kern stéréo.



Figure 17 - Optique stéréo complète FED pour 24 x 36 mm.

Supposons un capteur numérique de pas interpixels 10 microns, soit 10^{-5} m. On obtient alors les valeurs suivantes d'hyperfocale stéréo:

modèle	focale f	base b	hyp. stéréo h_s
Panasonic MFT	$12.5 \cdot 10^{-3}$	10^{-2}	12.5 m
Loreo MFT	$40 \cdot 10^{-3}$	$9 \cdot 10^{-2}$	360 m
Kern	$12.5 \cdot 10^{-3}$	$64 \cdot 10^{-3}$	80 m

A la profondeur 4 m, la précision en profondeur sera 1.3 m avec le Panasonic, 20 cm avec le Kern, 45 mm avec le Loreo. En réalité, le Kern en raison de sa grande ouverture (et donc faible diffraction) pourra être utilisé avec un pas interpixels plus faible, de l'ordre de 4 microns, ce qui donne alors une hyperfocale stéréo de 200 m et une précision de 8 cm à 4 m.

Notons qu'avec les résolutions vidéo classiques, le pas interpixels réel peut être plus grand et l'hyperfocale stéréo plus faible, donc l'estimation de profondeur moins précise.

7 Conclusion: quelques idées en germe

7.1 Robocop

Faire l'économie de l'une des caméras (et de l'acquisition synchrone) en la remplaçant par une source (ou plusieurs sources) de lumière, était l'idée du projet Robocop (repérage d'obstacles par observation et calcul des ombres portées). Une brève étude de faisabilité fut l'objet du stage de J.B Hayet en 1998 à l'ENSTA. Le sol étant supposé plan, la détection des terminateurs permet de calculer la position des obstacles; un chenillard composé de 2 sources de chaque côté du robot permet de lever la plupart des ambiguïtés.

7.2 Bascule (tilt)

Il ne s'agit pas ici d'une technique de vision, mais d'un raffinement qui peut permettre d'améliorer les performances ou de réduire les contraintes d'un système de vision. Prenons l'exemple d'un robot mobile devant inspecter le sol plan devant lui. Si l'on applique le principe de Scheimpflug, le problème de profondeur de champ ne se pose plus: il suffit pour cela (1) d'orienter l'optique de façon que son champ inclue la zone à explorer (qui peut aller jusqu'à l'horizon), et (2) de placer le capteur de façon que la droite d'intersection du plan du capteur avec le plan du sol, soit contenue dans le plan passant par l'optique et orthogonal à l'axe optique. Ceci suppose une intervention mécanique dans la caméra (sauf à utiliser par exemple un adaptateur à bascule du commerce) et quelques ajustements.

7.3 Déconvolution de bokeh

Une autre voie de recherche consisterait à faire de la vision stéréo mono-objectif, en utilisant une optique à grande ouverture munie d'un diaphragme spécial. Un diaphragme torique donnerait un bokeh torique et des images similaires à une optique Cassegrain: un système qui reconnaîtrait le diamètre des cercles, par exemple à l'aide de techniques spectrales, donnerait immédiatement la profondeur des objets. Il est aussi possible d'utiliser un diaphragme à deux trous ou basé sur une forme plus complexe, permettant d'obtenir un autostéréogramme (anaglyphe), au besoin en utilisant des filtres colorés, à l'aide d'une seule caméra.

Bibliographie

- [1] Andréani (Robert), *L'objectif photographique*, publications Photo-Revue, Paris, 1970.
- [2] Larmore (Lewis) *Introduction to photographic principles*, Dover, New York, 1965.
- [3] Louchet J., *Choisir et comprendre son appareil photographique*, 2011, ISBN 978-1-4477-3997-5.
- [4] Louchet J., *Les familles d'appareils photo numériques*, éd. lulu.com, 2011.
- [5] Mark S. K. Lau, C. P. Kwong, *Analysis of Echoes in Single-Image Random-Dot Stereograms*, Journal of Mathematical Imaging and Vision, 2002: 69-79.