Traitement de l'image
TP 1 : Formation d'une image

L'évolution de la vision par ordinateur a suivi un chemin semblable à de nombreux domaines scientifiques. Le postulat de base étant que l'être humain sait parfaitement percevoir et interagir avec le monde 3D qui l'entoure, les premières caméras ont un système de construction de l'image algorithmiquement semblable à celui de l'œil humain. Le modèle mathématique d'une caméra est l'ensemble des lois géométriques décrivant la façon dont un point de l'espace à trois dimensions se projette sur le plan image à deux dimensions. Il existe différents modèles permettant de représenter géométriquement une caméra. Le modèle de projection se rapprochant le plus de celui de l'œil humain est le modèle perspectif (ou modèle sténopé ou à projection centrale).

I. Modèle interne (projection perspective)

La projection perspective est le type de projection le plus communément utilisé en vision par ordinateur. La géométrie de transformation du monde 3D vers l'image 2D d'une caméra perspective est similaire à la formation d'une image dans une chambre noire (camera obscura en latin). Une chambre noire est constituée d'une pièce plongée dans l'obscurité dont l'une des parois est percée d'un minuscule trou. Les rayons lumineux réfléchis dans toutes les directions par les objets du monde et entrant dans la chambre sont contraints à entrer par l'unique trou de celle-ci. Par conséquent, chaque point de la surface du mur à l'intérieur de la chambre faisant face au trou ne reçoit la lumière que d'un seul rayon (dans le cas idéal) projeté par un point précis d'un objet. Il se forme donc sur le mur une image inversée de la scène extérieure faisant face au trou de la chambre. Une caméra perspective suit les mêmes principes que la chambre noire, le trou de convergence des rayons lumineux est appelé centre de projection et le mur où se forme l'image est un capteur photosensible.

Un modèle est caractérisé par un certain nombre de paramètres permettant de calculer, à partir des coordonnées cartésiennes d'un point de l'espace, les coordonnées du pixel qui lui correspond sur l'image acquise par la caméra. On distingue les paramètres intrinsèques et les paramètres extrinsèques d'une caméra. Les paramètres intrinsèques d'une caméra modélisent ses caractéristiques internes et ne dépendent donc pas de sa position ni de son orientation dans l'espace. Ce sont par exemple la distance focale, la taille des pixels, la position de l'axe de visée dans l'image etc. Les paramètres extrinsèques, quant à eux, relient le système de coordonnées de la caméra au système de coordonnées de la scène. Ils représentent donc la position et l'orientation de la caméra par rapport à un repère fixe. La projection d'un point de l'espace sur une image dépend donc de sa localisation par rapport à la caméra d'une part et des caractéristiques physiques du canal d'acquisition (objectif de la caméra) d'autre part.

I.I Projection perspective

La transformation d'un point 3D (en coordonnées homogènes) ${}^{\mathbf{c}}\mathbf{X} = ({}^{c}X, {}^{c}Y, {}^{c}Z, 1)^T$ dans le repère caméra ${}^{c}\mathcal{R}$ vers un point 2D $\mathbf{x} = (x, y, 1)^T$ dans le repère capteur photosensible

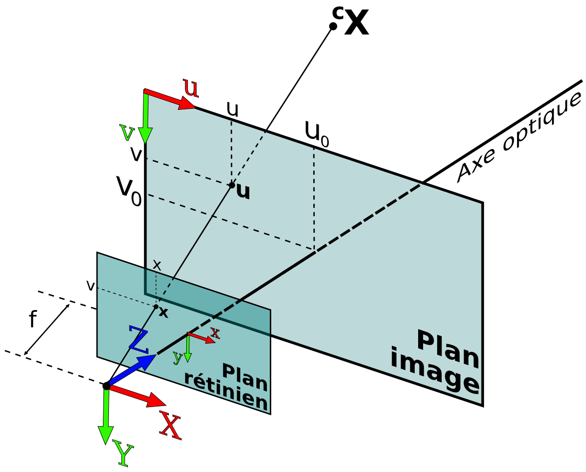

Figure 1: Schéma de la projection perspective d'un point du monde 3D dans l'image 2D

${}^{r}\mathcal{R}$ suit la projection perspective $pr_p()$ suivante :

$$\mathbf{x} = pr_p({}^{\mathbf{c}}\mathbf{X}) \quad avec \quad \begin{cases} x = \dfrac{{}^{c}X}{{}^{c}Z} \\ y = \dfrac{{}^{c}Y}{{}^{c}Z} \end{cases} . \tag{1}$$

Le point $\mathbf{x}$ appartient au plan rétinien exprimé dans le repère capteur ${}^{r}\mathcal{R}$ en unité métrique.

I.II Transformation capteur/image

Une transformation supplémentaire est nécessaire pour obtenir une image numérique d'une scène telle que nous la connaissons ordinairement (sous la forme de pixels). En effet, une image digitale est le fruit de la quantification spatiale de l'image formée sur le plan rétinien par un échantillonnage régulier (voir section II). Cet échantillonnage est composé de $M \times N$ pixels (picture elements) avec M la largeur de l'image digitale et N sa hauteur. La transformation permettant de passer du repère capteur ${}^{r}\mathcal{R}$ au repère image ${}^{i}\mathcal{R}$ pour un point $\mathbf{x}$ est :

$$\mathbf{u} = \mathbf{K}\mathbf{x} \tag{2}$$

où $\mathbf{u} = (u, v)$ est la position du point $\mathbf{x}$ dans le plan image en coordonnées pixeliques. La matrice de passage $\mathbf{K}$ est la matrice des paramètres intrinsèques de la caméra :

$$\mathbf{K} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

dans laquelle u_0 et v_0 désignent les coordonnées du point principal dans l'image. Ce point principal correspond à l'intersection de l'axe optique avec le plan image. α_u et α_v contiennent k_x et k_y le nombre de pixels par unité de longueur suivant respectivement les directions x et y du capteur, tout en prenant en compte la focale f de la caméra ($\alpha_u = fk_u$ et $\alpha_v = fk_v$). La focale étant la distance entre le centre optique et le capteur de la caméra. Les paramètres intrinsèques d'une caméra varient d'une caméra à l'autre. L'estimation de la valeur de ces paramètres pour une caméra est réalisée par calibration. Une caméra est dite calibrée lorsque ses paramètres intrinsèques sont connus. La symétrie sphérique des lentilles des caméras peut être à l'origine de distorsions radiales et tangentielles. La première est due à l'asymétrie des lentilles, la seconde à un mauvais alignement des lentilles. Les paramètres de distorsion peuvent également être estimés et utilisés afin d'enrichir le modèle de projection.

Finalement, le modèle de projection perspectif complet permettant de projeter un point 3D exprimé dans le repère caméra ${}^c\mathcal{R}$ au plan image pixelique dans le repère ${}^i\mathcal{R}$ peut s'écrire :

$$\mathbf{u} = \mathbf{K}pr_p(\mathbf{{}^cX}) \tag{4}$$

I.III Modèle externe

Dans la modélisation interne présenté précédemment, nous avons représenté les points 3D comme appartenant au repère caméra ${}^c\mathcal{R}$. Cependant, les points 3D d'une scène sont généralement représentés dans un repère orthonormé qui est propre à la scène elle-même et que l'on notera ${}^s\mathcal{R}$. La figure 2 illustre une scène attachée à son repère ${}^s\mathcal{R}$ dans laquelle évolue une caméra attachée à son repère ${}^c\mathcal{R}$.

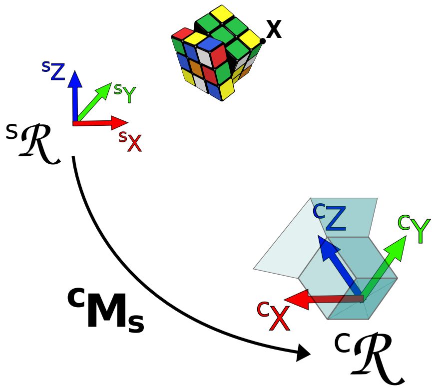

Figure 2: Représentation d'une scène 3D : relation géométrique entre la scène et une caméra

L'état relatif entre la scène et la caméra dans l'espace peut être décrit à partir de transformations 3D rigides entre leurs repères orthonormés respectifs. Nous notons ${}^c\mathbf{R}_s$ l'orientation relative entre le repère de la scène et le repère de la caméra. Cette transformation appartient au groupe spécial des matrices orthogonales $\mathbf{SO}(3)$. Le vecteur de translation ${}^c\mathbf{t}_s$ représente la position de la caméra dans le repère scène exprimée dans $\mathbb{R}^3$.

${}^c\mathbf{R}_s$ est une matrice orthogonale de taille (3×3) dont les colonnes sont les coordonnées des vecteurs $({}^s\vec{X}, {}^s\vec{Y}, {}^s\vec{Z})$ de ${}^s\mathcal{R}$ exprimées dans le repère ${}^c\mathcal{R}$. Les rotations de la matrice ${}^c\mathbf{R}_s$ peuvent être représentées sous différents formalismes. Dans ce TP, nous utiliserons les angles d'Euler. Les angles d'Euler décrivent l'orientation de la scène par rapport à la caméra (ou inversement) sous la forme de trois angles de rotation autour des axes ${}^sX$, ${}^sY$ et ${}^sZ$ (ou ${}^cX$, ${}^cY$ et ${}^cZ$). Dans la suite, nous noterons α la rotation autour de ${}^c\vec{X}$, β la rotation autour de ${}^c\vec{Y}$ et γ la rotation autour de ${}^c\vec{Z}$. Sous forme matricielle :

$${}^c\mathbf{R}_s({}^cX, \alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha) & -sin(\alpha) \\ 0 & sin(\alpha) & cos(\alpha) \end{bmatrix} \tag{5}$$

$$ {}^c\mathbf{R}_s({}^cY,\beta) = \begin{bmatrix} cos(\beta) & 0 & sin(\beta) \\ 0 & 1 & 0 \\ -sin(\beta) & 0 & cos(\beta) \end{bmatrix} \tag{6} $$

$$ {}^c\mathbf{R}_s({}^cZ,\gamma) = \begin{bmatrix} cos(\gamma) & -sin(\gamma) & 0 \\ sin(\gamma) & cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7} $$

Ces 3 rotations usuelles s'effectuent les unes après les autres, l'ordre dans lequel elles sont réalisées est donc important. Dans ce TP nous utiliserons ce formalisme :

$$ {}^c\mathbf{R}_s = {}^c\mathbf{R}_s({}^cZ,\gamma)\ {}^c\mathbf{R}_s({}^cY,\beta)\ {}^c\mathbf{R}_s({}^cX,\alpha) \tag{8} $$

Ainsi, un point 3D ${}^s\mathbf{X} = ({}^sX, {}^sY, {}^sZ, 1)^T$ appartenant au repère de la scène s'exprime dans le repère de la caméra par :

$$ {}^c\mathbf{X} = \begin{pmatrix} {}^cX \\ {}^cY \\ {}^cZ \\ 1 \end{pmatrix} = {}^c\mathbf{M}_s\ {}^s\mathbf{X} \tag{9} $$

où

$$ {}^c\mathbf{M}_s = \begin{pmatrix} {}^c\mathbf{R}_{s(3\times3)} & {}^c\mathbf{t}_{s(1\times3)} \\ \mathbf{0}_{(3\times1)} & 1 \end{pmatrix}_{(4\times4)} \tag{10} $$

Inversement, le passage d'un point 3D appartenant au repère caméra vers le repère de la scène se calcule à l'aide de ${}^c\mathbf{M}_s{}^{-1}$.

I.IV Modèle complet

En combinant la modélisation de la caméra et la modélisation de son positionnement dans l'espace, il est désormais possible de modéliser le processus complet de génération d'une image 2D d'une scène 3D. Par exemple, les coordonnées images $\mathbf{u}$ d'un point 3D ${}^s\mathbf{X}$ appartenant à la scène se trouvent par :

$$ \mathbf{u} = \mathbf{K}\, pr_p({}^c\mathbf{M}_s\ {}^s\mathbf{X}). \tag{11} $$

I.V Travail à réaliser

Écrire un programme permettant de simuler une caméra dont voici les paramètres intrinsèques : $\alpha_u = 700$, $\alpha_v = 700$, $u_0 = 400$ et $v_0 = 300$.
Plus précisément :

- Choisir N points 3D exprimés dans le repère attaché à la scène (${}^s\mathcal{R}$)
- Choisir les paramètres extrinsèques (position et orientation) de la caméra dans le repère attaché à la scène (${}^s\mathcal{R}$)
- Déterminer R (resp. T) matrice de rotation (resp. vecteur de translation) du repère attaché à la scène (${}^s\mathcal{R}$) au repère attaché à la caméra (${}^c\mathcal{R}$)

- Implanter le modèle de projection permettant de déterminer les coordonnées pixels de la projection dans le plan image des N points 3D à partir des paramètres intrinsèques et extrinsèques de la caméra.
- Générer une image de taille 800×600 contenant la projection des N points 3D

II. L'image numérique

Comme mentionné précédemment une image numérique est fruit de la quantification spatiale de l'image formée sur le plan rétinien par un échantillonnage régulier. En effet, vue du capteur photosensible de la caméra, une image optique est un signal continu généralement représenté par une fonction bidimensionnelle $f(i, j)$ représentant l'image en chaque point (i, j) de son espace (ex: l'intensité). L'image discrète, c'est-à-dire la représentation numérique du signal continu, est obtenue par l'échantillonnage des coordonnées spatiales de ce signal dans les deux dimensions de l'image, et par la quantification de la fonction continue, la transformation de la mesure de l'énergie lumineuse continue vers une valeur numérique. On peut alors représenter le signal par une matrice 2D pour une image monochrome et par une matrice 3D pour une image couleur. Chaque élément de la matrice est un pixel (qui vient de l'expression anglo-saxonne 'picture element'). La fidélité de la représentation fournie par l'image numérique par rapport à l'image modèle analogique dépend de nombreux paramètres très liés entre eux : la résolution, la définition, l'échantillonnage, la qualité de stockage etc.

II.I Quantification

L'opération de quantification consiste à attribuer un nombre binaire à toute valeur prélevée au signal lors de l'échantillonnage. C'est le CAN (convertisseur analogique numérique) qui réalise cette opération. On peut coder une image sur autant de niveaux de gris qu'on souhaite. Dans la majeure partie des cas, une image monochrome est codée sur 8 bits qui fournit $2^8 = 256$ niveaux (valeurs comprise entre 0 à 255 en format non signé). C'est ce qu'on appelle la dynamique de l'image.

II.I.1 Travail à réaliser

- Ouvrez l'image "lena_gray.tif".
- Changez le pas de quantification de l'image pour obtenir 128, 64, 32, 16, 8, 4 et 2 niveaux de gris.
- Sur une même figure (à 8 fenêtres), affichez l'image originale et les images en niveau de gris sur 128, 64, 32, 16, 8, 4 et 2 niveaux afin de visualiser la dégradation de l'image en cas de sous-quantification trop importante.
- Affichez l'histogramme de l'image en niveaux de gris pour chaque quantification.

II.II Espaces chromatiques

Une image couleur est définie, non plus par une fonction 2D mais 3D. Pour un système RVB, la dimension supplémentaire est égale à 3 car l'image est codée suivant 3 composantes : le rouge, le vert et le bleu.

Si on code chacune des 3 composantes couleurs d'une image, on utilise $3 \times 8 = 24$ bits ce qui donne $2^24 = 16777216$ couleurs. Une image couleur a 3 composantes qui dépendent de l'espace chromatique utilisé. Le plus commun est l'espace RVB basé sur la synthèse additive qui est composé d'une composante rouge, d'une composante verte et d'une composante bleue. Néanmoins, il existe de nombreux autres espaces de couleurs :

- les espaces YUV et YCbCr : le premier terme est l'information de luminance Y et les deux autres sont dits termes de chrominance,
- l'espace HSV signifie Hue, Saturation, Value : la teinte correspond au type de couleur (rouge, jaune, violet, etc.), la saturation correspond à l'intensité de la couleur et la valeur correspond à la brillance de la couleur,
- l'espace Lab : L représente la luminance, a représente la gamme de l'axe rouge/vert et b la gamme de l'axe jaune/bleu.

II.II.1 Travail à réaliser

- Ouvrez puis affichez l'image "lena_color.tif".
- Transformez puis affichez cette image en niveau de gris.
- Affichez sur trois figures différentes à 4 fenêtres l'image source ainsi que ses 3 composantes dans les systèmes RGB, YCbCr et HSV.

II.III Échantillonnage

Nous avons étudié la question de la discrétisation de la luminosité. Intéressons-nous à présent à la discrétisation de l'espace spatial appelé échantillonnage.

II.III.1 Travail à réaliser

- Nous allons dans un premier temps sous-échantillonner une image en ne prenant qu'un point sur deux dans les deux directions. Ainsi, le nombre de lignes et de colonnes vont être divisés par deux et par conséquent la taille de l'image sera divisée par quatre. Pour cela, ouvrez l'image "lena gray.tif". Créez une nouvelle image sous-échantillonnée.
- Refaire la même chose avec un facteur 4, ...
- Ensuite, utilisez une fonction d'interpolation pour sur-échantillonner par interpolation l'image précédemment sous-échantillonnée.