PREUVE QUE LES TRANSFORMATIONS ENTRE DEUX IMAGES SONT HOMOGRAPHIQUES

HYPOTHESE FONDAMENTALE : Dans ce raisonnement, on considère **deux photos** d'un même paysage, qui ont été prises au même endroit (c'est-à-dire que l'appareil photo était rigoureusement au même point). La seule différence est l'orientation de l'appareil : la 2e fois que l'appareil a été utilisé, la photo a été prise avec l'appareil dirigé un peu plus vers le ciel (ou un peu plus vers la terre) et un peu plus vers la gauche (ou au contraire un peu plus vers la droite) que la première fois.

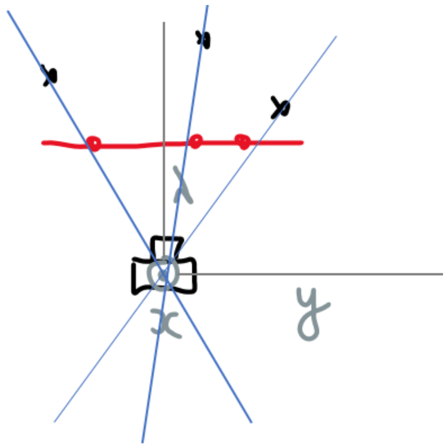
OBJECTIF : Trouver la relation entre la première image et la deuxième image. Autrement dit, si on considère un point $u \in \mathbb{R}^3$ et u_1 sa position sur l'image 1 et u_2 sa position sur l'image 2, on cherche la fonction mathématique f telle que $u_2 = f(u_1)$. En pratique, on cherche une fonction f_a à paramètres (avec une dizaines de paramètres $a_1, a_2, \ldots$) de sorte que lorsque si on connait une dizaine de couples (u_1^i, u_2^i) on puisse expliciter ces paramètres a_k pour ensuite « prédire » u_2 à partir de f_a et u_1 pour tous les autres points de l'image. On verra qu'il s'agit d'une transformation homographique, ce qui correspond à une transformation linéaire (matrice 3x3) dans l'espace homogène.

REMARQUE : En pratique, les points (u_1^i, u_2^i) utilisés pour calculer les paramètres seront des points remarquables : des bords, des coins, etc. Bref, c'est comme quand on fait un puzzle : les pièces de puzzle avec une seule couleur ne sont pas très utiles et en général placées à la fin, alors que les pièces représentant des coins ou des pointes (donc riches en couleurs différentes) sont très intéressantes. Mais on n'y est pas : pour l'instant, on cherche à trouver la forme générale de f_a et de ses paramètres.

RESOLUTION DU PROBLEME : On munit la photographie n°1 d'un repère orthogonal : l'origine au centre de la photo, un vecteur e_x vertical vers le haut et un vecteur e_y horizontal vers la droite. De sorte que chaque point de la photo est maintenant repéré par des coordonnées (x, y). Imaginons qu'on se tienne à l'emplacement exact où la photo a été prise. En plaçant la photo 1 à une certaine distance (dans la direction de $e_x \wedge e_y$) l'effet de perspective fait que le paysage est parfaitement remplacé par la photo. Dans $\mathbb{R}^3$, on définit donc un vecteur e_λ allant de la position de l'appareil au centre de la photo. Les points de la photo sont définis en coordonnées 3D par $(x, y, 1)$. Les lois de la perspective font que tout point $(\lambda x, \lambda y, \lambda)$ (avec $\lambda > 0$) est confondu avec le point $(x, y, 1)$ du point de vue de l'observateur placé en $(0,0,0)$. On dit que les coordonnées 3D sont homogènes. Par contraste, les coordonnées (x, y) sont dites inhomogènes. Le fait que deux points alignés avec O sont représentés par le même point sur l'image peut être vu comme une relation d'équivalence dont les classes d'équivalence sont les droites de $\mathbb{R}^3$ passant par O. Ces droites de perspective sont illustrées sur la **FIGURE 1**. On rappelle **l'hypothèse fondamentale** : la 2e photo a été prise au même point origine $(0,0,0)$. Seule variation : l'orientation de l'appareil. Il existe donc une transformation linéaire $L : \mathbb{R}^3 \to \mathbb{R}^3$ envoyant la nouvelle position de l'appareil (définie par trois vecteurs e'_x, e'_y, e'_λ) sur l'ancienne : $e_x = L e'_x$ (et ainsi de suite). Remarque : elle envoie une base sur une base, donc L est un isomorphisme. Il revient au même de considérer que l'appareil photo n'a pas bougé de place mais que ce sont les points de l'image qui ont été transportés par L. Cela tombe bien, car on considère des « droites d'équivalence » et les images par toute application linéaire de deux points d'une même droite appartiennent à la même droite. Ensuite, le passage en coordonnées inhomogènes permet de reconstruire la photo n°2. Par conséquent, le lien entre les photos 1 et 2 est simple : en coordonnées homogènes, il existe une matrice $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ telle que pour tout u_1, $\quad u_2 = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} u_1$. Dans le cas le plus général, il s'agit d'une **homographie**. Ensuite, il existe des cas particuliers de transformations homogènes que nous allons étudier.

ILLUSTRATIONS ET REMARQUES AVANCEES

<u>FIGURE 1 :</u> Les droites d'équivalence = droites de perspective



Les droites bleues sont des droites de perspective : ce sont des droites passant par l'origine du repère 3D (là où se tient l'observateur / l'appareil photo). Si deux points sont sur cette droite, ils correspondent au même point pour l'observateur. La ligne rouge symbolise la photo placée en $\lambda = 1$ (de sorte qu'elle coïncide avec la perspective pour se fondre dans le paysage). Les mouvements étant relatifs, il revient au même de changer la base de l'appareil photo (pour prendre en compte sa nouvelle orientation) ou de transporter tout le paysage par la même application linéaire. Fort heureusement, si deux points appartiennent à une même droite de perspective pour la photo n°1, alors leurs images par L appartiennent aussi à la même droite de perspective (pour la photo n°2). C'est important car avec l'appareil photo, on perd l'appréciation de la distance : on ne sait pas où sur la droite de perspective se trouve le point de paysage d'origine (voir illustration d'Escher page suivante) (*).

<u>REMARQUE AVANCEE :</u> Bien sûr, il n'y a aucune raison que la position parfaite pour superposer la 2^{e} image au paysage soit $\lambda = 1$: la 2^{e} image peut très bien être imprimée en plus grand ou en plus petit que la 1ère. On peut même imaginer que l'image est étirée horizontalement et/ou verticalement après impression. Heureusement, toutes ces distorsions sont linéaires et peuvent être prises en compte dans l'application linéaire L. Cela permet de faire comme si la 2^{e} image se confondait par effet de perspective avec le paysage en la position $\lambda = 1$.

<u>UNE TRANSFORMATION LINEAIRE QUELCONQUE DANS L'ESPACE HOMOGENE CORRESPOND BIEN A UNE TRANSFORMATION HOMOGRAPHIQUE DANS L'ESPACE 2D :</u>

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} u_{1x} \\ u_{1y} \\ 1 \end{pmatrix} = \begin{pmatrix} au_{1x} + bu_{1y} + c \\ du_{1x} + eu_{1y} + f \\ gu_{1x} + hu_{1y} + i \end{pmatrix}$$

En passant en coordonnées inhomogènes, cela donne :

- $u_{2x} = \frac{au_{1x}+bu_{1y}+c}{gu_{1x}+hu_{1y}+i}$
- $u_{2y} = \frac{du_{1x}+eu_{1y}+f}{gu_{1x}+hu_{1y}+i}$

Cela correspond bien à une transformation homographique dans l'espace 2D des images.

****

<u>(*) ESCHER ILLUSTRE TRES BIEN QU'EN 2D, ON PERD TOUTE NOTION DE PROFONDEUR :</u> M. C. Escher a souvent joué sur le fait que la perception de profondeur disparaît en 2D pour créer des formes impossibles en 3D. M. C. Escher joue notamment sur le fait que les « coins » du cube ci-dessous peuvent être « interprétés » par le cerveau comme étant à la fois « intérieurs » et « extérieurs ». **La notion de distance n'existe pas en perspective.**

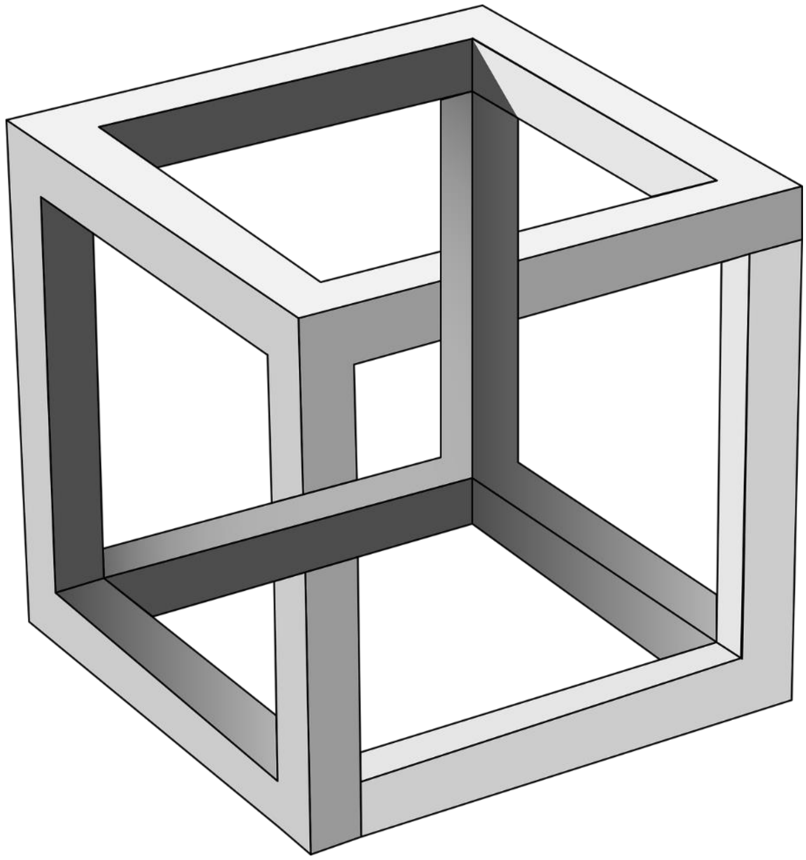


Image téléchargée sur [Cube Escher Pente Mc - Images vectorielles gratuites sur Pixabay]

<u>COMPLEMENTS SUR LA FIGURE N°1 / LIMITES DES HYPOTHESES :</u> Sur la figure 1, on a placé la photo parfaitement selon la coordonnée λ de sorte que la photo se fonde dans le paysage (par les lois de la perspective / les droites d'équivalence) pour un observateur placé au lieu de l'appareil. On peut se demander quel est le sens de ce raisonnement, et sous quelles hypothèses on peut accepter de remplacer l'appareil photo par un observateur et vice versa. En effet, pour être exact, on suppose que le centre optique de la lentille de l'appareil photo coïncide avec l'origine O du repère 3D. On suppose de plus que l'image se forme en « inversé » sur un tapis de capteurs placé quelque part en un $\lambda_{capteur} < 0$. Le prolongement des droites d'équivalence pour les lambdas négatifs permettrait alors de conclure que l'image formée sur le tapis de capteurs est reliée proportionnellement (selon λ) à la photo que l'on tire. En acceptant cette hypothèse, le raisonnement ci-dessus est correct. Si on suppose que tout le système est approximativement stigmatique et que l'appareil photo est constitué d'une seule lentille, cette hypothèse est vraisemblable. La réalité est plus compliquée parce que a) les lentilles ne sont pas des systèmes stigmatiques (voir le dossier « Stigmatism » sur GitHub) b) les appareils photos ont plusieurs lentilles et des traitements plus complexes qu'une simple relation de proportionnalité. **Le raisonnement présenté ci-dessus présente donc des limites, mais il reste intéressant et les relations homographiques sont très utilisées pour fusionner plusieurs images.**

CONCLUSION : Malgré ces limites dans les hypothèses, les notions de coordonnées homogènes et d'homographie ne sont pas du tout anodines. Elles sont largement utilisées dans des livres de référence en computer vision comme "Algorithms and Applications" de Richard Szeliski (https://szeliski.org/Book/)

Cette méthode (avec les hypothèses présentées dans ce document) s'appelle le « modèle du trou d'épingle » (Pinhole Model).