Quelques problèmes géométriques en vision par ordinateur

Frédéric SUR
École des Mines de Nancy / LORIA

Juillet 2009

Préambule. Ce document est la version (très) longue d'un texte écrit pour une épreuve de recrutement aux Grandes Écoles (niveau Bac+2). Le « bon » cadre pour étudier ces problèmes est celui de la géométrie projective, que l'on évite ici afin de rester à un niveau élémentaire.

Sommaire

Notations utilisées.

Dans ce document A^T désigne la transposée d'une matrice A, I_3 est la matrice identité d'ordre 3, et si M et N sont deux matrices ayant le même nombre de lignes, alors $\begin{pmatrix} M & N \end{pmatrix}$ est la matrice formée des « blocs » M et N (idem pour les colonnes). Deux matrices A et B de mêmes dimensions sont proportionnelles (ce qui sera noté $A \simeq B$) s'il existe $\lambda \neq 0$ tel que $A = \lambda B$. Cette définition s'applique en particulier à des vecteurs lignes ou colonnes.

D'autre part, $x \cdot y$ désigne le produit scalaire canonique de deux vecteurs x et y de $\mathbb{R}^3$, et $x \wedge y$ leur produit vectoriel. L'annexe 1 à la fin du document donne quelques résultats sur le produit vectoriel. Pour deux points A et B d'un espace affine, on notera $\overrightarrow{AB}$ le vecteur associé à A et B.

1 Introduction

La vision par ordinateur est un domaine de recherche très actif situé au croisement de l'informatique, des mathématiques, et, par certains aspects, de la robotique. L'objectif est d'extraire de l'information présente dans des images ou des séquences vidéo ; les applications vont de la télésurveillance aux effets spéciaux pour l'industrie cinématographique en passant par la localisation de robots autonomes et le multimédia. Ce texte présente quelques concepts et techniques géométriques utilisés dans certaines applications de vision par ordinateur.

2 Modélisation d'une caméra

On distinguera dans la suite d'une part le monde physique tridimensionnel, et d'autre part l'image bidimensionnelle que peut en produire un appareil d'acquisition. Cet appareil est appelé *caméra*, mais il peut tout aussi bien s'agir d'un appareil photographique car on ne considère ici que des images statiques[1]. La partie du monde physique vue par une (ou plusieurs) caméra(s) est appelée *scène*.

L'objectif est d'extraire de l'information sur la scène tridimensionnelle à partir d'images, il est donc nécessaire de commencer par modéliser la caméra. On présente dans la suite un modèle très courant : le modèle de *sténopé*, dit aussi de la *chambre noire*.

2.1 Modèle de sténopé

L'appareil photographique le plus simple à concevoir est sans doute celui à *sténopé* : une « boîte » ne laissant pénétrer la lumière que par un petit trou percé sur une face (trou d'épingle, d'où *pinhole camera* en anglais). L'image inversée de la scène se forme sur la face opposée au trou. Dans la pratique il faut déterminer assez soigneusement le diamètre du trou : trop gros la focalisation ne se fait pas correctement, trop petit la diffraction entre en jeu. On négligera ces problèmes dans le document.

La caméra à sténopé est réduite à sa plus simple expression : les images se forment sur un *plan-image* (noté $\mathcal{P}$) situé derrière un *centre optique* (noté C). Le plan-image correspond de nos jours à un capteur numérique, et le centre optique est l'« objectif » de la caméra, ici le trou d'épingle. Le plan-image est doté d'un repère naturel $(\Omega, \mathbf{U}, \mathbf{V})$ qui est celui des coordonnées en pixels : un point situé en ligne u colonne v dans l'image résultante a pour coordonnées (u, v). On suppose que les pixels sont rectangulaires, de sorte que $(\Omega, \mathbf{U}, \mathbf{V})$ est orthogonal (mais pas nécessairement orthonormé).

Tout point de la scène observé par la caméra se projette sur le plan-image par projection centrale par rapport au centre optique. Les coordonnées (x, y, z) d'un point M

[1] par opposition à des séquences vidéo.

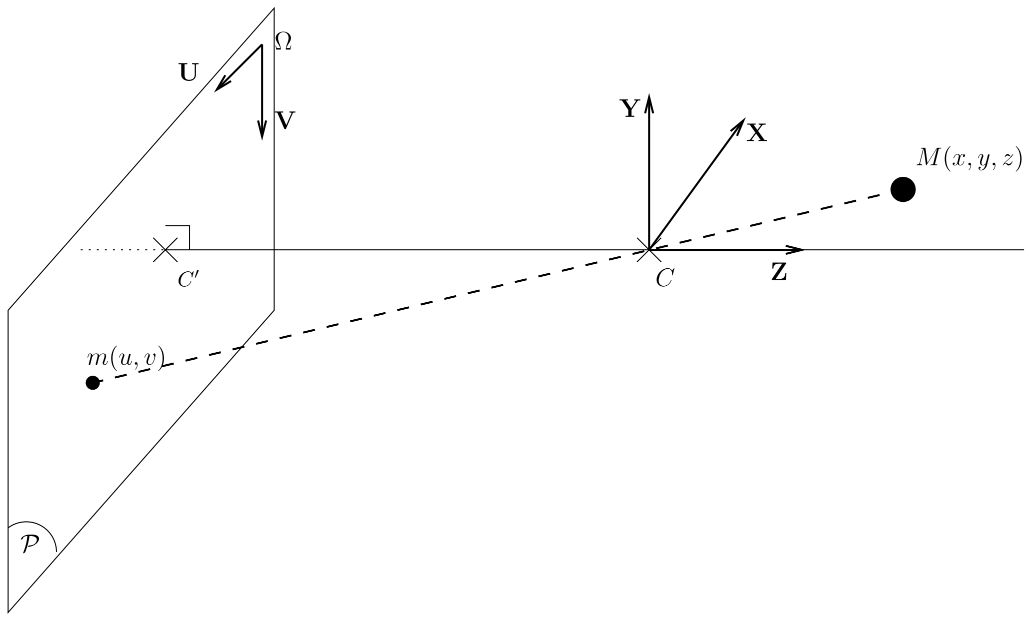

FIG. 1 – **Le modèle sténopé.** *On note C' la projection orthogonale du centre optique C sur le plan-image $\mathcal{P}$. La droite (CC') est appelée* axe optique *de la caméra. L'image du point M de la scène est le point m du plan-image.*

de la scène sont données dans un repère orthonormé direct $(C, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ attaché à la caméra : ce repère est centré sur le point C, l'axe $\mathbf{Z}$ est dirigé selon l'axe optique, et $\mathbf{X}$ et $\mathbf{U}$ sont colinéaires, de directions opposées.

La figure 1 illustre ces notions.

Remarque. Dans ce texte les coordonnées pixels sont des réels alors que « dans la réalité » ils sont des entiers. Les problèmes de discrétisation (ou d'*échantillonnage*) sous-jacents sont négligés, le lecteur intéressé consultera un cours de traitement du signal.

2.2 Matrice de projection de la caméra

Dans cette partie on établit les relations entre les coordonnées (x, y, z) dans le repère $(C, \mathbf{X}, \mathbf{Y}, \mathbf{Z})$ d'un point M de la scène et les coordonnées (u, v) dans le repère $(\Omega, \mathbf{U}, \mathbf{V})$ de son image m. On raisonne sur la figure 1.

2.2.1 Paramètres intrinsèques

Soit f la distance entre les points C et C' (appelée *focale* de la caméra). Alors d'après le théorème de Thalès :

$$\frac{1}{f}\,\overrightarrow{C'm} = -\frac{x}{z}\,\mathbf{X} - \frac{y}{z}\,\mathbf{Y}. \tag{1}$$

Construisons ensuite la base orthonormée directe du plan-image $(\mathbf{U}', \mathbf{V}')$ telle que $\mathbf{U} = -1/\alpha\;\mathbf{U}'$ et $\mathbf{V} = -1/\beta\;\mathbf{V}'$, avec α et β positifs. Ainsi, $\mathbf{U}' = \mathbf{X}$ et $\mathbf{V}' = \mathbf{Y}$.

Notons (u_0, v_0) les coordonnées de C' dans le repère $(\Omega, \mathbf{U}, \mathbf{V})$.

Comme $\overrightarrow{C'm} = (u - u_0)\mathbf{U} + (v - v_0)\mathbf{V}$, on peut écrire :

$$\overrightarrow{C'm} = -(u - u_0)/\alpha\;\mathbf{U}' - (v - v_0)/\beta\;\mathbf{V}'. \tag{2}$$

Des équations (1) et (2) on déduit :

$$\begin{cases} u = u_0 + \alpha\frac{fx}{z} \\ v = v_0 + \beta\frac{fy}{z} \end{cases} \tag{3}$$

D'où l'équation matricielle :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f\alpha & 0 & u_0 \\ 0 & f\beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x/z \\ y/z \\ 1 \end{pmatrix}. \tag{4}$$

Remarquons que les coordonnées (u, v) de la projection m du point M sur le plan-image ne dépendent des coordonnées (x, y, z) que par les rapports x/z et y/z. En effet, tout point de la droite (le « rayon lumineux ») joignant les points C et M se projette en m. Il est donc possible d'écrire :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \simeq \begin{pmatrix} f\alpha & 0 & u_0 \\ 0 & f\beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{5}$$

avec la notation $\simeq$ introduite au début du document.

La matrice triangulaire d'ordre 3 dans l'équation (5) est notée K et est appelée matrice des *paramètres intrinsèques* à la caméra. Ces paramètres sont la focale de la caméra, la taille de ses pixels, et les coordonnées pixels de C'. Remarquons que si les pixels sont carrés, alors $\alpha = \beta$.

Remarque importante. Pour simplifier l'écriture des équations dans la suite, on identifiera le point m de coordonnées pixels (u, v) et toute matrice-colonne proportionnelle à $(u, v, 1)^T$. L'équation (5) s'écrit alors :

$$m \simeq K \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{6}$$

et m est donné à un coefficient multiplicatif près. Les coordonnées (u, v) s'obtiennent en normalisant à 1 la dernière coordonnée de m.

2.2.2 Paramètres extrinsèques

L'équation (5) fournit les coordonnées de l'image d'un point de la scène à partir des coordonnées de ce point dans un repère orthonormé attaché à la caméra. Néanmoins, les points de la scène ne sont pas habituellement repérés dans le repère de la caméra (noté $\mathcal{R}_C$), mais plutôt dans un repère orthonormé indépendant « attaché » à la scène (noté $\mathcal{R}_\mathcal{S}$). Ceci est illustré par la figure 2. Notons R et t les matrices de rotation et translation pour passer du repère de la scène au repère de la caméra (R est une matrice de rotation d'ordre 3 et t une matrice-colonne à trois lignes). La relation suivante permet de passer des coordonnées (x', y', z') dans le repère $\mathcal{R}_\mathcal{S}$ aux coordonnées (x, y, z) dans le repère $\mathcal{R}_C$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + t. \tag{7}$$

Autrement écrit, sous forme d'un produit matriciel par blocs :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} R & t \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix}. \tag{8}$$

La matrice $\begin{pmatrix} R & t \end{pmatrix}$ d'ordre 3×4 est appelée matrice des *paramètres extrinsèques* à la caméra : R et t correspondent à l'orientation et à la position de la caméra dans le repère $\mathcal{R}_\mathcal{S}$.

Les équations (5) et (8) permettent ainsi d'écrire :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \simeq K \begin{pmatrix} R & t \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix}. \tag{9}$$

La matrice $\mathbf{P} = K \begin{pmatrix} R & t \end{pmatrix}$ d'ordre 3×4 est appelée *matrice de projection* de la caméra.

Remarque. La matrice $\mathbf{P}$ se décompose sous la forme :

$$\mathbf{P} = KR \begin{pmatrix} I_3 & R^{-1}t \end{pmatrix}. \tag{10}$$

D'après l'équation (7), le vecteur $\widetilde{C}$ des coordonnées du centre optique dans le repère de la scène vérifie $R\widetilde{C} + t = 0$. Donc :

$$\mathbf{P} = KR \begin{pmatrix} I_3 & -\widetilde{C} \end{pmatrix}. \tag{11}$$

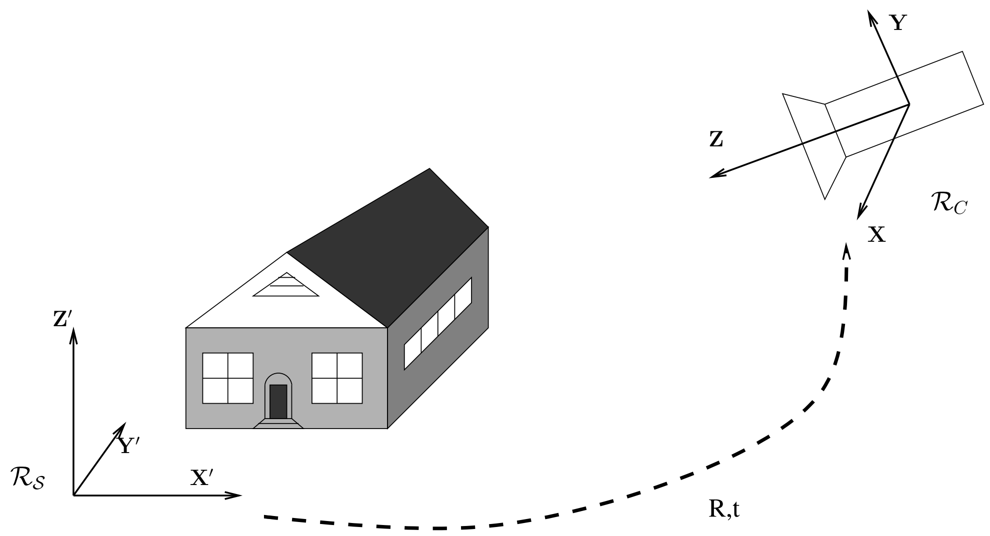

FIG. 2 – **Passage du repère de la scène au repère de la caméra.** *Il suffit de connaître la position et l'orientation de la caméra dans le repère de la scène.*

Cette dernière formule permet de déduire :

$$\mathbf{P} \begin{pmatrix} \widetilde{C} \\ 1 \end{pmatrix} = 0 \tag{12}$$

Remarquons que comme la matrice $\mathbf{P}$ est d'ordre 3×4, elle admet généralement un noyau de dimension 1, d'où l'unicité de $\widetilde{C}$.

2.3 Calibrage de la caméra

Le problème de la détermination des paramètres intrinsèques et extrinsèques d'une caméra est le problème du *calibrage*. On peut utiliser la méthode suivante, basée sur la décomposition RQ.

Proposition 1 (Décomposition RQ) *Pour toute matrice carrée A, il existe une matrice triangulaire supérieure R et une matrice orthogonale Q telles que : $A = RQ$. Si de plus A est inversible, il existe une unique décomposition RQ telle que les éléments diagonaux de R soient positifs.*

Remarque. Il existe des algorithmes pour déterminer effectivement R et Q. Parmi les outils du programme de premier cycle, le procédé d'orthonormalisation de Gram-Schmidt en est un.

Première étape : détermination de la matrice de projection $\mathbf{P}$. Étant donné (au moins) six correspondances entre des points M_i de la scène dont on connaît les coordonnées (x'_i, y'_i, z'_i) et des points m_i de l'image (coordonnées (u_i, v_i) également connues), la matrice de projection $\mathbf{P}$ peut être déterminée à partir des relations données par l'équation (9) : $m_i \simeq \mathbf{P} \begin{pmatrix} M_i^T & 1 \end{pmatrix}^T$. Pour ce faire, on peut fabriquer un système linéaire dont les inconnues sont les 12 coefficients de $\mathbf{P}$. Il faut 6 correspondances car chaque correspondance $m_i \leftrightarrow M_i$ fournit 3 équations, mais ajoute un paramètre (le coefficient de proportionnalité).

Concrètement, on utilise un objet (une *mire de calibrage*) sur lequel figurent des points facilement repérables dans l'image et dont on connaît la position tridimensionnelle avec précision.

Deuxième étape : détermination de K et R. D'après l'équation (11), on cherche K et R tels que $\mathbf{P} = KR \begin{pmatrix} I_3 & -\widetilde{C} \end{pmatrix}$.

1. $\widetilde{C}$ est déterminé en résolvant l'équation (12).
2. La proposition 1 appliquée à la matrice carrée formée des trois premières colonnes de $\mathbf{P}$ fournit alors les matrices K et R.

Remarque. Après décomposition RQ, la matrice K des paramètres intrinsèques a un coefficient en ligne 1 colonne 2 a priori non nul, alors que dans la définition de K donnée par l'équation 5, ce coefficient est nul. Il est (significativement) non nul dans le cas de pixels non rectangulaires.

Remarque. Si les paramètres intrinsèques de la caméra ne changent pas au cours du temps, ils sont déterminés par calibrage une fois pour toutes.

3 Problème à deux caméras

Intéressons-nous à présent aux problèmes soulevés dans le cas où l'on dispose de deux caméras (à des emplacements différents) qui regardent le même objet, ou de manière équivalente d'une caméra passant d'une position à une autre. Cette situation est illustrée par la figure 3, où les premières définitions sont données.

On cherche ici des informations sur la position des caméras et les coordonnées du point M de la scène en connaissant uniquement les « coordonnées pixels » des points m_1 et m_2 images de M. On sait uniquement que M est sur la droite $(C_1 m_1)$, et donc que m_2 est contraint à être sur l'image de cette droite par la deuxième caméra (qui est la droite l_2 dans la figure 3).

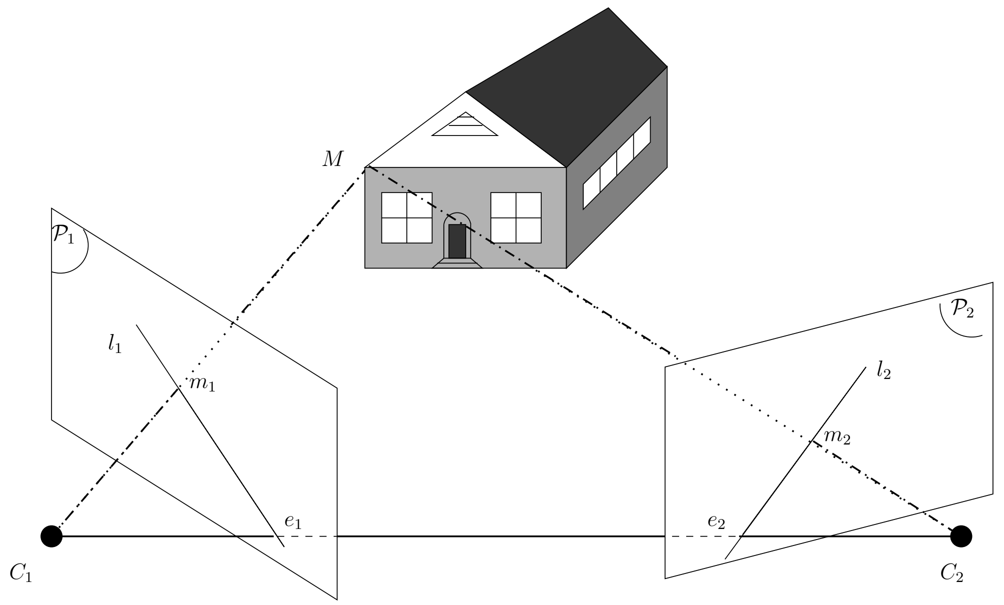

FIG. 3 – **Problème à deux caméras.** *Les deux caméras ont pour centres optiques respectifs C_1 et C_2 et pour plans-images $\mathcal{P}_1$ et $\mathcal{P}_2$. Pour des raisons de lisibilité, les plans-images sont placés dans ce schéma* devant *les centres optiques. Un point M de la scène se projette sur le plan-image de la caméra 1 (resp. 2) en m_1 (resp. m_2). Le point e_1 (resp. e_2) est l'image de C_2 (resp. C_1) sur $\mathcal{P}_1$ (resp. $\mathcal{P}_2$). Les points e_1 et e_2 sont appelés* épipoles. *La droite l_1 (resp. l_2) joignant les points e_1 et m_1 (resp. e_2 et m_2) est appelée* droite épipolaire *associée à m_2 (resp. m_1).*

3.1 Géométrie épipolaire

Le schéma de la figure 3 illustre une remarque fondamentale : l'ensemble des points de $\mathcal{P}_2$ qui peuvent être en correspondance avec un point m_1 de $\mathcal{P}_1$ (i.e. qui correspondent au même point M inconnu de la scène) est la droite épipolaire l_2, et vice-versa dans l'autre plan-image. C'est ce qui est appelé la *contrainte épipolaire*. Dans la suite, nous formalisons cette contrainte de manière algébrique.

3.2 Matrice fondamentale

3.2.1 Définition

Notons les matrices de projection des deux caméras comme dans la section précédente : $\mathbf{P}_1 = K_1 \left(R_1 \; t_1\right)$ et $\mathbf{P}_2 = K_2 \left(R_2 \; t_2\right)$. À ce stade cette décomposition ne peut pas être déterminée car les matrices de projection sont inconnues.

Notons également $\widetilde{C_1}$ et $\widetilde{C_2}$ les coordonnées des centres optiques dans le repère de

la scène. Comme $\mathbf{P}_1 \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} = 0$ (cf équation (12)), on déduit : $K_1 \begin{pmatrix} R_1 & t_1 \end{pmatrix} \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} = 0$.
D'où :

$$\widetilde{C_1} = -(K_1 R_1)^{-1} K_1 t_1. \tag{13}$$

Notons $\widetilde{M}$ les coordonnées du point M dans le repère de la scène. Comme le point M est sur la droite $(C_1 m_1)$, il existe un réel (non nul) α_1 tel que

$$\widetilde{M} = \widetilde{C_1} + \alpha_1 \mathbf{d}_1 \tag{14}$$

où $\mathbf{d}_1$ est un vecteur directeur de la droite $(C_1 M)$.

Comme m_1 est l'image de M par la caméra 1, on peut écrire[2] :

$$m_1 \simeq \mathbf{P}_1 \begin{pmatrix} \widetilde{M} \\ 1 \end{pmatrix} \simeq \mathbf{P}_1 \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} + \alpha_1 \mathbf{P}_1 \begin{pmatrix} \mathbf{d}_1 \\ 0 \end{pmatrix} \simeq 0 + \alpha_1 K_1 \begin{pmatrix} R_1 & t_1 \end{pmatrix} \begin{pmatrix} \mathbf{d}_1 \\ 0 \end{pmatrix} \simeq K_1 R_1 \mathbf{d}_1. \tag{15}$$

Rappelons en effet que $\mathbf{P}_1 \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} = 0$ d'après l'équation (12).

Par conséquent $\mathbf{d_1} \simeq (K_1 R_1)^{-1} m_1$.

Considérons à présent la droite épipolaire l_2. Elle passe par e_2 et m_2, image de M par la caméra 2. Or, par des calculs similaires à ceux de l'équation (15) :

$$m_2 \simeq K_2 \begin{pmatrix} R_2 & t_2 \end{pmatrix} \begin{pmatrix} \widetilde{C_1} + \alpha_1 \mathbf{d}_1 \\ 1 \end{pmatrix} = e_2 + \alpha_1 K_2 R_2 \mathbf{d}_1. \tag{16}$$

En effet $K_2 \begin{pmatrix} R_2 & t_2 \end{pmatrix} \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} = e_2$ par définition de l'épipole e_2.

Avec les explications de l'annexe 2, la droite l_2 est représentée par le vecteur :

$$l_2 = e_2 \wedge m_2 \simeq e_2 \wedge K_2 R_2 (K_1 R_1)^{-1} m_1. \tag{17}$$

D'autre part, par définition de e_2 :

$$e_2 \simeq \mathbf{P}_2 \begin{pmatrix} \widetilde{C_1} \\ 1 \end{pmatrix} = K_2 \begin{pmatrix} R_2 & t_2 \end{pmatrix} \begin{pmatrix} -(K_1 R_1)^{-1} K_1 t_1 \\ 1 \end{pmatrix}, \tag{18}$$

l'expression de $\widetilde{C_1}$ venant de l'équation (13).

Finalement, à l'aide des équations (17) et (18) :

$$l_2 \simeq \left(-K_2 R_2 (K_1 R_1)^{-1} K_1 t_1 + K_2 t_2\right) \wedge \left(K_2 R_2 (K_1 R_1)^{-1} m_1\right). \tag{19}$$

[2]En utilisant la notation de l'équation (6) dans la « remarque importante » page 5.

Avec la notation $[\dots]_\times$ introduite en annexe 1, on peut écrire $l_2 \simeq Fm_1$ où

$$F = [-K_2R_2(K_1R_1)^{-1}K_1t_1 + K_2t_2]_\times K_2R_2(K_1R_1)^{-1}. \tag{20}$$

La matrice F carrée d'ordre 3 est appelée *matrice fondamentale*. Elle est telle que la droite épipolaire l_2 dans l'image 2 associée à un point m_1 de l'image 1 est représentée par Fm_1.

3.2.2 Propriétés

Les résultats précédents permettent de donner quelques propriétés de la matrice fondamentale :

- Si m_1 et m_2 sont des points-images correspondant au même point de la scène, la contrainte épipolaire s'écrit
$$m_2^T F m_1 = 0. \tag{21}$$
En effet, m_2 est sur la droite épipolaire l_2 représentée par Fm_1, donc $m_2 \cdot l_2 = 0$ d'après l'annexe 2, et $l_2 \simeq Fm_1$.
- Symétriquement $l_1 = F^T m_2$ représente la droite épipolaire dans l'image 1 associée à un point m_2 de l'image 2.
- F est définie à un coefficient multiplicatif près, dans le sens où $F' = \lambda F$ vérifie $l_2 \simeq F'm_1$ ainsi que la contrainte épipolaire (21).

Dans le cas où les caméras ont été préalablement calibrées, les matrices K_1 et K_2 des paramètres intrinsèques sont connues. Supposons pour alléger les notations que le repère de la scène coïncide avec le repère de la caméra 1, de sorte que l'expression des matrices de projection devienne : $\mathbf{P}_1 = K_1 \begin{pmatrix} I_3 & 0 \end{pmatrix}$ et $\mathbf{P}_2 = K_2 \begin{pmatrix} R & t \end{pmatrix}$. La matrice fondamentale F s'écrit alors après simplification sous la forme :

$$F = [K_2 t]_\times K_2 R K_1^{-1}. \tag{22}$$

Comment s'interprête alors la contrainte épipolaire $m_2^T F m_1 = 0$? L'équation précédente permet d'écrire :

$$\left(K_2^{-1} m_2\right)^T \cdot t \wedge \left(R K_1^{-1} m_1\right) = 0. \tag{23}$$

Les vecteurs $K_2^{-1}m_2$ et $K_1^{-1}m_1$ sont les coordonnées des points images m_1 et m_2 dans les repères orthonormés attachés aux caméras. Sur la figure 3, le vecteur t dirige la droite (C_1C_2). Donc $t \wedge \left(RK_1^{-1}m_1\right)$ est normal au plan $(C_1C_2m_2)$. Ainsi, la contrainte épipolaire ne fait qu'exprimer la **coplanarité des points $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{m}_1$ et $\mathbf{m}_2$**.

D'autre part, avec l'équation 22 et en utilisant la proposition 2 (page 18, annexe 1) :

$$F \simeq (K_2^T)^{-1} [t]_\times R K_1^{-1}. \tag{24}$$

Remarque. De cette dernière équation on déduit que la matrice F est de rang 2 (sauf cas dégénéré), car K_1, K_2 et R sont inversibles et $[t]_\times$ est de rang 2 (comme toute matrice antisymétrique non nulle de taille 3, à faire en exercice en montrant par exemple qu'une valeur propre est nulle à l'aide du polynôme caractéristique).

On remarque que e_1 est dans le noyau de F : géométriquement, on voit que e_1 correspond à une situation dégénérée car e_1 se projette en e_2 dans la deuxième image et n'a donc pas de droite épipolaire associée. On peut aussi faire le calcul explicitement à partir de l'équation 20 et montrer que $Fe_1 = 0$. De même $F^T e_2 = 0$.

3.3 Matrice essentielle

On introduit alors la matrice $E = [t]_\times R$, appelée *matrice essentielle*. Remarquons qu'elle dépend uniquement de la position de la caméra 2 par rapport à la caméra 1. La contrainte épipolaire s'écrit :

$$(K_2^{-1} m_2)^T E (K_1^{-1} m_1) = 0. \tag{25}$$

Supposons à présent que l'on soit capable de déterminer E. Comment en déduire t et R ?

Détermination de t. Tout d'abord, on vérifie aisément[3] que : $\left([t]_\times\right)^2 = tt^T - (t \cdot t) I_3$ et que : $\text{Trace}([t]_\times^2) = -2\, t \cdot t$.

Donc :

$$tt^T = \begin{pmatrix} t_1^2 & t_1 t_2 & t_1 t_3 \\ t_1 t_2 & t_2^2 & t_2 t_3 \\ t_1 t_3 & t_2 t_3 & t_3^2 \end{pmatrix} = \frac{1}{2}\text{Trace}(EE^T) I_3 - EE^T, \tag{26}$$

car, $[t]_\times$ étant antisymétrique, $[t]_\times^2 = -[t]_\times [t]_\times^T$, d'où[4] $[t]_\times^2 = -EE^T$.

L'équation (26) donne alors deux solutions t_+ et t_-, de signes opposés.

Détermination de R. Pour déterminer R, on écrit :

$$[t]_\times E = [t]_\times^2 R = tt^T R - (t \cdot t) R. \tag{27}$$

Ensuite on exprime $tt^T R$ en fonction de la matrice essentielle E.

Si on note e_1, e_2 et e_3 les colonnes de la matrice $E = [t]_\times R$, et r_1, r_2, r_3 les colonnes de la matrice de rotation R (qui forment une base orthonormée directe), on voit que : $e_i = t \wedge r_i$ pour $i \in \{1, 2, 3\}$. On calcule alors :

$$e_1 \wedge e_2 = (t \wedge r_1) \wedge (t \wedge r_2) = (t \cdot r_3) t. \tag{28}$$

[3] Il suffit de décomposer $[t]_\times$ comme dans l'équation (35) et d'effectuer le calcul.
[4] On rappelle que la matrice R est orthogonale : $R^{-1} = R^T$.

Pour établir cette dernière égalité, il suffit de décomposer t dans la base (r_1, r_2, r_3) et de calculer les produits vectoriels en les décomposant dans cette base.

Ainsi, on obtient les égalités successives suivantes :

$$\begin{aligned} tt^T R = \begin{pmatrix} tt^T r_1 & tt^T r_2 & tt^T r_3 \end{pmatrix} &= \begin{pmatrix} (t \cdot r_1)t & (t \cdot r_2)t & (t \cdot r_3)t \end{pmatrix} \\ &= \begin{pmatrix} e_2 \wedge e_3 & e_3 \wedge e_1 & e_1 \wedge e_2 \end{pmatrix}. \end{aligned} \tag{29}$$

Il est donc possible de déterminer R à partir de E et t à l'aide des équations (27) et (29) :

$$R = \frac{1}{t \cdot t} \left(\begin{pmatrix} e_2 \wedge e_3 & e_3 \wedge e_1 & e_1 \wedge e_2 \end{pmatrix} - [t]_\times E \right). \tag{30}$$

Résumé. Si on cherche t et R tels que $E = [t]_\times R$, l'équation (26) fournit deux solutions pour t de signes opposés, notées t_+ et t_-. Néanmoins, E est connue à un coefficient multiplicatif près. D'après (26), si E est multipliée par $\lambda \neq 0$, t_+ et t_- sont multipliés par $|\lambda|$. Connaissant E et t, la matrice de rotation R est alors déterminée par l'équation (30). On remarque que si E est multipliée par $\lambda \neq 0$, R devient alors :

$$R' = \frac{1}{t \cdot t} \left(\begin{pmatrix} e_2 \wedge e_3 & e_3 \wedge e_1 & e_1 \wedge e_2 \end{pmatrix} - \text{signe}(\lambda)\ [t]_\times E \right). \tag{31}$$

Ainsi, la connaissance de E donne deux choix possibles pour t (t_+ et t_-), et deux choix possibles pour R :

$$R = \frac{1}{||t||^2} \left(\begin{pmatrix} e_2 \wedge e_3 & e_3 \wedge e_1 & e_1 \wedge e_2 \end{pmatrix} \pm\ [t]_\times E \right) \tag{32}$$

(avec t égal à t_+ ou t_-).

On obtient donc quatre choix possibles pour la position relative des caméras. De plus, si E n'est connue qu'à un facteur près, il en est de même pour le vecteur de translation t.

Ces quatre choix peuvent s'interpréter géométriquement : ils correspondent aux quatre possibilités pour les positions respectives des plans images, centres optiques et scène observée. Une seule de ces possibilités correspond à une scène située « devant » les deux caméras. Ceci est illustré par la figure 4.

3.4 Problème de la reconstruction

Dans le *problème de la reconstruction*, une scène est vue par plusieurs caméras (deux dans notre cas), et on cherche à déterminer la position respective des caméras, ainsi que les coordonnées des points de la scène observés dans les images.

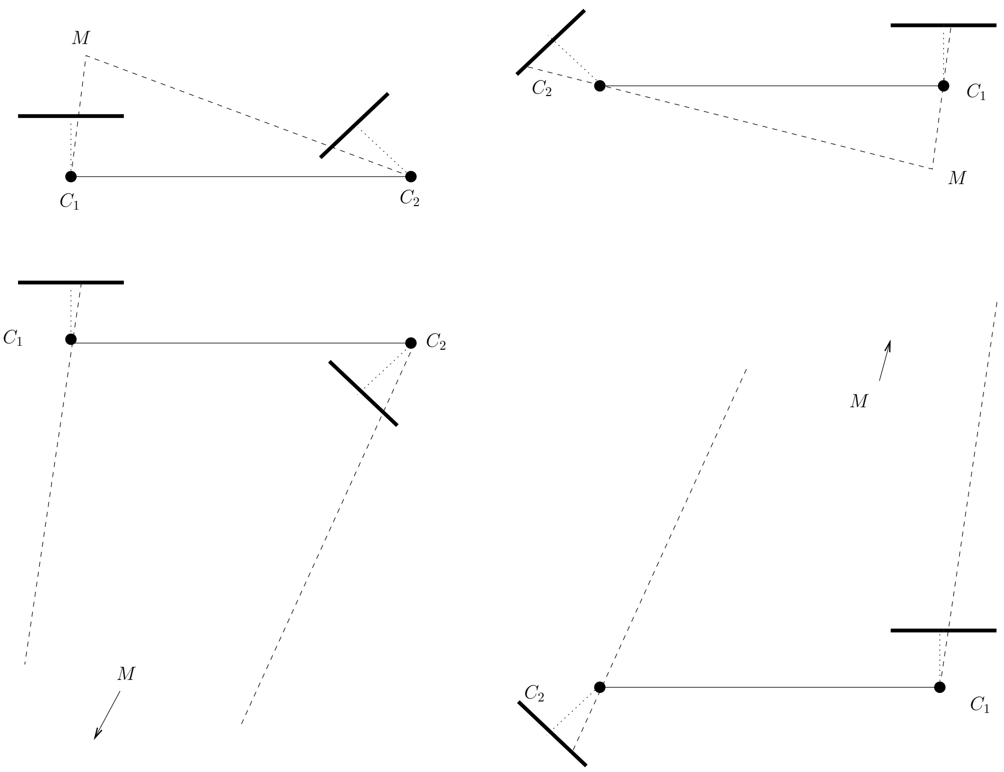

FIG. 4 – **Positions relatives des caméras.** *La matrice essentielle fournit quatre possibilités. Ici les plans images (segments gras) sont devant les centres optiques (gros points). Les directions (C_iM) sont fixées par rapport à l'axe optique des caméras, les quatre cas sont obtenus par symétrie par rapport à (C_1C_2) et échange des positions respectives des caméras. Dans les quatre cas, la scène se projette de la même manière sur les plans images. En haut à gauche, scène devant C_1 et devant C_2 ; en haut à droite, scène derrière C_1 et derrière C_2 ; en bas à gauche, scène derrière C_1 et devant C_2 ; en bas à droite, scène devant C_1 et derrière C_2. Une seule position est possible « physiquement ».*

Grâce aux outils introduits, il est possible de résoudre (en grande partie) ce problème. On dispose des deux images prises par les deux caméras, dont on suppose connus les paramètres intrinsèques obtenus par un calibrage préalable. Voici un algorithme possible :

1. **Mise en correspondance** des points images du même point de la scène. Dans l'exemple de la figure 3, il s'agit d'être capable de mettre en correspondance les points m_1 dans la première image et m_2 dans la deuxième image, en détectant qu'ils correspondent tous les deux au coin supérieur gauche de la façade de la maison. La maison n'ayant pas le même aspect dans les deux images, la mise en

correspondance est un problème en soi...

2. Avec au moins 8 correspondances, on peut effectuer le **calcul de la matrice fondamentale** F grâce aux contraintes épipolaires (cf équation (21)). En effet chaque contrainte épipolaire fournit une équation linéaire en les coefficients de F qui a 9 paramètres, mais est définie à un coefficient multiplicatif près. Reste à imposer ensuite que le rang de F est 2 (la solution du système n'a pas de raison de vérifier cette condition)...
3. Comme les paramètres intrinsèques des caméras sont connus, il est possible de **déterminer la matrice essentielle** E[5].
4. **Calcul des positions relatives des caméras** par estimation des matrices R et t associées à la décomposition de E. Comme la matrice $E = [t]_\times R$ est connue à un facteur multiplicatif près (de la même façon que F), il en est de même pour t.
5. **Reconstruction de la scène** par triangulation. Pour chaque couple de points (m_1, m_2) en correspondance dans les deux images, on cherche M qui se projette en m_1 par la caméra 1 et m_2 par la caméra 2. Le point M est donc solution des équations suivantes (où $\mathbf{P}_1$ et $\mathbf{P}_2$ sont maintenant connues) :

$$m_1 \simeq \mathbf{P}_1 \begin{pmatrix} \widetilde{M} \\ 1 \end{pmatrix} \quad \text{et} \quad m_2 \simeq \mathbf{P}_2 \begin{pmatrix} \widetilde{M} \\ 1 \end{pmatrix} \tag{33}$$

   Ce système admet en général une solution unique si m_1 et m_2 sont correctement appariés.En effet, le point M est à l'intersection des deux rayons lumineux passant par les centres optiques et les points images m_1 et m_2. Chaque égalité fournit trois équations, et il faut estimer les trois coordonnées de M ainsi que les deux coefficients de proportionnalité. On voit que si le mouvement entre les deux caméras est faible (comparativement à la distance de la scène), l'estimation est imprécise car les rayons lumineux sont quasiment parallèles.

Remarque. Comme la translation t est connue à un facteur près, la matrice de projection $\mathbf{P}_2$ est de la forme $K_2 \begin{pmatrix} R & \lambda t \end{pmatrix}$. Les solutions des équations (33) sont alors de la forme $\begin{pmatrix} \widetilde{M} \\ 1/\lambda \end{pmatrix} \simeq \begin{pmatrix} \lambda \widetilde{M} \\ 1 \end{pmatrix}$.

La reconstruction des points de la scène comme la position relative des deux caméras est donc connue au même facteur d'échelle près. Ceci correspond à l'intuition : une photographie ne donne pas l'échelle des objets qu'elle représente. La seule possibilité est de connaître la taille réelle d'un des objets de la scène, qui fixe alors l'échelle globale.

[5]Remarquons que si les paramètres intrinsèques sont connus, il est possible de déterminer directement E sans passer par F.

4 Quelques commentaires

4.1 Reconstruction par triangulation en pratique

Les problèmes pratiques d'implantation informatique ne sont pas du tout abordés ici. Les difficultés sont de deux ordres (en dehors des problèmes classiques de conditionnement des systèmes linéaires à inverser). Tout d'abord la position des points images n'est connue qu'à 0.1 - 0.5 pixel près (avec une interpolation convenable des images). Donc par exemple l'étape 5 de l'algorithme de la section 3.4 ne peut pas consister simplement en la résolution de ces équations, qui n'ont pas de solution en pratique pour ces raisons d'imprécision numérique. D'autre part, la recherche des appariements dans l'étape 1 est difficile, et les algorithmes utilisés génèrent souvent des faux appariements, c'est-à-dire des appariements entre points-images qui ne correspondent pas au même point de la scène. Pour résoudre ce problème on recherche les appariements sous contrainte épipolaire par des algorithmes robustes comme RANSAC.

4.2 Détermination de F à partir de correspondances

Il s'agit du point 2 du problème de la reconstruction (section 3.4). On suppose disposer de 8 correspondances entre deux vues.

Donnons quelques notations :

- $F = (F_{ij})_{1 \leqslant i,j \leqslant 3}$ est la matrice fondamentale ;
- f est le vecteur colonne $(F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T$;
- (m_j, m'_j) est la j-ème correspondance $(1 \leqslant j \leqslant 8)$;
- $m_j = (x_j, y_j, 1)$ est le vecteur des coordonnées homogènes du point m_j (idem pour m'_j), avec (x_j, y_j) les coordonnées pixels ;
- M est la matrice d'ordre 8×9 :

$$M = \begin{pmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ x'_2 x_2 & x'_2 y_2 & x'_2 & y'_2 x_2 & y'_2 y_2 & y'_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_8 x_8 & x'_8 y_8 & x'_8 & y'_8 x_8 & y'_8 y_8 & y'_8 & x_8 & y_8 & 1 \end{pmatrix}$$

La contrainte épipolaire (c'est-à-dire : $\forall j \in [1, n],\ m'_j F m_j = 0$) mène à l'équation matricielle : $Mf = 0$. La solution $f = 0$ n'est bien sûr pas acceptable, ainsi f est définie comme un vecteur non nul du noyau (donc à un coefficient multiplicatif près). D'autre part, F a pour rang 2, ce qui doit encore être imposé après coup, par exemple en annulant la plus petite valeur singulière dans la décomposition SVD de F. Cette méthode est connue sous le nom *méthode des 8 points*.

4.3 Notes bibliographiques

Pour aller plus loin, on pourra consulter par exemple :

- R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press 2000.
- O. Faugeras, Q.-T. Luong, T. Papadopoulou, *The Geometry of Multiple Images*, MIT Press 2001.

Annexe 1 : produit vectoriel dans $\mathbb{R}^3$

On rappelle que si u et v sont deux vecteurs de $\mathbb{R}^3$, de coordonnées respectives dans une base orthonormée (u_1, u_2, u_3) et (v_1, v_2, v_3), alors le produit vectoriel $u \wedge v$ a pour coordonnées dans cette base : $(u_2v_3 - u_3v_2,\ u_3v_1 - u_1v_3,\ u_1v_2 - u_2v_1)$.

Le produit vectoriel définit une application bilinéaire et antisymétrique.

Si (u, v) est une famille libre, alors $u \wedge v$ est orthogonal au plan engendré par u et v. Si la famille est liée, alors $u \wedge v = 0$. En particulier $u \wedge u = 0$.

On rappelle aussi que pour tout triplet de vecteurs (u, v, w) de $\mathbb{R}^3$,

$$\det(u, v, w) = (u \wedge v) \cdot w \tag{34}$$

où $\cdot$ représente le produit scalaire canonique et $\det$ est le déterminant d'un triplet de vecteurs.

À tout vecteur $u = (u_1, u_2, u_3)$ de $\mathbb{R}^3$, on peut associer une matrice antisymétrique $[u]_\times$:

$$[u]_\times = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix}. \tag{35}$$

Ceci permet d'écrire le produit vectoriel comme un produit entre matrices : quel que soit le vecteur v,

$$u \wedge v = [u]_\times v = -\left(v^T [u]_\times\right)^T. \tag{36}$$

Dans le document il est fait référence à la proposition suivante.

Proposition 2 *Si M est une matrice inversible d'ordre 3, alors*

$$Mu \wedge Mv = \det(M) \left(M^T\right)^{-1} (u \wedge v). \tag{37}$$

Démonstration. Pour tout vecteur w de $\mathbb{R}^3$, d'une part $\det(Mu, Mv, Mw) = (Mu \wedge Mv) \cdot Mw$ (d'après l'équation (34)), et d'autre part $\det(Mu, Mv, Mw) = \det(M) \det(u, v, w)$.
Donc $(Mu \wedge Mv) \cdot Mw = \det(M)(u \wedge v) \cdot w$.
Or $(Mu \wedge Mv) \cdot Mw = \left(M^T(Mu \wedge Mv)\right) \cdot w$.
Ainsi, pour tout w : $\left(M^T(Mu \wedge Mv) - \det(M)(u \wedge v)\right) \cdot w = 0$.
On conclut : $M^T(Mu \wedge Mv) = \det(M)(u \wedge v)$, d'où la proposition si la matrice M est inversible. $\square$

Annexe 2 : droite passant par deux points du plan-image

On cherche à caractériser la droite l passant par deux points A et B du plan-image $\mathcal{Q}$. Celui-ci a pour équation $w = 1$ dans le repère affine $(O, \mathbf{u}, \mathbf{v}, \mathbf{w})$. (Cf figure 5.)

Le vecteur $\overrightarrow{OA} \wedge \overrightarrow{OB}$ est orthogonal au plan (O, A, B), donc en particulier à la droite (AB). Ainsi, la droite $l = (AB)$ peut être représentée par le vecteur (encore noté l) :

$$l = A \wedge B \tag{38}$$

où on a identifié A à $\overrightarrow{OA}$ et B à $\overrightarrow{OB}$). En effet, un point P du plan-image appartient à cette droite si et seulement si $l \cdot P = 0$ (c'est-à-dire $\overrightarrow{OP}$ et $l = A \wedge B$ sont orthogonaux).

Autrement dit, si les composantes de $l = A \wedge B$ sont (l_1, l_2, l_3) et celles de P sont $(u, v, 1)$, alors l'équation de la droite (AB) est obtenue par le produit scalaire $l \cdot P = 0$:

$$l_1 u + l_2 v + l_3 = 0. \tag{39}$$

Il s'agit bien de l'équation d'une droite affine du plan $\mathcal{Q}$.

Remarquons que le vecteur l est connu à un coefficient multiplicatif près (comme les coordonnées de A et B d'après la « remarque importante » page 5).

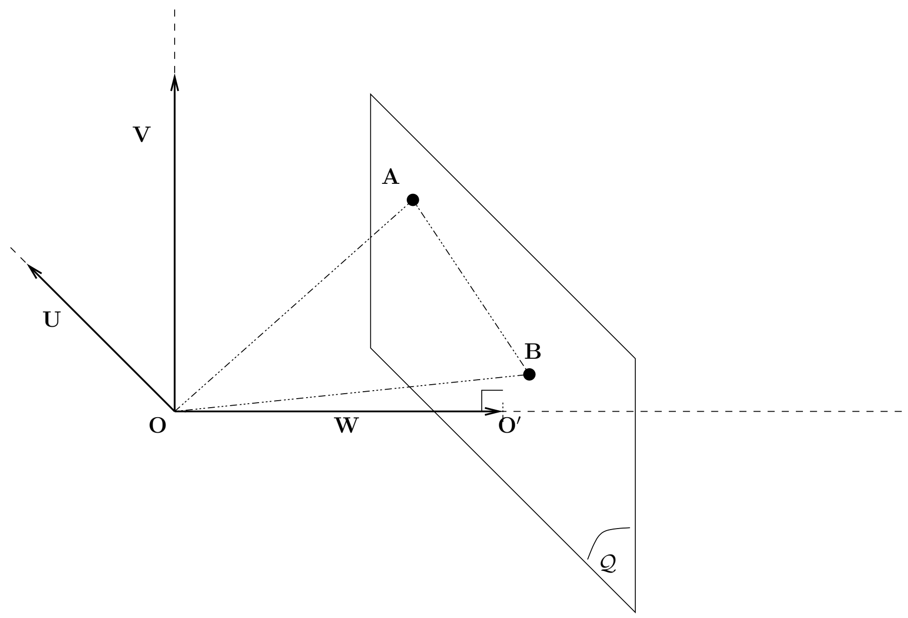

FIG. 5 – **Équation de la droite passant par deux points du plan-image.**

Remarque. Si les coordonnées de A et B sont respectivement (a_1, a_2, a_3) et (b_1, b_2, b_3), les coordonnées $(u, v, 1)$ d'un vecteur orthogonal au plan (OAB) vérifient :

$$\left\{ \begin{array}{l} a_1 u + a_2 v + a_3 = 0 \\ b_1 u + b_2 v + b_3 = 0 \end{array} \right. \tag{40}$$

Avec les formules de Cramer on déduit les solutions du système :

$$u = \frac{\begin{vmatrix} -a_3 & a_2 \\ -b_3 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}} \quad \text{et} \;\; v = \frac{\begin{vmatrix} a_1 & -a_3 \\ b_1 & -b_3 \end{vmatrix}}{\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}}. \tag{41}$$

Il est possible de les écrire sous la forme :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \simeq \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \wedge \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{42}$$

On retrouve le résultat établi géométriquement dans ce qui précède.