

ROB317 - Analyse et Indexation d'Images

TP2 - Découpage et indexation de vidéos

Bastien Hubert et Michel Bitar

3A - Parcours Robotique

1 Histogramme de couleurs

1.1 Question 1

Yuv est un espace colorimétrique qui est principalement utilisé sur un téléviseur analogique. Il était particulièrement utile lorsque la couleur a été introduite pour la première fois à la télévision, car il pouvait décoder et afficher à la fois la couleur et le noir et blanc. Essentiellement, un système Yuv est un système de codage mathématique comme les autres, mais qui intègre à la fois la luminosité et la couleur. Y signifie luma, ou luminosité, qui représente le niveau de gris, u signifie projection bleue, qui représente le contraste Bleu/Jaune et v signifie projection rouge, qui représente le contraste Rouge/Cyan.

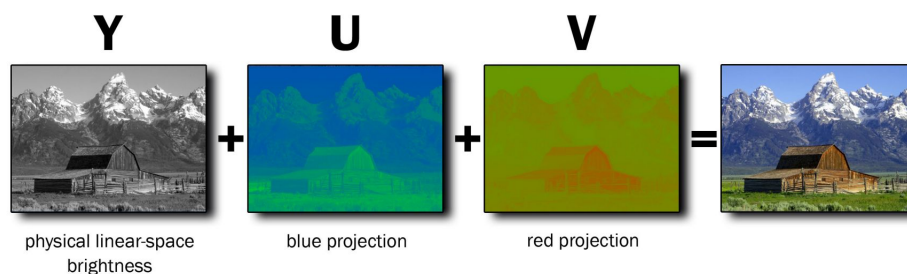


Figure 1: Yuv

RGB sont généralement simples: rouge, vert et bleu avec une taille de pixel donnée. Le problème avec RGB , cependant, est que ce n'est pas le meilleur pour représenter la perception visuelle. Les espaces colorimétriques YUV sont un codage plus efficace et réduisent davantage la bande passante que la capture RGB .

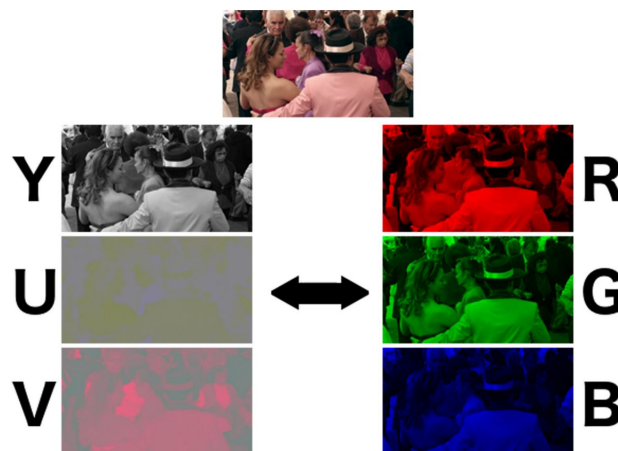


Figure 2: Différence entre Yuv et RGB

cv2.VideoCapture et *cv2.cvtColor* nous ont permis de lire chaque image de la vidéo en la transformant à l'espace *Yuv*. Ensuite, on a utilisé *cv2.calHist* pour tracer l'histogramme correspondant à la probabilité jointe des composantes chromatiques (u, v) du codage *Yuv* des images couleur.



Figure 3: Histogramme 2d

On remarque qu'il y a un changement significatif dans l'histogramme quand un changement de plan se passe. Pour cela, on a utilisé *cv2.comapareHist* pour calculer la corrélation entre les histogrammes. Les réductions rapides de la corrélation peuvent indiquer des changements de plans dans la video.

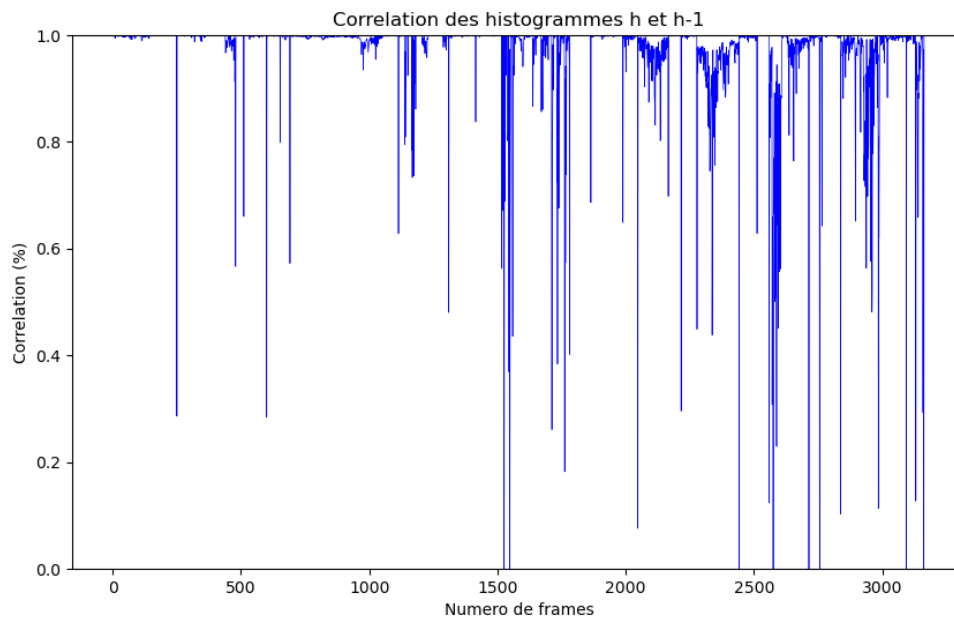


Figure 4: Corrélation entre les histogramme2d

Dans le cas de videos monochromes, il n'y aura plus d'histogramme 2d. L'absence des canal u et v introduit un histogramme 1d de video gris.

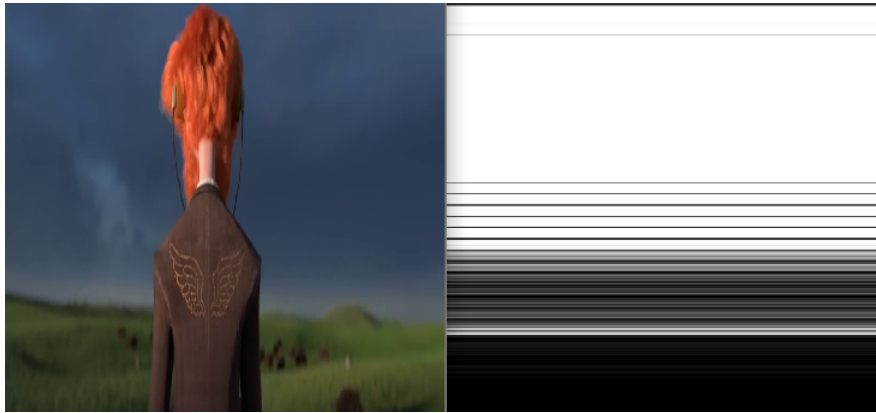


Figure 5: Histogramme 1d

2 Flot optique et histogramme de vitesses

2.1 Question 2

Le flot optique dense tente de calculer le vecteur de flot optique pour chaque pixel de chaque image. Bien qu'un tel calcul puisse être plus lent, il donne un résultat plus précis et un résultat plus dense adapté à des applications telles que l'apprentissage de la structure à partir du mouvement et de la segmentation vidéo. Cette méthode permet de déduire la vitesse et la direction du mouvement de l'objet en détectant les changements d'intensité des pixels de l'image au cours du temps. Il existe différentes implémentations de flux optique dense. Nous utiliserons la méthode Farneback, l'une des implémentations les plus populaires, avec OpenCV, pour l'implémentation.

Farneback a proposé une technique efficace pour estimer le mouvement d'éléments intéressants en comparant deux images consécutives. Tout d'abord, la méthode se rapproche des fenêtres des trames d'image par des polynômes quadratiques via une transformation d'expansion polynomiale. Deuxièmement, en observant comment le polynôme se transforme sous translation (mouvement), une méthode pour estimer les champs de déplacement à partir des coefficients d'expansion polynomiaux est définie. Après une série de raffinements, le flot optique dense est calculé. Il calcule l'amplitude et la direction du flux optique à partir d'un réseau à 2 canaux de vecteurs de flux (dx/dt et dy/dt), le problème du flux optique. Il visualise ensuite l'angle (direction) du flux par teinte et la distance (magnitude) du flux par valeur de représentation couleur HSV. La force du HSV est toujours fixée à un maximum de 255 pour une visibilité optimale.



Figure 6: Image et champs de vitesses (Farneback)

Dans le script *Dense – Optical – Flow.py*, on a utilisé *cv2.calcOpticalFlowFarneback* qui retourne les composantes (V_x, V_y) du mouvement optique. Les paramètres principaux de cette fonction sont *prvs*, qui est la première image d'entrée, *next*, qui est la deuxième image d'entrée de même taille et type que la précédente, *pyr_scale*, qui est le taux de réduction pyramidal de l'image, *levels*, qui est le nombre de niveaux de la pyramide incluant l'image initiale, *winsize*, qui est la taille de fenêtre de lissage (moyenne) des coefficients polynomiaux, *iterations*, qui est le nombre d'itérations effectués par l'algorithme à chaque niveau de la pyramide, *poly_n*, qui est la taille voisinage pour approximation polynomiale et *poly_sigma*, qui est l'écart type de la Gaussienne.

2.2 Question 3

Dans cette partie, on va analyser la forme de l'histogramme 2d (V_x, V_y) pour des différents mouvements de videos: plan fixe, Rotation OY: Pan, Rotation OX: Tilt, Rotation OZ (Roll), Travelling OX, Travelling OZ et Zoom in. L'histogramme de plan fixe est un point au centrale car il n'y a aucun mouvement.



Figure 7: Plan fixe

L'histogramme de rotation OY (Pan), présenté dans la figure ci-dessous, est un cône au centre de l'histogramme, d'axe horizontal vers la droite. C'est raisonnable car dans le mouvement de panoramique horizontal (vers la droite), la moitié du pixel en haut se déplace vers la gauche et le haut et la moitié du pixel en bas se déplace vers la gauche et le bas, cela forme un cône d'axe horizontal vers la droite.

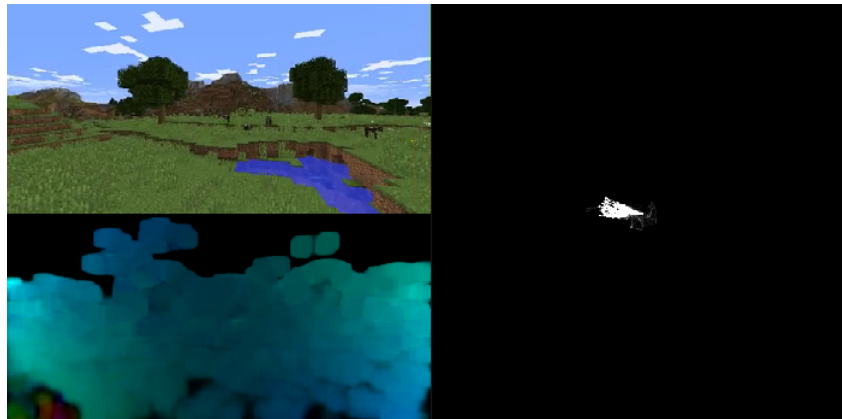


Figure 8: Rotation OY: Pan

L'histogramme de rotation OX (Tilt), présenté dans la figure ci-dessous, est un cône au centre de l'histogramme, d'axe vertical. C'est raisonnable car dans le mouvement de panoramique vertical, la moitié du pixel à gauche se déplace vers la gauche et le bas et la moitié du pixel à droite se déplace vers la droite et le bas, cela forme un cône d'axe vertical vers le haut.

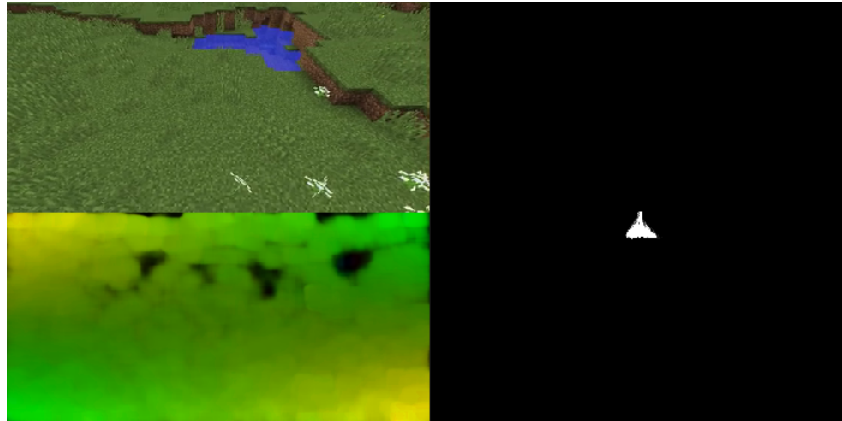


Figure 9: Rotation OX: Tilt

L'histogramme de rotation OZ (Roll), présenté dans la figure ci-dessous est un nuage de points au centre de l'histogramme. Pour un pixel dans l'image, plus loin il est du centre, plus grande sa vitesse de retournement. Puisque le déplacement se fait vers toutes les directions, l'histogramme est aussi quasi-circulaire.

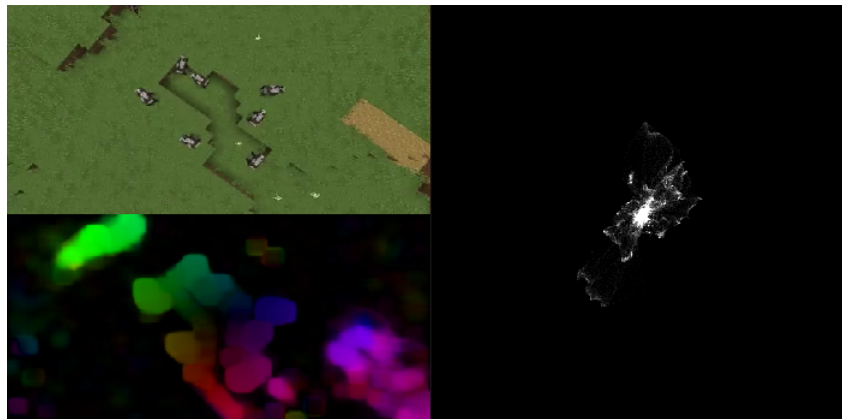


Figure 10: Rotation OZ: Roll

Le mouvement Translation OX de la caméra nous rend un histogramme présenté dans la figure ci-dessous. La cohérence entre le mouvement de caméra et le mouvement du pixel a été vérifiée dans cet histogramme.

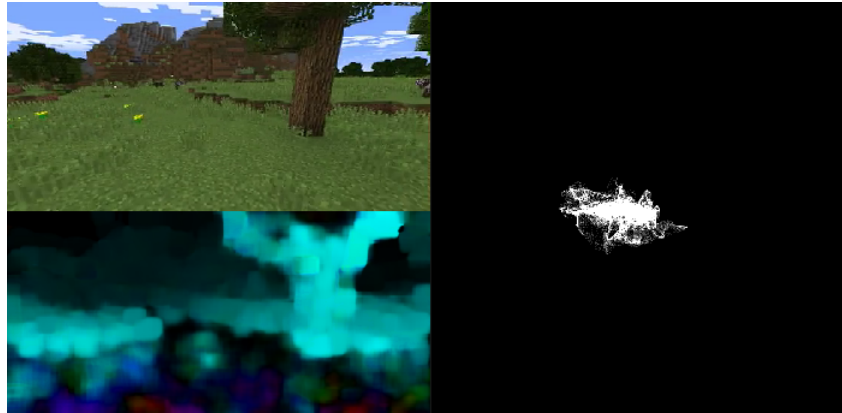


Figure 11: Translation OX

Le mouvement Translation OZ de la caméra nous rend un histogramme présenté dans la figure ci-dessous. La cohérence entre le mouvement de caméra et le mouvement du pixel a été vérifiée dans cet histogramme.

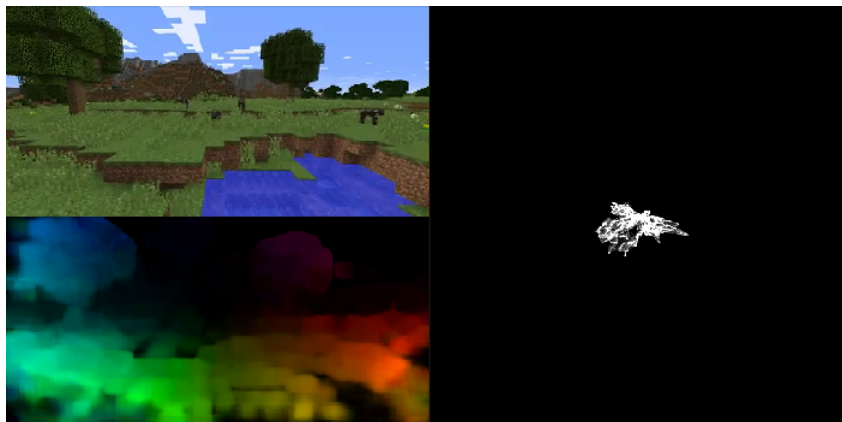


Figure 12: Translation OZ

Le mouvement zoom in de la caméra nous rend un histogramme présenté dans la figure ci-dessous, qui est similaire à l'histogramme de translation OZ. La rayon de la partie blanche est plus courte qu'avant car la vitesse de zoom in dans ce video est plus petite.



Figure 13: Translation OZ

3 Découpage et Indexation

Pour les questions suivantes, la vidéo utilisée est *Extrait5–Matrix–Helicopterscene(280p).m4v*.

3.1 Question 4

D'après la question 1, il est possible de calculer le pourcentage de corrélation entre les histogrammes d'une image h et l'histogramme de l'image précédente $h - 1$. Une manière d'identifier simplement un changement de plan d'observer un creux dans la fonction de corrélation, qui correspond à un moment où l'image courante est très différente de l'image précédente.

La première étape de notre découpeur en plan est donc de sélectionner tous les indices pour lesquels la fonction de corrélation des histogrammes passe en dessous d'un certain seuil. Empiriquement, la valeur du seuil a été choisie à 60%, ce qui conduit à la figure 14

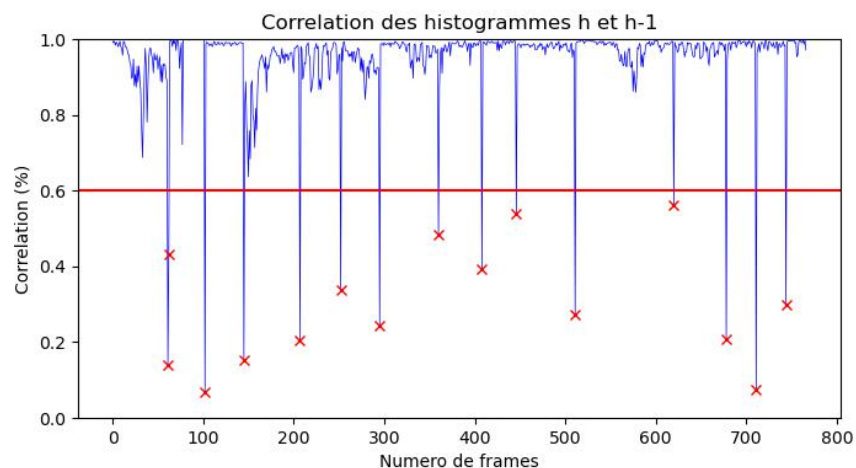


Figure 14: Première étape du détecteur de changement de plans

Cependant, on remarque la présence d'un faux positif au niveau du premier pic, et certaines images auraient pu être ajoutées si le seuil avait été plus élevé, comme le cluster d'images entre 150 et 180. Pour éviter ceci, on impose de ne pouvoir ajouter une nouvelle image tant que le niveau de corrélation n'est pas remonté au-dessus d'un second seuil, montrant que le plan s'est suffisamment stabilisé. Expérimentalement, on fixe ce nouveau seuil (haut) à 85%, et on augmente le seuil bas à 70% pour prendre en compte plus de changement de plans. Cette nouvelle version permet d'obtenir la figure suivante 15

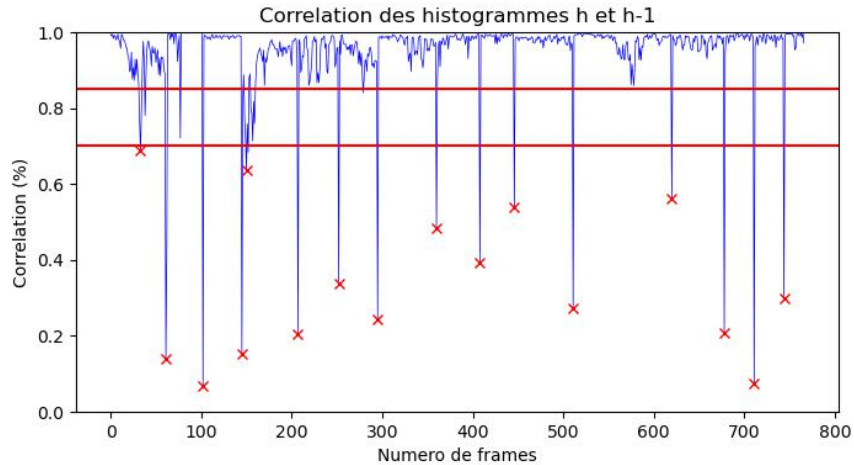


Figure 15: Deuxième étape du détecteur de changement de plans

Enfin, on ajoute les indices de début et de fin de la vidéo à la liste des indices calculés à l'étape 2 pour obtenir une série de plans formé par toutes les images entre les indices de début $b_k := e_{k-1}$ et de fin $e_k := b_{k+1}$.

3.2 Question 5

Une technique simple mais peu efficace d'extraction d'une image représentative d'un plan un consiste à choisir un indice r_k barycentrique entre l'indice du début du plan et celui de la fin du plan d'indice k . Le coefficient barycentrique choisi pour cette méthode est 0.2 (ie $r_k := 0.2 b_k + (1 - 0.2) e_k$) mais ce choix est arbitraire et on pourrait donner des arguments pour le fixer à 0.5 ou même à 0. Le principal défaut de cette méthode est son aspect statique : on ne prend pas en compte la structure de l'image ni son intégration dans le plan, et on n'a aucune information quantitative sur le degré de représentativité de l'image choisie pour tout le plan.

Une autre méthode, plus adaptative, consiste à définir la notion de représentativité d'une image f_i comme étant le degré de ressemblance de cette image avec toutes les autres du plan. Avec cette définition, on calcule la matrice de ressemblance M définie par : $\forall (i, j) \in [[0, e_k - b_k]]^2$, $M_{(i, j)} := d(f_i, f_j)$, où d est la distance euclidienne entre deux images. Cette matrice est symétrique par

symétrie de d .

De cette matrice, on peut définir la fonction de représentativité R par $R : f_i \mapsto \sum_{j=0}^{e_i-b_i} M_{(i,j)}$. Enfin, l'image choisie pour représenter le plan est l'image dont l'indice correspond à la plus petite valeur de R (ie. $r_k := \operatorname{argmin}\{R(f_i), i \in \llbracket 0, e_k - b_k \rrbracket\}$). L'image f_{r_k} est effectivement l'image dont la somme des différences avec les autres images est la plus petite, ce qu'on peut approximer comme étant l'image qui est la plus proche de toutes les autres images du plan, et donc la plus à même de représenter ce plan.

3.3 Question 6

Pour déterminer le type de plan parmi :

- Plan fixe
- Panoramique horizontal (pan) ou vertical (tilt)
- Rotation / axe optique
- Travelling horizontal, vertical, avant, arrière...
- Zoom avant ou arrière

nous allons avoir besoin de déterminer la direction globale de déplacement des images du plan, ce qui peut se faire en analysant la moyenne des histogrammes des différentes images du plan. Afin de détecter le type de plan dont il est question, la forme de l'histogramme va être étudiée et répartie en différentes catégories :

- histogramme fortement décentré : il s'agit d'un déplacement dans en direction du vecteur formé par la barycentre des particules et le centre du repère
 - histogramme avec une proportion importante autour de l'origine (de type cône) : il s'agit d'un **panoramique**, car toutes les particules ne se déplacent pas à la même vitesse, et sa direction est donnée par le barycentre de l'histogramme moyen
 - histogramme avec une faible proportion autour de l'origine : c'est un **traveling**, car peu de particules se déplacent à faible vitesse, et sa direction est donnée par le barycentre de l'histogramme moyen
- histogramme centré sur l'origine : il n'y a pas de direction prépondérante dans le flot, soit car les vecteurs vitesses se compensent, soit car ils sont faibles
 - histogramme circulaire de diamètre important : le plan présente des mouvements importants dans toutes les directions et à des vitesses différentes
 - * vecteur moyen correspondant au 2ème 9ème de cadran (la position de la touche 2 sur un clavier numérique de téléphone à touches) principalement horizontal : il s'agit d'une **rotation**

- * vecteur moyen correspondant au 2ème 9ème de quadrant principalement vertical : il s'agit d'un **zoom**
- histogramme circulaire de faible diamètre : il s'agit d'un **plan fixe** car l'histogramme ne montre que des déplacements faibles sans direction prédominante

Les figures 16a et 16b illustrent l'intérêt de l'étude des vecteurs du 2ème 9ème de quadrant :

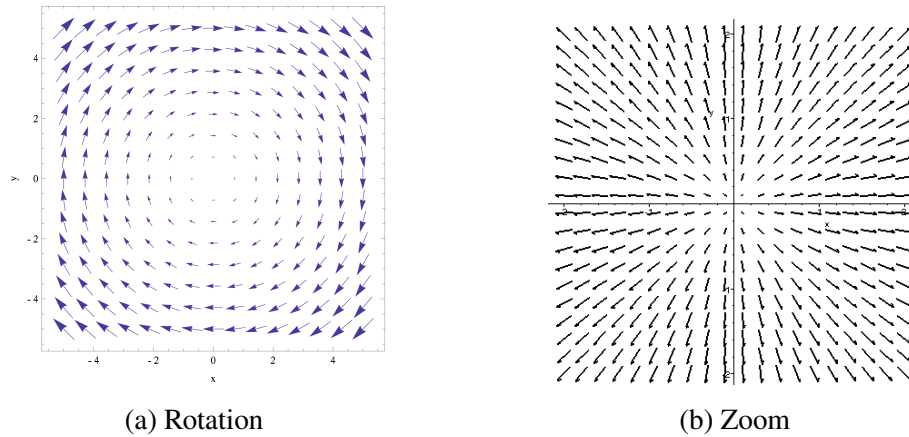


Figure 16: Champ vectoriel des vecteurs vitesses pour différents plans