

Réalité virtuelle et réalité augmentée

Suivi de personnes et évaluation

MBIAYA KWUITE Franck Anael
Promotion : SIM 24

Le travail expliqué dans cette présentation est disponible [ici](#)

Février 2021



1. Origine de la vidéo utilisée

- La vidéo utilisée dans ce travail est nommée **PETS09-S2L1** provient de la plate forme **motchallenge.net**
- Il s'agit d'une séquence très utilisée montrant jusqu'à 8 piétons marchant, en partie dans des schémas inhabituels.



2. Caractéristiques de la vidéo

- Cette séquence vidéo possède les saractéristiques suivantes
 - Nom : **PETS09-S2L1**
 - FPS : **7**
 - Résolution : **768x576**
 - Longueur : **01:54**
 - Nombre de frames : **795**
 - Pistes : **19**
 - Boîtes : **4476**
 - Densité : **5.6**
 - Nombre de personnes : **8**

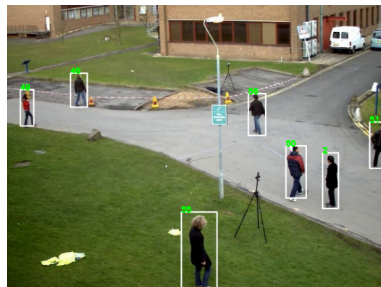


1. Fonctionnement

- Le suivi de personnes consiste ici à suivre plusieurs personnes dans une vidéo en se servant d'algorithmes.
- On utilise un détecteur et un traqueur pour suivre les personnes
- Le détecteur est un algorithme d'intelligence artificielle entraîné pour détecter les personnes
 - Nous avons utilisé ici le détecteur **YOLO V3** sur un CPU
 - Nous avons appliqué la détection de personnes sur tous les frames de la vidéo
- Le traqueur permet de traquer une personne, c'est à dire suivre son déplacement dans la vidéo
 - Nous avons utilisé ici le traqueur **Deep sort**

2. Résultat

Après exécution de la détection, nous avons obtenue le résultat dont l'un des frames est dans la figure suivante :



Généralités sur les mesures d'évaluation (MOTA et MOTP)

1. Formules

- **MOTA** est la précision globale du suivi en termes de faux positifs, de faux négatifs et d'erreurs de non-concordance (confusion d'identité)
- **MOTP** est la précision de suivi globale en termes de chevauchement de la boîte englobante entre la vérité terrain (Objet réel) et l'emplacement signalé (Hypothèse de présence d'objet)
- Si on considère l'ensemble d'objets visibles dans la vidéo $\{O_1, O_2, \dots, O_n\}$ et l'ensemble d'hypothèses de présence d'objets $\{h_1, h_2, \dots, h_m\}$, on a

$$\text{MOTP} = \frac{\sum_{i,t} d_t^i}{\sum_t c_t}$$

$$\text{MOTA} = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$

Avec :

- t : Intervalle de temps (correspond aux différents frames)
- c_t : Le nombre de correspondance entre les objets et les hypothèses à l'instant t
- d_t^i : La distance entre l'objet O_i et son hypothèses correspondante à l'instant t
- m_t : Le nombre d'objets manqués ou faux négatif (objets n'ayant pas d'hypothèses correspondantes) à l'instant t
- fp_t : Le nombre de faux positif (hypothèses qui se réfèrent à un objet qui n'existe pas) à l'instant t
- mme_t : Le nombre d'erreurs de non-concordance (hypothèses modifiées par rapport aux images précédentes à l'instant $t-1$) à l'instant t
- g_t : Le nombre total d'objets visibles à l'instant t

2. Algorithme

pour chaque intervalle de temps i :

● Etape 1 : Initialisation :

- $mme_t = 0$
- $fp_t = 0$
- $m_t = 0$

● Etape 2 :

- g_t = nombre total d'objets présents à ce moment t
- Pour chaque O_i, h_j au moment $t-1$, s'assurer que O_i, h_j est toujours valide
- Si O_i est toujours visible et h_j existe toujours au temps t , et si leur distance $<$ seuil, faire correspondance O_i, h_j pour le pas de temps t

● Etape 3 : Pour les objets dont aucune correspondance n'a été établie

- Trouver une hypothèse d'appariement en utilisant le problème d'attribution de poids minimum
- Si une correspondance O_i, h_k est faite qui contredit une correspondance O_i, h_j au moment $t-1$, remplacer O_i, h_j par O_i, h_k et incrémenter mme_t

● Etape 4 : Toutes les correspondances faites au moment t peuvent être décrites comme c_t

- Pour chaque match, calculer la distance d_t^i entre O_i et son hypothèse de correspondance
- Pour les hypothèses restantes (faux positifs), incrémenter fp_t en conséquence
- Pour les objets ratés (faux négatifs), incrémenter m_t en conséquence

● Etape 5 : Après tous les moments t (c'est à dire après avoir passé toutes les frames), calculer **MOTP** et **MOTA**

Calcul de MOTP et MOTA

1. Méthode de calcul

- Nous avons calculé MOTP et MOTA automatiquement en utilisant la librairie **motmetrics**
- Nous avons utilisé la vérité terrain (localisation des personnes dans la vidéo) disponible en téléchargeant la vidéo que nous avons utilisé. Il s'agit d'un fichier CSV dont le contenu ressemble à ceci

```
1,1,415,449,129,269,1,-1,-1,-1
1,2,581,443,97,269,1,-1,-1,-1
1,3,1347,415,153,377,1,-1,-1,-1
1,4,1429,423,161,339,1,-1,-1,-1
1,5,621,459,79,193,1,-1,-1,-1
1,10,473,453,103,261,1,-1,-1,-1
1,16,541,459,43,99,1,-1,-1,-1
```

Chaque ligne comporte 10 valeurs correspondant à :

- Numéro du frame (correspond aux différents moments décrit plus haut)
 - Id de l'identité de la personne
 - Coordonnée en largeur et en hauteur du coin haut-gauche de la boîte englobante
 - largeur de la boîte englobante
 - hauteur de la boîte englobante
 - confiance de détection de la personne
 - position x, y et z de la personne dans le monde réel
- Pendant le suivi avec le détecteur **YOLO V3** et le traqueur **Deep sort**, nous avons récupéré les informations sur les hypothèses à chaque moment (pour chaque frame de la vidéo) et nous avons mis à jour les résultats dans un objet **MOTAccumulator**. Les différentes étapes sont:
 - récupérer la liste des personnes et leurs Boîtes englobantes du frame en cours dans le fichier de vérité terrain
 - récupérer la liste des hypothèses donnée par le traqueur **Deep sort** et leurs boîtes englobantes
 - calculer la distance entre chaque personne et toutes les hypothèses (Nous avons utilisé une distance seuil de **50 cm**)
 - mettre à jour les valeurs de $mmet_t$, fp_t , mt

2. Résultats obtenus

- Après avoir parcouru tous les 795 frames, nous obtenons le tableau suivant :

IDF1	IDP	IDR	Rc11	Prcn	GT	MT	PT	ML	FP	FN	IDs	FN	MOTA	MOTP	IDt	IDa	IDm	
acc	51.4%	53.0%	49.9%	79.6%	84.4%	19	11	8	0	657	915	65	199	63.4%	0.289	37	16	3

- Tel que décrit dans la documentation de la librairie **motmetrics**, pour obtenir la valeur en pourcentage de **MOTP**, nous devons appliquer la formule suivante : $(1 - MOTP) * 100$. D'où on a $(1 - 0.289) * 100 = 71.1$
- Nous obtenons alors les résultats suivants pour MOTP et MOTA :

- **MOTA = 63.40 %**
- **MOTP = 71.10 %**

