IMT Atlantique

Dépt. Mathematical & Electrical Engineering Technopôle de Brest-Iroise - CS 83818

29238 Brest Cedex 3

Téléphone: +33 (0)2 29 00 13 04 Télécopie: +33 (0)2 29 00 10 12 URL: www.imt-atlantique.fr



Rapport Final PRONTO

GOARIN Maxime, maxime.goarin@imt-atlantique.net
MARBOEUF Kéwan, kewan.marboeuf@imt-atlantique.net
GHARSALLI Hanine, hanine.gharsalli@imt-atlantique.net
ROUSSEAU Edgar, edgar.rousseau@imt-atlantique.net
MOUSSADEK Mohamed-Aymane, mohamed-aymane@imt-atlantique.net
COLLOT François-Xavier, fx.collot@imt-atlantique.net

Enseignant Chercheur: PHAM Mai Quyen

Date d'édition : 26 septembre 2025

Version: 2.3



Sommaire

1.	Intro	Introduction		
2.	Orga	nnisation	2	
3.	Déma	arche et conception	3	
		Sous-groupe 1 (language grounding)	3	
		3.1.1. Contexte et objectif du sous-groupe 1	4	
		3.1.2. Avancées	4	
		3.1.3. modification des objectifs	4	
	3.2.	Sous-groupe 2 (object-centric-motion-generation)	5	
		3.2.1. Contexte et objectif du sous-groupe 2	5	
		3.2.2. Choix de l'environnement de développement	5	
		3.2.3. Étude de la base de données AMASS	5	
		3.2.4. Difficultés rencontrées avec la base Humanize	5	
		3.2.5. Redéfinition des axes de travail	6	
		3.2.6. Génération de translations et visualisation.	6	
		3.2.7. Utilisation du modèle SMPL-X	6	
		3.2.8. Traitement des données et animation finale	6	
		3.2.10. Intégration d'un modèle GRU pour la génération de poses	6 7	
			,	
4.	Résu	ltats (Validation, discussion)	8	
	4.1.	Cloture du projet	8	
		4.1.1. Tableau des livrables	8	
		4.1.2. Validation:	9	
		·	10	
	4.2.		10	
		*	10	
		4.2.2. Gestion des risques :	11	
5.	Conc	clusion et Perspectives	12	
6	Anne	exe	12	
٠.		Maxime Goarin : Décrivez une situation où vous avez pris le lead pendant ce projet.		
		Hanine GHARSALLI :Qu'avez-vous fait quand vous deviez motiver un groupe ou		
			13	
		Edgar ROUSSEAU : Racontez une situation dans ce projet où vous étiez confronté	10	
			13	
		Mohamed Aymane Moussadek: Dans ce projet, quand avez-vous pris un risque,		
			14	
		François-Xavier Collot: Racontez une situation dans ce projet où vous étiez	•	
			14	
			14	
_			1=	
1.	Bibli	ographie	15	

1. Introduction

Allier génération de scène 3D interactive et description textuelle est crucial pour améliorer le réalisme visuel et physique de l'interaction scène-corps humain. Dès lors ce sujet est le point d'intérêt d'un nombre croissant de projets de recherche visant à rendre plus efficient et facile d'utilisation la génération de scène 3D interactive. En effet, la modularité et la facilité d'utilisation de la description textuelle en font un outil puissant de création. Décrire par le texte est accessible à tous (pas de connaissance techniques nécessaires), offre une large variété de possibilités et permet d'inclure la subjectivité et la sensibilité de l'utilisateur qu'il est plus difficile d'exprimer à l'aide de Language de programmation classiques.

Ce sujet est le coeur de notre projet ingénieur PRONTO. L'objectif du projet est ambitieux mais essentiel : générer une scène 3D complète et un body capable de se mouvoir dans la scène à partir d'une description textuelle. Pour répondre à cet objectif, il nous a fallu, découvrir des outils nouveaux, et, faire preuve d'ingéniosité. En effet, le projet a nécessité des compétences en informatique (codage et simulation) nouvelles pour nous ainsi qu'une capacité à trouver des solutions simples à des problèmes complexes.

Notre étude s'intéressant à l'incorporation de la description textuelle à la génération de scène 3D interactive, notre travail nous permis de répondre deux problématiques en particulier; comment extraire les informations essentiel du texte? Comment générer une scène 3D et les modéliser les mouvements d'un body dans cette scène pour un consigne donnée?

Pour répondre à ces problématiques nous avons choisi une organisation de travail particulière. En effet nous avons scindé notre groupe de 6 en deux sous groupes de 3 chacun affecté à une des problématiques de notre travail. Chaque sous groupe avait des objectifs clairs et spécifiques ce qui a permis un travail plus efficace étant donné la variété des notions abordées dans le projet. Le sous groupe de la partie analyse du texte descriptif avait pour but d'extraire les informations essentielles du texte et les convertir en une consigne simple et compréhensible par nos outils de simulation. Le sous groupe chargé de la partie scène avait pour but de modéliser les mouvements du body dans la scène 3D (générée préalablement) sous la consigne donnée pour l'autre sous groupe. Etant donné la différence d'outils utilisés, la structuration de notre équipe (en deux sous groupes) a été un facteur clé de notre avancement. Le premier sous groupe a pu concentré ses efforts dans la maitrise des outils d'intelligence artificiels permettant l'extraction des données alors que l'autre sous groupe a pu maitriser au mieux les outils de simulation.

A l'état de notre avancement dans notre projet, ce rapport est divisé en 3 majeures parties. D'abord nous verrons la structure organisationelle de notre groupe et discuterons de son importance de l'avancée de notre projet. Par la suite, nous discuterons de la démarche choisie pour répondre aux objectifs du projet ainsi que de la conception des méthodes y répondant. Enfin, suit à cette partie plus technique nous mettrons en avant nos résulats en perspectives vis à vis des objectifs temporels et techniques fixés en début de projet. Nous conclurons enfin notre rapport avec une ouverture globale sur le projet.

2. Organisation

Comme précisé dans l'introduction, nous avons décidé de scinder notre groupe en deux sous groupes de 3 personnes. Le premier sous groupe constitué de Edgar, Mohamed-Aymane et Maxime avait pour objectif le traitement de la description textuelle dans le but de la convertir en une consigne de mouvement pour le body. Le second sous groupe constitué de Hanine, François-Xavier et Kéwan avait pour objectif la modélisation des mouvements du body dans la scène 3D sous la consigne donnée. Cette organisation de l'équipe a été choisie dans le début du projet et l'affectation des groupes à chacun des membres a été arbitraires sans influence des capacités techniques particulières de chacun.

Tout au long du projet, nous avons donc définis des objectifs clairs et distincts à chaque sous groupe et la majorité du travail se faisant en sous groupe. Cependant, dans des cas particuliers, nous travaillions tous ensemble notamment dans la réalisation des rendus d'équipe ne nécessitant pas une approche technique. Par ailleurs, malgré la structure de notre équipe, les échanges entre sous groupe étaient très fréquent car nous avons fait le choix de toujours travailler ensemble (les deux sous groupes) dans la même salle lors de séance dédiées à l'avancée technique du projet. Ce choix a permis, malgré la structure de l'équipe, de fluidifier dans une certaine mesure les échanges entre les sous groupes.

Les échanges avec l'encadrante se faisait via les outils de communications usuels (mail et Discord) et lors des réunions prévues avec elle. Lors de ces réunions, l'encadrante prenait le temps d'aider chaque groupe séparément à part au début du projet où nous avons tous participer à une réunion introductive du projet.

A la vue de notre avancement sur le projet nous avons pu tirer des conclusions sur le choix de notre structure organisationnelle et des pistes d'amélioration.

Tour d'abord, scinder le groupe en deux sous groupes a permis de simplifier la communication et l'organisation du travail étant donné que le nombre de personnes à coordonner était réduit. Par ailleurs, le choix organisationnel a permis à chaque équipe de concentrer leurs efforts sur un aspect technique du projet. Cet aspect a été essentiel étant donné la complexité du projet dans la variété des sujets abordés. En effet, le projet par nature a une structure divisée en termes d'objectifs et d'outils techniques nécessaires. Scinder l'équipe a donc permis de travailler simultanément sur les deux aspects du projet; traitement textuel et modélisation 3D. Cela a été un gain de temps pour nous car chacun des deux aspects demandent des outils différents et une approche globalement différente. Choisir de scinder notre groupe en deux a donc permis d'aborder au mieux le projet. Pour ce qui du rôle de chaque membre dans les sous groupes, nous avons fait le choix au début du projet de ne pas définir de tâche spécifique à chacun (rôle). En effet nous avons décidé de toujours travailler tous ensemble sur les même tâches de façon collaborative. Le travail personnel était tout de même là puisque lorsque le projet demandait une séparation des tâches (ce qui était plus courant pour le second groupe) nous n'avions aucun mal à le faire. Notamment lors de la réalisation des rendus demandés, nous définissions des tâches bien définies à chacun pour finaliser rapidement le projet. Ce choix de travailler ensemble sous les tâches (hors rendus) a permis à chacun de s'investir pleinement dans la réalisation du projet et de ne pas limiter les membres dans leur curiosité. Par ailleurs, cela a permis de garder une atmosphère ouverte et riche en échanges.

Nos choix organisationnels nous ont permis d'avancer relativement efficacement dans le projet. Cependant à la vue de nos réalisations, des leçons peuvent être tirées. D'abord, la structure scindée par nature a limité les échanges entre les membres (d'un sous groupe à l'autre). Par ailleurs, la structure a limité la compréhension et l'implication d'un sous groupe dans le travail de l'autre. Le sous groupe 1 n'a pas pus découvrir et comprendre les outils et méthodes permettant la modélisation des mouvements dans la scène 3D (et inversement) si ce n'est que très globalement. Pour ce qui est de l'implication des membres et l'assiduité, nous avons noté des hauts et des bas. Au début du projet le groupe a connu un manque de motivation et d'efficacité dans le travail. Ces défauts, principalement dus à une mauvaise communication entre les membres et avec l'encadrante ont pus être corrigés par la suite. En effet, la simplification des attendus par l'encadrante et un effort de tous dans nos échanges ont permis d'instaurer une atmosphère plus ouverte et productive. Par ailleurs, voir les premiers résultats de notre travail a relancé la motivation de chacun et dynamisé le groupe. Les outils de communications (groupes WhatsApp et Discord) ont été essentiels pour coordonner les membres du groupe et travailler au mieux, surtout lors de moment de doutes.

En conclusion, on notera l'importance de la communication dans le projet. En effet, dans certains moments, malgré des capacités techniques claires, le projet a été ralentit à cause d'échanges peu clairs entre les membres du groupe d'abord mais aussi entre le groupe et l'encadrante. A l'avenir, on pourra penser à introduire des activités de groupes hors travail technique permettant de renforcer les liens entre les membres, principalement lors que ceux-ci ne se connaissent pas. Par ailleurs, on notera qu'il est essentiel de dire les choses, aussi évident que cela soit-il. Dans notre cas ils nous a fallu plusieurs semaines pour mettre ds mots sur une idée partagée par tous intérieurement; la mauvaise communication avec l'encadrante. Mettre des mots sur ce soucis a permis de débloquer l'avancée du projet et d'améliorer l'atmosphère globale du groupe.

3. Démarche et conception

3.1. Sous-groupe 1 (language grounding)

3.1.1. Contexte et objectif du sous-groupe 1

Pour notre projet, notre système est entièrement basé sur du code, nous avons utilisé nos ordinateurs personnels, avec Vscode (langage de programmation) ainsi que de données pour la scène 3D (ScanNet) et pour la modélisation du body (Humanize). Par conséquent, nous avons pas utilisé de composant ou d'architecture particulière dans notre projet. Nos protocoles sont essentiellement sur notre manière d'approcher un problème afin d'y trouver une solution. De manière général pour le sous-groupe 1 (analyse de texte et scène 3D), la démarche était d'abord de comprendre l'objectif, puis de se renseigner sur les diverses possibilités (essentiellement grâce aux github donnée par l'enseignante-encadrante et quelques recherches internet) et d'en parler entre-nous pour décider de la meilleur solution à prendre face à notre problème rencontré ou à notre objectif. Cela passait souvent par un schéma visuel afin de mieux appréhender le point. Enfin, nous nous répartissions les différentes étapes du process et nous aidions ceux qui bloquait en parallèle ou bien une fois que le travail personnel était réalisé.

Pour rappel lors de notre dernier rendu, le sous-groupe 1 (basé sur l'analyse de texte et la gestion de la scène 3D) avait un code capable de trouver les target object et les anchor object (permettant de mieux situer la cible à partir du texte). Puis à partir de ça, à récupérer les segments des différents objets, c'est-à-dire les ensembles de points qui composent un meuble. L'étape suivante était d'ensuite d'obtenir les coordonnées de ces points puis de faire le trajet du body dans la scène 3D.

3.1.2. Avancées

Pour ce qui est des avancées qui ont été réalisées depuis, l'étape suivante à été accompli puisque nous avons pu acquérir les coordonnées des points qui compose un objet. Cela nous as permis de définir les objets bien plus précisément (connaître ses contours et rebords), notamment par l'ajout d'un centre géométrique calculé via ses coordonnées. Ce qui nous permet de donner une direction général vers la cible. Nous avons ensuite pu avancer vers la création du chemin pour le body. Pour modéliser la trajectoire, nous devions donc faire une liste de vecteurs qui nous permettrait d'aller du point de départ jusqu'à l'objet cible, tout en évitant les obstacles (les autres objets de la scène 3D), pour ne pas avoir un body qui traverserait des meubles sur son passage. Cette étape nous as beaucoup bloquée surtout sur l'objectif d'éviter les autres objets. En effet, la scène 3D comportait énormément d'objets plus ou moins grand et disposés un peu partout. Même avec les coordonnées de chaque objets, cela faisait énormément de données à traiter et à analyser pour un seul chemin, sachant que selon le texte entré, le chemin pouvait changer totalement se qui compliquait davantage un objectif déjà très compliqué.

3.1.3. modification des objectifs

Devant cet obstacle, les problèmes rencontrés par l'autre sous-groupe et avec le peu de temps qu'il nous restait, notre enseignante-encadrante à décider de revoir les objectifs. Pour la partie du sous-groupe 1 (analyse de texte et scène 3D), les points ont été grandement modifié. En effet, le projet ne se déroulera plus dans une scène 3D et le chemin sera donné par le mouvement prédéfini du body. Cela veut dire qu'il n'y a plus d'objet, et notre nouvel objectif était de créer les objets cités dans le texte sous forme de points qui serait sur le chemin du body, ainsi que renvoyer plusieurs informations qui seront données au deuxième sous-groupe (modélisation des mouvements du body) à savoir : le nom de l'objet, sa position, les vecteurs déplacements jusqu'au prochain objet et le temps (en nombre de d'image) que ce chemin prendrait. Pour y arriver nous avons tout d'abord, modifier notre code afin que ce dernier puisse trouver plusieurs objets cibles. Ensuite, nous avons récupérer les points de la trajectoire, pour créer des emplacements qui seront attribué au objets, cela permet de donner les coordonnées, les vecteurs et le temps à un objet, sans même en connaître la nature pour avoir un process plus flexible. Puis nous avons récupérer les vecteurs entre chaque points pour faire le chemin du départ à l'arrivée tout en passant par chacun des objets. Nous avons calculé le temps initial et final en fonction de la distance, et ceux pour chaque déplacement. Enfin, nous avons mis

toutes ces informations sous la forme d'un Json, composé de listes (pour accueillir tous les objets). De cette façon, le deuxième sous-groupe disposait des informations nécessaire pour faire les mouvements du body correctement.

3.2. Sous-groupe 2 (object-centric-motion-generation)

3.2.1. Contexte et objectif du sous-groupe 2

Notre sous-groupe a été chargé de concevoir et de mettre en œuvre la génération du mouvement en 3D d'un corps humain dans une scène fournie par un autre sous-groupe du projet. Cette scène contenait des objets et une trajectoire imposée que notre animation devait respecter. L'objectif principal était de produire une animation réaliste d'un personnage humain en mouvement, conscient de son environnement et capable de se déplacer selon un chemin défini.

3.2.2. Choix de l'environnement de développement

Nous avons mené notre travail principalement sur nos ordinateurs personnels, en utilisant Visual Studio Code et un environnement de développement basé sur WSL (Windows Subsystem for Linux). Ce choix s'est imposé après avoir rencontré plusieurs problèmes de compatibilité sous Windows, notamment avec les bibliothèques essentielles Torch et PyTorch3D. Le passage à un environnement Linux nous a offert une meilleure stabilité ainsi qu'un accès complet aux fonctionnalités de ces bibliothèques.

3.2.3. Étude de la base de données AMASS

La première phase du projet a été consacrée à l'analyse du support technique fourni par notre encadrante. Cette étape a permis de bien cerner les objectifs, planifier les tâches et installer les bibliothèques nécessaires. Nous avons ensuite exploré la base de données AMASS, et plus précisément son sous-ensemble CMU, afin de générer nos premiers maillages du corps humain.

Cette étude a nécessité une compréhension approfondie de certaines variables clés : betas (paramètres décrivant la morphologie du corps), verts (sommets du maillage 3D), et joints (positions des articulations). Ces expérimentations nous ont permis d'obtenir un maillage 3D réaliste.

Real Mesh

0.5 0.0 -0.5 -1.0 -1.0 -0.5 0.0 0.5 0.5

Figure 1 – visualisation d'un Human Body

3.2.4. Difficultés rencontrées avec la base Humanize

À la suite d'une proposition de notre encadrante, nous avons tenté de travailler avec la base de données Humanize. Toutefois, cette tentative s'est révélée peu fructueuse : la structure de la base était peu claire, et ses objectifs mal définis. Malgré trois semaines d'efforts, nous n'avons pas obtenu de résultats exploitables. Après concertation, nous avons décidé de réorienter notre projet vers des objectifs plus réalistes compte tenu du temps restant.

3.2.5. Redéfinition des axes de travail

Nous avons alors défini trois axes prioritaires :

- Génération de mouvements de translation suivant une trajectoire définie,
- Prise en compte des obstacles et objets présents dans l'environnement,
- Utilisation d'un modèle GRU (Gated Recurrent Unit) pour prédire ou simuler le mouvement.

3.2.6. Génération de translations et visualisation

Nous avons commencé par générer des mouvements de translation simples en utilisant les paramètres trans de la base CMU, représentant la position globale du personnage. Pour visualiser ces mouvements, nous avons développé une fonction Python visualize_seq_with_translation(verts, faces, trans) permettant d'animer le personnage image par image dans un notebook Jupyter.

Initialement, nous avons tenté d'adapter dynamiquement les axes de visualisation à chaque image afin de simuler un effet de suivi. Cependant, cette solution produisait un rendu peu naturel. Nous avons donc opté pour des axes fixes, ajustés à l'échelle maximale des coordonnées, afin de conserver une vue stable et cohérente de l'animation.

3.2.7. Utilisation du modèle SMPL-X

Sur le plan technique, nous avons utilisé le modèle SMPL-X (Skinned Multi-Person Linear Model – eXtended), une extension du modèle SMPL, intégrant la morphologie, les articulations et les expressions faciales. Nous avons utilisé la version masculine de ce modèle, disponible sous le fichier SMPLX_MALE.npz.

Les entrées du modèle SMPL-X incluent :

- pose_body : rotations articulaires (sous forme de matrices de rotation),
- betas : morphologie du corps,
- trans: position globale dans l'espace.

Le modèle génère les coordonnées des sommets (verts) et des articulations (joints), permettant ainsi de construire un maillage animé en 3D.

3.2.8. Traitement des données et animation finale

Les données exploitées proviennent de fichiers .npz issus de la CMU Motion Capture Database. Chaque fichier contient :

- les poses (poses),
- les paramètres morphologiques (betas),
- les translations globales (trans).

À chaque image, nous avons converti les poses en matrices de rotation, puis utilisé la fonction 1bs (Linear Blend Skinning) pour générer le maillage final en tenant compte de la translation.

3.2.9. Méthodes de visualisation et validation des résultats

La visualisation a été assurée via la bibliothèque matplotlib, plus précisément à l'aide du module FuncAnimation. Deux modes d'affichage ont été développés :

- un mode en nuage de points (affichage des vertices),
- un mode maillage triangulé (affichage de la surface du modèle).

L'analyse des trajectoires a confirmé la cohérence et le réalisme des translations extraites. L'animation obtenue montre une continuité fluide du mouvement, avec un déplacement harmonieux du personnage dans l'espace 3D.

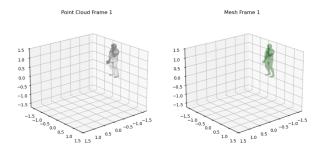


Figure 2 – le corps avant Translation

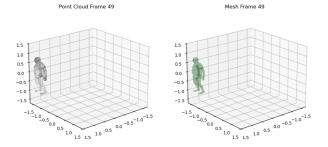


Figure 3 – le corps aprés Translation

3.2.10. Intégration d'un modèle GRU pour la génération de poses

Dans la continuité de nos objectifs révisés, et plus particulièrement pour répondre à l'axe concernant l'utilisation d'un modèle GRU (Gated Recurrent Unit), nous avons mis en place un système de prédiction automatique de mouvements corporels à partir d'informations de déplacement. L'objectif de cette approche était de générer des séquences de poses articulaires 3D cohérentes à partir de vecteurs décrivant le mouvement ou l'intention du personnage dans la scène.

Préparation des données d'entrée :

Les données exploitées proviennent de la base CMU, déjà utilisée pour la visualisation des maillages. Les vecteurs que nous avons nommés vecteurs représentent, pour chaque segment de trajectoire, une information directionnelle (par exemple, un vecteur de déplacement). À ces vecteurs sont associées des séquences de poses articulaires extraites des fichiers .npz, sous forme de matrices de taille ($N_{\rm frames}$, 66), où 66 correspond au nombre de paramètres décrivant les rotations articulaires du modèle SMPL-X.

Afin de rendre ces données exploitables par un réseau récurrent, nous avons segmenté puis restructuré les séquences de poses en blocs de taille uniforme, chacun correspondant à une portion de trajectoire. Cette étape a permis d'associer chaque vecteur d'entrée à une séquence de poses correspondante, sur laquelle le modèle pouvait apprendre une correspondance temporelle.

Conception du modèle GRU:

Le cœur de notre système est un modèle GRU conçu à l'aide de la bibliothèque PyTorch. Ce modèle, défini dans une classe personnalisée myGRU, est constitué :

- de deux couches GRU empilées, capables de traiter des séquences temporelles et de conserver la mémoire des états précédents;
- d'une couche fully-connected (linéaire) projetant les sorties du GRU vers un espace de dimension 66, correspondant à une pose complète du corps humain.

Le modèle reçoit en entrée un tenseur de forme (batch_size, sequence_length, 3) — chaque vecteur ayant trois composantes spatiales — et retourne un tenseur de forme (batch_size, sequence_length, 66), c'est-à-dire une séquence prédite de poses articulaires.

Le modèle a été entraîné sur GPU lorsque disponible, ce qui a permis d'accélérer les calculs.

Chaque itération d'entraînement comportait les étapes suivantes :

- 1. Chargement d'un batch de vecteurs d'entrée et de séquences de poses cibles;
- 2. Prédiction des poses à l'aide du modèle GRU;
- 3. Calcul de la perte (erreur quadratique moyenne) entre les prédictions et les poses attendues ;
- 4. Rétropropagation du gradient;
- 5. Mise à jour des poids du modèle.

Les pertes ont été enregistrées à chaque étape et ont montré une diminution progressive au fil des époques, signe d'un apprentissage efficace.

4. Résultats (Validation, discussion)

4.1. Cloture du projet

4.1.1. Tableau des livrables

Nom du livrable	Description	Format	Date de rendu
Cadrage du projet	Ce document a permis de saisir en détail l'état initial du projet et les sujets scientifiques et techniques qu'il couvre. De plus, c'est dans ce document qu'ont été donnés l'organisation initialement prévue du groupe et les échéances.	PDF	27/02/2025
Revue de projet et d'équipe	Le document permettait de mettre au courant notre professeur encadrant de nos avancements au niveau organisationnel et technique. De plus, il comportait nos perspectives à court-terme sur le projet pour la fin de ce dernier.	Présentation PDF	25/03/2025
Analyse de la dynamique de l'équipe	L'objectif principal du document était de fournir une analyse du fonctionnement de notre groupe au niveau de l'organisation et des problèmes rencontrés dans ce domaine.	PDF	15/04/2025
Rapport du projet	Il s'agit du document actuel, il présente globalement le projet en s'intéressant à tous les aspects de ce dernier.	Rapport PDF	20/05/2025

Table 1 – Tableau des livrables du projet

4.1.2. Validation:

Sous-groupe 1:

Dans la phase initiale du projet, le sous-groupe 1 avait pour mission de construire une chaîne complète partant de la compréhension d'un texte jusqu'à la génération d'une trajectoire 3D réaliste pour un corps humain dans une scène extraite de la base de données ScanNet. Le système devait identifier les objets cités dans le texte, les classifier comme anchor objects ou target objects, les localiser précisément dans la scène 3D, calculer leur centre géométrique, puis générer un chemin cohérent permettant au body de se déplacer jusqu'à la cible tout en évitant les obstacles présents dans l'environnement.

Cependant, cette approche s'est révélée trop complexe à mettre en œuvre dans le temps imparti, en particulier à cause de la densité d'objets dans les scènes 3D et de la difficulté à générer dynamiquement des trajectoires évitant les collisions. Face à ces difficultés, la professeure encadrante a décidé de revoir les objectifs pour recentrer le projet sur un cadre plus réalisable.

Dès lors, la scène 3D réaliste a été remplacée par un chemin prédéfini, le body ne traversant plus un environnement complexe mais suivant une trajectoire fixe. Le rôle du sous-groupe 1 a alors consisté à extraire plusieurs objets à partir du texte, et à associer à chacun d'eux une position sur le chemin, un vecteur de déplacement menant à l'objet suivant, et une estimation de la durée (en nombre d'images) de chaque déplacement. Toutes ces données ont été regroupées dans un fichier JSON structuré, destiné au sous-groupe 2, chargé d'animer les mouvements du body via la bibliothèque Humanize.

Afin de vérifier l'exactitude du programme remanié, le sous-groupe 1 a décidé de tester ce dernier sur des scénarii simples. Ainsi les différentes parties à valider étaient les suivantes :

- Les données du texte devaient être correctement extraites.
- La génération de points et de vecteurs représentant les objets devait être valide tout au long du trajet du body.
- Le temps calculé entre chaque objet devait être réaliste.
- Le fichier .json devait rester exploitable pour l'autre sous-groupe et fidèle aux données.

Le scénario utilisé pour tester le code est le suivants :

• "The boy goes to the sofa, then he sits on the chair and finally he lays on the bed." Dans ce scénario le pogramme renvoie pour les target objets: "sofa, chair, bed". Ce qui valide l'identification faite par IA dans le prompt envoyé. De plus, on observe que ces objets sont bien implémenter dans le .json en tant que points représentant les objets.

Sous-groupe 2:

Initialement, le sous-groupe 2 devait parvenir à générer un corps 3D des mouvements humains réalistes associés suivant une séquences d'actions données, dans une scène 3D donnée. Le système devait donc être capable d'interagir avec les objets afin de réaliser des actions basiques (s'asseoir). Pour cela, nous avions une ébauche d'un projet réalisé par un ancien étudiant ayant travaillé sur le même projet. Nous avions la bibliothèque python SIMPLX pour pouvoir générer des corps et dans animations et la base de donnée Humanize comme base de données d'entraînement pour notre modèle.

Néanmoins, l'utilisation de la base de données HUMANIZE s'est révélée trop complexe à exploiter dans le temps imparti, notamment de par la quantité de ressources qui ont rendu son exploitation moins naturelle.

Nous avons donc redéfini les objectifs en utilisant une base de données plus petite (CMU), plus simple à maîtriser et implémenter dans notre projet. Cette dernière fut bien intégrée dans le modèle GRU qui a ensuite été mis en place pour la génération des mouvements. Nous avons aussi adapté ce modèle afin qu'il puisse fonctionner avec les données renvoyées par le sous-groupe 1.

Nous avons ensuite testé le modèle sur différents cas simples afin de vérifier si les résultats que l'on obtient sont cohérents et ceux attendus ou non.

- Nous avons d'abord tester la capacité du modèle à apprendre sur des données fictives et plus simples.
 Test face auquel le modèle a montré une capacité d'apprentissage.
- Nous avons ensuite testé le modèle sur des données tirées de la base de données CMU (1 animation). Test face auquel le modèle a montré une capacité d'apprentissage, mais aussi des limites dûes à la faible quantité de données d'entraînement.
- Enfin nous l'avons testé de nouveau sur la base de données CMU, cette fois-ci avec un format plus adapté au problème. Test face auquel le modèle a répondu avec succès puisqu'il par

4.1.3. Amélioration du système :

Sous-groupe 1:

Concernant ce sous-groupe, l'axe d'amélioration est évident, il faut repasser au système qui était envisagé précédemment au remaniement d'objectif fait en accord avec la professeure-encadrante. Ainsi, le travail de l'IA serait bien plus important et utile dans des situations professionnelle dans le domaine de l'animation. En effet, il est rare d'avoir une scène et un mouvement prédéfini et de seulement devoir ajouter des objets anchor ou target. Là où sera utile l'IA est lorsque les objets sont présents et le body doit les éviter pour rester un réaliste un maximum. Pour améliorer le système il faudrait donc en théorie poursuivre les recherches faites avant le changement d'objectif et voir s'il est possible de coordonner celle-ci avec l'autre sous-groupe.

Sous-groupe 2:

Le projet fut un succès puisque nous avons réussi à implémenter le modèle, le relier au travail réalisé par le second sous-groupe et lui faire apprendre ce qui était l'objectif redéfini. Dans les perspectives d'amélioration, nous pourrions augmenter la quantité de données d'entraînement avec l'utilisation de la base de données HUMANISE. Cela permettrait de rendre notre modèle plus performant mais prolongerait aussi la durée de son entraînement et pourrait demander des équipements plus adaptés que nos ordinateurs personnels. Nous pourrions aussi développer d'autres modèles (RNN) afin de comparer les résultats renvoyés par les deux modèles.

4.2. Retour d'expérience sur la gestion de projet :

4.2.1. Planification temporelle:

Revenons sur les diagrammes de Gantt qui sont un bon moyen d'observer la planification et la réalisation temporelle du projet. Voici donc nos deux diagrammes de Gantt réalisé au debut et au milieu du projet que nous allons analyser par la suite :

- Diagramme de Gantt initial
- Diagramme de Gantt mi-projet

Finalement, comme les objectifs ont été remaniés complètement, un diagramme de Gantt final n'aurait pas été dans la continuité des deux derniers ce qui aurait été insensé.

Toutefois, les deux diagrammes de Gantt et les objectifs finaux expliqués plus tôt dans ce rapport, permettent de mettre en valeur les écarts entre nos perspectives initiales et ce qui a vraiment été produit lors de ce projet. En effet, on observe un avancement lent, trop lent comparé à ce qui était prévu.

On peut notamment observé que certains objectifs qui étaient censés finir pour mi-mars, donc au moment de rendre le diagramme de milieu de projet, n'étaient pas finis. Notamment, l'objectif : "essayer de retrouver les coordonné d'un chemin a prendre dans une scene 3D predefinie a partir de prompt uttilisateur" n'était qu'à 70% de réalisation.

De plus la modification des objectifs finaux montrent également un retard vis-à-vis des objectifs initialement fixés mais montrent surtout la difficulté à leur sujet.

Pour conclure, en observant la planification temporelle, on observe que les objectifs n'ont pas su être atteints à temps et ont forcé une modification de ces derniers pour rendre le sujet possible. Malgré tout, le retard au niveau de ces objectifs montrent également une certaines difficulté globale assez élevée du sujet.

Il reste cependant important de revenir sur certaines erreurs qui int été commises lors du projet qui ont pu ralentir ce dernier, la difficulté du sujet n'excuse pas tout. Tout d'abord je pense que la communication, aussi bien organisée soit elle n'était pas optimal. Effectivement il devient vite difficile de devoir gérer plusieurs sous-groupes et les conversations avec la professeure en parallèle. Même si la division du travail était une force pour notre groupe, elle nous a éloigné de certains objectifs et sujets que nous ne pouvions pas comprendre car non assigné pour nous.

Finalement, on peut retenir plusieurs choses de ces écarts sur les manières dont les choses auraient pu mieux s'organiser. Déjà, il faut rappeler et garder en tête qu'avec plus de réunions et de séances tous réunis entre les séances dédiées au PRONTO auraient pu aider. Ceci peut être un point d'amélioration. Un autre se situe dans la communication, il aurait sûrement été mieux si nous avions plus communiqué avec la professeure qui ne manquait pas d'envie de nous aider. Lorsque certains d'entre nous avaient des questions pour cette dernières il n'était pas évident d'en parler. De plus le fait de se réunir plus souvent entre membres du groupe auraient pu dénouer les langues et faciliter la compréhension inter-groupes.

4.2.2. Gestion des risques :

Pour assurer la réussite de notre projet, un aspect important à prendre en compte est la gestion des risques. En effet, la majorité du travail se faisant de façon collaborative et via des outils informatiques, le risque de perdre le travail ou de mal gérer le partage des données est important. Pour cela, nous avons choisis d'utiliser différents outils collaboratif permettant de sauvegarder le travail. En effet, pour le sous groupe 1, nous avons créer un google-colab pour y placer l'ensemble de nos codes python ce qui permet, un travail plus collaboratif, et, de sauvegarder en continue le travail. Cela permet d'éviter de perdre les données en cas de malfonction d'un des ordinateur des membres du sous groupe. Le sous groupe 2 a choisi de partager leur travail sur un GitHub permettant là encore de faciliter la collaboration et d'éviter les risques de pertes de données. Ce choix est plus judicieux en comparaison à un google-colab car l'utilisation d'un Github permet encore d'éviter les erreurs en conservant toujours la dernière version non modifiée en mémoire alors que le google-colab modifie en temps réel les données. Outre les outils de partage, les membres de chaque sous groupe conservaient une version plus ou moins ancienne du travail dans leur mémoire interne ce qui rajoute un nouveau de sécurité quand à la perte de données.

De plus, un autre domaine de risque était le domaine pédagogique, dans le sens ou il était possible que nous ne comprenions pas tous les objectifs attendus. Pour pallier à cela, en fin de chaque séance, un retour du travail effectué était envoyé à l'encadrante pour nous assurer de la pertinence de notre travail. Ce temps d'échanges permettait de nous assurer que nos méthodes étaient justes et éviter des erreurs.

Toujours dans l'aspect organisation, la mauvaise communication était un autre risque de haute importance car il pouvait, à cause de notre division en 2 sous-groupes rendre flou la compréhension globale du projet par les membres. Notre communication n'a pas été parfait mais nous avons essayé de communiquer un maximum via whatsapp et se tenir informer de nos avancées.

Pour ce qui de l'aspect technique, nous avons pris l'habitude au fur et à mesure du projet de bien relire nos codes python en vérifiant bien les entrées et les résultats obtenus (en comparaison aux attentes). Cette

relecture constante permet d'éviter de baser le travail suivi sur des erreurs et de saboter le projet. Cette phase de relecture se faisait de préférence en groupe afin d'échanger sur la justesse du code et éviter toutes mauvaise interprétation.

5. Conclusion et Perspectives

Ce projet final est certes loin de ce que nous voulions atteindre au départ, mais il présente tout de même des points forts et des points faibles.

Point faible: les principaux points faibles de notre solution viennent du fait qu'il soit trop minimaliste par rapport à l'idée de départ. En effet, le déplacement, qui devait se faire dans une scène 3D fournie en entrée du programme, a été remplacé par un environnement matplotlib sans texture. Au lieu de se déplacer entre des objets placés dans une pièce réelle, avec des positions prédéfinies, nous avons dû placer artificiellement les objets sur le chemin pour aller de l'un à l'autre.

De plus, l'utilisation de la base de données Humanize Dataset nous aurait été très utile pour entraîner notre modèle. Cette base de données contient de nombreuses animations correspondant à différentes actions et grâce à elle il aurait été beaucoup plus facile de récupérer toutes les animations de marche par exemple. Dans notre projet final, nous avons dû récupérer manuellement les animations pour l'entraînement. Cela a limité fortement la quantité de données utilisées lors de l'entraînement du modèle, ce qui l'empêche d'apprendre correctement.

Point fort: nous avons tout de même réussi à atteindre certains objectifs importants du projet. Le sous-groupe 1, a réussi à récupérer une entrée textuelle, l'envoyer à ChatGPT et à récupérer les objets importants dans le texte. Le sous-groupe 2, a réussi à faire fonctionner la librairie SMPL-X avec la base de données CMU. Nous avons pu extraire les données [poses], [betas] et [trans], puis construire un GRU avec PyTorch et le faire fonctionner correctement.

Ces travaux nous ont permis de mieux comprendre les librairies et leur utilisation et d'acquérir de l'expérience, notamment sur la manière d'utiliser une base de données pour entraîner un modèle. Même si le résultat final est plus simple que ce que nous avions prévu au départ, l'objectif principal qui était d'apprendre et d'expérimenter a bien été atteint.

6. Annexe

6.1. Maxime Goarin : Décrivez une situation où vous avez pris le lead pendant ce projet.

Notre projet Pronto porte sur la modélisation de mouvements humains dans une scène 3D à partir d'un texte. Nous sommes divisés en deux groupes, l'un sur l'analyse du texte et la scène 3D, l'autre sur la modélisation des mouvements. Les deux sous-groupes travaillent en autonomie et ont peu d'échange entre-eux. Par conséquent, il est rare d'avoir des situations qui permettent de prendre le lead sur les deux groupes. Cependant, quelques occasions se sont présentées lors de la remise des livrables, où j'ai souvent pris le lead.

J'ai pris les devants quand le groupe manquait de motivation, j'ai fait la liste des différents objectifs à remplir pour les rendus au tableau. J'ai expliqué les attendus, proposé les personnes qui pouvaient être les plus adaptés à la tâche, en essayant de répartir de la façon la plus équitable possible. Nous avons ensuite discuté ensemble de mes propositions selon l'avis et l'envie de chacun, j'ai essayé de réguler la discussion pour que tout le monde puisse avoir son temps de parole afin de donner son opinion. Enfin, j'ai posé une deadline avant la date du rendu pour que tout le monde soit au clair sur son travail et sur le temps qu'il a pour le réaliser. Une fois le travail fait et que nous étions regroupés, j'ai fait le tour des tâches réalisées pour que tout le monde puisse faire des retours constructifs sur le travail effectué. Une fois que tout était mis en commun, nous passions à la mise au propre du rendu.

Les tâches distribuées ont toutes été réalisé à la date prévue, ce qui nous as permis d'être bien plus efficace sur les rendus et de pouvoir utiliser notre temps ensemble pour mettre en commun. Cela nous as

permis d'améliorer les différentes parties en nous regroupant dessus, pour faire un compte rendus plus clair et complet. Sans ça, nous aurions quand même rendu le rapport dans les temps mais nous aurions été beaucoup plus pressé et le rendu en aurait souffert.

6.2. Hanine GHARSALLI :Qu'avez-vous fait quand vous deviez motiver un groupe ou une personne dans ce projet?

Pour motiver le groupe tout au long du projet, j'ai essayé d'adopter une attitude bienveillante et proactive. Dès le début, j'ai proposé que l'on définisse ensemble des objectifs clairs et un planning de travail réaliste, afin que chacun sache ce qu'il avait à faire et puisse visualiser l'avancement global. Cela a permis de donner un cadre au travail et de responsabiliser chacun.

Au fil du projet, j'ai veillé à maintenir une bonne ambiance dans le groupe, en valorisant régulièrement les efforts de chacun. Par exemple, après l'accomplissement des tâches demandées, je prenais le temps de souligner ce qui avait bien été fait, même si ce n'était qu'un petit progrès. Ce type de reconnaissance contribue à renforcer la motivation et à créer un climat de confiance.

Lorsqu'une personne du groupe montrait moins d'implication ou semblait découragée, j'ai essayé de proposer des solutions adaptées : réorganiser certaines tâches, travailler en binôme pour ne pas laisser quelqu'un isolé, ou recentrer le travail sur des aspects qu'elle trouvait plus intéressants ou motivants.

Enfin, pour garder une dynamique d'équipe, j'ai aussi suggéré de faire des points réguliers, où chacun pouvait s'exprimer librement, que ce soit pour partager des idées ou soulever des problèmes. Cette écoute collective a renforcé l'engagement de chacun et nous a permis de rester soudés, même en cas de difficultés.

6.3. Edgar ROUSSEAU : Racontez une situation dans ce projet où vous étiez confronté à un problème qui avait plusieurs solutions possibles.

Lors de notre projet, la partie s'occupant de l'analyse du texte et de la scène 3D, dont je fais partie, a dû choisir une clé API d'une IA. API signifie en français interface de programmation d'application. Cette clé permet d'utiliser une IA dans des codes grâce à des fonctionnalités comme la compréhension d'images et de textes.

Ainsi, afin de choisir une clé API, nous avons utilisé openrouter.ai. Il s'agit d'un site Internet regroupant de nombreuses clés API permettant de générer des clés API personnelles pour utiliser des IA dans nos codes.

Le problème étant que sur ce site, diverses IA sont proposées possédant pour la plupart les mêmes fonctionnalités, notamment celles qui nous intéressent l'analyse et la reconnnaissance d'objets dans des images associées à des mots dans des textes. Donc plusieurs solutions étaient possibles. Il a donc fallu que nous fassions un choix et je m'en suis occupé. Pour ce faire, j'ai comparé les différentes intelligences artificielles en fonction de leur niveau d'entrainement. J'ai finalement opté pour l'IA d'OpenAI, qui était bien assez entrainé pour ce que nous souhaitons faire.

Pour résumer, le choix fait n'a pas eu une grande influence car comme plusieurs solutions étaient possibles, si nous n'arrivions pas avec l'une il y avait toujours de l'espoirs. Faire un choix de cette manière demande de prendre une forme de leadership certes, mais ne demande pas de prendre énormément de risques ce qui est une bonne chose pour s'initier à prendre des décisions et plus tard peut-être gérer une équipe.

6.4. Mohamed Aymane Moussadek : Dans ce projet, quand avez-vous pris un risque, fait une erreur ou échoué?

Au cours du projet nous avons du faire preuve d'ingéniosité pour dépasser des situations complexes. Cependant, et malgré un travail de réflexion poussé, dans certaines situations nous n'avons pas pu conclure à une méthode résolvant le problème. En effet, en guise d'exemple, pour le sous groupe 1, la consigne à extraire dans la description textuelle devait prendre la forme d'un chemin décrivant les déplacements du body dans la scène. Ce chemin doit être délivré sous forme d'une liste de vecteurs de déplacements du body dans la scène à chaque frame. Cependant, pour rendre le mouvement du body le plus naturel, ce dernier doit pouvoir suivre un chemin en évitant tout obstacle possible; là était le problème rencontré. Pour y répondre nous avions accès aux segments de coordonnées définissant les limites géométriques de chaque objet appartenant la scène. Il nous a fallu alors penser à une méthode permettant de calculer la distance du body à chaque objet de la scène à chaque frame et modifier le chemin lorsque le body rentre en contact avec un objet. Pour calculer la distance nous avons pensé à travailler avec le barycentre du body. Cependant, cette méthode n'est pas optimale car elle implique des situations où une partie du body risque de traverser un objet. On a alors pensé à utiliser les extrémités du body plutôt que le barycentre, cependant, pour notre part, nos connaissance des outils de simulation du body n'étaient pas suffisantes pour conclure. Finalement, ce problème a été résolu par intervention de l'encadrante qui a finalement décidé de simplifier le problème ce qui nous a permis d'avancer dans le projet. Cette étape du projet a été en partie un échec pour nous car nous n'avons pas pu conclure quand aux attentes initiales du projet. Cependant cette situation a été enrichissante pour nous. En effet, on s'est rendu compte qu'un échec (à première vue) lors d'une étape n'était pas synonyme d'échec de l'entièreté du projet et qu'il faut parfois envisager d'autres situations pour avancer dans le projet.

Pour ce qui est des risques pris par le groupe 1, nous n'avons pas eu souvent à en prendre. Cependant, l'analyse de la description textuelle nécessitant un modèle d'intelligence artificielle permettant de traiter le texte, le choix du modèle a été un risque pour nous. En effet, le choix du modèle définit en grande partie la performance de notre méthode et donc de nos résultats. Il existe une multitudes de modèles répondant aux attentes. Notre choix s'est basé sur les avis (trouvé en ligne) d'utilisateurs et aux correspondance aux attentes de la méthode. Choisir un modèle peu performant aurait été un frein énorme à notre projet. Fort heureusement, les modèles choisis se sont avérés suffisamment performants et nous ont permis d'obtenir des résultats cohérents.

6.5. François-Xavier Collot: Racontez une situation dans ce projet où vous étiez confronté à un problème qui avait plusieurs solutions possibles.

Dès le début du projet nous avons eu un souci avec l'utilisation de bibliothèques python. En effet, certaines bibliothèques n'étaient pas utilisables de façon native sur Windows. C'était un problème inattendu auquel nous avons été confrontés et le projet reposant sur cette bibliothèque, nous devions trouver une solution .

Face à cela, nous avons eu plusieurs choix : Installation d'une distribution Linux, utilisation de WSL ou utilisation d'une machine virtuelle. Dans ce cas-ci, le choix de la solution n'a aucune importance sur la suite du projet, donc chaque membre du sous-groupe avait la liberté de choisir une solution personnelle. Nous avons donc chacun déterminé la solution qui nous était le plus adaptée en fonction de ce que nous avions déjà ou des méthodes avec lesquelles nous étions confiants. Nous avons ainsi utilisé les trois différentes méthodes!

Puis, nous avons chacun eu accès à un environnement où la bibliothèque était disponible et native, nous permettant de pouvoir faire avancer le projet et cela, sans problème de compatibilité.

6.6. Kewan Marboeuf: Avez-vous pris des risques, fait des erreurs ou échoué?

Le déroulement du projet n'étant pas linéaire, il est naturel que nous ayons commis des erreurs et rencontré des échecs. Le principal risque dès le départ était de viser un projet trop ambitieux. En effet, nous avions réparti le travail en deux sous-groupes : l'un chargé de transformer une entrée textuelle en

prompt pour ChatGPT, l'autre de convertir ces données en un fichier JSON exploitable pour entraîner un modèle prédictif des positions humaines dans une scène 3D. La réussite des deux sous-groupes étant interdépendante, une erreur dans l'un d'eux pouvait avoir un effet boule de neige sur l'autre.

Je vais maintenant me concentrer sur le travail réalisé et les difficultés rencontrées au sein du sous-groupe dont je faisais partie, car il est plus facile pour moi d'en parler.

Le début du projet avançait bien. Nous avons d'abord installé les bibliothèques nécessaires pour générer un humain dans une scène 3D et appris à modifier leurs paramètres. La difficulté est apparue lors de l'intégration de la bibliothèque *Humanize Dataset*, qui répertorie différents déplacements humains et leurs "labels" associés. Nous comptions l'utiliser pour entraîner notre modèle uniquement sur la marche, afin de simplifier la tâche. Cependant, il n'était pas simple de comprendre ce qu'il fallait faire; nous nous sommes retrouvés devant un GitHub avec un *README* que l'on peinait à comprendre.

En regardant en arrière, je peux dire que le plus gros problème était que nous ne nous étions pas bien concertés avec la professeure encadrante sur ce que nous devions faire exactement.

Cette incompréhension a entraîné un retard d'environ un mois, car nous tentions de faire fonctionner une bibliothèque dont on ne comprenait pas l'objectif. Ce retard a obligé tout le groupe, y compris l'autre sous-groupe, à revoir les objectifs initiaux et à abandonner certaines parties du projet, même celles qui fonctionnaient.

7. Bibliographie

Les sources utilisés sont :

-Le GitHub donné par notre professeure encadrante : Generating Human Motion in 3D Scenes from Text Descriptions, 2024, CVPR

https://github.com/zju3dv/text_scene_motion

-l'article : Generating Human Motion in 3D Scenes from Text Descriptions, 2024, CVPR https://zju3dv.github.io/text_scene_motion/

-Notre Github avec tous nos codes commentés : https://github.com/Hanine22/PRONTO

-OpenrouterIA qui nous as permis d'utilisé notre IA pour le texte :

https://openrouter.ai/cognitivecomputations/dolphin3.0-r1-mistral-24b:free

OUR WORLDWIDE PARTNERS UNIVERSITIES - DOUBLE DEGREE AGREEMENTS





3 CAMPUS

IMT Atlantique Bretagne-Pays de la Loire - http://www.imt-atlantique.fr/

Campus de Brest

Technopôle Brest-Iroise CS 83818 29238 Brest Cedex 3 France T +33 (0)2 29 00 11 11 F +33 (0)2 29 00 10 00

Campus de Nantes

4, rue Alfred Kastler CS 20722 44307 Nantes Cedex 3 France T +33 (0)2 51 85 81 00 F +33 (0)2 99 12 70 08

Campus de Rennes

2, rue de la Châtaigneraie CS 17607 35576 Cesson Sévigné Cedex France T +33 (0)2 99 12 70 00 F +33 (0)2 51 85 81 99









