

GUIDE

2024

Rédigé par
Vincent
Clement,
Marion
Cossin et
Isabelle
Domens

POUR UNE DÉMOCRATISATION
DE L'UTILISATION DE LA
CAPTURE DE MOUVEMENT
DANS LES ARTS VIVANTS



LES **7Doigts**

HUPR ✱

merci !

Ce guide n'aurait pu être réalisé sans le soutien et les contributions de nombreuses personnes et organisations. Nous tenons à exprimer notre gratitude aux expertes et experts qui ont généreusement offert de leur temps pour nous aider à monter ce document :

Chris Gatti et Pierre Schmidt de HUPR
Vincent Ladouceur des 7 Doigts de la main
Manuel Bolduc de la Société des arts technologiques
Raphael Dupont de Supply & Demand
Lisa Jamhoury
Topher Maraffi de NCSU MAD Tech
Ruby-Maude Rioux
Hugues Kir
Matthieu Thoer de Pink Cloud
Casey Côtes-Turpin de l'UQAT

Ainsi que tous ceux et celles qui n'ont pas désiré être nommés.

Nous souhaitons également reconnaître les Fonds de recherche du Québec (FRQ) pour le soutien financier accordé qui a permis de financer cette initiative et de soutenir la personne ayant travaillé à la réalisation de ce guide.



Genèse du document

La création de ce guide est née d'un besoin de la communauté des arts vivants de mieux comprendre les différentes technologies de capture de mouvement. HUPR et le LAB7 ont exploré un grand nombre de technologies de capture de mouvement, beaucoup par essai-erreur et en discutant avec d'autres professionnels du domaine ou d'autres domaines. Toutes ces explorations ont démontré le manque de document de ce type et l'envie de faire profiter à l'ensemble de la communauté des connaissances accumulées par nos équipes, mais aussi d'autres organisations. C'est dans ce contexte que HUPR et le LAB7 ont pris l'initiative de concevoir un guide pratique destiné aux artistes et créateurs en arts vivants. L'objectif est de fournir des solutions concrètes et des conseils adaptés aux réalités des arts vivants.

Le centre de recherche sur le potentiel humain (HUPR) développe des projets de recherche appliquée en collaboration avec de nombreux partenaires. Les travaux s'articulent autour de quatre sphères d'innovation : la performance humaine, le design d'équipement et d'environnement de performance, l'innovation sociale et les technologies numériques. HUPR offre également de l'aide technique, de l'information et des formations aux personnes qui œuvrent dans le milieu du cirque, des arts vivants et de la performance humaine.

Le LAB7, laboratoire technologique de la compagnie de cirque Les 7 doigts de la main, a pour mission d'explorer l'intégration des technologies de pointe aux réalités du spectacle vivant. Dédié au prototypage et à la recherche appliquée, le LAB7 propose une approche collaborative pour développer des langages artistiques novateurs et expérimenter de nouveaux processus créatifs. Ses activités comprennent l'accompagnement au prototypage et la médiation technologique : recherche appliquée, partage de connaissance, vitrine technologique.

METHODOLOGIE

1 - Recherche documentaire :

Une vaste revue de la littérature académique, non académique et technique a été réalisée pour identifier les pratiques actuelles et se familiariser avec les technologies de capture de mouvement.

2- Entretiens et questionnaires avec des experts :

Nous avons mené des entretiens et envoyé un questionnaire en ligne à remplir auprès de 14 expertes et experts ayant utilisé des technologies de capture de mouvement. Nous avons ciblé un large éventail de domaines (cinéma, biomécanique, arts vivants, arts numériques et géographique) afin de recueillir des perspectives pratiques et des retours d'expérience concrets.

3 - Analyse et rédaction :

Les données ont été analysées et synthétisées, pour structurer et rédiger les sections du guide.

4 - Validation :

Une première version du guide a été révisée par plusieurs experts afin de valider son contenu.



POINTS IMPORTANTS CONCERNANT CE DOCUMENT

Nous souhaitons mettre de l'avant que dans le contexte actuel dans lequel les technologies évoluent à un rythme effréné, il est fort probable que ce guide ne soit plus à jour passé sa date de publication (fin 2024). Nous vous conseillons donc de prendre les informations de ce document comme une photographie à un temps donné. Nous vous recommandons donc de pousser vos recherches si vous êtes intéressés par ce sujet.



Nous avons pris quelques libertés quant à la classification que nous proposons. Comme toute représentation, cette classification contient des limites et nous sommes bien conscient qu'il est possible de l'améliorer. Au-delà de cette classification, ce qui est important de retenir est que nous avons répertorié le plus de technologies possibles et essayer de les évaluer en fonction de certains critères.

Nous n'avons malheureusement pas pu incorporer toutes les technologies existantes. Cela peut avoir plusieurs raisons. Il est possible que certaines technologies nous aient paru moins pertinentes dans le contexte des arts vivants. Il est également possible que nous n'ayons pas pu obtenir assez d'information, par manque d'experts ou de réponse d'experts.



Dans ce document, le genre masculin est utilisé comme générique, dans le seul but de ne pas alourdir le texte.

En complément de ce guide axé sur les technologies de capture de mouvement, nous avons également publié un autre guide intitulé "Guide pour l'intégration des technologies numériques dans les arts vivants".

Table des matières

p. 01

Introduction

p. 05

Tableau récapitulatif

p. 07

Types de caméras

p. 20

Méthodes avec capteurs

p. 25

Méthodes avec marqueurs

p. 33

Vision par ordinateur et IA

p. 77

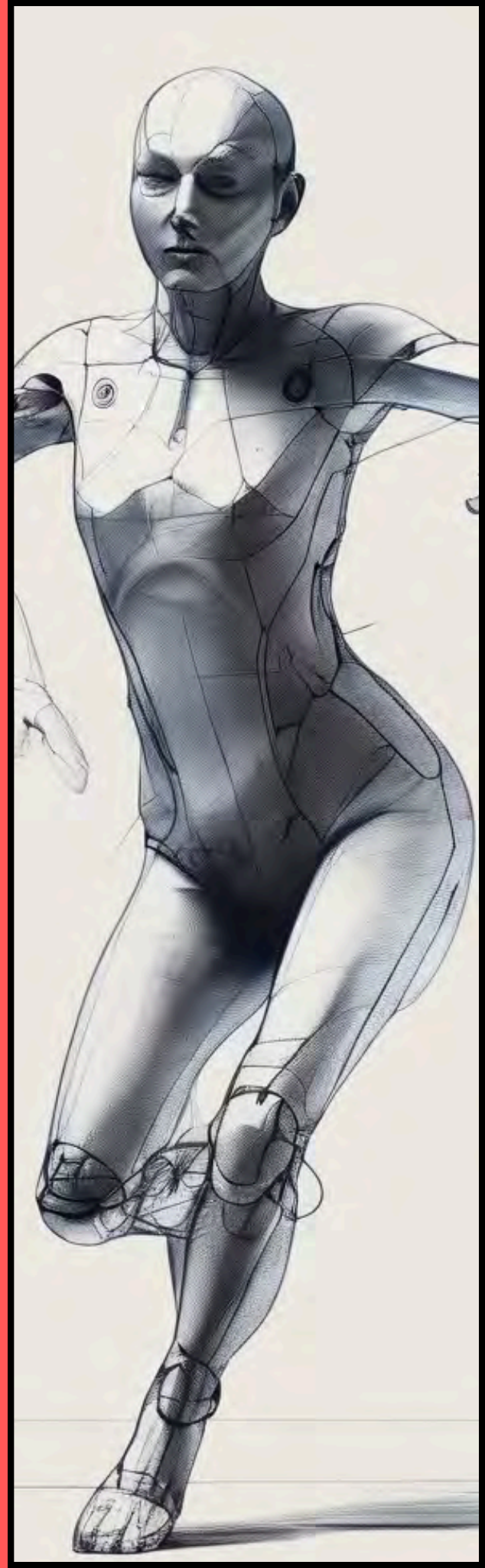
Autres méthodes

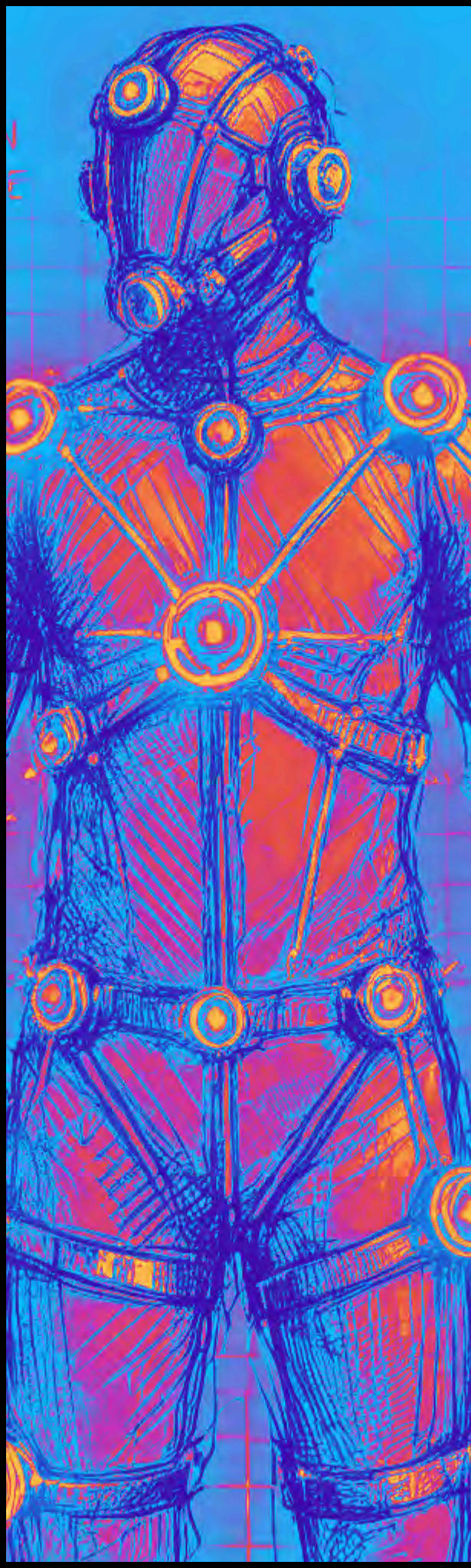
p. 84

Après la captation

p. 88

Conclusion





Liste des solutions dans le document

Caméra RGB	p. 09
Caméra Infrarouge	p. 11
Caméra à profondeur de champ	p. 13
Système de capture du mouvement inertiel	p. 20
Systèmes optiques avec marqueurs	p. 25
Système intégré haut de gamme	p. 36
Système intégré accessible	p. 38
Estimation de pose (Pose Estimation)	p. 41
Solution intégrée dans un SDK	p. 44
Solution à implémenter soi-même	p. 47
Estimation de pose par caméra RGB	p. 50
Estimation de pose par caméra infrarouge	p. 52
Estimation de pose par caméra de profondeur	p. 54
Algorithme de segmentation d'image/vidéo	p. 56
Soustraction de l'arrière-plan	p. 61
Algorithme de classification de mouvements/objets	p. 63
Algorithme de reconnaissance	p. 66
Systèmes volumétriques	p. 69
Nuage de points	p. 74
Face tracking	p. 77
Hand tracking	p. 80
Ultrasons	p. 82
Électromagnétique	p. 83

Introduction

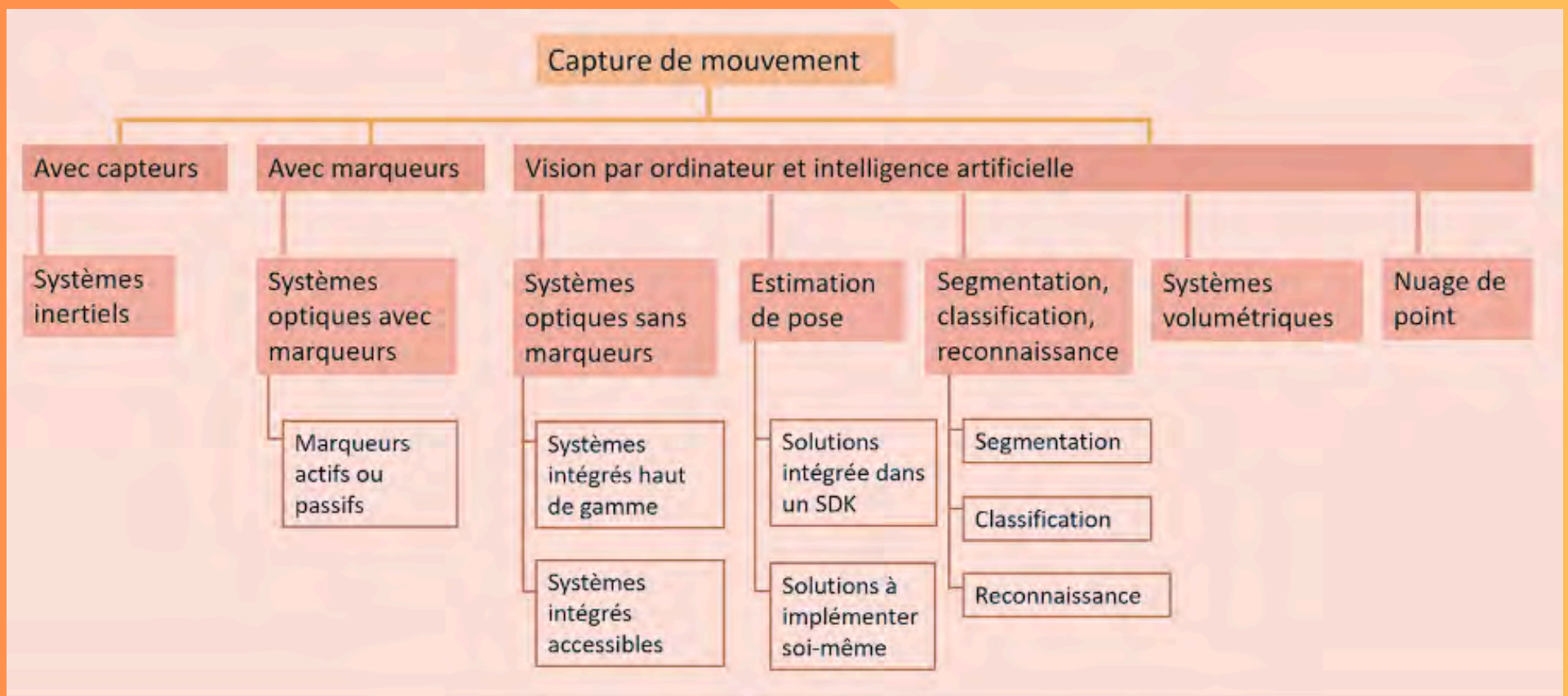
Les technologies de capture de mouvement ont un même but : capter, reproduire, et modéliser le mouvement d'une entité non statique.

Chacune des technologies de capture de mouvement utilise des méthodes différentes d'enregistrement du mouvement dont le but commun réside dans la précision de la captation.

Ces technologies permettent de transposer le réalisme du mouvement de la réalité physique sur une entité numérique à des fins aussi diverses que l'étude du mouvement ainsi décomposé ou l'animation d'un double fictif.

L'évolution technologique rapide que nous avons connue avec le numérique pose également la question de l'accessibilité de ces outils qui se montrent parfois difficile d'accès par leur prix, leurs complexités, ou tout simplement par leurs difficultés d'usage. Dans cette perspective, le présent guide a pour objectif d'aider les professionnels des arts vivants qui auraient l'intention d'étendre leurs pratiques à de nouveaux horizons artistiques, en offrant une cartographie des différents dispositifs de capture de mouvement. Étant donné la multiplicité des solutions à disposition, ce guide est pensé pour orienter les artistes et les créateurs dans le choix du dispositif adéquat qui saura concrétiser les velléités créatrices propre à chaque pratique.

Nous avons conçu ce guide à la fois comme un outil d'aide à la décision et comme un support pour mieux comprendre le fonctionnement de ces technologies. C'est pourquoi nous combinons des vues d'ensemble synthétiques avec des explications plus poussées. Ainsi, pour chaque technologie, nous en présenterons les points clés, les points positifs et les limitations, ainsi que des explications techniques plus poussées. Des tableaux comparatifs offriront également une vue d'ensemble qui devrait vous aider à choisir les outils les plus adaptés pour servir vos projets. N'hésitez pas à faire des allers-retours dans ce guide, à utiliser les sommaires pour cibler les technologies que vous voudrez explorer plus en détails.



Arbre des différentes familles de technologies de capture de mouvement

Dans cette optique, ce guide est organisé de telle sorte que vous puissiez naviguer entre les différentes composantes des dispositifs de capture de mouvement. Après avoir décrit les différents types de caméras disponibles sur le marché, nous allons distinguer trois grandes familles de capture de mouvement. Ces familles représentent des méthodes différentes de capture de mouvement : les méthodes avec capteurs, les méthodes avec marqueurs, les méthodes de vision par ordinateur et intelligence artificielle. Nous finirons par évoquer quelques autres systèmes, plus nichés ou aux usages plus spécifiques.

De manière plus formelle, on entendra la capture de mouvement comme « le processus d'enregistrement d'un événement de mouvement en direct et sa traduction en termes mathématiques utilisables en suivant un certain nombre de points clés dans l'espace au fil du temps et en les combinant pour obtenir une représentation unique en trois dimensions (3D) de la performance » (Menache, 2011).

L'ensemble des technologies de capture de mouvement sont composées d'un dispositif matériel (caméra, capteur, ordinateur...) et logiciel, dont le rôle est d'extraire les données brutes enregistrées afin de les traiter (étiqueter, filtrer...) ou les stocker en temps réel à partir d'une application de visualisation.

Comme vous le verrez, les possibilités de combinaisons matériels et logiciels sont nombreuses. Aussi, pour un projet de capture de mouvement optimale, nous vous conseillons de réfléchir en amont aux besoins de votre captation pour tirer parti des avantages et des inconvénients de chaque dispositif. Une scène avec beaucoup de mouvement en l'air par exemple, ne demandera pas les mêmes conditions de captation qu'une scène au sol plus statique. Le processus de capture de mouvement commence avec votre idée et termine avec les données capturées, mais entre les deux il y a une relation de cocréation et de concrétisation de votre idée par le dispositif choisi. Nous pensons que c'est en ayant connaissance des différents éléments des dispositifs de capture de mouvement que vous serez le plus à même de faire un choix éclairé.



OVERVIEW

OVERVIEW TABLE

OVERVIEW



TABLEAU

RECAPITULATIF

BLUISH

32

NEUTRAL

32

BLUISH

32

ORANGE

32

BLUISH

32

TABLEAU RÉCAPITULATIF DES FAMILLES DES TECHNOLOGIES DE CAPTURE DE MOUVEMENT

	1. Méthodes avec capteurs	2. Méthodes avec marqueurs	3. Vision par ordinateur et IA				
	1.1 Systèmes inertiels	2.1. Systèmes optiques avec marqueurs	3.1. Systèmes optiques sans marqueurs		3.2 Estimation de pose (Pose Estimation)		3.3 Segmentation, classification et reconnaissance
			3.1.1 Système intégré haut de gamme	3.1.2 Système intégré accessible	3.2.1 Solutions intégrées dans un SDK	3.2.2 Solutions à implémenter soi-même	
Exemple	Perception Neuron, Xsens, Rokoko	Vicon, Optitrack, Qualisys	Theia, Captury, Simi	MoveAI, RokokoStudio	SDK de la Kinect	OpenPose, MediaPipe	SegmentAnything, Mask R-CNN
Coût							
Capteurs/ Marqueurs							
Caméra							
Ordinateur nécessaire							
Facilité de prise en main							
Intérieur/ Extérieur							
Précision							
Capacité acrobatique							
Luminosité							
Nombre de personnes							

	3. Vision par ordinateur et IA		4. Autres méthodes			
	3.4 Systèmes volumétriques	3.5 Nuages de points	4.1 Face tracking	4.2 Hand tracking	4.3 Ultrasons	4.4. Méthodes électromagnétiques
Exemple	4DViews	Lidar	Rokoko's Face Capture, Optitrack	Rokoko Smart Gloves, Xsens	Chirp Microsystems, Pozyx	Polhemus, Ascension Technology
Coût						
Capteurs/ Marqueurs						
Caméra						
Ordinateur nécessaire						
Facilité de prise en main						
Intérieur/ Extérieur						
Précision						
Capacité acrobatique						
Luminosité						
Nombre de personnes						

LÉGENDE

Bas coût

Avec marqueurs

Pas de caméra

Basse Puissance

User Friendly

Haute précision

Moyen coût

Avec capteurs

Une caméra

Moyenne Puissance

Capacité de programmation nécessaire

Besoin de lumière

Haut coût

Sans capteurs ni marqueurs

Plusieurs caméras

Haute Puissance

Basse précision

Peu de participants

Intérieur

Extérieur

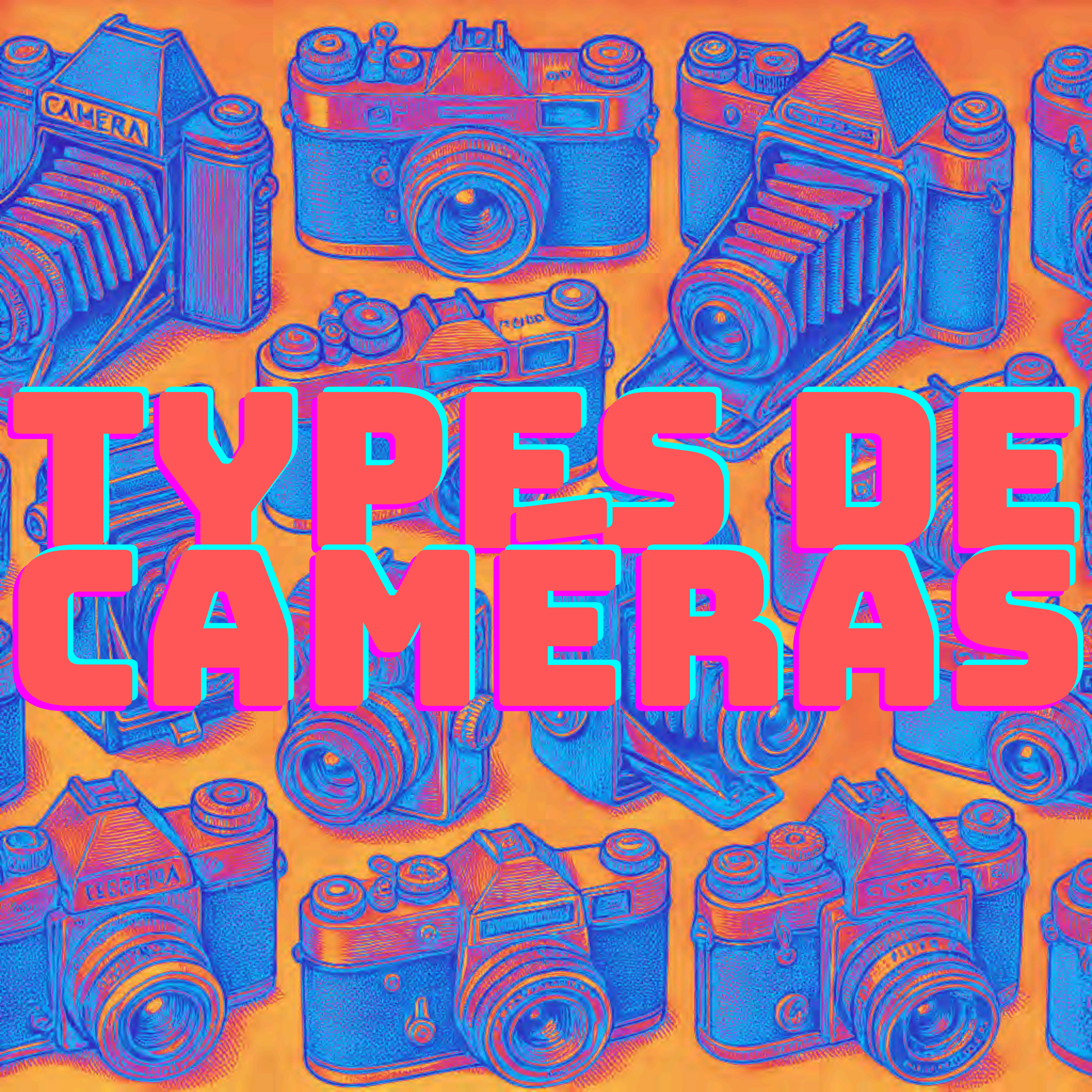
Bonne capacité de captation acrobatique

Capacité de captation acrobatique faible

Bonne précision

Multiple participants

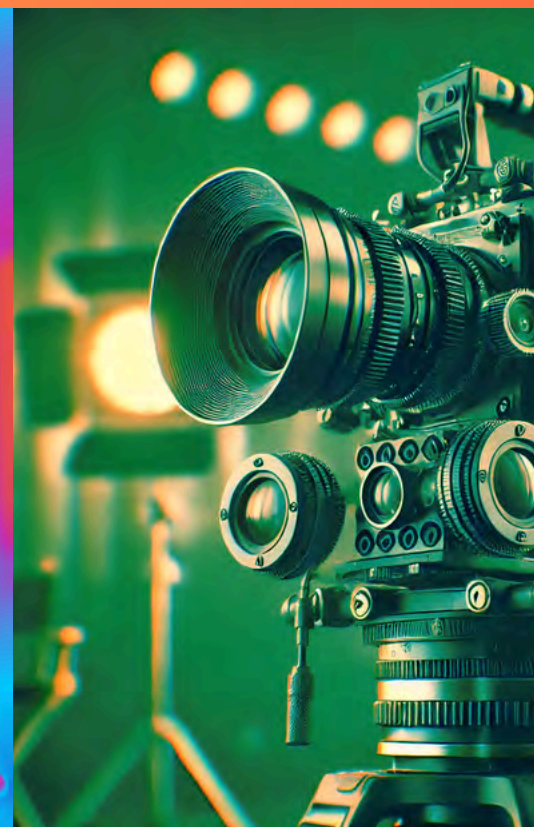
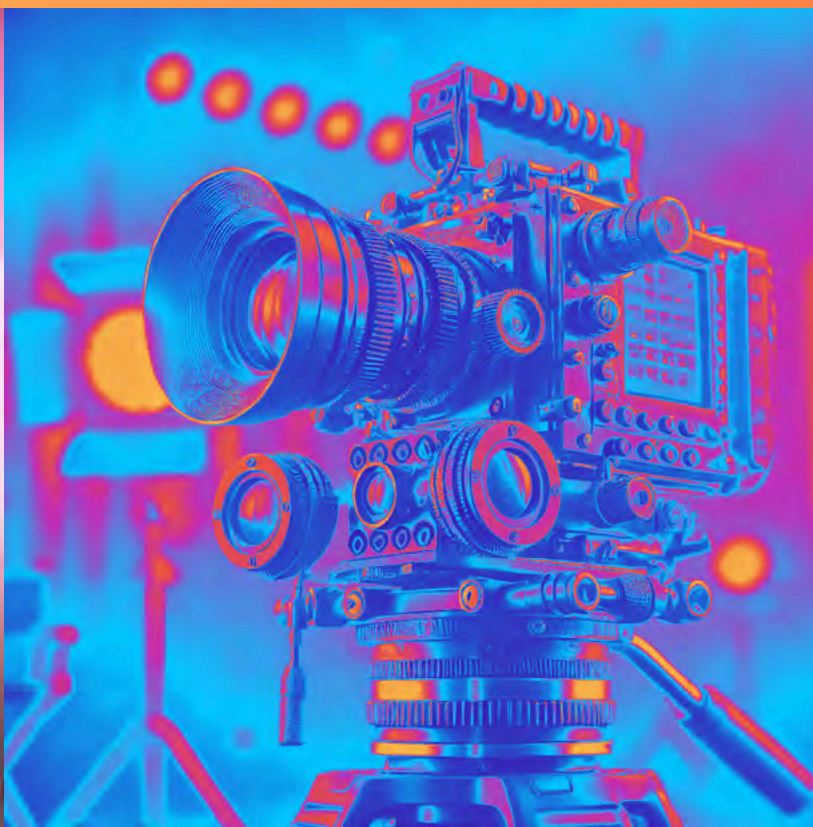
6



TYPES DE CAMERAS

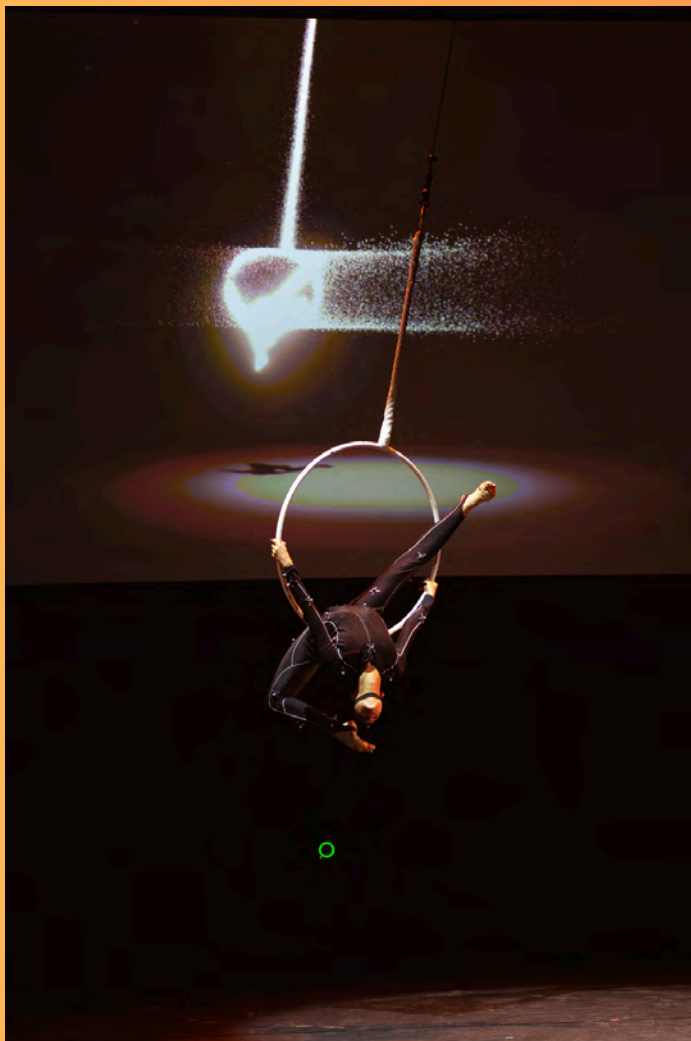
PRÉAMBULE

Nous allons présenter ici les types de caméras qui permettent une capture de mouvements sans marqueurs, sans nous préoccuper pour le moment du traitement du mouvement en tant que tel. Disons que nous allons nous intéresser ici à l'œil qui capte les données, plutôt qu'au cerveau pour les traiter. La partie logicielle complète les capacités des caméras dédiées : ce sont les logiciels qui vont permettre d'analyser et exploiter les données saisies. Nous les passerons en revue un peu plus loin.



Il existe une infinité de possibilités d'agencement logiciel et matériel des technologies de capture de mouvement sans marqueurs, selon le genre de caméra pour capter les données et le type de logiciel pour les traiter, chacun répondant à des objectifs différents. Nous verrons qu'il existe aussi des solutions qui intègrent les composantes logiciel et matériel dans une même solution, l'œil et le cerveau dans un même dispositif.

Nous allons explorer les différents types de caméras : les caméras RGB, les caméras infrarouges, et les caméras à profondeur de champ.



*Crédit photo :
Anaïs
Charbonneau-
Meilleur*

*Exemples de
performances
en capture de
mouvement.*

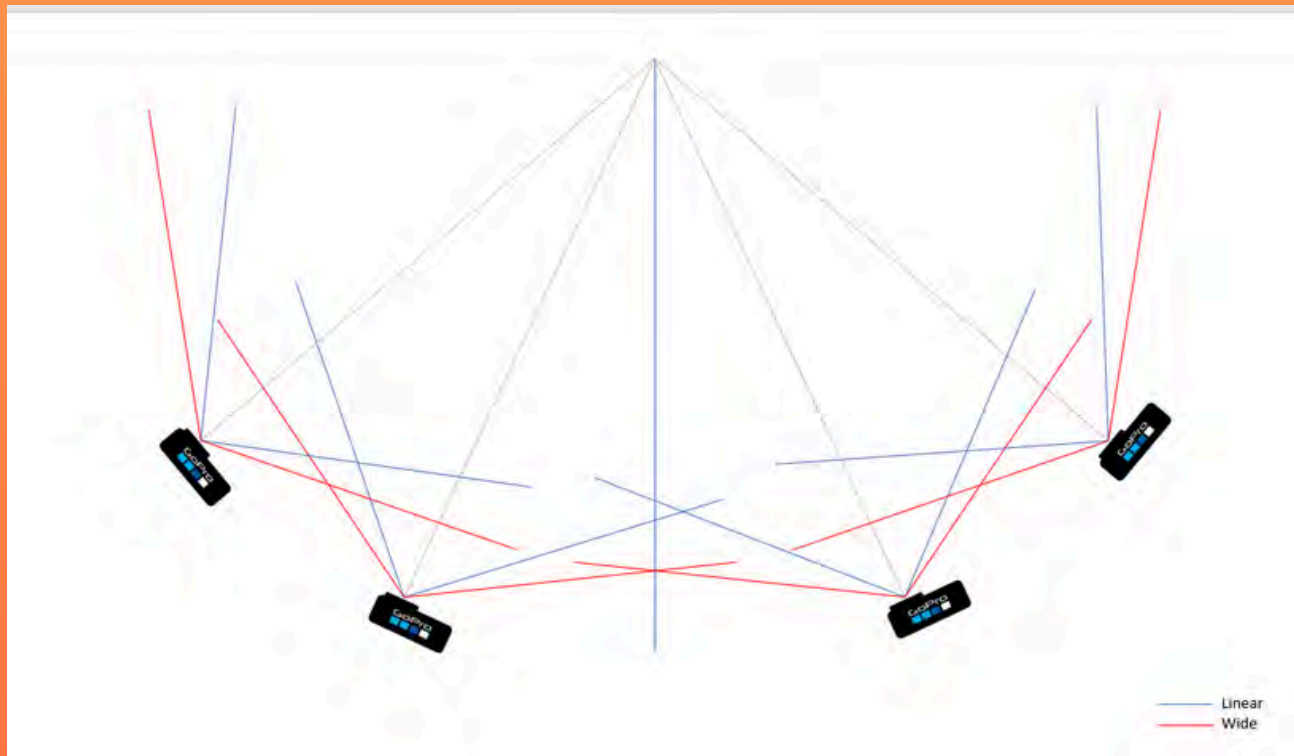
*Crédit photo :
Francisco Cruz,
Les 7 Doigts*



CAMÉRA RGB

POINTS À RETENIR

- Captation colorimétrique
- Image vidéo
- Compatible avec beaucoup de systèmes
- Possibilité d'une bonne résolution
- Sensible à la luminosité
- Absence de profondeur
- Possibilité de (très) hauts FPS
- Différentes gammes de prix (très économique à plus dispendieux)
- Sujet à la déformation radiale de la lentille, correction logicielle possible



Triangulation des caméras RGB sur la zone à capter

❁ PRINCIPE

C'est la caméra standard que l'on retrouve dans les smartphones, les webcams, les appareils photos numériques ou les caméras professionnelles. Une caméra RGB est utilisée pour fournir des images en couleur de personnes et d'objets en captant la lumière dans les longueurs d'onde rouge, verte et bleue (RGB pour Red, Blue, Green en anglais). Couplée à des algorithmes décrits à la section 2, elle peut permettre la reconnaissance faciale et corporelle. La fluidité de la captation est affectée par le FPS (Frame per Seconds) qui détermine le nombre d'images par secondes captées par la caméra. Un plus grand nombre équivaut à une acuité plus précise sur les captations faites.



Visualisation de la synchronisation de plusieurs GoPro
Crédit photo : Francisco Cruz, Les 7 Doigts

CAMÉRA INFRAROUGE

☀ POINTS À RETENIR

- Captation de lumière infrarouge
- Peut fonctionner dans l'obscurité totale si présence d'illuminateur infrarouge directement sur la caméra (ex. caméra de sécurité)
- Sensible à la lumière contenant de l'infrarouge (soleil, éclairages non-LED, couleurs chaudes...)
- Possibilité de visualiser l'émission de chaleur d'un objet ou d'une personne
- Sensible à la fumée
- Pour capter la réflexion de lumière infrarouge, un émetteur infrarouge est souvent nécessaire

☀ PRINCIPE

Une caméra infrarouge est un type de caméra qui capture les radiations infrarouges plutôt que la lumière visible. L'infrarouge fait partie du spectre électromagnétique et se situe juste au-delà de la lumière visible, ce qui permet à ces caméras de détecter la chaleur émise par des objets ou des êtres vivants, ainsi que la réflexion des ondes infrarouges. Les caméras infrarouges ne capturent pas les couleurs visibles, donc les images qu'elles produisent seulement des informations de réflectance des objets. Il existe également des filtres qui permettent de transformer une caméra RGB en caméra infrarouge (solution souvent beaucoup moins efficace que d'avoir la bonne caméra).

Les systèmes optiques avec marqueurs utilisent des caméras infrarouges avec des marqueurs réfléchissants mais il est possible d'utiliser ces caméras sans marqueurs. Les caméras infrarouges peuvent être utilisées en conjonction avec des algorithmes de vision par ordinateur et d'intelligence artificielle pour suivre les mouvements d'une personne (voir Section 2). Notons qu'un éclairage (illuminateur) infrarouge est nécessaire (et peut être pour certains systèmes déjà inclus dans la caméra), afin que la caméra puisse percevoir quelque chose.



Exemple d'une caméra infrarouge pour le système Optitrack.

Crédit photo : Francisco Cruz, Les 7 Doigts



CAMÉRA À PROFONDEUR DE CHAMP

Dans la famille des technologies sans marqueur, nous retrouvons aussi les caméras à profondeur de champs. Une caméra à profondeur de champ, également appelée caméra de profondeur ou *depth camera* en anglais, est un dispositif capable de capter la distance entre la caméra et les objets ou les personnes, en plus de l'image 2D. Dans beaucoup de cas, une caméra à profondeur de champ sera constituée d'une caméra RGB et ajoute un autre système, par exemple une caméra infrarouge ou une deuxième caméra RGB, pour capter la profondeur.

POINTS À RETENIR (CAMÉRA À PROFONDEUR)

- Captation de la profondeur
- Captation 3D : image 2D + profondeur
- Plusieurs technologies existantes : la lumière structurée (projection de motifs infrarouges), la stéréoscopie (deux caméras pour simuler la vision binoculaire) ou le temps de vol (TOF, Time of Flight, où le temps de retour de la lumière infrarouge est mesuré).
- Peut fonctionner dans diverses conditions d'éclairage, y compris dans des environnements sombres.
- Précision et portée qui varient selon les modèles
- La qualité des caméras à profondeur dépend largement des algorithmes utilisés pour traiter les données de profondeur. Un algorithme performant peut réduire le bruit et améliorer la clarté des données captées, mais des algorithmes moins robustes peuvent générer des artefacts (comme des flous ou distorsions) qui limitent la qualité des estimations de pose.

CAMÉRA RGB + INFRAROUGE

PRINCIPE

Pour trianguler le mouvement désiré, la sous-famille des caméras de profondeur, tel que la Kinect développée par Microsoft (et maintenant reprise par Orbbec), ne nécessite qu'une seule caméra grâce à la double action d'une caméra infrarouge, comme pour les systèmes optiques, et d'un projecteur de points infrarouges qui fonctionne selon le principe de la « lumière structurée » (la projection d'un motif connu sur la scène permettant de déduire la profondeur à partir de la déformation de ce motif) ou du calcul du temps de vol ou ToF (la mesure du temps nécessaire à un objet, une particule ou une onde pour parcourir une certaine distance). Les caméras à lumière structurée utilisent un émetteur infrarouge pour éclairer la scène avec un motif prédéfini de bandes, de barres ou même de points lumineux invisibles à l'œil nu. En observant les distorsions du motif réfléchi, ces caméras peuvent calculer la profondeur et les contours des objets de la scène. La première génération de Kinect utilise la lumière structurée (motifs infrarouges) tandis que la deuxième génération utilise la technologie ToF.

Les caméras à profondeur avec caméra infrarouge peuvent être affectées par des sources de chaleur ambiantes et ne pas fonctionner aussi bien en extérieur ou en plein soleil, car la lumière naturelle peut interférer avec les signaux infrarouges. Également, la distance constitue souvent une contrainte majeure, car son fonctionnement optimal se situe généralement entre 0,6 m et 5 m.

CAMÉRA CONSTITUÉE DE DEUX CAMÉRAS RGB (STÉRÉOVISION)

❁ PRINCIPE

Une autre catégorie de caméra de profondeur, comme la série « Zed » de la compagnie Stereolabs, fonctionne selon le principe de la vision stéréoscopique (ou stéréo) que l'on retrouve déjà dans la famille des caméras optiques avec marqueurs. Une caméra stéréo repose sur le même principe que le regard humain : la vision binoculaire. La vision binoculaire humaine utilise ce que l'on appelle la disparité stéréo pour mesurer la profondeur d'un objet. La disparité stéréo est la technique qui permet de mesurer la distance d'un objet en utilisant la différence entre l'emplacement d'un objet vu par deux capteurs ou caméras différents (les yeux dans le cas de l'homme). Dans le cas d'une caméra stéréo, la profondeur est calculée à l'aide d'un algorithme qui s'exécute sur l'ordinateur.



*Caméra à
profondeur
de champ,
ZED.
Crédit photo
:Francisco
Cruz, Les 7
Doigts*

Toutefois, pour que la caméra fonctionne efficacement, les deux images doivent avoir suffisamment de détails et de texture. En extérieur, les objets tels que les bâtiments, les arbres, les routes et les rochers fournissent souvent ces caractéristiques. C'est pourquoi les caméras stéréo sont adaptées pour les applications extérieures avec un grand champ de vision. Comme elles utilisent la lumière ambiante naturelle, ça leur permet de fonctionner sans problème en plein soleil ou dans des conditions bien éclairées. Un fond contrasté et simple (non chargé de détails ou d'objets) améliore la capacité de la caméra stéréo à différencier les objets et le sujet capté.

PRINCIPE

Quand on parle de LiDAR (Light Detection And Ranging), en réalité, cela désigne la technique de télédétection utilisant les faisceaux laser pour mesurer les distances, et cela englobe une partie matérielle et une partie logicielle. Elle utilise des impulsions laser pour mesurer la distance entre la caméra et l'environnement, en calculant le temps que met la lumière à revenir après avoir rebondi sur ces objets. Cela permet de créer une carte tridimensionnelle de l'environnement, souvent sous forme de nuage de points.

Les capteurs LiDAR peuvent être classés dans la catégorie des caméras à profondeur, mais elles se distinguent par leur technologie spécifique, qui est différente des caméras à profondeur de type infrarouge ou stéréoscopiques, par exemple. Nous évoquerons plus en détail cette technologie dans la section Nuage de points.

Le LiDAR 2D est une solution simple et économique pour des tâches où une vue plane (carte 2D) est suffisante, tandis que le LiDAR 3D est indispensable pour des applications exigeant une compréhension complète de l'environnement, avec une cartographie 3D. On peut donner comme exemple de modèle pour le LiDAR 2D le Hokuyo UTM-30LX. Pour la partie matérielle 3D, on peut nommer les capteurs Ouster OS1 ou le Velodyne Puck qui capturent les données brutes. La partie logicielle visualise et analyse les points dans l'espace pour en tirer des informations utiles. Certains fabricants, comme Velodyne et Ouster, fournissent des SDK (comme dans certaines caméras à profondeur) qui permettent de lire les données brutes directement depuis leurs capteurs et de les transformer en nuages de points exploitables.

AUTRES CAMÉRAS À PROFONDEUR DE CHAMP

Il existe aussi d'autres types de caméras à profondeur de champ, par exemple des caméras à lumière structurée ou à temps de vol (ToF) qui ne possèdent pas de capteur RGB. Ces dernières sont couramment utilisées dans des domaines où seule la profondeur est pertinente. Nous ne rentrerons pas dans le détail de ces technologies à l'intérieur de ce guide parce que ces technologies sont moins utilisées dans notre contexte.

Tableau récapitulatif des caméras à profondeur de champ

Type de caméra à profondeur	Exemples	Méthode utilisée	Avantages	Inconvénients
RGB + infrarouge	Kinect	RGB + lumière structurée ou ToF	Capture simultanée des couleurs et de la profondeur	Sensible aux interférences lumineuses
Stéréo	Zed	Vision stéréoscopique (2 caméras)	Indépendance de la lumière projetée	Dépend des textures dans la scène
LiDAR	Ouster OS1 ou Velodyne Puck	LiDAR (Light Detection and Ranging)	Idéal pour la cartographie 3D de précision	Coût élevé pour le LiDAR 3D
ToF	Basler ToF Camera	Temps de vol (signal infrarouge)	Spécialisé dans les données de profondeur	Moins précise sur les détails
Lumière structurée	Orbbec Astra S (versions sans capteur RGB)	Projection et analyse de motifs	Très précise en intérieur	Fonctionne mal en extérieur ou avec surfaces lisses

CORRESPONDANCE SYSTÈME DE CAPTURE DE MOUVEMENT ET CAMÉRAS					
	Caméras				
	RGB				
	Webcam et caméras USB (Logitech C920, Razer Kiyo Pro Ultra, ...)	Caméras vidéos standard (Canon XA11, Sony FDR-AX100, Panasonic HC-X2000,...)	Caméras d'action/compactes (GoPro HERO12 Black, Sony RX0 II, ...)	Caméra de smartphone et tablette (iPhone, iPad, Samsung, ...)	Caméras de vision industrielle (Basler ace 2, FLIR Blackfly S, IDS uEye CP, ...)
Systèmes					
Systèmes optiques avec marqueurs (Vicon, Optitrack, Qualisys, Motion Analysis, ...)					
Système intégré haut de gamme sans marqueur (Theia Markerless, Captury, Simi Motion Reality, SWRI ENABLE, ...)	4	4	4		4
Système intégré accessible sans marqueur (MoveAI, Rokoko Video, ...)	2	2	2	2	
Solution intégrée dans un SDK (ZED SDK, Kinect SDK, Orbbec SDK, ...)					
Modèle d'estimation de pose open source (OpenPose, MediaPipe, YOLO, VITPose, ...)	1	1	1	1	2
Algorithme de vision par ordinateur: segmentation, classification et reconnaissance (OpenCV, DeepLab, SVM, ...)	1	1	1	1	2
Systèmes volumétriques (Depthkit, 4Dviews, ...)					2
Nuage de point(Augmenta, PCL (Point Cloud Library),Open3D, ...)					
Système de création multimédia(Notch, VVVV, Touchdesigner, ...)	1	1	1		1

	Profondeur		Infrarouge	
	Caméra à profondeur(Intel RealSense D455, Azure Kinect, Orbbec, Zed2i, ...)	Caméra nuage de point(Velodyne Puck, Ouster OS1,...)	Caméra infrarouge associé à un système optique(caméra Vicon, caméra Optitrack, caméra Qualisys, ...)	Caméra infrarouge(FLIR Lepton, FLIR C5, Seek Thermal Compact, ...)
Systèmes				
Systèmes optiques avec marqueurs (Vicon, Optitrack, Qualisys, Motion Analysis, ...)			4	
Système intégré haut de gamme sans marqueur (Theia Markerless, Captury, Simi Motion Reality, SWRI ENABLE, ...)			6	
Système intégré accessible sans marqueur (MoveAI, Rokoko Video, ...)				
Solution intégrée dans un SDK (ZED SDK, Kinect SDK, Orbbec SDK, ...)	1			
Modèle d'estimation de pose open source (OpenPose, MediaPipe, YOLO, VITPose, ...)	1	1		1
Algorithme de vision par ordinateur: segmentation, classification et reconnaissance (OpenCV, DeepLab, SVM, ...)	1	1		1
Systèmes volumétriques(Depthkit, 4Dviews, ...)	2	4		
Nuage de point(Augmenta, PCL (Point Cloud Library),Open3D, ...)	1	1		1
Système de création multimédia(Notch, VVVV, Touchdesigner, ...)	1			1

Note pour le tableau : Le nombre de caméra dépend du volume à capter, du système et des ressources informatiques, c'est souvent plus de caméras qui est requis. C'est à valider avec les besoins. Toutes les informations sont à vérifier en ligne ou avec le fournisseur. Les fournisseurs donnent un nombre minimum de caméras à utiliser mais le nombre réel de caméras nécessaires dépend de la zone de capture (superficie mais aussi hauteur) et parfois du nombre de personnes que l'on veut capter. Quand on veut utiliser plus d'une caméra, il faut s'assurer qu'on a bien la possibilité de les connecter et synchroniser entre elles pour fusionner leurs données.



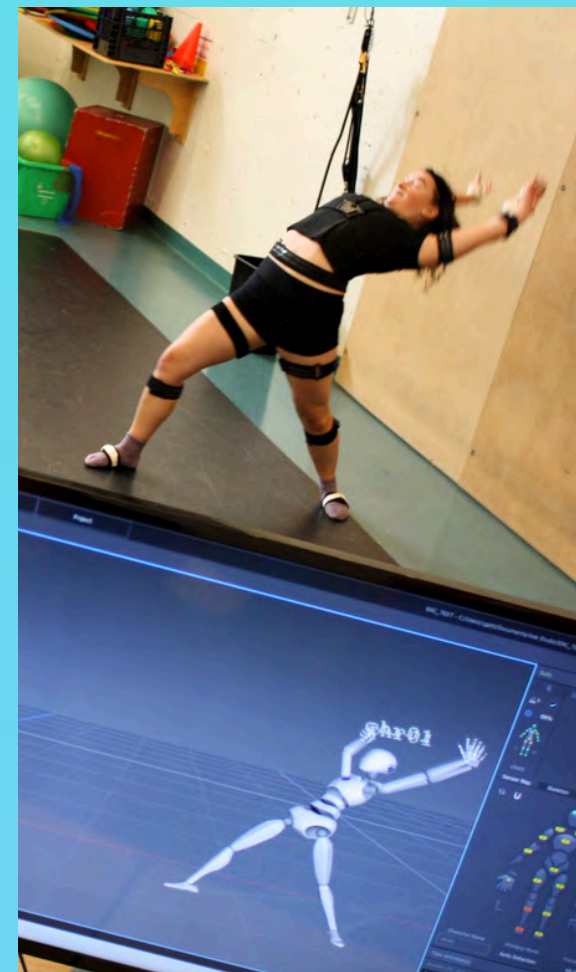
**MÉTHODES AVEC
CAPTEURS**

SYSTÈME DE CAPTURE DU MOUVEMENT INERTIEL

Les méthodes basées sur des capteurs utilisent, comme leur nom l'indique, des dispositifs (capteurs) portés directement par le sujet ou l'objet pour capter sa position et ses mouvements.

POINTS A RETENIR

- Exemples : Perception Neuron, Xsens, Rokoko
- Données de sortie : positions 3D des articulations, orientations des segments, données brutes (accélération et rotation)
- Capteurs sur le corps (en général au nombre de 17) fixés avec une combinaison ou des sangles
- Pas de caméra mais un récepteur branché à l'ordinateur
- Logiciel dédié et payant (parfois inclus dans le prix du système)
- Coût entre 3000\$ et 12000\$ pour une combinaison
- S'utilise autant à l'intérieur qu'à l'extérieur
- En général, fonctionne dans un rayon de 30 m ; certains systèmes peuvent aller jusqu'à 150 m
- Équipement informatique : un processeur (CPU) Intel Core i7 ou équivalent, un GPU NVIDIA GeForce GTX 1060 ou supérieure. Un SSD est recommandé pour des performances optimales lors de l'enregistrement et du traitement des données



Exemple de système inertiel
Crédit photo : Marion Cossin,
HUPR

☀ POINTS POSITIFS

- Bonne précision pour une application artistique, surtout sur les systèmes haut de gamme
- Possibilité de temps réel
- Possibilité d'être connecté avec d'autres logiciels
- Possibilité limitée de capter des objets, un capteur pourrait être fixé sur un objet
- La mise en place est assez simple et rapide
- Aucunement sensible aux occlusions puisque cela fonctionne sans caméra

☀ LIMITATIONS

- Présente une dérive importante (drift), notamment pour les disciplines aériennes ou les sauts. Certains systèmes peuvent avoir un dispositif comme le Coil Pro de Rokoko qui vise à améliorer la précision du suivi en fournissant un positionnement absolu.
- La calibration s'avère souvent difficile ce qui affecte considérablement la qualité du mouvement
- Le port de capteurs peut être gênant pour les artistes et interférer avec les accessoires/appareils. En particulier les systèmes filaires avec batteries et transmissions, qui limitent considérablement les mouvements
- La qualité et la précision de la capture sont souvent limitées aux mouvements en contact avec le sol
- Les méthodes de fixation des capteurs peuvent être améliorées car les capteurs peuvent se détacher lors de mouvements rapides. La sangle/ceinture pour le capteur au bas du dos peut bouger lors de mouvements acrobatiques et affecter la précision de l'avatar
- Très sensible aux interférences magnétiques

● PRINCIPE

La capture inertielle n'a besoin d'aucune caméra pour fonctionner, mais de capteurs qui mesurent la pose du corps humain en temps réel et un récepteur pour recevoir les données émises. L'absence de dispositif optique pour « voir » le mouvement et l'enregistrer est ce qui distingue cette famille de la capture de mouvements "optique" (qui utilise des caméras). À la place des caméras et des marqueurs, la technologie inertielle utilise des IMU (inertial measurement units) avec des capteurs intégrés dans une combinaison ou fixés par des sangles pour détecter la position et le mouvement dans l'espace et dans le temps. Il s'agit généralement de gyroscopes, d'accéléromètres et parfois de magnétomètres.

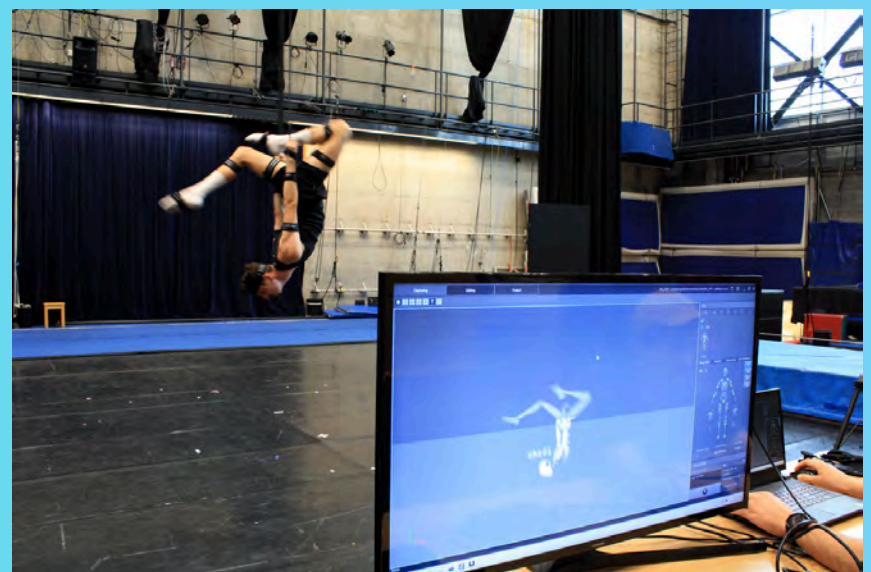
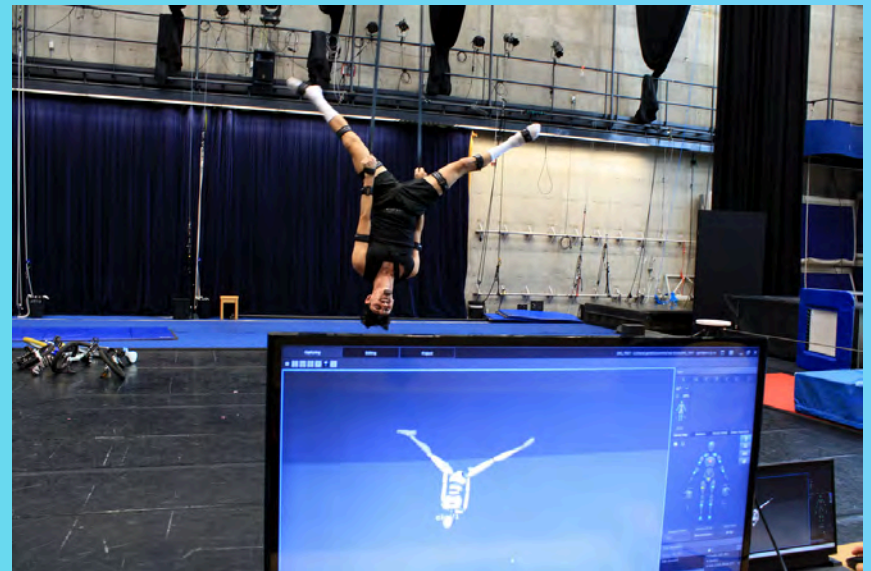
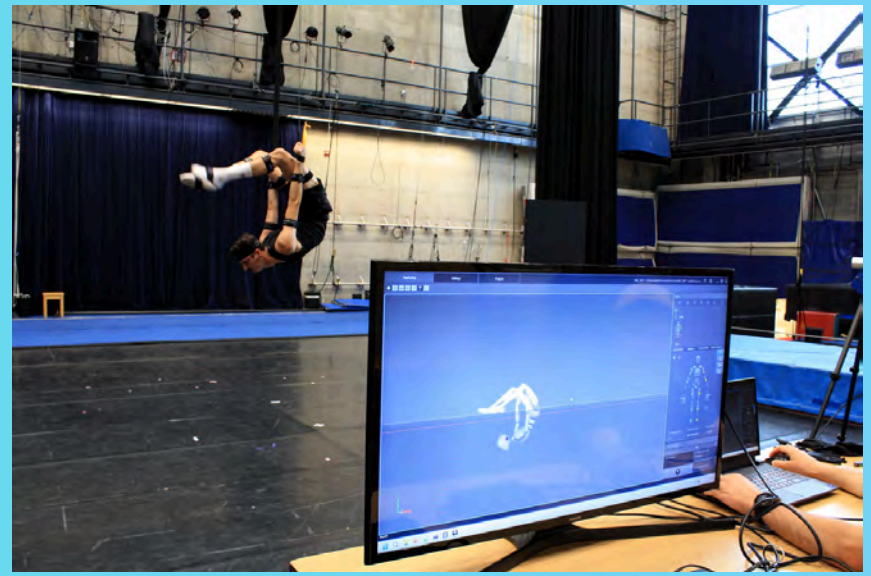
Le **gyroscope** mesure la vitesse angulaire. Il est utilisé pour déterminer l'orientation rotationnelle de l'unité de mesure inertielle. **L'accéléromètre** détecte l'accélération et la force gravitationnelle. Il est ensuite utilisé pour calculer le changement de position par rapport à la gravité (inclinaison) ainsi que le mouvement de l'IMU sous l'effet de l'accélération dans n'importe quelle direction. **Le magnétomètre** mesure le champ magnétique terrestre ou un champ magnétique créé artificiellement, il est utilisé pour orienter l'IMU. Chaque capteur a besoin d'une position de base à partir de laquelle il se déplace. Cela permet de traduire les mouvements détectés en données significatives (mouvement par rapport à une position dans l'espace).

La mobilité et la convivialité de la capture inertielle de mouvement figurent parmi ses principaux avantages. Elle ne nécessite pas d'espace clos avec des conditions d'éclairage particulières contrairement aux dispositifs optiques ; elle peut donc être utilisée à l'extérieur, avec des éclairages de scène et peut facilement être cachée sous un costume.

Cependant, en raison de la dérive des capteurs, la capture inertielle de mouvement éprouve plus de difficultés à enregistrer la position et l'orientation absolues pendant de longues périodes de captation. En effet, comme le capteur calcule la position depuis l'accélération, il peut y avoir un décalage qui devient de plus en plus grand au fur et à mesure du temps, ce qui veut dire que le capteur mesure bien la position d'un segment corporel par rapport à un autre, mais peut induire un décalage dans la position absolue (la personne captée est décalée par rapport à sa position dans l'espace).

Il peut donc être difficile de capter l'interaction de plusieurs personnes entre elles ou avec des objets, et des décalages de position peuvent apparaître durant une longue durée de captation.

Ces systèmes sont aussi parfois limités aux mouvements réalisés avec les pieds au sol, ce qui peut être contraignant pour certaines pratiques. Le mouvement sera toujours capté avec précision, mais la personne captée pourra se retrouver décalée par rapport à sa position dans l'espace. Il est donc recommandé de découper la captation en séquences courtes ou bien de combiner à un système dédié à suivre la position de la personne dans l'espace.



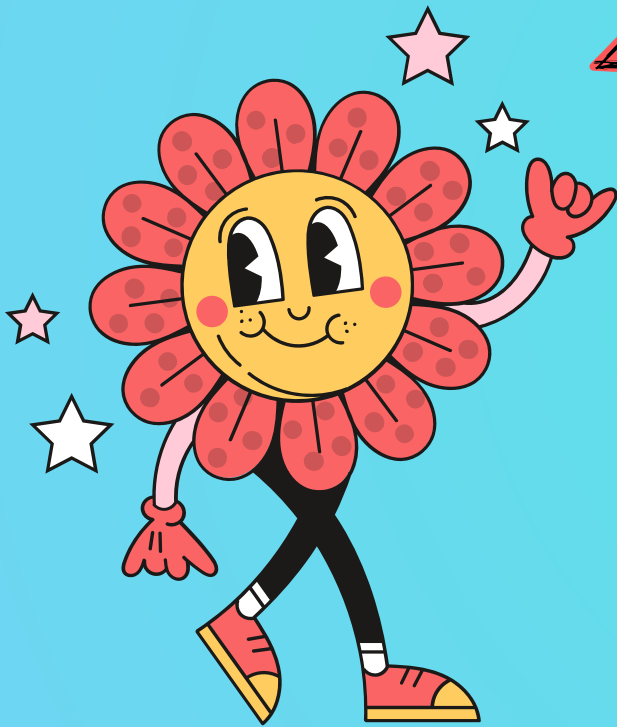
Exemples d'un artiste portant une combinaison inertielle.

Crédit photo : Marion Cossin, HUPR

POUR ALLER PLUS LOIN

Delusional World, un exemple de performance qui utilise la capture de mouvement inertielle. De l'artiste multimédia chinoise Lu Yang mettant en vedette la danseuse et chorégraphe Louise Lecavalier.

Lien : https://www.youtube.com/watch?v=yphCsHMaKqI&ab_channel=Radio-CanadaInfo



A futuristic laboratory scene with a robot and a digital human figure. The robot, in the foreground, is dark with glowing yellow circular markers on its joints and head. It stands on a floor with yellow geometric patterns. In the background, a translucent digital human figure is visible, also with yellow markers. The room features large windows, various cables, and equipment, all bathed in a blue and yellow light.

NÉTHODES AVEC MARQUEURS

SYSTÈMES OPTIQUES AVEC MARQUEURS

Les méthodes basées sur la vision utilisent des caméras pour capturer le mouvement d'un sujet ou d'un objet dans l'espace. Elles se basent principalement sur des images visuelles ou des vidéos pour analyser la position et les mouvements dans l'espace. Les systèmes optiques peuvent utiliser des marqueurs visibles ou invisibles sur le sujet ou l'objet pour suivre ses mouvements ou, dans le cas des systèmes sans marqueurs, des algorithmes avancés pour interpréter directement les images.

● POINTS À RETENIR

- Exemples : Vicon, Optitrack, Qualisys
- Données de sortie : Coordonnées 3D des marqueurs, squelette 3D et trajectoire des objets
- Choix entre marqueurs actifs et passifs
- Minimum 8 à 10 caméras dans un espace de capture restreint. En réalité 20 à 80 caméras si les mouvements ne sont pas simples et pour éviter les occlusions
- Logiciel dédié et payant
- Matériel informatique : un processeur multicœur haut de gamme est souvent recommandé, un GPU est optionnel mais recommandé pour les applications en temps réel ou pour l'exportation des données en visualisation 3D
- Cout : Entre \$40K-\$200K (dépend de la configuration). Une caméra peut coûter de 2000\$-16000\$ et le logiciel entre 1000\$-11000\$
- Utilisation à l'intérieur recommandée, bien que ce soit possible à l'extérieur avec une calibration adaptée

● POINTS POSITIFS

- Très grande précision
- Possibilité de temps réel
- Possibilité de connexion à d'autres logiciels (plug-in)
- Possibilité de capter le mouvement d'objets
- Certains systèmes peuvent relativement bien gérer l'occlusion



Acrobatie à deux avec marqueurs.
Crédit photo : HUPR et CDRIN

● LIMITATIONS

- L'installation peut être longue, cela implique un placement et un calibrage minutieux des caméras et de l'éclairage
- La prise en main du logiciel peut être difficile, le dispositif requiert différentes connaissances techniques. Il est conseillé de suivre une formation avec le fournisseur
- Le confort des personnes captées est acceptable mais les combinaisons peuvent être très chaudes à l'effort, et certains marqueurs peuvent gêner ou occasionner des douleurs dans certaines positions ou mouvements (contact au sol ou impact directement là où sont placés les marqueurs)
- Les méthodes de fixation des marqueurs ne sont pas idéales et ceux-ci peuvent se détacher lors de certains mouvements ou de contacts entre les individus ou les accessoires/appareils.
- Les marqueurs actifs nécessitent des fils et/ou des batteries.
- L'intégration des marqueurs dans des costumes n'est pas simple car les marqueurs doivent rester bien visibles aux caméras.



*Exemple d'artiste portant une combinaison avec marqueurs.
Crédit photo : Anaïs Charbonneau-Meilleur,
Les 7 doigts*

● PRINCIPE

La capture de mouvement avec marqueurs s'appuie sur la reconnaissance de marqueurs disposés à des endroits prédéfinis du corps humain que plusieurs caméras infrarouges placées à différents angles de vue suivent. Dans cette configuration, les caméras sont utilisées pour mesurer de petits points lumineux dans un espace de capture dédié, émis ou réfléchis par des marqueurs soigneusement attachés à une personne ou à un objet. La lumière réfléchie ou émise par les marqueurs est reçue sous forme d'image bidimensionnelle par chaque caméra. L'ordinateur peut alors déterminer l'emplacement exact des marqueurs et surtout la relation entre les marqueurs dans l'espace en combinant les images de chaque caméra.

Ce processus repose sur des principes de photogrammétrie utilisant la stéréoscopie (vision binoculaire). Ainsi, les dispositifs optiques avec marqueurs sont capables de créer un modèle tridimensionnel à partir d'au moins deux images, mais la plupart des systèmes optiques avec marqueurs requiert que chaque marqueur soit vu par au moins trois caméras pour plus de précision.

Il faut bien préciser que ce n'est pas l'image de la personne ou de l'objet qui est capté (la chair ou son aspect) mais bien les marqueurs, à partir de laquelle le logiciel déduit « l'ossature », les joints, ou encore les points d'articulation des membres qui permettent le mouvement du corps. C'est pour cette raison que bien souvent, les marqueurs seront disposés à des endroits précis du corps qui permettent d'extrapoler les mouvements des articulations.



Exemple d'une pose de calibration en A
Crédit photo : Les 7 doigts

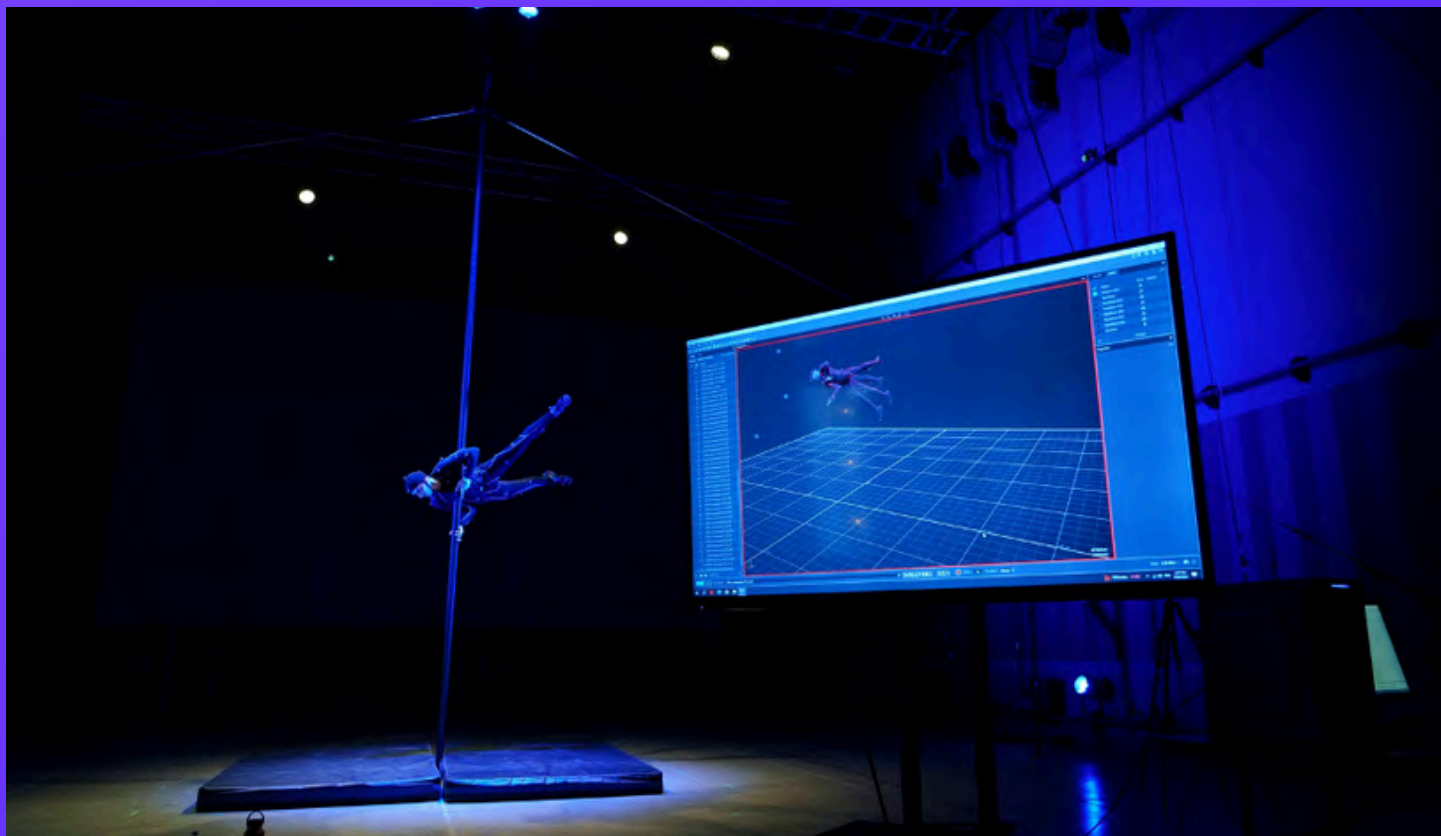


Exemple d'une pose de calibration en T.
Crédit photo : Francisco Cruz, Les 7 doigts et HUPR

Un logiciel spécialisé utilise les relations trigonométriques entre les marqueurs dans les images capturées et les emplacements des caméras pour calculer en temps réel les positions et les orientations des articulations du corps. Le logiciel doit aussi traquer les marqueurs, c'est-à-dire : les inscrire dans une liste de suivi d'identification à partir d'une pose initiale de l'acteur (les plus communes étant la position en T ou "T-pose" et la position en A ou "A-pose"); gérer les occlusions (quand un membre, un objet présent dans l'espace de capture, ou un autre acteur dans le cas de systèmes à acteurs multiples, vient cacher un marqueur à la vue d'une ou plusieurs caméras) ; enfin gérer le swapping (problème de confusion ou d'échange accidentel dans l'identification de deux marqueurs, par exemple deux mains qui passent près l'une de l'autre).

La majorité des systèmes de capture de mouvement optique utilisent des marqueurs passifs qui font "rebondir" la lumière émise par des diodes électroluminescentes (DEL) infrarouges entourant les objectifs des caméras. Mais la lumière peut aussi provenir des marqueurs : on parlera de marqueurs "actifs" lorsque la source de lumière provient des marqueurs eux-mêmes, et "passifs" lorsque c'est la caméra qui projette de la lumière sur les marqueurs qui réfléchissent la lumière en retour pour donner leurs positions. Cette technologie, basée sur l'utilisation de lumière infrarouge émise ou reflétée par des marqueurs, a pour avantage une très grande précision puisque ce sont les seules données captées et interprétées par les caméras.

Pour une utilisation optimale, il faut cependant s'assurer qu'il n'y a pas d'autres lumière infrarouge émise dans l'espace (donc travailler en mode "boite noire" sans lumière naturelle et choisir des lampes qui n'émettent pas d'infrarouges) et qu'il n'y a pas d'objets réfléchissants dans l'espace de capture (ceux-ci pouvant être interprétés comme des marqueurs par le système et donc causer du "bruit" dans les données).



Un artiste de cirque en train d'être capté par un système optique avec marqueurs. Crédit photo : Francisco Cruz, Les 7 doigts

Pour minimiser l'impact de l'occlusion, les systèmes avec marqueurs utilisent généralement plusieurs caméras disposées autour de la zone de capture, afin que chaque marqueur soit visible d'au moins trois angles. Cependant, pour des mouvements complexes impliquant des acrobaties, des positions recroquevillées ou des rotations, les occlusions sont inévitables et peuvent causer des pertes de données. Certains systèmes ont des algorithmes de suivi et des techniques de prédiction pour « deviner » la position temporaire d'un marqueur lorsqu'il est masqué, mais ces prédictions peuvent introduire des erreurs.

Les différentes étapes lors de l'utilisation de ce genre de systèmes sont :

Installation : Cette étape peut nécessiter 1 à 2 journées de travail selon le volume à couvrir. Il s'agit d'installer les caméras afin de couvrir la zone de captation, puis d'effectuer toutes les calibrations et ajustements.

Équiper les personnes : Il s'agit de placer les marqueurs réfléchissants sur les points d'articulation selon un modèle spécifique. La durée peut aller de 30 minutes à 1h par personne selon le modèle et le type de marqueurs.

Calibration des personnes : Avant d'effectuer la captation, il faut calibrer le système avec par exemple la pose en A ou en T (voir photos plus haut).

Capture de mouvements : Enregistrements des séquences de mouvements. Plusieurs prises sont souvent nécessaires pour obtenir la meilleure qualité.

Traitement des données : Cela consiste à corriger les erreurs (marqueurs manquants ou mal attribués au fil de l'enregistrement), d'attribuer les marqueurs aux segments corporels (manuellement ou automatiquement), d'effectuer d'autres traitements si nécessaire puis d'exporter les données. Cette étape peut être très longue selon la qualité de la captation et la qualité souhaitée des données.

POUR ALLER PLUS LOIN

Exemple de deux projets artistiques mobilisant la méthode avec marqueurs. :

Carry me home, Les 7 Doigts

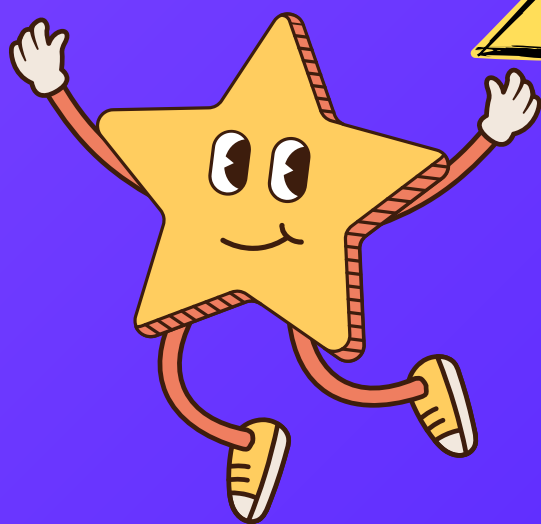
Lien : <https://7doigts.com/spectacles/creations/carry-me-home>

et

« Cosmogony » de Cie Gilles Jobin.

Lien du making of :

<https://vimeo.com/manage/videos/723708045/7422cb3949>





Vision par ordinateur et intelligence artificielle

PRÉAMBULE

Nous allons maintenant discuter d'un autre élément de la capture de mouvement optique sans marqueurs, la vision par ordinateur. Tout comme les dispositifs vus plus haut, la clef de la vision par ordinateur réside dans son « cerveau », si on veut prolonger la métaphore anthropomorphique, ou plutôt dans l'algorithme qui permet l'analyse des mouvements captés par « l'œil » de la caméra.



La vision par ordinateur est un domaine de la science de l'informatique qui entraîne les ordinateurs à interpréter et comprendre le monde visuel par différents moyens. Comme les autres types d'IA, la vision par ordinateur cherche à exécuter et à automatiser des tâches dévolues aux capacités humaines. Dans ce cas, la vision par ordinateur cherche à répliquer la manière dont les êtres humains donnent un sens à ce qu'ils voient.

Les applications de vision par ordinateur utilisent les entrées visuelles des appareils de détection pour capter leur environnement, puis l'intelligence artificielle, le machine learning (apprentissage automatique) ou le deep learning (apprentissage profond) remplaceront le rôle du « cerveau » dans l'analyse de ce qui est vu. Les applications de vision par ordinateur s'exécutent sur des algorithmes qui sont entraînés sur de grandes quantités de données visuelles ou d'images. Elles reconnaissent les modèles à partir de l'apprentissage de ces données visuelles et utilisent ces modèles pour déterminer le contenu d'autres images similaires.

De manière générale, la capture de mouvement sans marqueurs ou sans capteurs est le moyen privilégié pour capturer le mouvement sans avoir recours à des combinaisons, des marqueurs, du matériel spécialisé ou tout autre vêtement.

À noter que les mouvements capturés par caméras RGB ou infrarouges peuvent aussi être analysés manuellement pour en extraire les informations sur la position ou le mouvement d'une personne, sans recours à des algorithmes de vision par ordinateur ou d'estimation de pose, mais nous ne détaillerons pas ici ces techniques, qui sont fastidieuses et complexes.

SYSTÈME INTÉGRÉ

● PRINCIPE

Il existe des solutions intégrées « tout en un » et faciles à prendre en main (sans avoir à coder). C'est-à-dire que ces solutions offrent à la fois des suggestions de matériel et surtout le logiciel pour pouvoir effectuer une captation de mouvements, sans la difficulté de devoir ajuster les différentes composantes du dispositif avec des compétences de programmation.

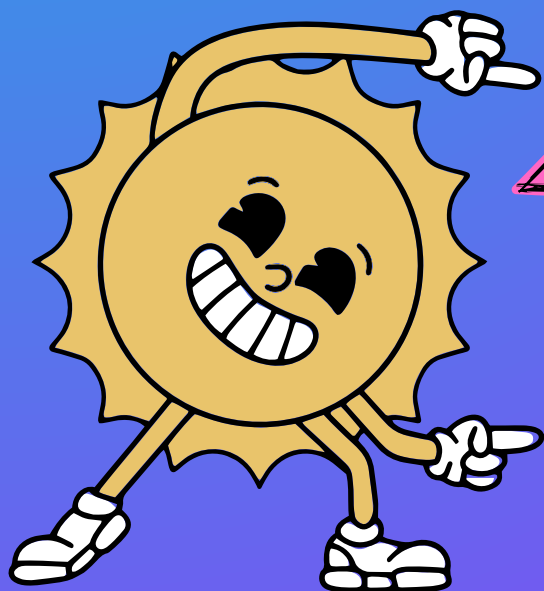
Il s'agit de solutions multi-caméras capables de capturer les mouvements avec précision, en temps réel et sans marqueurs. Ces solutions sont utilisées dans des contextes professionnels ou de recherche car ils s'intègrent bien dans des workflow (flux de travail). Il faut noter toutefois que si ces solutions sont plus faciles d'accès par leur facilité d'usage, elles n'en restent pas moins plus onéreuses et moins modulables.

Il existe aussi des systèmes intégrés haut de gamme, de qualité professionnelle, comme Theia ou Captury, qui associent des caméras (RGB haute résolution, caméras infrarouges, ou à profondeur de champ) avec des algorithmes capables de suivre et analyser le mouvement capté. Certaines entreprises développent des caméras spécifiques à leur utilisation et optimisées pour l'intégration avec leur logiciel.

SYSTÈME INTÉGRÉ HAUT DE GAMME

☀ POINTS A RETENIR

- Exemples : Theia, Captury, Simi
- Données de sortie : Squelette 3D et trajectoire de chaque segment corporel
- 8 à 10 caméras est probablement le minimum recommandé. Les espaces de capture plus grands nécessitent probablement plus de 20 caméras.
- Logiciel dédié et payant
- Extérieur et intérieur
- Coût : entre 35 000\$ à 150 000\$ en fonction de la configuration (caméras, ordinateur, logiciels)
- Pas de marqueurs, pas de capteurs sur le corps des personnes captées
- Presque tous les systèmes doivent être ajustés pour gérer les arts du spectacle et les mouvements de cirque
- Les configurations multi-caméras requièrent des connaissances techniques. Pour travailler avec un logiciel, il faut souvent comprendre les réglages et les conditions qui peuvent affecter les résultats. Nous recommandons les formations ou un suivi avec le fournisseur.
- Certaine capacité à gérer l'occlusion
- Le volume de captation dépend du nombre de caméras (16 à 150 m2)
- Seul Captury semble pouvoir faire du temps réel
- Possibilité d'être connecté avec d'autres logiciels (plug-in)
- Certaine capacité à capter les objets
- Capacité informatique : un processeur multicœur, une carte graphique (GPU) NVIDIA avec support CUDA, un SSD est recommandé



POUR ALLER PLUS LOIN

Le concert de Jean Michel Jarre en réalité mixte dans la galerie des glaces du chateau de Versailles. La musique est associée à une technologie de capture de mouvement sans marqueur.

Lien : https://www.youtube.com/watch?v=SaLHTsAH9Dc&ab_channel=JeanMichelJarre

SYSTÈME INTÉGRÉ ACCESSIBLE

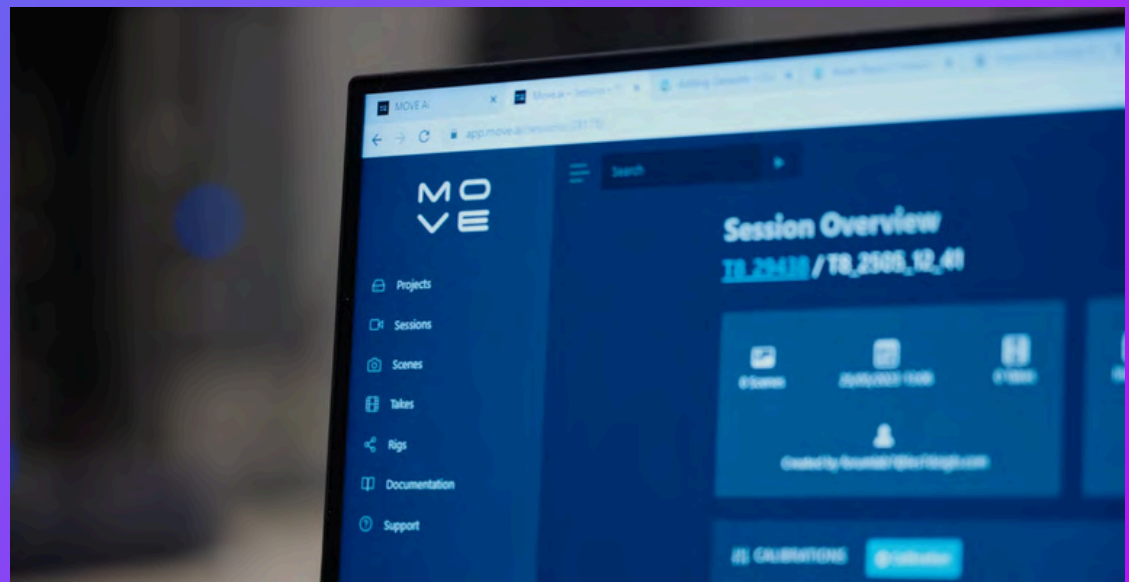
☀ POINTS A RETENIR

- Exemples : MoveAI, Rokoko Video, DeepMotion, Plask
- Données de sortie : squelettes 3D estimés, souvent sans profondeur réelle
- Fonctionne avec 1 ou 2 caméras pour Rokoko Video, DeepMotion et Plask, et jusqu'à 12 caméras pour MoveAI
- La capture multi-caméras est probablement nécessaire pour des mouvements complexes
- Fonctionne avec des webcams, GoPro et smartphones
- Logiciel ou application dédiée
- Coût : offre une courte durée gratuite, puis premier plan à 15-20\$/ mois, autres plans sur demande
- Extérieur et intérieur
- Pas de capteurs ni de marqueurs sur la personne captée
- Les systèmes ne sont pas conçus pour les arts du spectacle et le cirque
- Capacité à capter des mouvements acrobatiques très limitée
- Espace restreint et préférablement intérieur : 20 x 20 m dépendamment du nombre de caméras, 5 m x 5 m avec une caméra seule
- Facile et rapide à installer
- Éclairage uniforme et un fond contrasté améliorent la précision
- Pas de capacité à capter des objets
- MoveAI offre des capacités de capture de mouvement en temps réel avec sa solution Move Live, ainsi que le traitement de vidéos préenregistrées. Rokoko Video, DeepMotion et Plask se concentrent sur le traitement de vidéos préenregistrées
- Capacité informatique : processus multicœur, carte graphique (GPU) NVIDIA GeForce GTX 1080 ou supérieure, SSD de 512 Go ou plus

PRINCIPE

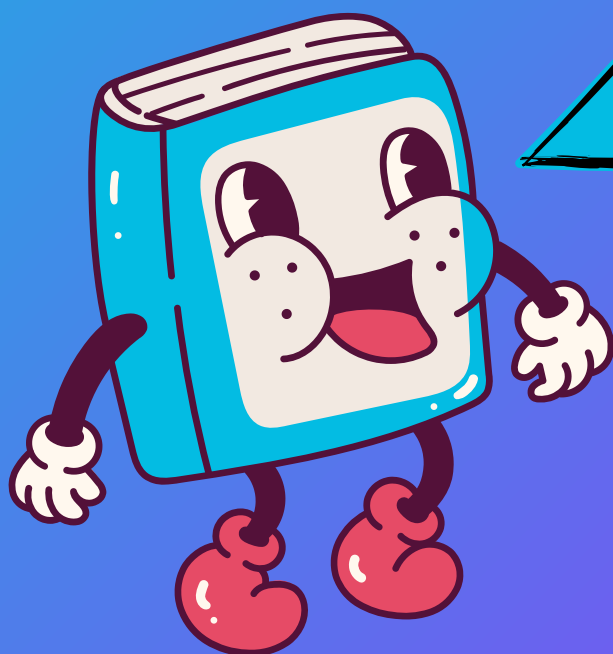
Les systèmes intégrés plus accessibles, comme MoveAI ou Rokoko Video, offrent un algorithme adaptable à la plupart des caméras. Les systèmes intégrés accessibles sont des solutions similaires au système intégré haut de gamme mais conçus pour rendre la capture de mouvement accessible à un plus grand nombre. Leur coût est plus abordable et ils s'utilisent avec du matériel grand public comme des caméras RGB classiques ou des smartphones. Bien que leur précision et fiabilité n'atteindra pas celles des systèmes intégrés haut de gamme, ils offrent une solution équilibrée entre précision, coût et facilité d'utilisation.

Un autre avantage de ce type de solution est leur utilisation sur le serveur cloud (nuage), c'est-à-dire qu'il n'est pas nécessaire d'installer et de gérer localement les modèles qui alimentent ces solutions et donc il n'est pas nécessaire d'avoir du matériel informatique puissant.



Interface de MoveAI.
Crédit photo :
Francisco Cruz, Les 7
Doigts

Exemple d'un dispositif intégré utilisant un cellulaire.
Crédit photo : Francisco Cruz, Les 7 Doigts



POUR ALLER PLUS LOIN

Grimes' XR Performance à Coachella 2024.
La prévisualisation en temps réel et la capture de mouvement de Grimes ont permis de donner vie au concept de Grimes AI dans XR live, pour la première fois à Coachella. Le jumeau numérique de Grimes s'est produit avec un éclairage virtuel, brouillant les frontières entre la réalité et l'imagination.

Lien : https://www.youtube.com/watch?v=WtxuobyzLSI&ab_channel=MoveAI

ESTIMATION DE POSE (POSE ESTIMATION)

● PRINCIPE

Dans le champ de la capture de mouvement, la vision par ordinateur servira essentiellement à l'estimation de pose, un sous-ensemble du domaine plus vaste de la vision par ordinateur et de l'apprentissage automatique.

L'estimation de pose est une technique de traitement des données qui prend en entrée des images ou des vidéos et qui utilise des algorithmes de vision par ordinateur et d'intelligence artificielle pour détecter automatiquement la position des articulations ou du corps entier. L'algorithme est utilisé pour identifier des points sur le corps humain et extraire le mouvement de la séquence vidéo capturée par un appareil photo numérique ou une caméra spécialisée. La SDK de Kinect utilise par exemple des caméras à profondeur de champ, et OpenPose ou Mediapipe, des caméras RGB. Certains algorithmes fonctionnent même à partir de la simple captation d'une caméra de téléphone.



Captation acrobatique avec estimation de pose
Crédit photo : HUPR

L'estimation de pose détermine la position et l'orientation d'un corps humain dans une image ou une trame vidéo par la reconnaissance et la catégorisation des positions et des articulations par un algorithme qui aura été entraîné à exécuter cela. Il s'agit essentiellement pour ce dispositif d'un processus d'identification de points clés connectés en "paires valides". Un point clé est une coordonnée (2D ou 3D) décrivant l'emplacement d'une articulation anatomique qui peut être utilisée pour décrire une pose. Ces points se connectent pour former des paires, qui sont considérées comme "valides" lorsqu'elles correspondent à un élément saillant du corps en mouvement (tête, avant-bras, bassin, etc.).

Il est possible d'aller encore plus loin et d'utiliser un algorithme de classification de mouvement pour classer et donc déterminer le type d'action. On pourrait donc avoir un algorithme capable de reconnaître par exemple une personne en train de marcher, courir ou sauter, si on prédéfinit ces trois catégories.

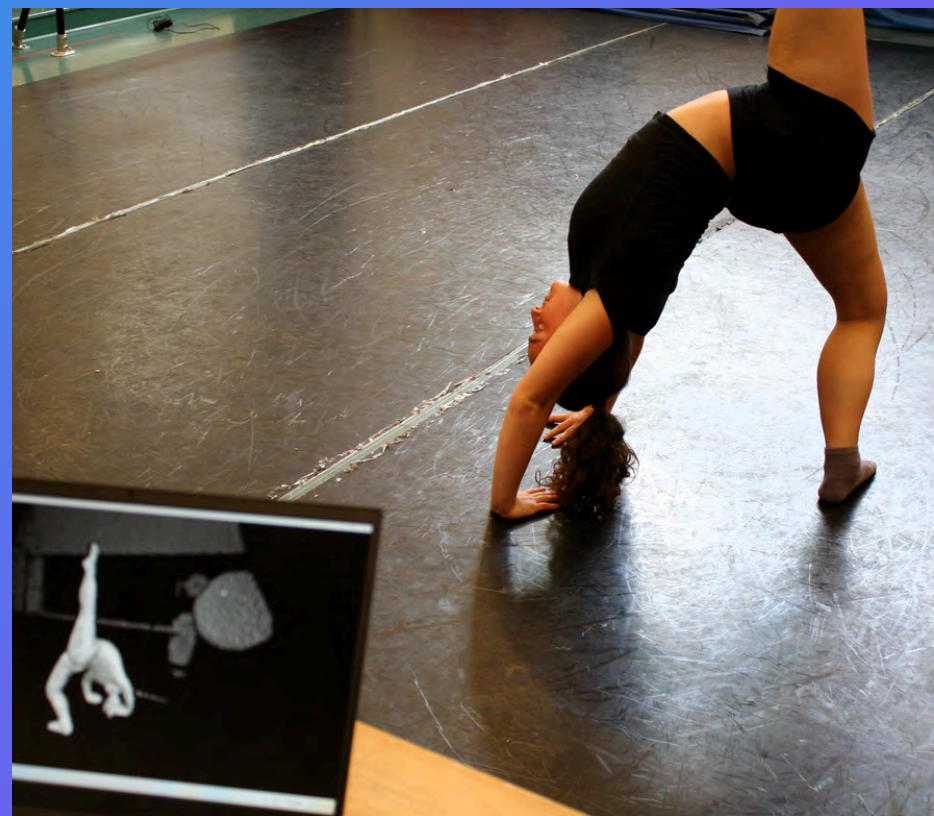


Figure 4: Captation acrobatique avec estimation de pose.

Crédit photo : HUPR



Captation acrobatique avec estimation de pose.

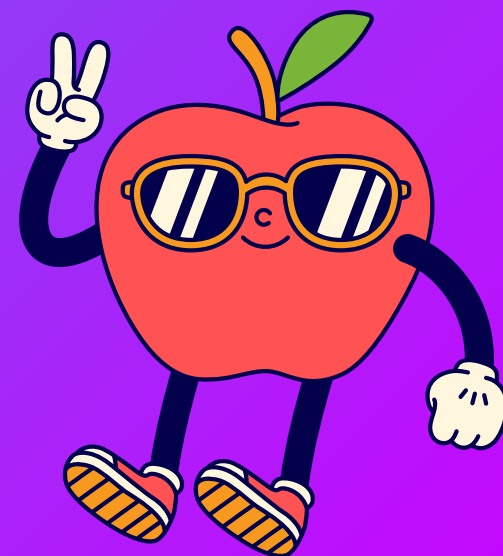
Crédit photo : HUPR

Il existe plusieurs façons de mettre en place une solution avec un modèle d'estimation de pose et cela réside dans le niveau d'efforts nécessaires. Il existe des modèles intégrés dans une solution SDK (software development kit) ou des solutions à implémenter soi-même. Pour ces deux solutions, il est nécessaire d'avoir des compétences en programmation comme Python.

POUR ALLER PLUS LOIN

Projet Mégaceta avec Creo et SAT pour une exposition.
Mégaceta invite les visiteurs à interagir avec des baleines en utilisant une technologie d'estimation de pose.

Lien : <https://creo.ca/fr/megaceta/>

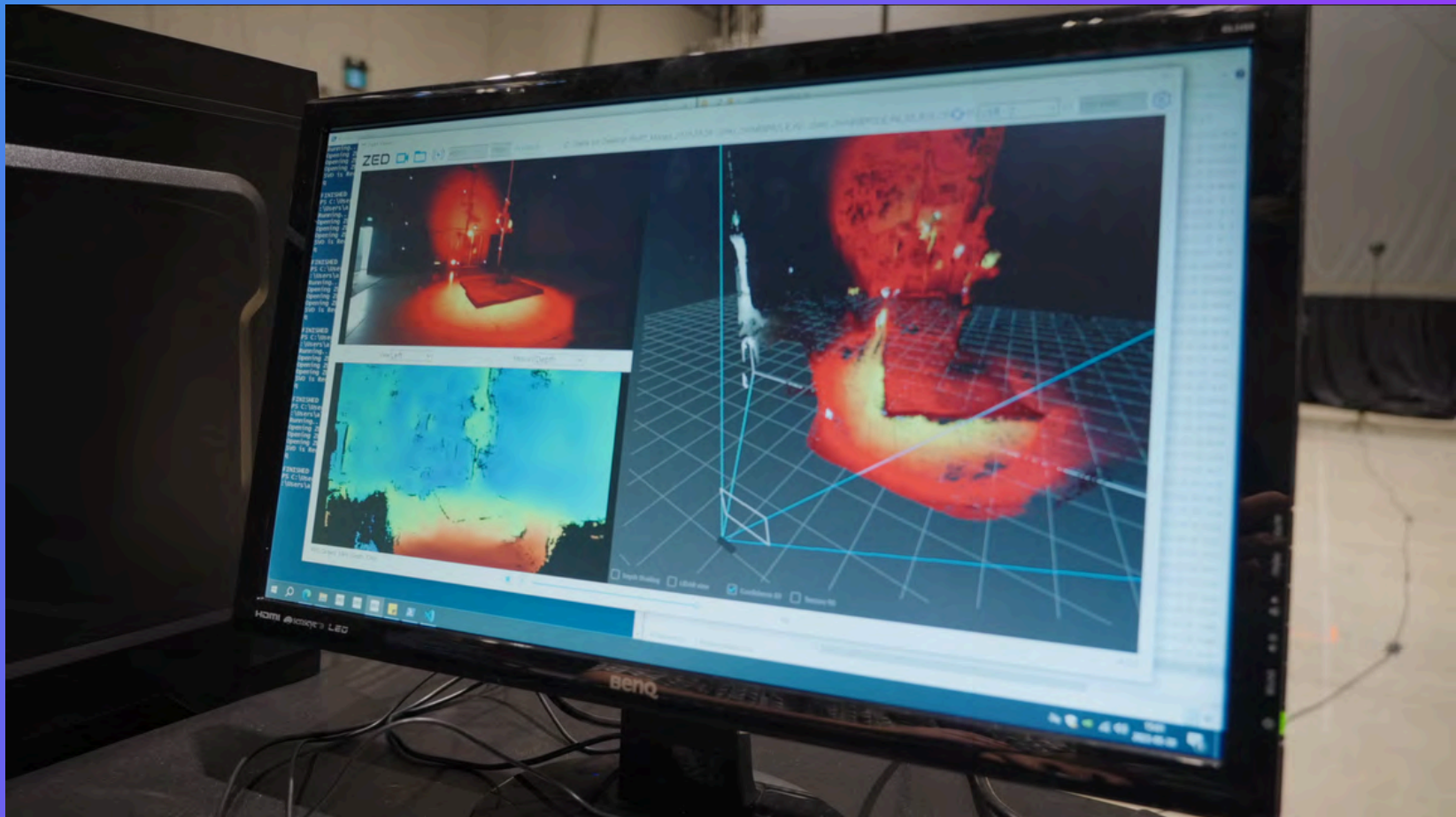


SOLUTION INTÉGRÉE DANS UN SDK

☀ POINTS A RETENIR

- Exemples : SDK de la ZED, SDK de la Kinect, SDK de la Orbecc, SDK de la Intel RealSense
- Données de sortie : squelettes 3D, coordonnées 3D des articulations, flux vidéo RGB, carte de profondeur et/ou nuages de points 3D selon les caméras
- Une ou plusieurs caméras. La capacité d'utilisation multi-caméras dépend du système de SDK
- Solution incluse avec l'achat de la caméra (pas de logiciel à acheter en plus)
- Le coût d'une caméra avec SDK est entre 500\$ et 1000\$
- La SDK est spécifique à la caméra
- Solution clé en main qui nécessite très peu de compétences de programmation
- Intérieur et extérieur
- Pas de capteurs ni de marqueurs sur le corps de la personne à capter
- Les algorithmes d'estimation de la pose rencontrent des difficultés avec les mouvements acrobatiques ou des accessoires comme des costumes, chapeaux, etc.
- Relativement facile à prendre en main si l'on sait comment travailler avec les SDK
- La capacité à gérer l'occlusion dépend du SDK
- Le volume de captation dépend du nombre de caméras, environ 5 m x 5 m avec une caméra

- Le contact entre personne ou avec des objets peut conduire à l'occlusion, ce qui peut être problématique en fonction de l'algorithme utilisé
- FPS relativement bas (max 60 ou 100 selon les modèles et la résolution)
- Possibilité de temps réel
- Possibilité d'être connectée avec d'autres logiciels (plugin)
- Certaines SDK peuvent faire du suivi d'objet
- Capacités informatiques : processeur multicœur, carte graphique (GPU) NVIDIA avec CUDA, un SSD est recommandé pour le stockage de données vidéo en temps réel.



Visualisation d'une captation SDK.
Crédit photo : HUPR

☀ PRINCIPE

Les solutions SDK sont celles proposées par le fabricant de caméras à profondeur comme la ZED, la Orbecc ou Azure Kinect. Ce sont des outils qui incluent des algorithmes d'estimation de pose prêts à l'emploi.

Une fois la caméra en main, il faut télécharger et installer le SDK fourni par le fabricant, qui inclut des bibliothèques, des exemples de codes et d'autres outils. Il faut ensuite utiliser les API (interface de programmation d'application) fournies par le SDK dont une documentation bien détaillée est fournie. Une fois déployées, ces solutions proposent un ajustement de certains paramètres. L'avantage de cette solution est que le fabricant a déjà fait une majorité du travail, ce qui permet une implémentation plus rapide, cependant, cela apporte moins de flexibilité dans les fonctionnalités (exemples : nombre d'articulations captées, découpe du corps, etc.) Également, il n'y a que le prix de la caméra à considérer, la SDK étant incluse dans le produit ou accessible en ligne. Notons quand même que la SDK de la Kinect et de la RealSense ne sont plus maintenues par leurs fournisseurs.



Captation acrobatique
Crédit photo : HUPR

SOLUTION À IMPLÉMENTER SOI-MÊME

☀ POINTS À RETENIR

- Exemples : OpenPose, MediaPipe, PoseNet, AlphaPose, BlazePose, VITPose, YoloV8
- Données de sortie : Coordonnées 2D ou 3D des articulations
- Mediapipe et YOLO sont les plus accessibles en termes de programmation. Il est possible de faire fonctionner MediaPipe directement sur le web[1].
- 1 caméra minimum, 2 ou 3 caméras améliorent la qualité, notamment pour une estimation en 3D
- Algorithme gratuit mais compétence en programmation nécessaire
- Possibilité d'être utilisé avec des caméras peu coûteuses comme des webcams ou des GoPro
- Intérieur et extérieur
- Pas de marqueurs ni de capteurs sur le corps de la personne captée
- Les algorithmes d'estimation de la pose rencontrent des difficultés avec les mouvements acrobatiques, cela dépend des algorithmes
- L'installation est rapide
- La prise en main ou la mise en place peut être plus compliquée et plus longue comme des compétences en programmation sont nécessaires
- La capacité à gérer l'occlusion dépend de l'algorithme
- Le volume de captation dépend du nombre de caméras, environ 5 m x 5m avec une caméra

- Le contact entre personnes ou avec des objets peut conduire à l'occlusion, ce qui peut être problématique en fonction de l'algorithme utilisé
- Capacité de temps réel dépend de l'algorithme
- FPS souvent bas dans un contexte dynamique
- Connexion avec d'autres logiciels (plug-in) possible avec une solution personnalisée
- Captation d'objet possible avec certains algorithmes (MediaPipe ou BlazePose par exemple), sinon il est possible d'associer un autre algorithme de détection d'objets aux algorithmes d'estimation de pose
- Capacité informatique : processeur multicœur, nécessite un GPU puissant et beaucoup de RAM pour des performances en temps réel. PoseNet, MediaPipe et YOLO roulent aujourd'hui à 30 FPS directement sur CPU.



*Captation
acrobatique
Crédit photo :
Joe Alvoeiro,
HUPR*

☀ PRINCIPE

Les solutions à implémenter soi-même permettent plus de flexibilité sur le choix du matériel et du modèle d'estimation de pose, mais cela implique de devoir développer votre propre pipeline.

Il faut d'abord choisir entre plusieurs modèles, comme OpenPose, MediaPipe, PoseNet ou HRNet qui sont des modèles pré-entraînés, mais vous pouvez aussi entraîner votre propre modèle. Cependant, entraîner son propre modèle est une tâche beaucoup plus complexe que d'utiliser un modèle pré-entraîné, surtout si l'on souhaite entraîner un modèle qui se compare un tant soit peu aux modèles d'estimation de pose les plus performants aujourd'hui. La plupart des modèles sont open source, et leurs spécificités varient en termes de précision, vitesse ou capacité informatique nécessaire.

Ensuite, il faut installer les dépendances nécessaires comme PyTorch ou des bibliothèques spécifiques. Il faut ensuite écrire le code pour intégrer l'algorithme dans l'application avec le traitement des images ou des vidéos. Mettre en place ce genre de solutions prend plus de temps et nécessite des compétences avancées en programmation, voire en vision par ordinateur ou apprentissage automatique. Le principal avantage de cette solution est la flexibilité totale dans le choix du matériel, de l'algorithme et du pipeline. Comme cette solution n'est pas liée à du matériel spécifique, il est possible de l'utiliser avec des caméras basiques comme des webcams.

SPÉCIFICITÉS PAR TYPE DE CAMÉRA

ESTIMATION DE POSE PAR CAMÉRA RGB

Les modèles d'estimation de pose, comme OpenPose ou MediaPipe utilisent le flux vidéo de caméras RGB pour détecter les articulations et les membres à partir de caractéristiques visuelles (textures, contours, contraste).

☀ AVANTAGES

- Beaucoup de modèles d'estimation de pose sont conçus pour fonctionner avec des caméras RGB standards.
- Les modèles détectent facilement les poses dans des contextes bien éclairés.
- Idéal pour un projet à faible coût

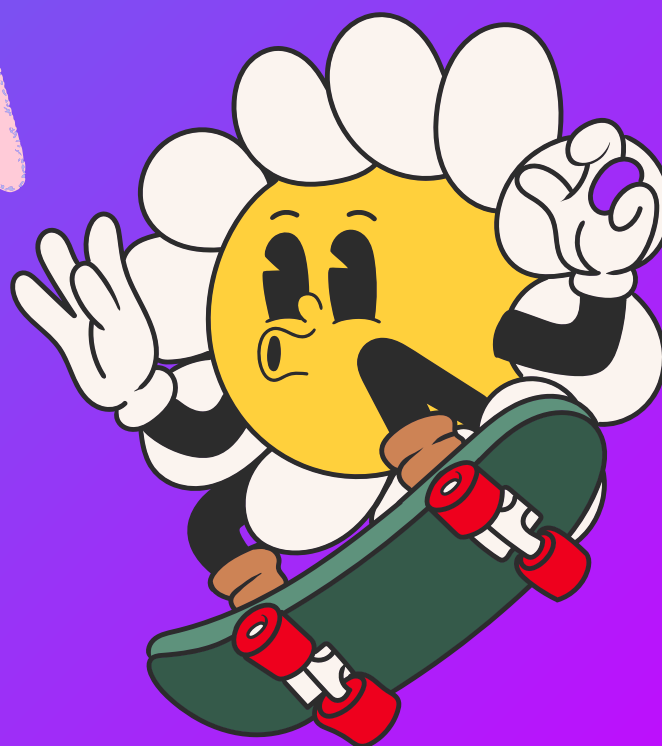
☀ LIMITATIONS

- Moins performants dans des environnements peu éclairés et des mouvements rapides ou acrobatiques
- Les mouvements en 3D (profondeur) s'ils sont captés sont estimés à partir d'images 2D et peuvent contenir des erreurs

BON À SAVOIR!!

Cette approche est une des plus courantes puisqu'elle utilise des caméras standards, largement disponibles et peu coûteuses. L'inconvénient d'utiliser des caméras RGB est que la qualité peut être affectée par la luminosité. Comme ce sont des caméras qui captent les couleurs, il est plus difficile de capter en cas de faible éclairage ou changement d'éclairage. La plupart des modèles d'estimation de pose, comme OpenPose ou MediaPipe sont compatibles et sont conçus pour fonctionner avec des caméras RGB. La performance dépendra ensuite de la résolution de la caméra, de l'éclairage et de l'adéquation du modèle avec ce qui est capté.

Il est possible d'améliorer la précision et la fiabilité en utilisant plusieurs caméras RGB au lieu d'une seule. Cependant, il est nécessaire de modifier le pipeline puisqu'il faut ajouter un module de synchronisation et de fusion des données issues des différentes caméras.



ESTIMATION DE POSE PAR CAMÉRA INFRAROUGE

Les modèles d'estimation de pose utilisent les données captées en infrarouge donc des niveaux de gris ou chaleur pour détecter la pose. Ces modèles nécessitent souvent des algorithmes adaptés à des images sans couleur.

■ AVANTAGES

- Fonctionnent dans des environnements avec peu ou pas de lumière visible
- Moins affectés par des arrière-plans visuellement complexes ou chargés

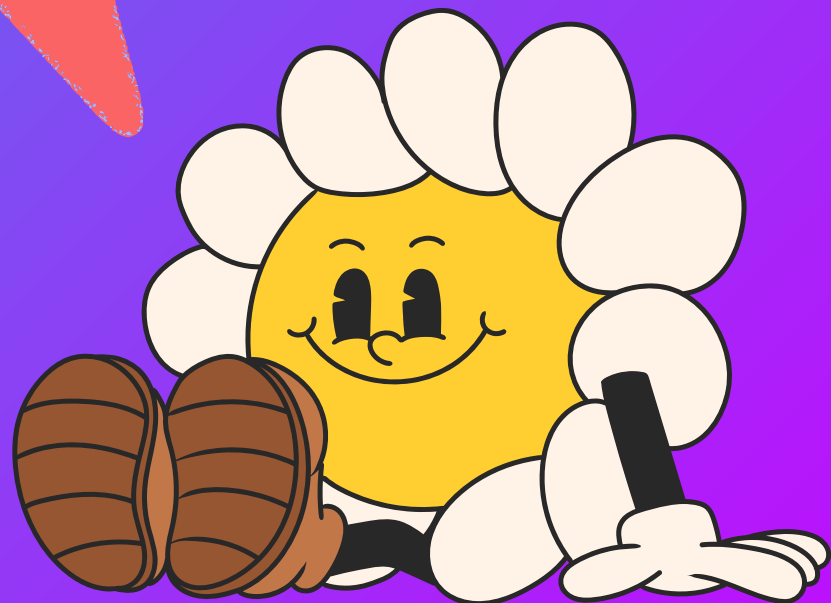
■ LIMITATIONS

- Les modèles d'estimation de pose entraînés avec des images RGB doivent souvent être réentraînés pour fonctionner avec des données infrarouges. Notons le modèle ThermalPose, conçu spécifiquement pour fonctionner avec des données thermiques/infrarouges
- Moins précis car manque de texture

BON À SAVOIR !

Les caméras infrarouges capturent des images basées sur la chaleur ou la réflexion infrarouge, ce qui peut être utile en cas de faible luminosité. À l'inverse, en cas de forte lumière du soleil, les caméras peuvent moins bien fonctionner. Également, ces caméras peuvent manquer des détails de couleur, ce qui peut affecter la précision de capture de mouvement.

Certains modèles d'estimation de pose pourraient moins bien fonctionner avec des caméras infrarouges s'ils ont été entraînés exclusivement avec des caméras RGB. Comme la caméra donne des niveaux de gris, il est possible qu'un modèle entraîné sur des caméras RGB fonctionne, mais il est nécessaire de le tester. Il est également possible de trouver des modèles spécialisés avec des caméras infrarouges, comme ThermalPose.



ESTIMATION DE POSE PAR CAMÉRA DE PROFONDEUR

Les modèles utilisent directement les données, souvent vidéo et profondeur, pour détecter la position des articulations et des membres en 3D.

■ AVANTAGES

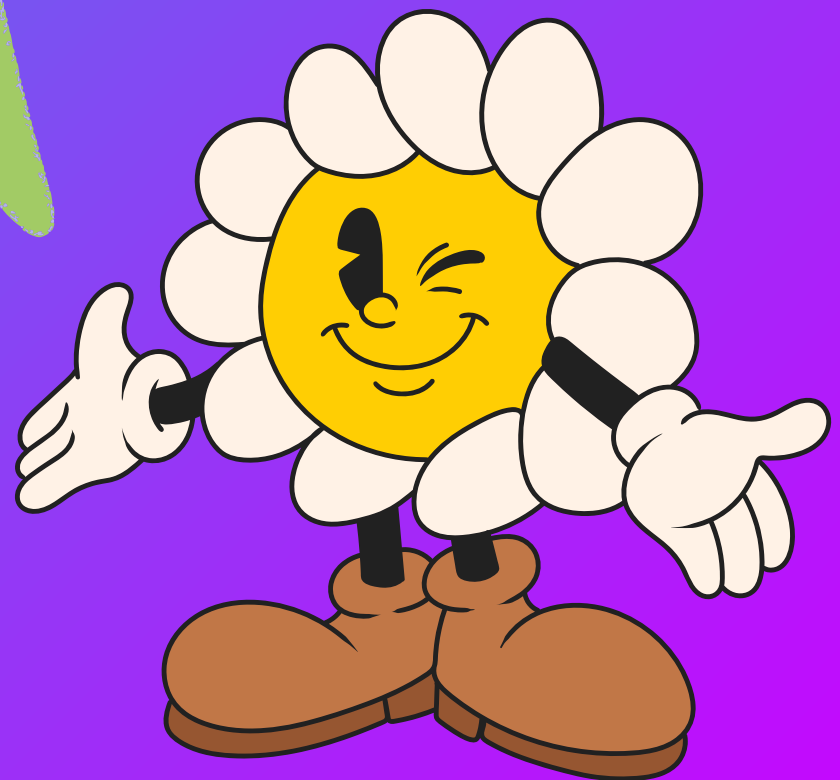
- Estimation précise en 3D
- Moins affectés par des arrière-plans visuellement complexes ou chargés
- Certaines caméras à profondeur incluent la détection d'objet

■ LIMITATIONS

- Il faut utiliser des modèles développés pour de la 3D, ou les modèles 2D existants (comme OpenPose) doivent être adaptés. Notons les modèles DepthPose, PoseNet 3D ou Open3D Pose, conçu spécifiquement pour des données de profondeur
- La précision du modèle dépend directement de la qualité de la caméra de profondeur

BON À SAVOIR !

L'information de profondeur détectée par les caméras à profondeur amène une information additionnelle par rapport à l'utilisation d'une caméra RGB ou infrarouge seule. Ainsi, ces caméras permettent une représentation en 3D. Similairement aux caméras infrarouges, si un modèle a été entraîné sur des images 2D RGB, alors il ne sera sûrement pas performant avec des caméras à profondeur. Il est donc souvent nécessaire de travailler avec des modèles spécifiquement conçus pour l'estimation de pose en 3D. Il est aussi possible d'adapter un modèle de pose 2D existant pour qu'il prenne en compte les informations de profondeur, mais cela nécessite plus de développement. C'est pourquoi les fabricants de caméras à profondeur incluent généralement leur propre SDK qui est adapté à leur caméra. Les modèles OpenPose et MediaPipe, par exemple, peuvent combiner des données de profondeur avec des données 2D pour estimer la coordonnée de profondeur.



SEGMENTATION, CLASSIFICATION ET RECONNAISSANCE

Dans le domaine de la vision par ordinateur, d'autres techniques que l'estimation de pose permettent l'analyse du contenu lié à la capture du mouvement : la segmentation, la classification et la reconnaissance. À noter que la vision par ordinateur englobe aussi d'autres concepts, comme la détection et le suivi d'objet, mais nous resterons concentrés sur des éléments proches de la capture de mouvement.

ALGORITHME DE SEGMENTATION D'IMAGE/VIDÉO

☀ POINTS À RETENIR

- Exemples : OpenCV, SimpleCV, DeepLab
- Données de sortie : contour des éléments détectés
- Solution gratuite, il n'y a que la ou les caméras à acheter
- Compétence en programmation nécessaire
- Ne fait pas de la capture de mouvement, mais du suivi de silhouette
- Intérieur et extérieur
- Pas de marqueurs, pas de capteurs sur le corps de la personne captée
- Possibilité de temps réel
- Connexion avec d'autres logiciels (plug-in) possible avec une solution personnalisée
- Capacité informatique : processeur multicœur, peuvent fonctionner avec ou sans GPU pour des applications de base, nécessite un GPU pour des tâches de segmentation complexes ou en temps réel

☀ PRINCIPE

Les algorithmes de segmentation aident à la détection d'objets en segmentant une image numérique en groupes discrets de pixels (segments d'image). En décomposant les données visuelles complexes d'une image en plusieurs segments de forme spécifique, la segmentation d'image permet un traitement plus rapide et plus avancé des images captées et analysées. Les algorithmes de segmentation d'images traitent les caractéristiques visuelles de chaque pixel, comme la couleur ou la luminosité, pour identifier les limites de l'objet et les régions de l'arrière-plan. Ils servent notamment à la conduite de voitures autonomes, à l'analyse d'images médicales et la détection de divers éléments spécifiques.

La **segmentation d'images** est une technique qui permet de diviser une image ou une vidéo en différentes régions, segments ou objets distincts. Cela facilite la détection d'objets et les tâches de classification. D'une manière générale, la segmentation d'image est utilisée pour trois types de tâches : la segmentation sémantique, la segmentation d'instance et la segmentation panoptique. La **segmentation sémantique** attribue à chaque pixel une classe d'objet ou une catégorie, par exemple 'personne', 'arbre', etc. La **segmentation d'instance** permet de distinguer chaque objet d'une même classe d'objet. Ainsi, si deux personnes sont présentes sur une image, elles seront classifiées en deux personnes même si elles appartiennent à la même classe 'personne'. La **segmentation panoptique** combine la segmentation sémantique et d'instance. Elle permet d'avoir en même temps la distinction de chaque objet comme objet unique et les non-objets, comme le ciel ou le sol sont segmentés via la segmentation sémantique. Cela permet d'avoir une segmentation complète.



Exemple d'un dispositif de capture de mouvement utilisant une caméra à profondeur et un algorithme de segmentation. Crédit photo : HUPR

Ainsi, on peut comprendre que la segmentation seule peut être utilisée pour suivre ou détecter des objets, des personnes ou des parties du corps, mais elle ne permet pas d'obtenir une analyse précise des articulations ou des poses, comme c'est le cas avec les modèles d'estimation de pose. Dans ce cas, la segmentation peut être utilisée pour suivre la silhouette d'une personne en mouvement ou pour suivre la position d'un objet. En fonction des besoins, il est tout à fait possible que cette solution soit suffisante. Par exemple, dans le contexte de jeux interactifs avec de la projection vidéo au sol, il est peut-être simplement nécessaire de connaître la position d'une personne dans l'espace pour déclencher des effets de jeu mais pas sa pose.

OpenCV est l'une des bibliothèques les plus populaires et complètes pour la vision par ordinateur qui permet la segmentation par soustraction de fond (séparer la personne du fond) et la segmentation par contours (détection des contours de la silhouette humaine). SimpleCV est une autre bibliothèque de vision par ordinateur qui propose des outils plus faciles à utiliser pour faire de la segmentation. Pour utiliser des modèles de segmentation plus avancés comme DeepLab ou des modèles de segmentation basés sur des réseaux de neurones, des cadres d'apprentissage profond open source comme PyTorch ou TensorFlow seront utiles.

Il est en revanche, possible d'utiliser un modèle d'estimation de pose décrit à la section précédente. L'ajout de la segmentation avant un modèle d'estimation de pose peut être utile si l'environnement est dynamique ou non structuré, par exemple une fluctuation de l'arrière-plan ou un arrière-plan chargé. Cela permet d'améliorer la précision et d'éviter des erreurs de l'algorithme d'estimation de pose.

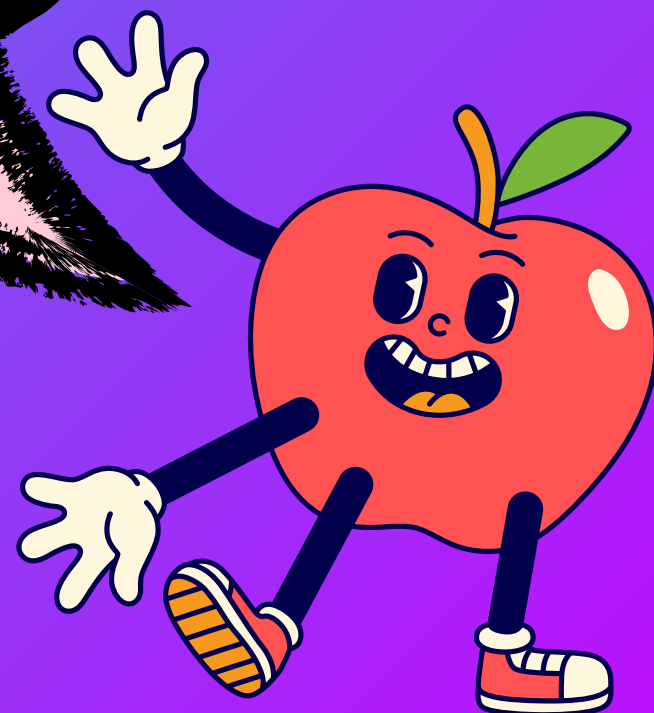
POUR ALLER PLUS LOIN

Mojo (2008) by Christian Moeller est une installation d'art cinétique qui intègre la détection d'objets et l'interactivité en temps réel. Il est possible que Mojo utilise un système basé sur la vision pour détecter la présence et les mouvements des personnes à proximité.

Lien : <https://www.publicartinpublicplaces.info/mojo-2008-by-christian-moeller>

Rafael Lozano-Hemmer a deux oeuvres : **Tape Recorders**, **Subsculpture 14** ou **Vanishing Points**, les deux utilisent du suivi de personnes, l'un détecte le temps passé dans une salle, l'autre le suivi de la tête des gens.

Lien : <https://www.wired.com/2012/02/tape-recorders-art-installation/>



SOUSTRACTION DE L'ARRIÈRE-PLAN

☀ POINTS À RETENIR

- Exemples : MOG2 (Mixture of Gaussians) et KNN (K-Nearest Neighbors) dans OpenCV, OpenCV for Unity, Running Average
- Données de sortie : contour des éléments détectés
- Solution gratuite, il n'y a que la caméra à acheter
- Accessibilité : utilise une caméra RGB standard (webcam, caméra HD, etc.).
- Ne fait pas de la capture de mouvement, mais du suivi de silhouette
- Connexion avec d'autres logiciels (plug-in) possible avec une solution personnalisée
- Des outils prêts à l'emploi sont disponibles (comme OpenCV for Unity), sinon un minimum de compétences en programmation est nécessaire pour des solutions personnalisées
- Pas de marqueurs, pas de capteurs sur le corps de la personne captée
- Possibilité de temps réel
- Sensibilité aux changements de lumière
- Le fond doit rester relativement constant pour de meilleurs résultats.
- Pas aussi précis que les caméras infrarouges pour des suivis complexes

☀ PRINCIPE

La soustraction de l'arrière-plan ou Background Subtraction est une méthode spécifique de segmentation basée sur le mouvement. Elle utilise une caméra RGB standard pour détecter les mouvements en comparant les variations entre l'image actuelle et une image de référence (souvent appelée "background"). Cette méthode soustrait une image de fond statique (ou mise à jour dynamiquement) de l'image en temps réel pour isoler les objets ou les personnes en mouvement. Elle permet de détecter les contours ou les formes d'objets dans la scène sans avoir besoin d'une caméra infrarouge ou de profondeur. On peut nommer les algorithmes MOG2 (Mixture of Gaussians) et KNN (K-Nearest Neighbors) utilisé dans OpenCV, ou Running Average.



*Dispositif de capture de mouvement interactif
Crédit photo : Francisco Cruz, les 7 Doigts*

ALGORITHME DE CLASSIFICATION DE MOUVEMENTS/OBJETS

☀ POINTS À RETENIR

- Exemples : ResNet, VGG, MobileNet ou SVM (Support Vector Machine)
- Données de sortie : labels ou classes détectés
- Solution gratuite, il n'y a que la ou les caméras à acheter
- Compétence en programmation nécessaire
- Ne fait pas de la capture de mouvement, mais de la classification de mouvement
- Intérieur et extérieur
- Pas de marqueurs, pas de capteurs sur le corps de la personne captée
- Possibilité de temps réel mais cela dépend de la complexité, de la capacité informatique disponible et des besoins en précision et rapidité
- Connexion avec d'autres logiciels (plugin) possible avec une solution personnalisée
- Capacité informatique : processeur multicœur, peuvent fonctionner avec ou sans GPU pour des applications de base, nécessite un GPU pour des tâches de classifications complexes ou en temps réel



Artiste dont le mouvement capté sert à générer des effets visuels
Crédit photo : HUPR et Les 7 Doigts

☀ PRINCIPE

Dans ce champ d'analyse à partir de la vision, nous avons aussi, les algorithmes de classification, qui contrairement aux algorithmes de segmentation ne s'attardent pas sur l'analyse de chaque pixel d'une image pour en donner un sens, mais cherchent à étiqueter et classer une image dans son ensemble. Ces algorithmes sont entraînés à faire la différence entre un « chien » et un « camion » à partir de la reconnaissance de forme sur l'image entière. Pour ce faire, les algorithmes de classification utilisent des données sur lesquelles ils ont été entraînés en entrée dans le but de prédire la probabilité que les données suivantes à reconnaître entrent dans l'une des catégories prédéterminées, « chien » ou bien « camion ». L'algorithme de classification est donc utilisé pour identifier la catégorie des nouvelles observations sur la base des données d'apprentissage.

La classification peut être utilisée pour reconnaître des mouvements spécifiques sur une séquence d'images ou de vidéos. Par exemple, après avoir segmenté une silhouette humaine en mouvement, un algorithme de classification peut identifier si la personne est en train de : marcher, courir, sauter, s'asseoir ou lever les bras. De même que la segmentation, la classification de mouvement ne permet pas de localiser chaque partie du corps, mais permet de classer différents types de mouvement. Il est également possible qu'en fonction de l'objectif du projet, cela soit suffisant. Si on reprend l'exemple du jeu interactif, il est possible de concevoir que le simple fait de reconnaître les types de mouvement soit suffisant pour activer des techniques de jeu.

Il est possible d'utiliser des modèles pré-entraînés comme ResNet, VGG, MobileNet ou un modèle plus simple comme un SVM (Support Vector Machine) pour classer des objets ou des actions[Ui1] . Mediapipe et YOLO ont également des modèles de classification d'image. Des compétences en programmation (Python), en architecture réseaux et en apprentissage automatique/apprentissage profond sont nécessaires pour mettre en place cette solution.

POUR ALLER PLUS LOIN

DeepDream est un programme de vision par ordinateur créé par Google fonctionnant sur les algorithmes de classification d'image. Il génère des images surréalistes en améliorant et en exagérant les motifs qu'il identifie dans une image.

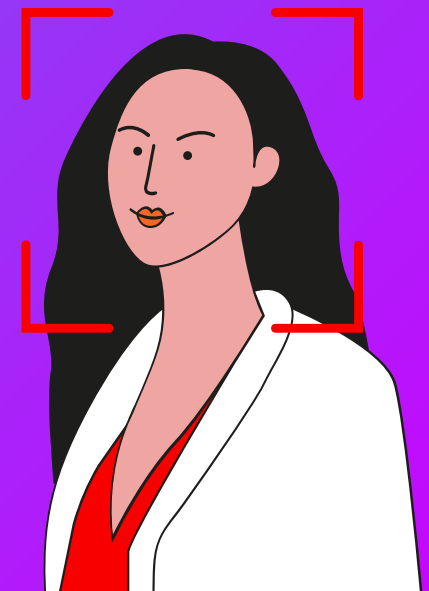
Lien : <https://research.google/blog/inceptionism-going-deeper-into-neural-networks/>



ALGORITHME DE RECONNAISSANCE

☀ POINTS À RETENIR

- Exemples : YOLO, Faster R-CNN, CNN+LSTM, 3D CNN
- Données de sortie : boîte englobante de l'élément détecté
- Solution gratuite, il n'y a que la ou les caméras à acheter
- Compétence en programmation nécessaire
- Ne fait pas de la capture de mouvement, mais de la reconnaissance de mouvement
- Intérieur et extérieur
- Pas de marqueurs, pas de capteurs sur le corps de la personne captée
- Possibilité de temps réel mais cela dépend de la complexité, de la capacité informatique disponible et des besoins en précision et rapidité
- Connexion avec d'autres logiciels (plug-in) possible avec une solution personnalisée
- Capacité informatique : processeur multicœur, peuvent fonctionner avec ou sans GPU pour des applications de base, nécessite un GPU pour des tâches de reconnaissances complexes ou en temps réel



☀ PRINCIPE

Il est possible d'avoir de la reconnaissance d'objets (par exemple, reconnaître une voiture parmi d'autres objets dans une image), de la reconnaissance faciale (identifier une personne en comparant son visage avec une base de données), de la reconnaissance d'actions (par exemple, reconnaître une personne qui court ou saute) et de la reconnaissance de gestes (par exemple, un signal de la main ou un signe de la tête). La reconnaissance vise à identifier un objet spécifique, une personne, un mouvement ou un geste qui se trouve dans l'image ou la vidéo, en se basant sur des modèles ou des bases de données existantes. La reconnaissance de gestes pourrait être utilisée dans les systèmes interactifs où les utilisateurs effectuent des gestes spécifiques qui sont ensuite reconnus par le système pour déclencher une action comme dans un jeu interactif.

Pour la reconnaissance d'objets ou d'actions ou de gestes, les modèles comme YOLO, Faster R-CNN, CNN+LSTM ou 3D CNN sont souvent utilisés. De même que pour la classification, des compétences en programmation (Python), en architecture réseaux et en apprentissage automatique/apprentissage profond sont nécessaires pour mettre en place cette solution.



Un autre exemple d'artiste dont le mouvement capté sert à générer des effets visuels, HUPR et Les 7 Doigts

En résumé, la différence entre segmentation, classification et reconnaissance peut se comprendre comme suit :

- La segmentation se concentre sur le découpage d'une image en plusieurs parties ou objets, en assignant un label à chaque pixel pour isoler des éléments spécifiques.
- La classification se concentre sur l'étiquetage général d'une image.
- La reconnaissance cherche à identifier un objet ou une personne spécifique dans une image ou une vidéo, en comparant avec des modèles ou des bases de données connus.

Ces trois techniques peuvent potentiellement identifier une personne, mais chacune utilise une méthode différente.



Exemple d'une utilisation de groupe de la capture de mouvement. Crédit photo: Kun Chang, projet LabO (Moment Factory & Les 7 Doigts)

SYSTÈMES VOLUMÉTRIQUES

☀ POINTS À RETENIR

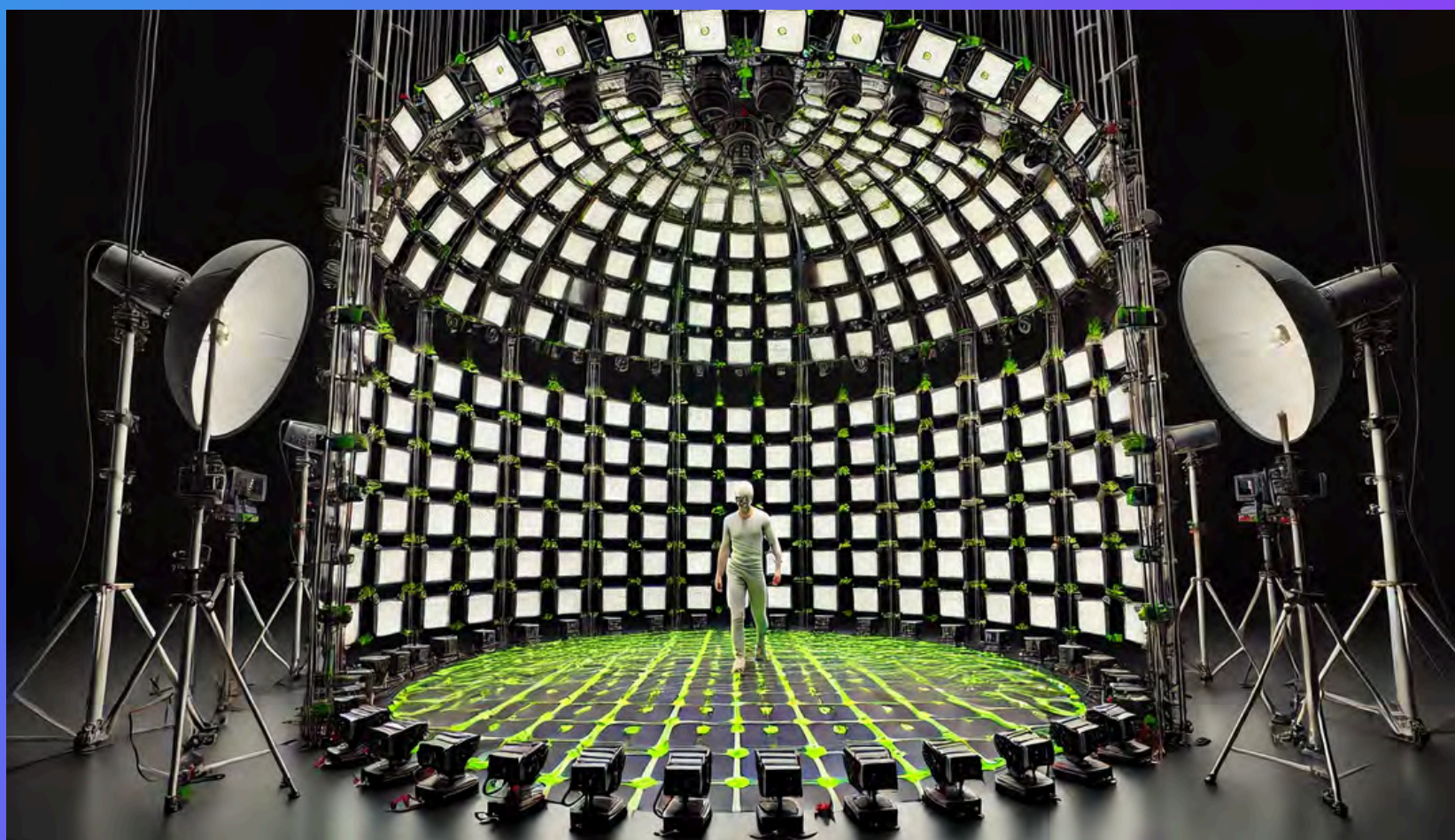
- Exemples : 4DViews, Dephtkit
- Données de sortie : modèles 3D volumétriques
- Minimum 8 à 10 caméras mais cela peut être un grand nombre
- Coût : minimum 100 000\$, ça peut aller à plusieurs centaines de milliers voire millions de dollars pour une installation de haute qualité
- Logiciel dédié et payant
- Intérieur
- Ne fait pas de capture de mouvement mais de la capture volumétrique
- Idéal pour des postures statiques et les détails réalistes
- Idéal pour les applications nécessitant un rendu à 360 degrés (réalité virtuelle ou augmentée par exemple)
- Compétences techniques avancées
- Nécessite un studio avec une installation de caméras synchronisées
- Nécessite un système de gestion des données robuste parce que cela produit beaucoup de données volumineuses
- Capacité informatique : processeur multicœur haut de gamme, carte graphique extrêmement puissante, stockage SSD de grande capacité et système de stockage additionnel

■ PRINCIPE

Autre que ces exemples d'application de vision par ordinateur, il existe d'autres technologies qui utilisent des algorithmes pour la reconnaissance de mouvement ou d'image mais se démarquent par leur dispositif, à l'instar de la vidéo volumétrique, aussi appelée capture volumétrique (Vocap). La Vocap est de plus en plus considérée comme la prochaine étape du développement dans le domaine de la production médiatique. Dans le contexte du développement rapide des marchés de la réalité virtuelle (RV) et de la réalité augmentée (RA), la vidéo volumétrique est en train de devenir une technologie décisive. Bien qu'à première vue les dispositifs de capture volumétrique partagent certaines similarités avec la capture de mouvement optique, notamment la nécessité d'un studio pour installer la multitude de caméras dédiées à capter et produire une énorme quantité de données dans un espace consacré afin d'émuler un objet en 3D, nous verrons que ces deux dispositifs diffèrent autant dans leur utilisation que dans les résultats obtenus.

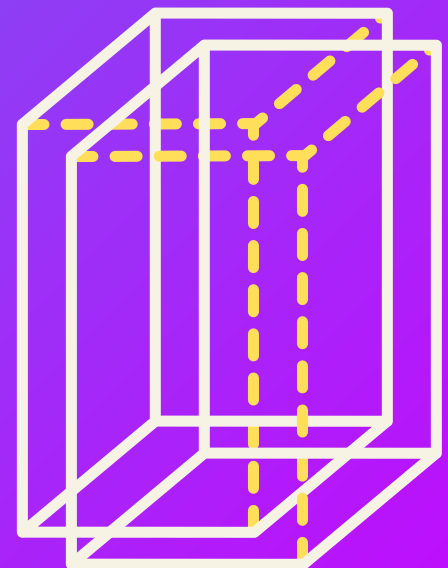
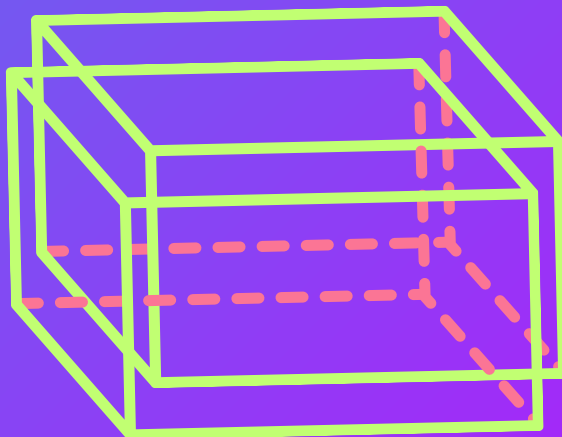
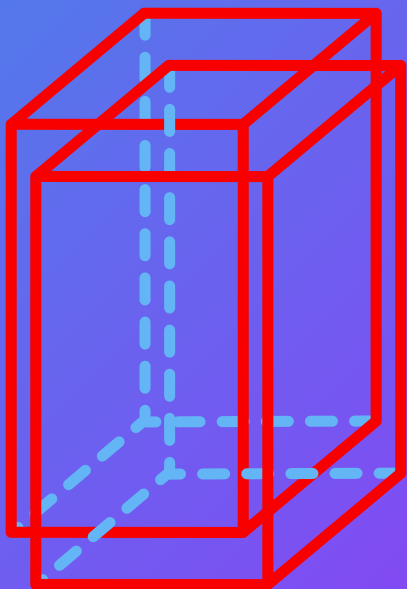
En effet, plus que le mouvement, la Vocap a pour vocation comme son nom l'indique, de capturer le volume d'une personne, d'un objet, ou d'un environnement réel dans un espace tridimensionnel qui pourra être visible sous n'importe quel angle après post-production et intégré à du contenu numérique visuel. Le processus consiste à scanner en 3D des objets à l'aide d'une matrice composée de plusieurs caméras disposées en « cercle » de façon à capter simultanément chacun des angles qui composent le volume d'une personne.

Ces caméras peuvent être de plusieurs types, une caméra numérique à haute définition peut suffire, bien que souvent les dispositifs de capture volumétrique privilégieront des caméras de profondeur, comme celles utilisées par la capture de mouvement, mais en plus grand nombre, afin de modéliser le volume et pas seulement la triangulation du mouvement par des capteurs. Par la nature de son fonctionnement, la capture volumétrique est encore limitée dans la captation de mouvement ample et privilégiera les postures plutôt statiques qui permettront un rendu plus réaliste.



Représentation d'un artiste par intelligence artificielle dans un système de captation volumétrique

C'est d'ailleurs là-dessus que réside la réelle différence avec la capture de mouvement « traditionnelle », dans l'utilisation qui peut être faite de la technologie de capture volumétrique. Contrairement à la mocap qui capture le mouvement pour pouvoir en transférer le réalisme sur une entité numérique, la Vocab est particulièrement intéressante lorsque l'on souhaite capturer des détails réalistes jusque dans l'apparence ou la texture d'une personne. La ressemblance d'un acteur, l'authenticité de sa performance et sa garde-robe peuvent être capturées au cours d'une même captation grâce à un rendu très photoréaliste de l'être humain capté. De plus, comme pour la vision par ordinateur, la capture volumétrique ne nécessite pas de matériels supplémentaires, seulement des caméras et un espace dédié qui obligera dans la plupart des cas à faire appel à une compagnie qui dispose de la place requise pour déployer le dispositif important de la technologie de capture volumétrique. Les données volumétriques brutes capturées nécessitent un traitement important en post-production, pour corriger les erreurs de captation, assembler les images et optimiser le rendu final. Cette étape requiert souvent des compétences techniques et un matériel informatique avancé.





POUR ALLER PLUS LOIN

"Queerskins: A Love Story"

Une expérience de réalité virtuelle immersive qui combine narration et capture volumétrique. Les acteurs ont été capturés par captation volumétrique, permettant aux spectateurs d'interagir avec des personnages volumétriques dans un environnement virtuel.

Lien : <https://vr.queerskins.com/a-love-story>

NUAGE DE POINTS

☀ POINTS À RETENIR

- Exemples : Augmenta, PCL, Open3D
- Données de sortie : Nuages de points 3D
- 1 à 2 caméras LiDAR 3D peuvent suffire pour capturer un espace
- Entre 15 000\$ et 40 000\$ la caméra LiDAR 3D avec la plateforme Augmenta (payante)
- Grande précision
- Grande distance de captation, maximum 100 m
- Permet une capture complète sur 360°
- Idéal pour la cartographie en temps réel et le suivi de mouvements dans des environnements de grande échelle
- Jour et nuit, intérieur et extérieur
- Moins performant sous la pluie, le brouillard ou dans des conditions de lumière extrêmes
- Ne fait pas de la capture de mouvement, mais une cartographie 3D de l'environnement[Ui1]
- Les données brutes issues des nuages de points requièrent souvent une analyse et un filtrage pour être exploitable
- Capacité informatique : processeurs multicœurs, une carte graphique puissante et beaucoup de RAM (64 à 128 Go), un système de stockage rapide et de haute capacité comme un SSD



Représentation d'un artiste par intelligence artificielle avec le nuage de point projeté



☀ PRINCIPE

Un nuage de points est un ensemble de points données dans un espace tridimensionnel, représentant la surface d'un objet, d'un environnement, ou d'une scène. Chaque point est défini par ses coordonnées (x, y, z), et parfois des informations supplémentaires comme la couleur ou l'intensité. La technique de nuages de points peut être générés par plusieurs technologies comme la photogrammétrie (analyse d'images 2D), certaines caméras à profondeur (Kinect, RealSense), les scanners 3D et le LiDAR. Il est possible d'associer la technologie de nuages de points avec des algorithmes de traitement comme la segmentation, mais aussi l'étiquetage d'élément comme des personnes ou des objets de manière très précise, puisque chaque micro-laser permettra de segmenter en milliers l'environnement. Les avantages de cette technologie sont la précision spatiale, la robustesse (fonctionne dans des environnements complexes), l'indépendance de l'éclairage et le fait qu'il n'y ait pas besoin de marqueurs ou de vêtements spécifiques. La technologie de nuage de point n'a pas été conçue pour l'estimation de pose (identification des positions spécifiques des articulations ou des structures squelettiques). Cependant, le suivi d'une silhouette est tout à fait possible.

En utilisant LiDAR (ou d'autres caméras), nous vous recommandons la technologie Augmenta qui offre une plateforme puissante pour le développement d'expériences interactives en temps réel. Augmenta permet de connecter et synchroniser plusieurs caméras avec nuage de points (LiDAR 2D/3D ou infrarouge ou à profondeur avec infrarouge). Il est possible de détecter et de suivre précisément des présences (personnes ou objets). Par la suite, il est possible d'y coupler un déclenchement d'effet, par exemple lors qu'une personne entre dans une zone spécifique, cela déclenche une vidéo.

Dans les solutions à programmer soi-même, pour faire le suivi de points, les grandes étapes pour mettre en place sont : le prétraitement des données, la segmentation de la silhouette humaine et le suivi de la silhouette. Des outils comme PCL, bibliothèque open-source pour le traitement, filtrage et segmentation des nuages de points 3D, et Open3D, bibliothèque pour visualiser et traiter des données de nuage de points et faire le suivi des objets, sont communément utilisés. Des compétences en vision par ordinateur, traitement des données 3D et programmation sont nécessaires. Des récentes recherches décrivent comment les nuages de points LiDAR peuvent être exploités pour suivre et détecter efficacement les mouvements humains.



*Projet LabO
(Moment
Factory & Les 7
Doigts) utilisant
le LiDAR pour
détecter la
position des
personnes.
Crédit photo :
Kun Chang.*



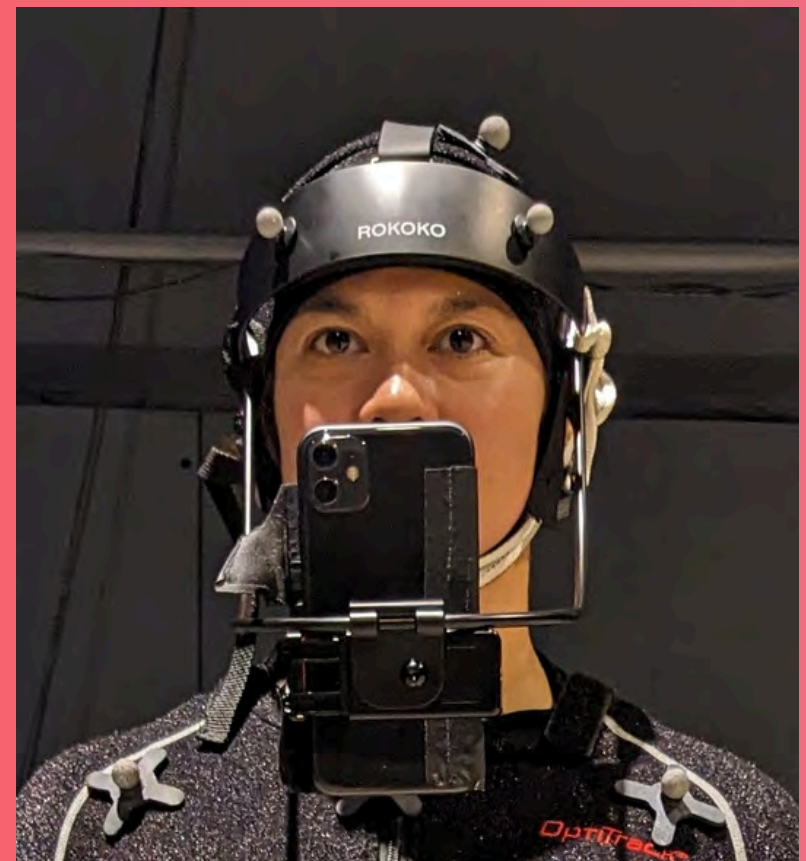
AUTRES MÉTHODES

FACE TRACKING

● PRINCIPE

Dans la continuité des technologies de capture de mouvement que nous avons pu voir tout au long de cette taxinomie, il existe des dispositifs qui permettent de prolonger et détailler l'expérience de la transposition du mouvement, notamment des mouvements complexes qui demandent une haute précision des points d'articulation que les dispositifs de capture de mouvement standards ne peuvent capter seuls. L'une d'entre elles est la capture de mouvement facial, qui comme son nom l'indique, permet de capturer avec fidélité les micromouvements du visage ; elle permet de saisir des nuances et des émotions subtiles qu'il est difficile de reproduire avec une animation numérique.

Plusieurs méthodes peuvent être utilisées selon le niveau de détails voulu. Les méthodes les plus courantes sont celles avec et sans marqueurs. La capture de mouvement facial à l'aide de marqueurs reprend le même procédé que pour la capture des mouvements du corps, elle consiste à fixer de petits points réfléchissants ou magnétiques sur le visage de l'acteur et à les suivre à l'aide de caméras et de capteurs. Pour la méthode sans marqueurs, le visage sera capté par des dispositifs de vision par ordinateur, dont l'algorithme aura été spécifiquement entraîné à reconnaître les mouvements du visage, plutôt que ceux d'un corps en mouvement.



*Exemple d'une captation de
Face Tracking
Crédit photo : Les 7 doigts*

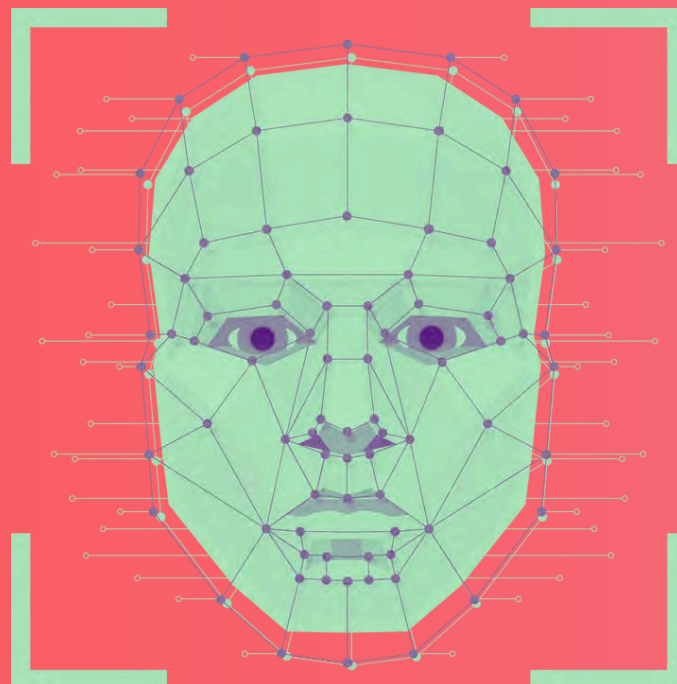
Si une bonne précision est nécessaire, la capture de mouvement facial sans marqueur nécessite généralement de capter le visage de façon très précise et stable, et donc d'équiper les personnes avec un casque ou autre support permettant de conserver une caméra toujours face à leur visage pendant la captation. Cependant, pour des applications moins exigeantes ou grâce aux avancées technologiques récentes, il est possible de capturer les mouvements faciaux sans casque ni support spécifique, à condition d'avoir un bon système de caméras et des algorithmes performants.

Exemples de solutions professionnelles :

Faceware Technologies, Dynamixyz, OptiTrack, Vicon – Cara, Cubic Motion

Exemples de solutions consommateurs :

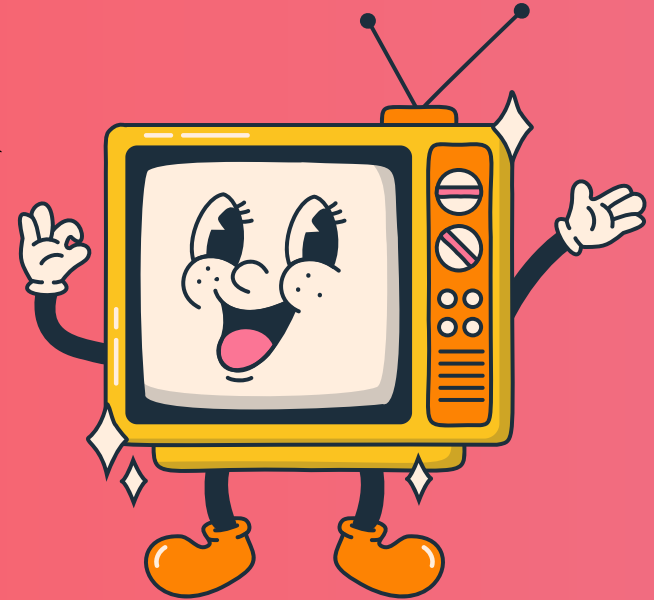
Apple - ARKit Face Tracking, Microsoft - Face Tracking SDK, FaceRig (maintenant Animaze par Holotech Studios, Kinect avec SDK, Visage Technologies – FaceTrack, DeepMotion - Animate 3D, Zappar - Face Tracking SDK, Nvidia - Nvidia Face Tracking SDK, Hyprsense, Zign Creations - Zign Track, Glassbox Technologies



POUR ALLER PLUS LOIN

"As We Are" par Matthew Mohr : Une très grande sculpture interactive en forme de tête humaine géante qui utilise des algorithmes de reconnaissance faciale pour afficher le visage du public sur la sculpture après avoir capturé leur image.

Lien : <https://dirdim.com/portfolio/as-we-are-interactive-sculpture/>



HAND TRACKING

● PRINCIPE

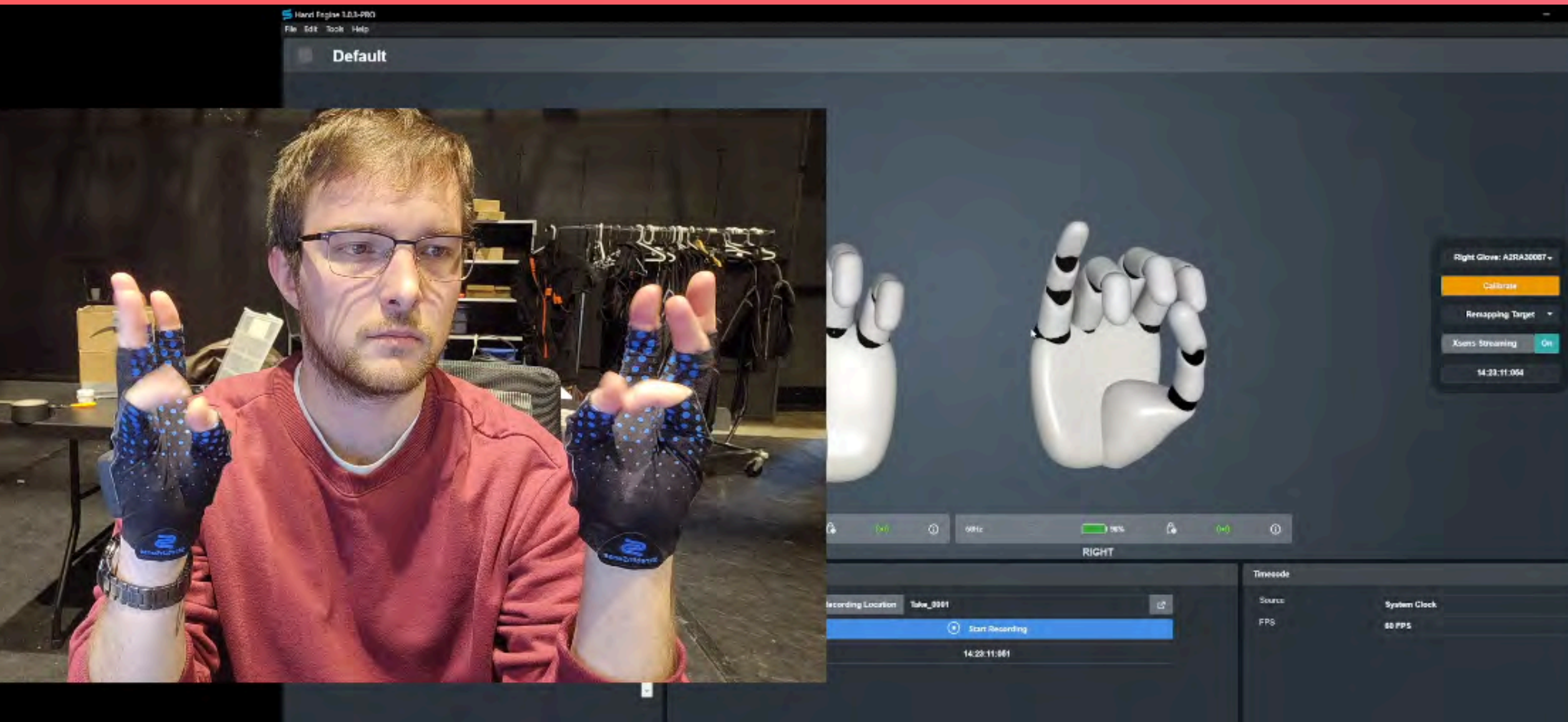
À ces technologies, nous pouvons ajouter les gants de capture de mouvement, qui auront pour tâche de reproduire les mouvements d'une ou plusieurs mains. Avec le visage, les mains font parties des éléments du corps qui sont les plus difficiles à modéliser, étant donné la complexité des articulations qui composent leur mouvement et les différentes poses qu'elles peuvent prendre. Aussi, tout comme la capture de mouvement facial, les gants offrent un rendu plus réaliste grâce à l'intégration des micromouvements dans la captation. Les systèmes inertiels sont souvent favorisés pour ce genre d'application, puisque les capteurs permettent une bonne incorporation des rotations de la main. Il est également possible de suivre et d'analyser les mouvements des mains en temps réel, souvent sans avoir besoin de marqueurs ou de dispositifs portés par l'utilisateur. Leap Motion (Ultraleap) est une technologie qui utilise des caméras infrarouges de haute précision pour le suivi des doigts et de la paume mais qui fonctionne bien dans des environnements contrôlés. MediaPipe, bibliothèque open source pour le traitement utilisant des caméras RGB, a également un modèle qui détecte les articulations de la main. Enfin pour les applications de réalités virtuelles, le Oculus Quest (Meta) utilise les caméras RGB intégrées du casque VR pour suivre les mains. La technologie est moins précise que Leap Motion mais est intégrée directement dans un casque VR.

Exemples à partir de caméra :

Leap Motion, MediaPipe, Oculus Quest, Theia, Microsoft - Azure Kinect D, ARKit (Apple), Stereolabs - ZED Camera

Exemples avec des gants :

Perception Neuron, Xsens, StretchSense, Manus - Manus Prime II, Manus Quantum Metagloves, Noitom - Hi5 VR Gants, Synertial - Gants Inertiels IntertiaXon, SenseGlove - SenseGlove Nova



*Exemple d'une application de la technologie de Hand Tracking
Crédit photo : Les 7 doigts*



ULTRASONS

● PRINCIPE

En dehors des trois grandes familles que nous avons présentées précédemment, il existe d'autres technologies que l'on peut associer à la capture de mouvement, à l'instar des systèmes basés sur l'ultrason, une méthode qui consiste à envoyer en continu une pluralité d'ondes ultrasoniques ayant des caractéristiques de spectre de fréquence différentes à l'objet ou l'environnement que l'on souhaite capturer, à partir d'une pluralité de positions différentes. Des capteurs disposés sur l'objet à différentes positions permettent de recevoir les ondes ultrasoniques par le dispositif de réception et convertir les ondes ultrasoniques en signaux électriques. La différence de temps de vol entre les ondes permettra d'obtenir la forme de l'objet et ses changements dans l'espace, son mouvement. Bien que la capture de mouvement par ultrasons soit réalisable et ait été utilisée avec succès dans certaines applications, ses limites en termes de précision et de portée sont telles qu'elle n'est pas adaptée pour la capture de mouvement plus complexe. La compagnie Polhemus propose des solutions de suivi de mouvement, y compris des systèmes basés sur les ultrasons combinés à des technologies électromagnétiques. Beaucoup des solutions existantes actuelles, comme Brigade ou Leuze, fournissent des capteurs industriels pour la détection d'objet.

Exemple :
Polhemus

Exemples pour des capteurs :
Brigade, Leuze, PING))) Ultrasonic Distance Sensor, Balluff

ÉLECTROMAGNÉTIQUE

☀ PRINCIPE

Une autre de ces technologies, la capture de mouvements électromagnétiques, utilise des capteurs placés sur le corps pour mesurer le champ magnétique à basse fréquence généré par une source émettrice. Les capteurs et la source peuvent être reliés par câble, mais aussi sans, à une unité de contrôle électronique qui établit une corrélation entre les positions signalées dans le champ. Les unités de contrôle électronique sont mises en réseau avec un ordinateur hôte qui utilise un pilote logiciel pour représenter ces positions et rotations dans l'espace 3D. Il est possible de placer plusieurs capteurs sur le corps d'une personne comme un système inertiel afin de capturer le mouvement d'une personne. Plusieurs limitations, interférences électromagnétiques, portée limitée, sensibilité aux objets métalliques et contraintes matérielles, rendent cette technologie idéale pour des environnements contrôlés, mais moins adaptée à des contextes complexes. La compagnie Polhemus offre par exemple des systèmes de suivi des mouvements par électromagnétisme.

Exemple :
Polhemus

Exemples pour des capteurs :
3D Guidance de NDI



*Illustrations d'un dispositif
électromagnétique pour les mains
généré par intelligence artificielle*





APRÈS LA CAPTATION



*Exemple de
visuel avec
Notch.
Crédit photo :
Francisco Cruz,
Les 7 Doigts*

Les données captées par les systèmes présentés dans les sections précédentes peuvent ensuite être traitées en post-production ou en temps réel par des logiciels pour créer des effets. Ces effets peuvent être opérés et/ou développés par des développeurs, des artistes 3D et des créateurs de contenu. Ils peuvent par exemple créer des personnages animés qui reproduiront les mouvements captés. Il s'agit souvent d'Unreal Engine, de Unity, de Maya d'Autodesk, de Blender ou de Houdini. Unreal Engine est un moteur de rendu en temps réel avec des outils de modélisation et d'animation, principalement utilisé pour le développement de jeux vidéo, mais est de plus en plus utilisé dans des expériences immersives ou interactives.

Unity est un moteur de développement multiplateforme gratuit permettant de créer des jeux vidéo, des expériences interactives en 3D/2D, des applications de réalité augmentée (AR) et virtuelle (VR), ainsi que d'autres contenus immersifs. Maya est un logiciel professionnel de modélisation, d'animation, et de rendu 3D, largement utilisé dans l'industrie du cinéma et des jeux pour créer des personnages ou des effets spéciaux. Blender est un logiciel open-source de création 3D qui propose des outils pour la modélisation, le texturage, l'animation, le rendu, et la composition. Houdini est un logiciel de création 3D spécialisé dans la création des effets spéciaux procéduraux, comme des explosions, des simulations de fluides et des animations complexes.



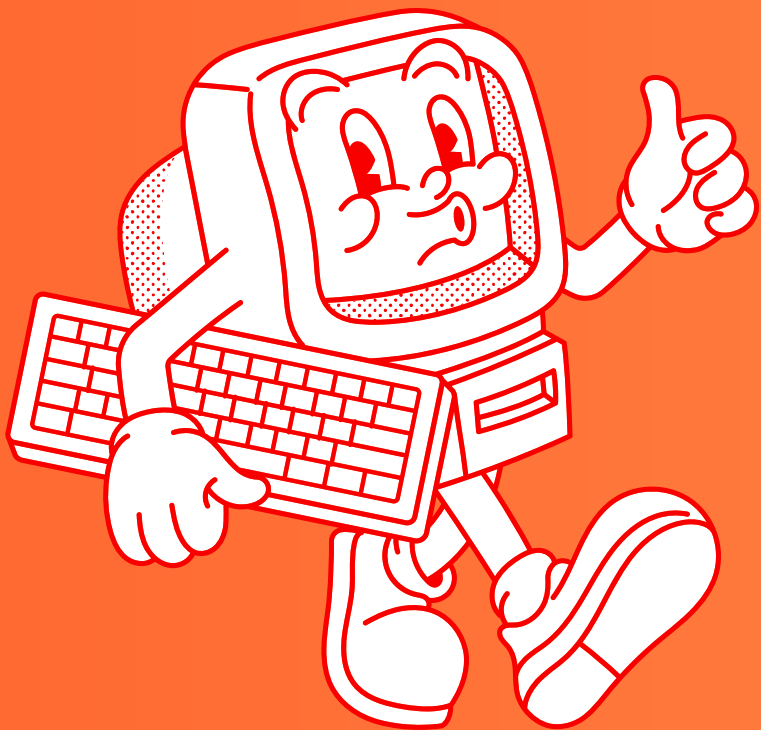
*Exemple de
visuel avec
Notch.
Crédit photo :
Les 7 Doigts.*

Les données captées peuvent aussi servir de données d'entrées pour des logiciels de création multimédia comme Notch, VVVV, TouchDesigner ou SMODE. Ce sont des outils puissants pour générer des contenus visuels, sonores ou scénographiques interactifs et des installations multimédias immersives. Ils offrent une gamme étendue de possibilités, comme la création de visuels, la génération de particules et d'effets spéciaux, la projection mapping et la scénographie interactive. Ils peuvent intégrer des données en temps réel provenant de diverses sources, comme des données de capture de mouvement, des flux vidéo, des nuages de points, et créer des animations visuelles en direct.



*Exemple d'une utilisation de Notch
Crédit photo : Francisco Cruz, Les 7 Doigts*

Notch est un logiciel de création visuelle en temps réel spécialement conçu pour produire des effets visuels, du contenu interactif et des animations immersives pour les spectacles en direct. VVVV est un logiciel de programmation adapté pour le prototypage et le contrôle d'installations interactives, de visuels génératifs et de performances multimédias en temps réel. TouchDesigner est une plateforme de création visuelle en temps réel qui permet de produire des installations interactives, du mapping vidéo et des visuels immersifs, et est compatible avec de nombreuses sources de données. Notch, VVVV et TouchDesigner utilise une interface en nodal (système visuel d'interaction permettant de manipuler des flux de données). SMODE est une plateforme de création et de diffusion visuelle en temps réel qui combine des outils de composition, de projection mapping et d'effets visuels interactifs.



CONCLUSION



CONCLUSION

À travers ce guide, nous avons révélé la diversité des technologies de capture de mouvement. Chaque caméra, système et logiciel a ses spécificités et son potentiel unique. Ces outils offrent aux artistes et créateurs un vaste éventail de possibilités pour enrichir et transformer leurs œuvres artistiques.

Ce guide a pour vocation de rendre ces outils accessibles et utilisables pour tous, des jeunes créateurs aux artistes expérimentés, afin que chacun puisse s'appropriier ces outils et en tirer pleinement parti. En démocratisant la capture de mouvement, on donne aux arts vivants la possibilité d'amplifier et de réinterpréter les corps, les gestes et les émotions dans le numérique.

Intégrer la capture de mouvement dans les performances, c'est aussi explorer de nouvelles relations entre le public et l'œuvre, où le corps humain et les effets numériques se rencontrent pour créer des expériences nouvelles. Ces technologies nous poussent à réfléchir : jusqu'où intégrer le numérique sans perdre l'humain et l'art ? En rendant la capture de mouvement plus accessible, on ouvre la porte à un dialogue entre le monde réel et virtuel, entre l'artiste et la machine. Quelles histoires pourrons-nous raconter avec ces nouveaux moyens technologiques ? Comment l'expérience du public évoluera-t-elle ? En poursuivant cette démocratisation, nous offrons aux artistes les moyens d'innover. Nous espérons seulement que les artistes et les créateurs puissent préserver leur capacité à émouvoir, toucher, surprendre ou inspirer, même avec ces nouveaux outils.



REFERENCES

Colyer, S., Evans, M., Cosker, D. (2018). A Review of the Evolution of Vision-Based Motion Analysis and the Integration of advanced Computer Vision Methods towards Developing a Markerless System. *Sports Med - Open* 4(24), 1-15.

Dower, J. et Langdale, P. (2022). *Performing for motion capture : A guide for practitioners*. Bloomsbury Publishing.

Galinium, M., Yapri, J., Purnama, J. (2019). 27 markerless motion capture for 3D human model animation using depth camera. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(3), 1300 - 1309.

Jürgens, S, Correia, N., Masu, R. (2020). Designing Glitch Procedures and Visualisation Workflows for Markerless Live Motion Capture of Contemporary Dance. *7th International Conference on Movement and Computing(MOCO '20)*, 1-8.

Kitagawa, M et Windsor, B. (2008). *Mocap for artist*. Elsevier Inc.

Ribeiro, C., Kuffner, R, Fernandes, C. (2022). Capturing and visualizing 3D dance data: Challenges and lessons learnt. *Dans, Dance Data, Cognition, and Multimodal Communication*.

Schreer, O., Feldmann, I., Renault, S. (2019). Capture and 3D Video Processing of Volumetric Video. Conference: *2019 IEEE International Conference on Image Processing (ICIP)*, 4310 - 4313.

Strutt, D. (2022). Motion capture and the digital dance aesthetic: Using inertial sensor motion tracking for devising and producing contemporary dance performance. *Dans, Dance data, cognition and multimodal communication*.

Sun, Y, Orgs, G., Greaves, D. (2023). Using Wearable Sensors to Measure Interpersonal Synchrony in Actors and Audience Members During a Live Theatre Performance. *Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies*, 7(1), 1 - 29.

Avatar

Représentation virtuelle d'une personne utilisée dans les jeux, un univers virtuel, les réseaux sociaux ou le métavers. On parle généralement d'avatar pour désigner un personnage 3D, même si cela peut aussi désigner une image 2D.

Algorithme

Séquence de règles opératoires dont l'application permet de résoudre un problème énoncé au moyen d'un nombre fini d'opérations.

Calibration

Ajustement et configuration des systèmes de capture pour garantir des mesures précises et fiables.

Capteur

Dispositif permettant de mesurer une grandeur physique comme la position, la température ou la pression et la convertit en un signal électrique ou numérique interprétable par un système.

Drift (dérive)

Erreur cumulative dans les données de position causée par les limitations des capteurs inertiels.

Frame Rate

Nombre d'images ou de mesures capturées par seconde par le système qui s'exprime souvent en fps (frame par seconde ou image par seconde).

Fitting

Alignement initial des données de capture sur un modèle ou un squelette spécifique. Le fitting est une étape préalable à la préparation des données.

Intelligence artificielle

Ensemble d'algorithmes et de méthodes de calculs visant à simuler certains traits de l'intelligence humaine

Interpolation

Création ou calcul de points de données intermédiaires entre des positions enregistrées à des instants spécifiques, utilisé notamment pour remplacer des données manquantes de données manquantes.

Latence (en capture de mouvement)

Délai entre le moment où un mouvement se produit dans le monde réel et le moment où ce mouvement est détecté et visualisé

Latence (en informatique)

Délai entre une action ou une requête effectuée par un système (ou utilisateur) et la réponse ou l'exécution effective de cette action par le système.

Marqueur

Point ou objet physique placé sur le corps pour en effectuer le suivi de mouvement. Les marqueurs sont souvent placés sur des articulations ou des parties spécifiques du corps et peuvent être passifs (réfléchissants la lumière infrarouge) ou actifs (émettant leur propre lumière).

Occlusion

Situation qui arrive lorsqu'une partie du corps suivi est masquée par un autre élément comme une autre partie de son corps, une autre personne ou un objet.

Noise (bruit)

Perturbations ou erreurs dans les données capturées.

Rigging

Processus d'attachement d'un squelette numérique à un personnage 3D pour l'animer à partir des données capturées.

Retargetting

Action qui consiste à transférer les mouvements capturés d'un squelette source (souvent celui d'une personne captée) à un squelette cible (un personnage virtuel). Cela implique d'adapter les proportions et les contraintes du squelette cible pour reproduire fidèlement les mouvements, tout en tenant compte des différences de morphologie.

Swapping

Situation qui arrive lorsque la trajectoire de deux marqueurs qui se croisent est mal interprétée par le système de capture et donc les marqueurs sont mal étiquetés. Exemple : le marqueur du poignet est interverti par celui de la hanche lorsque la main passe devant la hanche.

Triangulation

Processus permettant de déterminer les positions en 3D d'un point dans l'espace à partir de ses projections 2D captées par au moins deux caméras ou capteurs positionnés différemment.

Volume de capture

Zone dans laquelle les mouvements peuvent être détectés.

Workflow (flux de travail)

Ensemble structuré des étapes et des processus des opérations techniques et des tâches humaines pour enregistrer, traiter et intégrer les mouvements capturés.