

Etude exploratoire d'un assistant digital d'aide au montage basé sur la projection en réalité augmentée

An Exploratory Study of a Projection-based Augmented Reality Digital Assistant to Facilitate Assembly

Simon Ruffieux*

simon.ruffieux@hes-so.ch

University of Applied Sciences and Arts Western
Switzerland
Fribourg, Switzerland

Maurizio Caon

maurizio.caon@hes-so.ch

University of Applied Sciences and Arts Western
Switzerland
Fribourg, Switzerland

Samuel Torche*

samuel.torche@hes-so.ch

University of Applied Sciences and Arts Western
Switzerland
Fribourg, Switzerland

Omar Abou Khaled

omar.aboukhaled@hes-so.ch

University of Applied Sciences and Arts Western
Switzerland
Fribourg, Switzerland

ABSTRACT

This article presents *PMF-Vision*, a digital assistant for manual assembly based on the concept of augmented reality information projected on the workspace. This digital assistant aims at assisting operators with mild mental deficiencies in their assembly tasks by tracking their actions in real-time. In this research, we analyzed hand tracking algorithms based on the latest deep learning techniques and highlighted the best candidate for our context in terms of accuracy and processing time. We present different interaction concepts for the system and propose to evaluate two visualisation concepts to assist the operator during component placement. The results of the exploratory study ($N=6$) highlighted the excellent usability of the system and the interest, expressed by the target population, to use such tools in their daily tasks.

RÉSUMÉ

Cet article présente *PMF-Vision*, un assistant digital d'aide au montage manuel basé sur le concept de projection d'informations en réalité augmentée sur l'espace de travail. Cet assistant digital vise à assister des opérateurs en situation de handicap mental dans leurs tâches de montage en suivant leurs actions en temps-réel. Dans le cadre de cette recherche, nous avons analysé les algorithmes de suivi des mains basés sur les dernières techniques d'apprentissage profond et mis en valeur le meilleur candidat pour notre contexte en termes de précision et de temps de traitement. Nous présentons différents concepts d'interaction pour le système et proposons d'évaluer deux concepts de visualisation pour assister l'opérateur

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IHM '22, April 5–8, 2022, Namur, Belgium

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9189-4/22/04...\$15.00

<https://doi.org/10.1145/3500866.3516375>

lors du placement de composants. Les résultats de l'étude exploratoire ($N=6$) ont permis de mettre en valeur l'excellente utilisabilité du prototype et l'intérêt, exprimé par la population cible, d'utiliser de tels outils dans leurs tâches quotidiennes.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Accessibility technologies**; • **Computing methodologies** → **Tracking**.

KEYWORDS

manual assembly, assistive systems, augmented reality, projection, hand tracking, human-computer interfaces

MOTS CLÉS

assemblage manuel, systèmes d'assistance, réalité augmentée, projection, suivi des mains, interfaces humain-machine

ACM Reference Format:

Simon Ruffieux, Samuel Torche, Maurizio Caon, and Omar Abou Khaled. 2022. Etude exploratoire d'un assistant digital d'aide au montage basé sur la projection en réalité augmentée : An Exploratory Study of a Projection-based Augmented Reality Digital Assistant to Facilitate Assembly. In *IHM '22: Proceedings of the 33rd Conference on l'Interaction Humain-Machine (IHM '22)*, April 5–8, 2022, Namur, Belgium. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3500866.3516375>

1 INTRODUCTION

A la différence des grandes entreprises, pour de nombreuses petites et moyennes entreprises (PME) qui réalisent des montages manuels d'un nombre raisonnable de pièces, une pleine robotisation n'est pas intéressante en termes de coûts et de quantité de pièces par série. Dans ces PME, des opérateurs réalisent manuellement des tâches d'assemblage au quotidien. Dans de nombreuses entreprises, du personnel peu qualifié, parfois avec des handicaps légers, est chargé de réaliser les montages tandis que des encadrants sont responsables de les superviser et de réaliser les contrôles de qualité finaux permettant de garantir la conformité du produit.

La solution *PMF-Vision* vise à développer un assistant de montage manuel augmenté pour la réalisation de séries limitées. Cette solution se base sur les dernières avancées technologiques dans le domaine de la vision par ordinateur, de l'intelligence artificielle et de la projection en réalité augmentée pour assister l'opérateur en temps-réel et pour réaliser des contrôles continus de la conformité du montage. L'objectif final est de réduire le nombre d'erreurs de montage mais également de faciliter l'apprentissage des opérateurs et de les accompagner au travers d'instructions intuitives. La solution est développée en collaboration avec l'entreprise *PMF-System*¹ (*PMFch SA*) qui réalise différents types de mobiliers modulaires, dont des tables de montage.

Actuellement, plusieurs solutions sont déjà disponibles sur le marché. Cependant, elles sont généralement destinées aux grandes entreprises et sont basées sur des systèmes complexes et sur de nombreux capteurs disposés sur l'opérateur et dans l'environnement. Les récentes avancées sur le traitement de l'image et les algorithmes d'apprentissage profond permettent désormais d'envisager de réaliser le suivi de l'opérateur et de ses actions en temps-réel avec un minimum de capteurs non intrusifs et pour un coût raisonnable. Les derniers modèles de projecteurs permettent de projeter de l'information directement sur l'espace de travail depuis une distance réduite tout en garantissant une bonne qualité d'image. Les critères techniques sont donc réunis pour permettre de faire évoluer les assistants digitaux d'aide au montage.

Le projet *PMF-Vision* a deux objectifs principaux: rendre aux encadrants du temps pour accompagner socialement les opérateurs au lieu de passer du temps à inspecter et vérifier des montages, et permettre aux opérateurs de regagner confiance en eux en leur offrant une plus grande autonomie. Cependant, il reste encore de nombreuses interrogations sur l'applicabilité de tels systèmes en conditions réelles ainsi que sur les méthodes les plus adaptées pour permettre une interaction naturelle avec l'utilisateur. La perception des utilisateurs finaux face à de tels systèmes est également cruciale pour la suite du projet. Dans cet article, nous présentons le prototype du système *PMF-Vision* et les résultats d'une première étude exploratoire réalisée avec de futurs utilisateurs du système. Cette étude a deux objectifs principaux: évaluer l'utilisabilité et l'acceptabilité générale du prototype en se basant sur un questionnaire SUS [8] et évaluer les besoins des utilisateurs finaux en terme de projection en réalité augmentée en se basant sur un questionnaire personnalisé.

2 ÉTAT DE L'ART

Pour plus de clarté, nous avons divisé la revue de l'état de l'art en trois sections distinctes: un aperçu des assistants de montage digitaux développés dans le monde commercial et académique, une revue des techniques et algorithmes permettant de réaliser le suivi des mains et des actions de l'opérateur et une revue des interfaces humain-machine utilisées dans le domaine des assistants de montage.

2.1 Assistant de montages digitaux

Les systèmes d'aide au montage font partie des technologies émergentes de l'industrie 4.0 [44]. Ces systèmes, aussi dénommés *Next*

Generation Production Systems (NGPSs), visent à rendre les outils de production adaptés à la flexibilisation, la personnalisation et à la gestion de séries limitées et à la durée de vie raccourcie [5]. Différents travaux académiques ont été réalisés dans le domaine des systèmes d'aide au montage. Les différentes recherches sont principalement centrées sur les domaines suivants : reconnaissance visuelle et suivi de l'opérateur, interaction humain-machine, métriques, et ergonomie.

Il est intéressant de noter que différentes compagnies commercialisent déjà des systèmes d'aide au montage. Cependant, la majorité des produits disponibles requiert l'utilisation de capteurs sur l'opérateur ou dans l'environnement et cible le marché des grandes usines avec des installations complexes. Par exemple, l'entreprise Bosch Rexroth propose le système *ActiveAssist* [7] depuis 2017. Ce système permet d'assister un opérateur travaillant sur des séries impliquant des opérations sur des tables de montage disséminées dans l'environnement et impliquant des outils spécifiques. Cette solution vise principalement les ateliers de montage de grande envergure et nécessite l'installation de capteurs et/ou de marqueurs sur l'opérateur, sur les outils utilisés et sur les tables de montages. L'entreprise Tulip, basée aux Etats-Unis et en Angleterre, vend une solution plus proche de celle envisagée dans le cadre de ce projet [39]. Une fois installé sur une table de montage, leur kit modulaire permet d'assister les opérateurs dans leur montage et de suivre leur rythme de travail à distance au travers d'une plate-forme en-ligne. Ce kit modulaire qui est composé de capteurs, de marqueurs et de senseurs physiques peut-être proposé sous de nombreuses variantes en fonction des besoins du client; son utilisation nécessite cependant des connaissances avancées pour sa configuration, son utilisation et son adaptation à de nouvelles séries. La compagnie Soft2tec propose le système *Nexonar* [36]. Ce système permet de suivre en 3D les mains de l'opérateur ainsi que des outils 3D avec une précision millimétrique en se basant sur des marqueurs électroniques et des caméras placées dans l'environnement. La compagnie Build2Light propose le système *Lightning Pick* qui propose différents types d'affichage physiques (petits écrans, LEDs, etc.) à installer sur les tables de montage [23]. Finalement, la compagnie Arkite propose un produit proche de la solution *PMF-Vision* avec son produit, nommé "Human Interface Mate" (HIM), qui permet de suivre les gestes de l'opérateur à l'aide d'une caméra de profondeur et de s'intégrer avec un projecteur pour afficher l'information directement sur l'espace de travail [1]. Cependant, à notre connaissance, leur solution n'est pas pensée pour des personnes en situation de handicap et n'a pas été évaluée de manière scientifique.

Le travail de Wang propose une revue de la littérature des systèmes d'assistance au montage et de leur application se basant sur la réalité augmentée entre 1990 et 2015 [41]. Il met en lumière trois principales catégories abordées en recherche: assistance au montage, entraînement au montage et simulation pour le design et la planification des tâches d'assemblage. Le travail de Pimminger et al. [31] propose quand à lui une évaluation technique des approches et des technologies requises pour réaliser une table de montage intelligente. Ils mettent notamment en avant le potentiel des algorithmes d'apprentissage profond pour la tâche de reconnaissance des mains face aux technologies précédemment utilisées. Le travail de Ferrari et al. [14] se concentre sur le suivi des mouvements de l'opérateur dans l'espace pour quantifier le temps passé sur les

¹<https://www.pmf-system.ch/home/>

différentes zones de travail. Grâce à cette quantification, il est ensuite possible d'identifier les activités à faible valeur ajoutée, tel que déplacements, recherche d'outils, etc. Leur objectif final est de chercher à optimiser les tâches réalisées par l'opérateur. Dans leurs travaux, Bortolini [6] et Colim [11] cherchent à mesurer et améliorer l'ergonomie des mouvements effectués afin de réduire les risques de blessures musculaires pour les opérateurs effectuant des tâches très répétitives.

2.2 Suivi des mains et des actions

De nombreux travaux de recherche se concentrent sur le suivi des mains pour des cas applicatifs distincts [30]. Dans le cadre des assistants de montage, les caméras de profondeur et les caméras couleur, assistées par des techniques avancées de traitement d'images, ont petit à petit remplacé les technologies intrusives. En effet, dans les précédentes approches, en plus de caméras, il était généralement nécessaire d'avoir recours à des marqueurs physiques placés sur l'opérateur et/ou les objets ainsi qu'à des senseurs placés sur la table de montage [33]. Des travaux récents ont cependant montré qu'il est devenu possible de réaliser des solutions suffisamment robustes pour être utilisées en conditions réelles en utilisant uniquement des caméras de profondeur [9] et/ou couleur [41]. Ces dernières années, les avancées en terme d'algorithmes d'apprentissage profond (*Deep Learning*) ont permis un gain supplémentaire en terme de précision et de temps de traitement, permettant ainsi de réaliser le suivi des mains et des actions de l'opérateur en temps-réel [24, 31] ou d'identifier les outils utilisés à moindre coût [22]. Les dernières méthodes proposées en recherche, disponibles sous forme de bibliothèques *open-source*, promettent de simplifier leur mise en place et d'améliorer encore les performances; que ce soit la bibliothèque *HandTrack* proposée par V. Dibia basée sur les MobileNet SSD [40, 42], la bibliothèque *MediaPipe* proposée par Google et basée sur des modèles SSD personnalisés [45] ou encore la bibliothèque *TRTPoseHand*, optimisée pour les processeurs Nvidia et inspirée de *OpenPose*, proposée par Cao et basée sur le concept de "Part Affinity Fields" [10]. Ces techniques récentes n'ont cependant pas encore été évaluées scientifiquement dans un contexte d'assistant de montage, ayant comme particularité un placement de la caméra en vue de dessus (*top-down view*).

2.3 Interface Humain-Machine

Les systèmes d'aide au montage requièrent à la fois de pouvoir transmettre les informations à l'opérateur de manière naturelle, mais ce dernier doit également pouvoir interagir avec le système. Différentes méthodes permettant de transmettre de l'information à l'opérateur ont été proposées et évaluées dans le contexte de l'aide au montage; les instructions peuvent être présentées sur un écran traditionnel [22, 38], des lumières LEDS peuvent être installées sur les boîtes contenant les pièces (*pick-by-light*) [13] ou les instructions peuvent être projetées en réalité augmentée directement sur l'espace de travail à l'aide d'un ou de plusieurs projecteurs (*pick-by-vision*) [35]. Certaines études ont également exploré l'utilisation de lunettes de réalité augmentée portées par l'opérateur, avec des résultats mitigés [34].

Des études ont cherché à quantifier l'utilité et l'applicabilité de la réalité augmentée pour des tâches d'assemblage manuel. Dans

son étude, Blattgerste a notamment comparé des instructions dans un mode *in-view* et *in-situ* [4]. Dans le mode *in-view*, une image de l'instruction de montage est fournie à l'utilisateur, tandis que dans le mode *in-situ*, les instructions sont superposées à la réalité. Ces deux modalités étaient transmises au travers de différents médiums: papier, smartphone ou lunettes de réalité augmentée. Leur évaluation, basée sur l'approche General Assembly Task Model (GATM) proposée par Funk [16], a montré que les instructions *in-situ* étaient plus efficaces et généraient une charge mentale plus réduite. Les limitations techniques des lunettes de réalité augmentées réduisait cependant en partie le gain obtenu par les instructions *in-situ*. D'autres études ont également montré les avantages des instructions *in-situ* projetées directement sur l'espace de travail et plus particulièrement dans le cas de personnes handicapées [15, 17]. Ces travaux ont notamment permis de montrer la supériorité des instructions projetées *in-situ* de type "contour" et "image" pour indiquer la manière de placer les éléments.

Dans le cadre de ce travail nous proposons donc d'utiliser le concept de réalité augmentée tel que décrit par Dubois [12]: l'interaction se passe dans le monde réel et est augmentée au moyen d'informations projetées directement sur l'espace de travail (*in-situ*). Un exemple séminal de ce type de réalité augmentée est le *DigitalDesk* proposé en 1993 par Wellner. [43].

Les principales contributions scientifiques de ce travail peuvent être résumées en 3 points:

- (1) Une évaluation quantitative des performances des algorithmes d'apprentissage profond existants pour le suivi des mains en utilisant un capteur en position "vue de dessus".
- (2) une évaluation de la perception, par les utilisateurs finaux, de l'ajout d'information sur l'espace de travail (*in-situ*) lorsque l'information est également disponible sur un écran déporté (*in-view*).
- (3) une évaluation qualitative, dans un contexte écologique, de la solution d'assistance au montage en réalité augmentée par la population cible.

Bien que proche et inspirée des travaux de Funk et al [15–17], notre approche diffère en différents points. La possibilité de suivre les mains et leur conformation sur tout l'espace de travail apporte des possibilités nouvelles en terme d'interaction et de finesse du suivi. Elle permet notamment de suivre les opérations effectuées par l'opérateur de manière plus précise, par exemple en permettant de reconnaître certains gestes fins effectués par l'opérateur. Le déroulement du flux de travail diffère également avec l'introduction d'une opération de contrôle du montage qui permet de vérifier la validité d'une étape. Cette opération implique, en cas de problème, une gestion de l'erreur sous la forme d'une interaction de l'opérateur avec le système au travers de boutons virtuels placés sur l'espace de travail. Finalement, dans notre solution, les instructions de montage sont fournies de manière mixte: *in-view* sur l'écran déporté et *in-situ* au travers de projections sur l'espace de travail. Dans notre contexte applicatif, l'affichage *in-view* reste indispensable afin de permettre une compréhension fine des pièces qui ne possèdent pas un contour permettant d'identifier leur orientation.

3 CONCEPTION

Dans cette section, nous présentons la conception de la solution *PMF-Vision*: la table de montage et les capteurs, un aperçu de l'architecture logicielle, une description de la sélection de l'algorithme de suivi des mains et une description des concepts de l'interface projetée sur l'espace de travail.

3.1 Table de montage

La partie physique de la solution *PMF-Vision*, illustrée sur la figure 1, est composée d'une table ajustable en hauteur, d'une étagère de rangement modulaire et de différents composants électroniques qui sont précisés ci-dessous.

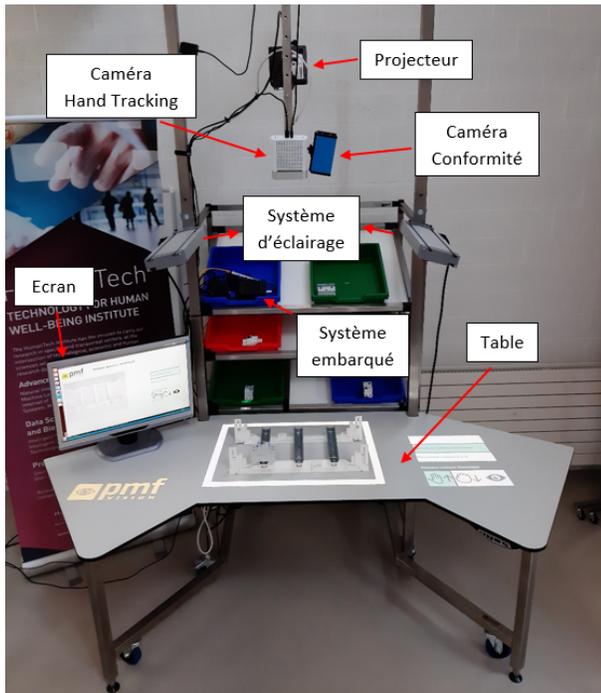


Figure 1: Prototype actuel de la table de montage avec différents éléments de l'interface utilisateurs projetés sur la table.

Le système embarqué, un NVIDIA Jetson AGX Xavier [27], offre la puissance de calcul nécessaire pour réaliser le traitement vidéo ainsi que les opérations complexes de traitement d'images requises par les algorithmes de suivi des mains et de détection des actions décrits dans la section 3.3. De plus, il offre une connectique complète, une nuisance sonore réduite ainsi qu'une consommation électrique raisonnable. La caméra de détection des mains, une Azure Kinect [26], permet de capturer les informations de couleur et de profondeur en temps-réel avec une haute résolution. Elle permet également de pouvoir capturer l'ensemble de la table de montage depuis une hauteur réduite grâce à son grand champ de vision (*FoV*). Une caméra de contrôle de conformité, la DFK Z30GP031 de la compagnie The Imaging Source [37] permet de réaliser les tâches de contrôle qualité grâce à sa très haute résolution et son zoom

optique. Le projecteur Optoma ML1050ST [29], compact et léger, permet de projeter les instructions avec une luminosité de 1000 lumens sur l'ensemble de la table depuis une hauteur réduite grâce à sa focale courte. Avec 22 décibel, ce projecteur offre également un nuisance sonore réduite. Finalement, le système d'éclairage, composé de deux lampes standards dans le domaine, permet d'obtenir une luminosité uniforme sur l'espace de travail et de garantir ainsi un confort visuel optimal pour les opérateurs.

3.2 Architecture logicielle

La partie logicielle de la solution *PMF-Vision* est décomposée en modules ayant chacun des fonctions spécifiques, comme illustré sur la figure 2. Les principaux modules sont:

- **PMF-Vision**: point d'entrée et contrôleur principal du programme. Il gère les communications entre les interfaces, le suivi du scénario de montage (voir section 3.4) et les modules de détection des mains et d'actions.
- **Hand Tracking**: détection des mains. Ce module reçoit en entrée une image couleur et profondeur venant de la caméra de détection des mains et retourne la position 3D des mains et leur conformation dans ces images.
- **Hand Action**: détection des actions. Ce module reçoit en entrée la position 3D des mains et leur conformation et retourne l'action détectée: *pick*, *place* ou *pas d'action*.
- **Scenario**: suivi d'un scénario de montage (voir section 3.4). Ce module reçoit en entrée la position 3D des mains, l'action détectée et les images couleur et profondeur venant de la caméra de détection des mains. Il retourne toutes les informations de suivi du scénario, parmi lesquelles lorsqu'un changement d'état a lieu, par exemple le passage d'une étape.
- **Conformity Control**: contrôle de la conformité du montage actuel. Ce module reçoit en entrée une image prise depuis la caméra de conformité et la compare avec une image de référence du montage pour déterminer si le montage actuel est valide ou non.

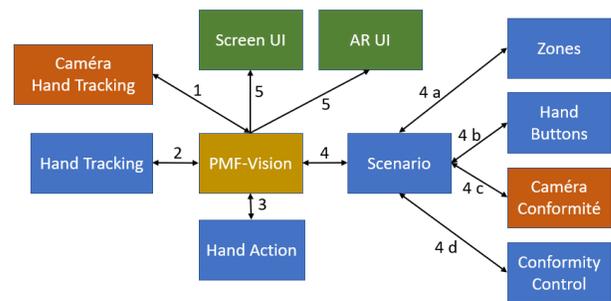


Figure 2: Schéma de l'architecture logicielle. En jaune le module principal, en rouge les modules liés aux capteurs, en vert les modules liés à l'affichage et en bleu les modules de traitement des données.

Le module *PMF-Vision* contient la boucle principale, exécutée à une fréquence de 20 images par seconde (*fps*), qui effectue les opérations suivantes:

- (1) Récupère les images couleur et profondeur en temps-réel de la caméra de détection des mains.
- (2) Fournit au module de *Hand Tracking* ces images.
- (3) Les positions des mains en 3D sont transmises au module *Hand Action*.
- (4) Les informations sur la position 3D des mains, les actions détectées, et les images couleur et profondeur sont transmises au module *Scenario*. Un scénario est lié à une définition de *Zones* (voir section 3.4) qui lui permettent de connaître la configuration physique actuelle de la table de montage (emplacement des bacs, emplacement de la zone d'assemblage). Un scénario est également lié à une configuration virtuelle de *Hand Buttons* qui contiennent les emplacements de boutons virtuels permettant à l'utilisateur d'interagir avec le système. Le module *Scenario* est également responsable de capturer une image en utilisant la caméra de conformité lorsque nécessaire et de la transmettre au module *Conformity Control*.
- (5) Lorsqu'un changement d'état a lieu dans le module *Scenario*, *PMF-Vision* communique les changements avec les interfaces *Screen UI* et *AR UI*. *Screen UI* correspond à l'interface affichée sur l'écran d'ordinateur et *AR UI* correspond à l'interface projetée sur la table.

3.3 Algorithme de suivi des mains

La robustesse du module de suivi des mains est critique pour le bon fonctionnement du système. En effet, une détection interrompue ou qui générerait des faux positifs rendrait le système inutilisable. Les approches disponibles ont évolué très rapidement ces dernières années au fil des avancées des algorithmes d'apprentissage profond. Les algorithmes permettant de réaliser un suivi des mains depuis une vue du dessus (top-view) sont cependant plus rares et ont été moins documentés dans la littérature. Afin de sélectionner l'algorithme le plus adapté, nous avons donc réalisé une évaluation et une comparaison pratique des algorithmes disponibles. La comparaison de quatre algorithmes, retenus après une première sélection, est résumée ci-dessous. La comparaison a été effectuée à l'aide de vidéos pré-enregistrées par le système tel que décrit dans la section 3.1. Ces vidéos ont été utilisées pour comparer les performances des algorithmes en termes de précision et de temps de traitement. Le temps de traitement a été mesuré, pour tous les algorithmes, en utilisant les capacités de calcul de la carte graphique (GPU) de l'ordinateur embarqué. Les critères suivants ont été retenus pour la comparaison:

- **Temps de traitement.** L'algorithme doit pouvoir être exécuté en temps réel. Selon nos spécifications, l'algorithme doit être capable de traiter au moins 10 images par seconde sur le Jetson AGX Xavier.
- **Précision.** L'algorithme doit être capable de détecter les mains de manière robuste. Selon nos spécifications, il doit détecter correctement les mains et leur position dans plus de 90% des images traitées.
- **Disponibilité.** L'algorithme doit être *open-source* et disponible sous une licence permettant une utilisation dans un cadre commercial.

La table 1 résume les résultats de notre analyse. Nous avons évalué 4 librairies: HandTrack [40] qui utilise un modèle SSD [25]

entraîné sur la base de données EgoHands [2], MediaPipe [45] qui est développé par Google et utilise un modèle SSD personnalisé [19], TRTPoseHand [28] qui est développé par NVIDIA et utilise un modèle ResNet [20], et enfin OpenPose [21] qui utilise un modèle CPM [42]. Deux algorithmes sont rapidement ressortis de cette analyse. TRTPoseHand et MediaPipe étaient en effet les seuls à satisfaire le premier critère: le temps de traitement. Suite à cette première évaluation, nous avons analysé la précision de la détection des mains pour ces deux algorithmes. Pour ce faire, nous avons enregistré des vidéos de 30 secondes, prises en *top-view* depuis la caméra de détection des mains. Dans ces vidéos, les deux mains d'une personne sont visibles en tout temps par la caméra. Nous avons ensuite traité la vidéo image par image, et compté le nombre de mains détectées par image. Ainsi, on obtient la formule suivante (p = précision, m = somme du nombre de mains détectées à chaque image, n = nombre d'images dans la vidéo): $p = \frac{m}{2 * n}$. Nous avons obtenu respectivement 90% et 99% pour TRTPoseHand et MediaPipe. Après une inspection visuelle, nous avons cependant remarqué que les résultats retournés par TRTPoseHand contenaient beaucoup de faux positifs, des mains fantômes ou des mains détectées au mauvais endroit, particulièrement lorsque les distances séparant les mains de la caméra étaient supérieures à 50cm. Nous avons donc sélectionné *MediaPipe* [45], car il est plus précis, plus facilement déployable sur le Jetson AGX Xavier, mieux documenté et disponible sous licence *Apache License 2.0*. De plus, selon ses spécifications, il garantit un fonctionnement indifféremment de la couleur de peau.

L'algorithme de détection des mains de *MediaPipe* est composé de deux modèles:

- (1) **Palm detector:** ce premier modèle reçoit une image couleur en entrée et la transforme afin d'être en résolution fixe de 128x128. Il est responsable de fournir le cadre de délimitation (*bounding box*) pour chaque main détectée dans l'image. Ce modèle est implémenté à l'aide d'une architecture SSD [19] personnalisée.
- (2) **Hand landmark model:** pour chaque main détectée par le *Palm detector*, le *Hand landmark model* reçoit l'image rognée selon le cadre de délimitation en entrée, transformé en résolution 224x224. Ce modèle est alors responsable de détecter le squelette de la main et retourne ainsi en sortie la conformation de la main sous forme de 21 repères localisés en 3D par rapport à la position du poignet. Ce modèle est implémenté à l'aide d'un modèle de régression.

3.4 Scénario de montage

Un scénario de montage est le plan qui décrit le déroulement dans lequel un assemblage doit être réalisé. Il contient ainsi une structure qui permet de décrire l'ensemble des étapes à réaliser. Un outil spécifique a été développé afin de faciliter la création et la gestion des scénarios. Cette application permet à l'utilisateur spécifier les zones d'une table de montage, d'éditer et de créer les étapes du scénario de manière visuelle, de capturer les images de référence du montage et d'y spécifier des régions d'intérêt. L'application permet finalement d'exporter le scénario complet de manière structurée sous forme de dossiers hiérarchisés et de fichiers structurés. Dans la solution *PMF-Vision*, un scénario de montage est principalement décrit par deux fichiers structurés au format *JSON*: *zones* et *scenario*.

Table 1: Comparaison des différents algorithmes

Nom	Technologie	Modèle	Images par seconde (résolution)	Précision
HandTrack [40]	TensorFlow	SSD [25]	5-8 (320x180)	-
MediaPipe [45]	TensorFlow	Custom SSD + Regression [19]	25-30 (128x128)	99%
TRTPoseHand [28]	TensorRT	ResNet [20]	25-30 (224x224)	90%
OpenPose [21]	PyTorch	CPM [42]	<1 (320x180)	-

zones: Ce fichier structuré contient la définition des zones auxquelles un scénario est attaché. Une zone correspond à un emplacement physique sur la table de montage et est donc définie par des coordonnées 3D. Ces zones sont spécifiées manuellement au travers d'une application de gestion de scénarios qui a également été développée dans le cadre du projet. Une zone peut être de type:

- **Assembly:** correspond à l'emplacement de la zone d'assemblage sur la table.
- **Collect:** correspond à l'emplacement des bacs contenant les pièces du montage dans l'étagère de rangement.
- **Display:** correspond à l'emplacement d'une zone dans laquelle des notifications pourront être projetées pour l'utilisateur.

scenario: Ce fichier structuré contient la liste des étapes qui constituent un scénario. Ces étapes et opérations sont spécifiées manuellement au travers de l'application de gestion de scénarios. Chaque étape est décomposée en une liste d'opérations, chaque opération étant attachée à une zone définie dans le fichier *zones*, dont l'action peut être de:

- **Pick:** prendre une pièce.
- **Place:** déposer une pièce.
- **Control:** contrôler le montage à partir d'une image de référence stockée en mémoire. Pour cette action, l'opération peut également contenir les positions de régions d'intérêts (*Region Of Interest - ROI*) liées à cette image de référence. Ces régions d'intérêts correspondent à des sous-zones qui doivent être spécifiquement contrôlées par le module de contrôle de conformité.

La figure 3 illustre une partie d'un exemple concret de scénario de montage d'un tableau électrique. La figure illustre les deux premières étapes du scénario, composées chacune de trois opérations. Lors de la première étape du scénario, l'opérateur prend le composant *châssis* depuis une zone de collecte, le dépose sur la zone d'assemblage et active l'opération de contrôle en retirant ses mains de la zone d'assemblage. Lors de la deuxième étape, l'opérateur prend le composant *A* depuis une zone de collection, dépose le composant à son emplacement prévu sur le châssis et active l'opération de contrôle. Lors de cette opération de contrôle, l'emplacement exact de la nouvelle pièce est définie dans le scénario comme une région d'intérêt et donc contrôlée en détail par le module de conformité.

3.5 Interface Humain-Machine

L'interface utilisateur est accessible au travers de deux médiums principaux; l'écran d'ordinateur et les éléments projetés sur l'espace de travail. L'écran d'ordinateur est utilisé pour démarrer le système, sélectionner le scénario et permet ensuite d'afficher l'image de

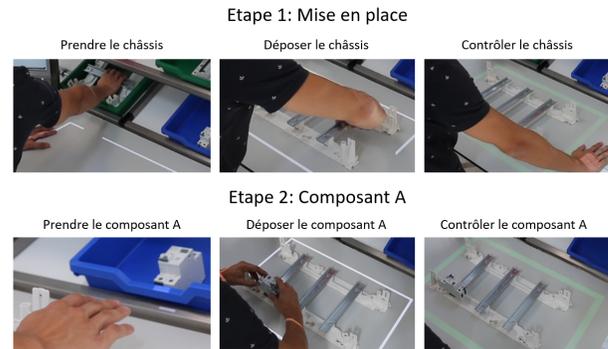


Figure 3: Exemple concret de scénario composé de deux étapes ayant chacune trois opérations à être réalisées.

référence *in-view* pour l'étape en cours; cette image de référence, qui correspond au montage attendu à la fin de l'étape en cours, est nécessaire afin de permettre à l'opérateur d'identifier correctement l'orientation de certaines pièces qui peut dépendre de détails trop fins pour une visualisation *in-situ*.

La projection de l'interface en réalité augmentée directement sur la table permet de fournir les instructions à l'opérateur *in-situ* et de maintenir son attention sur l'espace de travail. Comme décrits dans l'état de l'art [4, 15–17], ces projections permettent de notifier, de manière non-intrusive, l'utilisateur des régions d'intérêt sur lesquelles il doit se concentrer, de fluidifier son flux de travail en supprimant de potentiels doutes sur ses tâches et finalement de permettre de réaliser des notifications qui soient facilement détectables à distance par l'encadrant en cas de problèmes. Une zone spécifique, sur la droite de la table de montage, est constamment visible et résume l'avancement du montage. Cette interface, visible sur la droite de la figure 1, affiche le nom de l'étape actuelle et l'opération qui est actuellement en cours (pick, place, control) au travers de pictogrammes comme encouragé par les travaux de Funk et al.[15]. Ces pictogrammes sont illustrés sur la droite de la figure 4.

Diverses notifications sont considérées dans le prototype actuel et sont affichées en fonction de l'opération en cours. Ces notifications sont illustrées sur la figure 3 et décrites ci-dessous:

- **pick:** une notification de couleur blanche, oscillant de gauche à droite, est projetée sur le bord de la caisse contenant l'objet à prendre. Cette notification est illustrée sur l'image présente en bas à gauche de la figure 3.
- **place:** un cadre blanc, clignotant, met en valeur la zone d'assemblage. L'endroit exact ou la pièce doit être déposée est

mis en valeur à l'aide d'un rectangle bleu. Cette notification est illustrée sur les figures 6 et 7.

- control:** la phase de contrôle débute lorsque, après avoir placé la pièce, l'utilisateur sort ses mains de la zone d'assemblage. La notification consiste alors en un rectangle vert ou orange qui encadre la zone d'assemblage pour indiquer le résultat du contrôle. Si la pièce est détectée comme correctement placée, le rectangle est affiché en vert; sinon il est affiché en orange. Cette notification est illustrée sur l'image présente en bas à droite de la figure 3

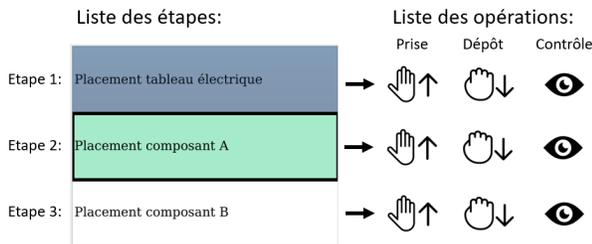


Figure 4: Liste des étapes et opérations. Mockup de l'interface projetée sur la droite de la table de montage et permettant de présenter l'étape et l'opération actuellement en cours à l'opérateur de manière simplifiée.

Dans différents cas de figure, l'utilisateur peut interagir directement avec le système au travers de boutons virtuels, comme illustré sur la figure 5. L'utilisateur maintient simplement sa main sur le bouton virtuel pour l'activer. Ainsi différentes interactions sont prises en compte dans l'interface actuelle. L'utilisateur démarre le scénario de montage en plaçant ses deux mains sur la table. Lorsqu'un contrôle est négatif l'utilisateur a la possibilité de corriger son montage puis de relancer l'opération de contrôle en plaçant sa main sur un bouton virtuel affiché de manière contextuelle sur la droite de la zone d'assemblage.

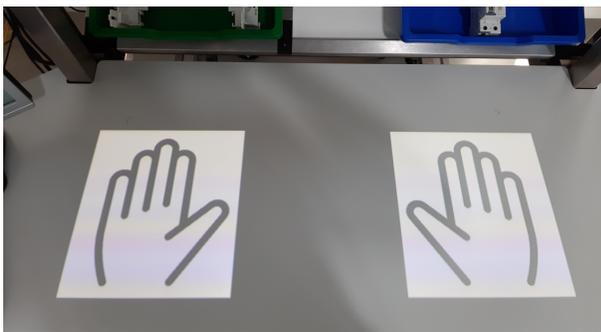


Figure 5: Exemples de boutons virtuels projetés sur la table permettant à l'utilisateur d'interagir avec le système.

Pour l'opération de placement (*place*), qui permet de spécifier de manière précise où la pièce doit être déposée, deux options distinctes ont été directement inspirées des résultats obtenus dans les travaux de Funk et al. [15, 17]. Ces études ont en effet montré que

les visualisations de type *contour* et *image* étaient les plus efficaces en terme de temps de compréhension et de réduction des erreurs de montage. Dans notre cas, ces visualisations servent à compléter la visualisation *in-view* proposée sur l'écran déporté. Dans notre contexte, nous espérons donc que le mode superposé permette de réduire le recours à la visualisation *in-view* (déporté sur l'écran) et donc d'améliorer l'utilisabilité du système en maintenant le focus de l'opérateur sur l'espace de travail. Ces deux modes, illustrés sur les figures 6 et 7, seront évalués dans le cadre de l'étude exploratoire avec les utilisateurs:

- (1) **Mode classique:** projection d'un cadre blanc indiquant la zone d'assemblage globale et un rectangle bleu précisant l'emplacement exact de la pièce.
- (2) **Mode superposé:** projection d'un cadre blanc indiquant la zone d'assemblage globale, de l'image de référence du montage sur la zone d'assemblage en superposition et un rectangle bleu précisant l'emplacement exact de la pièce.



Figure 6: Mode classique - affichage d'un rectangle bleu indiquant l'emplacement où déposer la pièce.

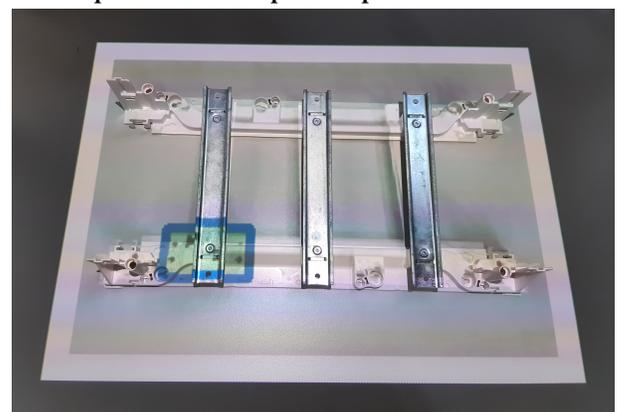


Figure 7: Mode superposé - affichage de l'image de référence superposée sur la zone d'assemblage avec un rectangle bleu indiquant l'emplacement où déposer la pièce.

4 ÉTUDE EXPLORATOIRE

Afin d'obtenir une première évaluation du système et de guider les futurs développements, nous avons réalisé une étude exploratoire. Cette section présente les principaux objectifs de l'étude, la méthodologie utilisée ainsi que les résultats obtenus.

4.1 Objectifs

Cette étude exploratoire vise à réaliser une première évaluation de l'utilisabilité et de l'acceptabilité du prototype avec la population cible. Cette étude vise également à évaluer l'interface utilisateur projetée sur l'environnement de travail et plus particulièrement à évaluer les deux différents modes de projection envisagés pour indiquer l'emplacement de dépôt d'un composant, illustrés sur les figures 6 et 7. Nous cherchons notamment à comprendre si la superposition apporte un bénéfice à l'opérateur étant donné que l'information fournie est redondante avec l'information affichée sur l'écran déporté.

Les objectifs peuvent être résumés ainsi:

- évaluer l'utilisabilité du système développé avec les utilisateurs cibles.
- évaluer la perception de l'acceptabilité d'un tel système par les utilisateurs cibles.
- évaluer l'interface projetée et notamment les deux indications possibles pour le placement d'une pièce sur la zone d'assemblage.

4.2 Méthodologie

L'étude exploratoire a été réalisée avec une population cible de 6 individus, 2 femmes et 4 hommes, présentant de légères déficiences intellectuelles et/ou des difficultés à réaliser des tâches simples. Ils présentent typiquement des difficultés de compréhension et de mémorisation, des troubles de l'attention et/ou un manque de confiance en soi. Ces individus sont placés par l'Assurance Invalidité Suisse dans une PME partenaire et réalisent des tâches simples d'assemblage au quotidien. Les participants de l'étude ont été sélectionnés, sur la base du volontariat, par un encadrant de cette PME. Sur les 6 participants, deux ont des difficultés prononcées lors de la réalisation de tâches de montage tandis que les quatre autres étaient plus à l'aise avec ce type de tâche. Parmi ces 4 derniers, seulement 2 étaient familiers avec les montages de type relais électriques, la tâche à réaliser dans le cadre de cette étude.

La procédure de test s'est déroulée en trois phases:

- (1) Découverte et prise en main du système (en groupe)
- (2) Utilisation du système selon les deux conditions (individuel)
- (3) Réponse aux questionnaires (individuel)

Lors de la première phase, l'encadrant a réalisé une présentation générale du système et de ses objectifs. Les principales fonctionnalités du système ont été présentées. Les participants ont ensuite pu se familiariser avec le système à tour de rôle en utilisant un scénario de démonstration composé de trois étapes. Ils pouvaient également poser des questions durant cette phase.

Lors de la deuxième phase, les participants ont été séparés afin de réaliser deux essais de montage sur le système. Une démonstration du scénario de montage était tout d'abord réalisée par une personne de l'équipe puis le participant réalisait ses deux essais. Le scénario

de montage comprenait 6 étapes; il consistait à placer le châssis sur l'espace de travail puis à placer cinq pièces (relais électriques) à des positions précises sur les trois différents rails du châssis, comme illustré sur la figure 3. Durant chaque essai, une condition spécifique de projection était testée: soit le mode *classique* ou le mode *superposé*. L'ordre de la condition a été alterné entre les utilisateurs. Lors des essais, une personne de l'équipe de développement était constamment présente afin d'observer le participant, de vérifier le bon fonctionnement du système et d'assister l'utilisateur en cas de nécessité absolue. A noter que, dans le cadre de cette première étude exploratoire, nous n'avons pas mesuré de données quantitatives lors de la réalisation des essais par les participants.

Finalement, lors de la troisième phase, chaque participant devait compléter deux questionnaires au format électronique. Une personne de l'équipe de développement était constamment présente afin de répondre aux potentielles questions. Le participant commençait par compléter les 10 questions du questionnaire SUS (en français) afin de mesurer l'utilisabilité du système [8, 18]. Ils poursuivaient ensuite avec un questionnaire personnalisé contenant des questions réparties selon 4 sujets: deux questions sur la familiarité de l'utilisateur avec les tâches de montage manuel, 12 questions centrées sur l'interface projetée, quatre questions centrées sur l'acceptabilité du système et finalement deux questions ouvertes pour des remarques générales. Toutes les réponses aux questions fermées étaient basées sur une échelle de *Likert* (de 1 à 5).

4.3 Résultats

Dans cette section, nous présentons les résultats de l'étude exploratoire. La première partie présente une description de l'observation visuelle des participants lors des tests et la deuxième partie présente les réponses aux deux questionnaires en reportant, entre parenthèse, les scores moyens obtenus pour les différentes questions. Une valeur de 1 indique que les participants n'étaient pas du tout d'accord avec l'affirmation, tandis qu'un score de 5 indique qu'ils étaient tout à fait d'accord.

Les participants ont tous réussi à effectuer les montages avec succès lors de leurs deux essais respectifs. La plupart des participants ont eu de 0 à 2 pièces à repositionner car elles ont été détectées comme incorrectement placées par le système lors d'une des opérations de contrôle de conformité. Généralement, les pièces étaient décalées de quelques millimètres par rapport à la position attendue ou mal orientées de 180 degrés. Tous les participants ont corrigé le problème d'eux-même suite à la notification et quitté l'erreur, entraînant par la suite un contrôle valide par le système. Pour comprendre le problème, ils avaient généralement recours à l'image de référence affichée sur l'écran d'ordinateur ou comprenaient naturellement leur erreur en inspectant le montage présent devant eux. A noter que, dans le mode superposé, la qualité de l'image projetée sur le table de montage ne permettait pas de visualiser aisément une erreur autre qu'une erreur grossière de placement. Avec ce mode, il est ainsi possible de comprendre qu'une pièce est placée sur le mauvais rail mais difficile de comprendre si une pièce est mal orientée ou mal placée de quelques millimètres.

Les résultats du questionnaire SUS étaient très bons. Avec un score SUS moyen de 94.2 ± 4.4 ($min = 87.5$, $max = 100$), l'utilisabilité

du système peut être considérée comme "Excellente" (A) selon les critères standards d'évaluation [3]. Comme illustré sur la figure 8, les participants ont trouvé le système facile à utiliser (5) et l'ont utilisé avec une grande confiance (4.83). La grande majorité des participants aimerait utiliser un tel système fréquemment pour leur travail au quotidien (4.67). Ils estiment qu'ils n'auraient pas besoin de l'aide d'un expert technique pour utiliser le système (1.84); cependant, les réponses à cette affirmation sont plus nuancés (2 sont totalement affirmatifs et 3 sont plutôt affirmatifs).

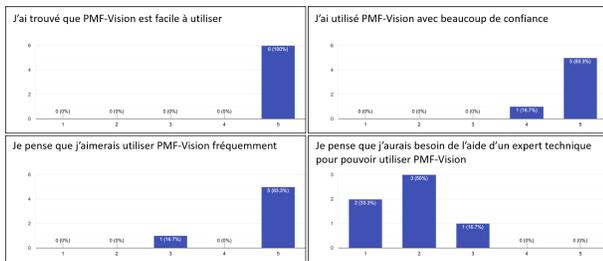


Figure 8: Distribution des résultats des 6 participants pour 4 questions spécifiques du questionnaire SUS. (1=pas du tout d'accord avec l'affirmation, 5=tout à fait d'accord avec l'affirmation)

Les résultats des questions sur l'interface utilisateur du questionnaire personnalisé ont permis de montrer que les différents interfaces et notifications sont considérées comme utiles mais que leur visibilité pourrait être améliorée. Les participants ont trouvé utile la liste des étapes et des opérations du montage en cours projetée sur la droite de la table (4.83), cette interface est visible sur les figures 4 et 1). La majorité des participants a également trouvé que la notifications sur les bacs étaient utiles (4.5). La figure 9 présente les réponses aux questions portant sur les notifications lors de la dépose de pièce en distinguant le mode classique sur le haut de la figure et le mode superposé sur le bas de la figure. Dans les deux modes, les utilisateurs ont reporté que la notification manquait de visibilité (4); un participant a même indiqué qu'il n'arrivait pas à la voir dans le mode "classique" et donc qu'elle n'était pas utile. Par contre, les autres participants ont trouvé les deux modes utiles, avec un très léger avantage pour le mode superposé (4.5) par rapport au mode classique (4.33). Finalement, les participants étaient plutôt d'accord avec l'affirmation "Le mode "superposé" permet de mieux comprendre comment déposer la pièce (orientation, position)" avec un score moyen de 4 (3 participants étaient tout à fait d'accord, 2 étaient partiellement d'accord et 1 pas du tout d'accord).

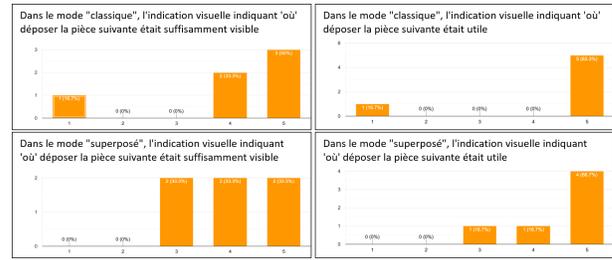


Figure 9: Distribution des résultats des 6 participants pour 4 questions spécifiques à l'interface projetée du questionnaire personnalisé. Comparaison entre le mode classique (haut de la figure) et le mode superposé (bas de la figure). (1=pas du tout d'accord avec l'affirmation, 5=tout à fait d'accord avec l'affirmation)

Les résultats des questions sur l'acceptabilité du système du questionnaire personnalisé sont illustrés sur la figure 10. Les résultats montrent que les utilisateurs seraient plus favorables à utiliser le système dans le contexte d'une phase d'apprentissage d'un montage (4.83) que dans une phase de production (4.33). La moitié des participants ne semblent pas se sentir plus surveillés que dans leur contexte habituel en utilisant un tel système et l'autre moitié des participants étaient indécis (2). La plupart des participants ne ressentiraient pas de stress supplémentaire en utilisant un tel système (1.16); plusieurs participants ont même reporté oralement que cela permettrait de réduire fortement leur stress de faire des erreurs et permettrait de réduire leur charge cognitive.

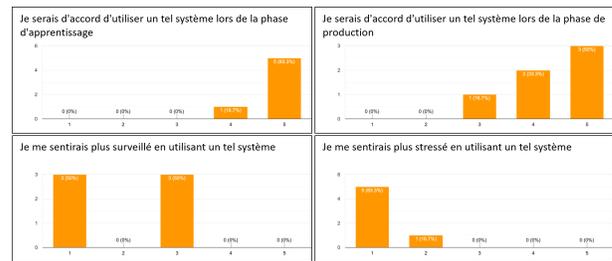


Figure 10: Distribution des résultats des 6 participants pour 4 questions spécifiques à l'acceptabilité du système du questionnaire personnalisé. (1=pas du tout d'accord avec l'affirmation, 5=tout à fait d'accord avec l'affirmation)

Les deux questions ouvertes à la fin du questionnaire ont permis de préciser certains points ou d'en mettre d'autres en valeur qui n'était pas directement abordés par les questionnaires. Un participant a notamment indiqué qu'un tel système lui permettrait d'avoir plus de confiance en lui car, n'ayant pas son encadrant sur le dos, il se sentirait plus autonome. De la même manière, un autre participant a mentionné qu'un tel outil serait très valorisant du fait de pouvoir apprendre à réaliser des montages de manière autonome. Plusieurs participants ont mentionné qu'un tel système serait bénéfique pour réduire leur stress et pallier à leur difficulté à se concentrer. Deux participants, ayant l'habitude de réaliser des

montages de relais électriques, ont mentionné que le système proposé serait peu utile pour des gens très familiers avec la tâche ou ayant peu de difficultés au travail. Ils ont quand même mentionnés qu'un tel système serait effectivement utile pour des montages plus complexes que celui réalisé lors de l'étude; typiquement des montages contenant plus de 15 pièces. La majorité des participants ont également rapporté qu'une luminosité et une résolution plus grande de la projection serait bénéfique de manière générale et plus spécifiquement que la notification affichée sur les bacs lors de la prise de pièce serait plus utile si elle était plus saillante.

5 DISCUSSION

Les résultats obtenus dans le cadre de l'étude exploratoire avec la population cible du projet sont très encourageants. Le nombre de participants limité ne permet malheureusement pas de réaliser une évaluation statistique des résultats mais permet déjà de mettre en valeur certains points et d'identifier différentes pistes de réflexion pour la suite du projet, que ce soit au niveau technique ou au niveau du design de l'interface.

Les résultats observés lors des tests et les réponses aux questionnaires montrent que le système développé est facilement utilisable et apprécié par les futurs utilisateurs. Les algorithmes développés permettent de réaliser le suivi des mains et des actions de manière robuste, de suivre les différentes étapes d'un scénario de montage et de détecter avec succès de potentielles erreurs de montage. Les mécaniques de contrôle et de gestion des erreurs développées pour ce projet ont également été facilement comprises et intégrées par les utilisateurs. Les participants de l'étude ont utilisé le système avec confiance, ce qui semble indiquer que le système est considéré comme suffisamment fiable pour être déployé dans un atelier, malgré la présence occasionnelle de faux-négatif lors de l'opération de contrôle de conformité. La majorité des participants ont également affirmé qu'ils apprécieraient d'utiliser un tel système dans leur travail au quotidien. Son utilisation dans le cadre d'une phase d'apprentissage de réalisation d'un montage est mieux acceptée que dans le cadre d'une phase de production. Bien que, selon les réponses au questionnaire, les participants resteraient d'accord de l'utiliser en production, plus particulièrement pour des montages comprenant de nombreuses étapes. Les principaux bénéfices perçus par les utilisateurs sont de gagner en autonomie, de réduire la charge mentale et de réduire le stress de faire des erreurs de montage. Des résultats proches de ceux obtenus dans de précédentes études similaires [15, 17]. Avec un score moyen de 94 au questionnaire SUS, l'utilisabilité du système est considérée comme excellente. Une future étude avec plus de participants permettra de confirmer ces premiers résultats.

Les résultats observés lors des tests et les réponses aux questionnaires ont permis d'identifier différents points positifs et négatifs de l'interface humain-machine et de comparer deux modes de notification distincts pour aider au placement des pièces. Selon les réponses aux questionnaires, le résumé des étapes et opérations, projeté sur la droite de la table, semble avoir été utile et apprécié; bien que selon nos observations visuelles lors de tests nous n'avons pas eu le sentiment que cette partie de l'interface ait été fréquemment utilisée. De la même manière, les participants avaient fréquemment recours à l'image de référence affichée sur l'écran d'ordinateur et il

sera donc indispensable de maintenir cette modalité dans le futur. Dans le cadre de futures études, il sera intéressant de quantifier plus précisément les interfaces utilisées par les utilisateurs en se basant par exemple sur des outils permettant de suivre le regard des utilisateurs (*Eye-tracking*). Lors de nos observations, nous avons eu l'impression que les utilisateurs travaillaient de manière plus fluide dans le mode superposé, mais nous n'avons pas de mesure pour le quantifier. Les résultats aux questionnaires semblent indiquer une légère préférence pour ce mode de notification de la part des utilisateurs. Plusieurs améliorations pourront être apportées aux modes *classique* et *superposé*: augmenter la luminosité, augmenter la résolution ou améliorer le positionnement de la projection par rapport à la réalité en étant capable de détecter précisément les modifications de position et d'orientation du montage sur l'espace de travail par rapport à l'image de référence. L'idée de projeter les informations en prenant en compte la 3D ("projection mapping") permettrait également de superposer les informations de manière précise sur des éléments 3D complexes. Des tests additionnels en se basant sur des mesures quantitatives seront nécessaires afin de démontrer scientifiquement la supériorité du mode mode *superposé* dans notre contexte. Ces mesures pourront être développées en s'inspirant du *General Assembly Taks Model* proposé par Funk pour quantifier le temps passé sur chacune des opérations [16, 32]. La gestion des erreurs de montage et la manière de notifier clairement à l'utilisateur l'erreur détectée est également un point très important pour la bonne utilisabilité du système; il sera nécessaire d'explorer différentes possibilités dans de futurs travaux.

Cette première étude a également permis de mettre en lumière la robustesse de l'approche basée sur les algorithmes d'apprentissage profond pour la reconnaissance et le suivi des mains dans notre cas applicatif. Tout en étant peu gourmand en terme de temps de traitement (<20ms) et précis (99%), l'algorithme implémenté permet de suivre distinctement les deux mains l'opérateur sur une large zone et de connaître leurs conformations en temps-réel. Dans le futur, les informations sur la conformation des mains pourra être exploitée afin d'améliorer la finesse du suivi des opérations et de développer des interactions plus avancées que celles proposées en se basant sur des gestes statiques (postures) ou des gestes dynamique. Les tests ont également permis d'identifier certains problèmes au niveau de la conception du système actuel. Dans certains cas, à cause du positionnement de la caméra de suivi des mains au dessus de l'espace de travail, la tête de l'opérateur peut venir obstruer le champ de vision de la caméra, ce qui interrompt le suivi des mains. Ce problème a parfois perturbé le système qui démarrait une opération de contrôle alors que l'utilisateur n'avait pas encore fini de placer la pièce. Le développement d'un module de détection de tête, se basant sur les données de profondeur de la caméra, est envisagé pour résoudre ce problème. Finalement, en de rares occasions, la prise de pièce dans un bac n'a pas été détectée et il sera nécessaire de gérer ce cas de figure, par exemple en notifiant l'utilisateur et en lui permettant de gérer l'erreur de manière autonome.

La présente étude permet de contribuer aux avancées scientifiques dans le domaine des assistants de montage principalement de trois manières distinctes. Premièrement, elle démontre que les derniers algorithmes d'apprentissages profonds permettent de réaliser, de manière robuste, le suivi des mains de l'opérateur

malgré la position "vue de dessus" du capteur. Cette approche permet d'intégrer facilement des interactions avancées au système et d'envisager des fonctionnalités nouvelles comme l'automatisation du modèle GATM proposé par Funk et al. [16]. Deuxièmement, d'après l'analyse des résultats des questionnaires et nos observations, l'ajout d'information superposée directement sur l'espace de travail permet de fluidifier les tâches de l'opérateur en lui évitant d'avoir recours à l'écran déporté lorsque ce n'est pas nécessaire. Les études de Funk et al. [15–17] avaient déjà permis de démontrer que des instructions projetées *in-situ* permettaient de réduire la charge cognitive et d'améliorer l'efficacité des opérateurs lors de la phase d'apprentissage comparé à des instructions *in-situ*; une solution mixte n'avait cependant pas été évaluée dans le cadre de leurs études. Nous estimons cependant que des tests additionnels, sur une longue période et réalisés avec des outils permettant de suivre le regard de l'opérateur, seront nécessaires afin de valider quantitativement cette assumption. Finalement, l'évaluation du prototype, réalisée par la population cible, a permis de mettre en valeur l'intérêt et l'utilisabilité de la solution dans un contexte écologique. D'après l'analyse des résultats des questionnaires, pour la population cible, les principaux bénéfices perçus sont de gagner en autonomie et d'être rassuré sur les actions préalablement effectuées, notamment grâce aux opérations de contrôle mises en place pour chacune des étapes d'un montage dans la solution PMF-Vision.

6 CONCLUSION

Dans cet article nous avons présenté *PMF-Vision*, un assistant digital d'aide au montage basé sur la projection d'informations en réalité augmentée sur l'espace de travail et sur les dernières techniques d'apprentissage profond de traitement d'images. Nous avons présenté l'architecture logicielle du système, le fonctionnement global du système ainsi que ses différentes interfaces humain-machine. Nous avons identifié différents algorithmes potentiels pour la tâche de suivi des mains et sélectionné le candidat le plus performant en se basant sur une analyse quantitative. Nous avons réalisé une première étude exploratoire afin d'évaluer l'utilisabilité et l'acceptabilité de la solution par six utilisateurs finaux et de comparer deux modes de notification potentiellement utilisables pour préciser l'emplacement de dépôt d'une pièce. Cette première étude a permis de montrer que le système fonctionnait bien avec la population cible et que ces derniers considèrent son utilisabilité comme excellente avec un score moyen de 94.2 au questionnaire SUS. Bien qu'une différence significative entre les deux modes de notification n'a pas pu être clairement démontrée, le mode *superposé* semble avoir été plus apprécié par les utilisateurs et a paru être plus efficace selon nos observations. Finalement, les retours obtenus dans le cadre de l'étude sont très encourageants, la majorité des participants ont exprimé un fort intérêt à utiliser un tel système au quotidien, pour apprendre de nouveaux montages mais également durant les phases de production. Ils ont également mis en lumière le potentiel de la solution *PMF-Vision* pour leur donner une plus grande autonomie dans leur travail et ainsi permettre de se sentir plus valorisé et de renforcer leur confiance en eux.

REMERCIEMENTS

Nous tenons à remercier l'Agence Suisse pour l'Innovation (Innosuisse) qui supporte ce projet dans le cadre de la subvention 45748.1 IP-ICT. Nous tenons également à remercier le Centre d'Intégration Socioprofessionnel de Fribourg (CISF) pour son support lors de la réalisation des tests et son aide dans la réalisation du projet. Enfin, nous remercions Lino Pevereda de la compagnie PMF-System (PMFch SA) pour son support et son accompagnement.

REFERENCES

- [1] Arkite. 2020. Arkite. <https://arkite.com/product/>
- [2] Sven Bambach, Stefan Lee, David J. Crandall, and Chen Yu. 2015. Lending A Hand: Detecting Hands and Recognizing Activities in Complex Egocentric Interactions. In *2015 IEEE International Conference on Computer Vision (ICCV)*. 1949–1957. <https://doi.org/10.1109/ICCV.2015.226>
- [3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies* 4 (2009), 114–123. <https://doi.org/10.5555/2835587>
- [4] Jonas Blattgerste, Benjamin Streng, Patrick Renner, Thies Pfeiffer, and Kai Essig. 2017. Comparing Conventional and Augmented Reality Instructions for Manual Assembly Tasks. In *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, New York, NY, USA, 75–82. <https://doi.org/10.1145/3056540.3056547>
- [5] Marco Bortolini, Francesco Gabriele Galizia, and Cristina Mora. 2018. Reconfigurable manufacturing systems: Literature review and research trend. *Journal of Manufacturing Systems* 49 (oct 2018), 93–106. <https://doi.org/10.1016/j.jmsy.2018.09.005>
- [6] Marco Bortolini, Mauro Gamberi, Francesco Pilati, and Alberto Regattieri. 2018. Automatic assessment of the ergonomic risk for manual manufacturing and assembly activities through optical motion capture technology. In *Procedia CIRP*, Vol. 72. Elsevier B.V., 81–86. <https://doi.org/10.1016/j.procir.2018.03.198>
- [7] Bosch. 2017. Bosch Active Assist. <https://www.boschrexroth.com/en/xc/products/product-groups/assembly-technology/news/activeassist-assistance-system/index>
- [8] J Brooke. 1996. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 207–212. <https://doi.org/10.1201/9781498710411-35>
- [9] Sebastian Büttner, Oliver Sand, and Carsten Röcker. 2017. Exploring Design Opportunities for Intelligent Worker Assistance: A New Approach Using Projection-Based AR and a Novel Hand-Tracking Algorithm. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 33–45. https://doi.org/10.1007/978-3-319-56997-0_3
- [10] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2021. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 01 (jan 2021), 172–186. <https://doi.org/10.1109/TPAMI.2019.2929257>
- [11] Ana Colim, Carlos Faria, Ana Cristina Braga, Nuno Sousa, Luís Rocha, Paula Carneiro, Nelson Costa, and Pedro Arezes. 2020. Towards an Ergonomic Assessment Framework for Industrial Assembly Workstations—A Case Study. *Applied Sciences* 10, 9 (apr 2020), 3048. <https://doi.org/10.3390/app10093048>
- [12] Emmanuel Dubois, Laurence Nigay, J Troccaz, O Chavanon, and L Carrat. 1999. Design Space for Augmented Surgery, an Augmented Reality Case Study. In *Proc. Interact99, Edinburgh, A. Sasse & C. Johnson Eds, [IFIP] [IOS] Press Publ.* 353–359.
- [13] Åsa Fash-Berglund and Johan Stahre. 2013. Cognitive automation strategy for reconfigurable and sustainable assembly systems. *Assembly Automation* 33, 3 (jul 2013), 294–303. <https://doi.org/10.1108/AA-12-2013-036>
- [14] Emilio Ferrari, Mauro Gamberi, Francesco Pilati, and Alberto Regattieri. 2018. Motion Analysis System for the digitalization and assessment of manual manufacturing and assembly processes. *IFAC-PapersOnLine* 51, 11 (jan 2018), 411–416. <https://doi.org/10.1016/j.ifacol.2018.08.329>
- [15] Markus Funk, Andreas Bächler, Liane Bächler, Oliver Korn, Christoph Krieger, Thomas Heidenreich, and Albrecht Schmidt. 2015. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. *8th ACM International Conference on Pervasive Technologies Related to Assistive Environments, PETRA 2015 - Proceedings* (jul 2015). <https://doi.org/10.1145/2769493.2769496>
- [16] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with augmented reality? A long-term analysis of in-situ instructions at the assembly workplace. *ACM International Conference Proceeding Series Part F1285*, November (2017), 222–229. <https://doi.org/10.1145/3056540.3056548>
- [17] Markus Funk, Sven Mayer, and Albrecht Schmidt. 2015. Using in-situ projection to support cognitively impaired workers at the workplace. *ASSETS 2015 - Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (oct 2015), 185–192. <https://doi.org/10.1145/2700648.2809853>

- [18] Meiyuzi Gao, Philip Kortum, and Frederick L Oswald. 2020. International Journal of Human-Computer Interaction ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/hihc20> Multi-Language Toolkit for the System Usability Scale Multi-Language Toolkit for the System Usability Scale. (2020). <https://doi.org/10.1080/10447318.2020.1801173>
- [19] Google. [n.d.]. *Model Card MediaPipe Hands*. Retrieved 2021-09-09 from <https://drive.google.com/file/d/1yiPfk4b4hSbXJZaSq9vDmhZ24XVZmxpL/preview>
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [21] Hzzone. 2021. *pytorch-openpose*. *GitHub repository* (2021). <https://github.com/Hzzone/pytorch-openpose>
- [22] Ze-Hao Lai, Wenjin Tao, Ming C. Leu, and Zhaozheng Yin. 2020. Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing. *Journal of Manufacturing Systems* 55 (apr 2020), 69–81. <https://doi.org/10.1016/j.jmsy.2020.02.010>
- [23] LightingPick. 2021. *Build2Light*. <https://build2light.com/>
- [24] Liang Liu, Yong Liu, and Jiangning Zhang. 2019. Learning-Based Hand Motion Capture and Understanding in Assembly Process. *IEEE Transactions on Industrial Electronics* 66, 12 (dec 2019), 9703–9712. <https://doi.org/10.1109/TIE.2018.2884206>
- [25] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. [n.d.]. SSD: Single Shot MultiBox Detector. In *Computer Vision – ECCV 2016*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.).
- [26] Microsoft. [n.d.]. *Azure Kinect DK*. Retrieved 2021-09-09 from <https://azure.microsoft.com/fr-fr/services/kinect-dk/>
- [27] NVIDIA. [n.d.]. *JETSON AGX XAVIER*. Retrieved 2021-09-09 from <https://www.nvidia.com/fr-fr/autonomous-machines/embedded-systems/jetson-agx-xavier/>
- [28] NVIDIA-AI-IOT. 2021. *Hand Pose Estimation And Classification*. *GitHub repository* (2021). https://github.com/NVIDIA-AI-IOT/trt_pose_hand
- [29] Optoma. [n.d.]. *ML1050ST Projecteur LED Courte Focale Ultra-Compact*. Retrieved 2021-09-09 from <https://www.optoma.fr/product-details/ml1050st>
- [30] Munir Oudah, Ali Al-Naji, and Javaan Chahl. 2020. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques. *Journal of Imaging* 2020, Vol. 6, Page 73 6, 8 (jul 2020), 73. <https://doi.org/10.3390/JIMAGING6080073>
- [31] Sebastian Pimminger, Werner Kurschl, Mirjam Augstein, Josef Altmann, and Johann Heindelreiter. 2019. Low-cost tracking of assembly tasks in industrial environments. In *ACM International Conference Proceeding Series*, Vol. 19. Association for Computing Machinery, New York, NY, USA, 86–93. <https://doi.org/10.1145/3316782.3321526>
- [32] Sebastian Pimminger, Werner Kurschl, Lisa Panholzer, Thomas Neumayr, Mirjam Augstein, Josef Altmann, and Johann Heindelreiter. 2020. Assembly task analysis using the General Assembly Task Model (GATM) on the shop floor. *Procedia CIRP* 93 (2020), 1109–1114. <https://doi.org/10.1016/J.PROCIR.2020.04.007>
- [33] Sajeed C. Puthenveetil, Chinmay P. Daphalapurkar, Wenjuan Zhu, Ming C. Leu, Xiaoqing F. Liu, Julie K. Gilpin-McMinn, and Scott D. Snodgrass. 2015. Computer-automated ergonomic analysis based on motion capture and assembly simulation. *Virtual Reality* 19, 2 (jun 2015), 119–128. <https://doi.org/10.1007/s10055-015-0261-9>
- [34] Patrick Renner and Thies Pfeiffer. 2017. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces, 3DUI 2017 - Proceedings*. Institute of Electrical and Electronics Engineers Inc., 186–194. <https://doi.org/10.1109/3DUI.2017.7893338>
- [35] Patrick Renner and Thies Pfeiffer. 2017. Evaluation of attention guiding techniques for augmented reality-based assistance in picking and assembly tasks. In *International Conference on Intelligent User Interfaces, Proceedings IUI*. <https://doi.org/10.1145/3030024.3040987>
- [36] Soft2tec. 2021. *Nexonar*. <https://www.nexonar.com/en/worker-assistance/pick-place-commissioning>
- [37] The Imaging Source. [n.d.]. *DFK Z30GP031*. Retrieved 2021-09-09 from <https://www.theimagingsource.com/products/zoom-autofocus-cameras/gige-color-zoom-cameras/dfkz30gp031/>
- [38] Wenjin Tao, Ze Hao Lai, Ming C. Leu, Zhaozheng Yin, and Ruwen Qin. 2019. A self-aware and active-guiding training & assistant system for worker-centered intelligent manufacturing. *Manufacturing Letters* 21 (aug 2019), 45–49. <https://doi.org/10.1016/j.mfglet.2019.08.003>
- [39] Tulip. 2021. *Tulip*. <https://tulip.co/platform/vision/>
- [40] Dibia Victor. 2017. *HandTrack: A Library For Prototyping Real-time Hand Tracking Interfaces using Convolutional Neural Networks*. *GitHub repository* (2017). <https://github.com/victordibia/handtracking/tree/master/docs/handtrack.pdf>
- [41] Yue Wang, Shusheng Zhang, Sen Yang, Weiping He, and Xiaoliang Bai. 2018. Mechanical assembly assistance using marker-less augmented reality system. *Assembly Automation* (2018). <https://doi.org/10.1108/AA-11-2016-152>
- [42] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. 2016. Convolutional Pose Machines. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4724–4732. <https://doi.org/10.1109/CVPR.2016.511>
- [43] Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. *Commun. ACM* 36, 7 (jul 1993), 87–96. <https://doi.org/10.1145/159544.159630>
- [44] Li Da Xu, Eric L. Xu, and Ling Li. 2018. Industry 4.0: State of the art and future trends. *International Journal of Production Research* (2018). <https://doi.org/10.1080/00207543.2018.1444806>
- [45] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. *Mediapipe hands: On-device real-time hand tracking*. *arXiv preprint arXiv:2006.10214* (2020).