# Introduction

The use of signature is generalized in our society. Used in every field where identifying someone is important, it is a mean difficult to fraud. However, when signing is not possible, for example in an environment where contact must be reduced, there isn’t many other choices for identifying someone. Such environments might be scarce overall, but in the advent of a pandemic where social distancing could have great benefit (Thunström L. et al. 2020), they would become omnipresent.

One solution to answer this contact barrier is a contactless mean of signing. Ideally, something easy to use, that would provide good feedback on the signing procedure. No solution is available, but some experiments were done already (Mascarelli A. et al. 2023). They’ve shown the potential of using a computer camera for hand tracking and signature recognition, using deep-learning algorithms.

This report explores the possible improvements that could be made upon the previous work, such as:

* Improving the signature placement relative to the camera. The imaginary plane on which one might be drawing its signature may be tilted relative to the camera. As such the 2D projection will also be tilted. Hence, the need for some transformations.
* Improving the hand recognition algorithm. The previous work had some trouble differentiating the hand with the background or the person’s face. Therefore, the necessity to improve the hand recognition algorithm.
* Improving the signature recognition algorithm. The previous work has shown the capability of their project but does not permit to determine the beginning and the end of the signature.
* Trying to give a better estimate of the depth of the signature. The previous work heavily relies on relative depth computed via a pre-trained deep-learning algorithm. It can slow the video feed, but also incorrectly predict the drawing area.

Some other little improvements on the quality of life of the usage of their work will be made.

The report is organized as follows:

[Introduction 1](#_Toc160941584)

[Related Work 2](#_Toc160941585)

[Envisioned implementation 2](#_Toc160941586)

[Expected timeline and management 3](#_Toc160941587)

[Conclusion 3](#_Toc160941588)

[Reference 3](#_Toc160941589)

# Related Work

Gesture recognition problems can be resolved using two methods, or with a combination of both: the ones based on sensors, and the ones based on computer vision. This report primarily focuses on computer vision. The use of deep-learning algorithms has shown efficacity in detecting gestures and recognizing sign language gesture (Mustafa Z. et al. 2023), (De Castro et al., 2023). Most of them rely on training a pre-trained model with specialized datasets to obtain useful results. One such study uses Mediapipe Holistic (Goyal, K. et al. 2023) to show live feed sign language get translated to words. Mediapipe Holistic is an upgrade of the pretrained model used in the previous work. It can recognize at the same time both the face and the hands, thus resolving the problem we could have in the original work. It can also recognize predefined basics gestures.

Une image contenant capture d’écran, intérieur, mode

Description générée automatiquement

Figure 1[Mediapipe holistic on two use case "silence" and "hello"](https://blog.research.google/2020/12/mediapipe-holistic-simultaneous-face.html)

Following the same logic, depth perception problems can be resolved using two methods: specialized equipment and computer vision. The studies have shown potentials in estimating the depth using infrared light (Sun, S. et al. 2023), or LiDAR technology (Shao, S. et al. 2023). The studies experiment new algorithms to enhanced otherwise heavy processing. There are also studies trying to resolve this problem using two cameras (Li, Z et al.2021) giving results in lightness over the LiDAR method. On the other hand, the computer vision assisted method rely on deep-learning algorithms such as MiDaS, or more recently ZoeDepth (Bhat, S.F. et al. 2023) which is an improvement from MiDaS.

# Envisioned implementation

This report briefly presents the envisioned strategy to solve some on the precedent problematics.

1. To reduce interruption in drawing a signature, it is envisioned to use the improved version of MediaPipe, which is MediaPipe Holistic, as it seemed to give better results when considering a hand in front of a face. The model appears to be only 10% slower than the one used in the previous work (Google 2020).
2. MediaPipe might also be used to better define the start and the end of a signature, by implementing specific gestures to use.
3. To replace and transform the drawing perpendicular to the camera, linear algebra will be used. However, to this stage, it is not yet clear how to optimize it best.
4. Finaly, some tests will be done using two cameras, trying to replicate ADAS stereo vision, with the aim of optimizing the video feed output. Whether it work will define its use in a later stage, otherwise, the MiDaS depth estimation method will stay.

Whether the future work realize all those objectives will depend on the planning.

# Expected timeline and management

Considering the different tasks at hand, and the team’s composition, a schedule was made.

Une image contenant texte, capture d’écran, ligne, nombre

Description générée automatiquement

Figure 2 Tasks Schedule

The project’s schedule considers some group meeting to check on advancement regularly.

To keep track of the different tasks one [planner](https://trello.com/invite/b/QR4xOhmA/ATTI4d967fbcbc94211c82c68063892cdf34E8248233/remotescan-recognition) is used. A second [planner](https://trello.com/invite/b/d19CRSw7/ATTIbd9311de3c8abd52d6122d286b78cb4aA303EFDD/signature-recogition) is used to keep an historic of the papers read. They will help in finding relevant information related to the tasks.

Finaly, the [project repository can be found on github.](https://github.com/AllouetteVB/RemoteScan-Recognition) It will help keep a track of all the files and group work.

# Conclusion

# References

Thunström L, Newbold SC, Finnoff D, Ashworth M, Shogren JF. The Benefits and Costs of Using Social Distancing to Flatten the Curve for COVID-19. Journal of Benefit-Cost Analysis. 2020;11(2):179-195. doi:10.1017/bca.2020.12

MASCARELLI A., BITTON Y., AMIEL Q., LEFEUVRE K., COGNEVILLE E., DEVE E. (2023) Identification of Finger-based Signatures Using Computer Vision Techniques, (last project) (<https://github.com/FallenElias/Finger_AirSigning_Project>)

Mustafa Z, Nsour H, Tahir SBud. 2023. Hand gesture recognition via deep data optimization and 3D reconstruction. PeerJ Computer Science 9:e1619 <https://doi.org/10.7717/peerj-cs.1619>

De Castro, G. Z., Guerra, R. R., & Guimarães, F. G. (2023). Automatic translation of sign language with multi-stream 3D CNN and generation of artificial depth maps. Expert Systems with Applications, 215, 119394.

<https://blog.research.google/2020/12/mediapipe-holistic-simultaneous-face.html>

Goyal, K. (2023). Indian sign language recognition using mediapipe holistic. arXiv preprint arXiv:2304.10256.

Sun, S., Liu, R., & Yang, X. (2023). Depth-hand: 3d hand keypoint detection with dense depth estimation. IEEE Signal Processing Letters.

Shao, S., Pei, Z., Chen, W., Liu, Q., Yue, H., & Li, Z. (2023). Sparse Pseudo-LiDAR Depth Assisted Monocular Depth Estimation. IEEE Transactions on Intelligent Vehicles.

Bhat, S. F., Birkl, R., Wofk, D., Wonka, P., & Müller, M. (2023). Zoedepth: Zero-shot transfer by combining relative and metric depth. arXiv preprint arXiv:2302.12288.

Li, Z., Liu, X., Drenkow, N., Ding, A., Creighton, F. X., Taylor, R. H., & Unberath, M. (2021). Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6197-6206).

https://learnopencv.com/adas-stereo-vision/